

DOI: 10.32517/0234-0453-2021-36-9-40-46

ПОВЫШЕНИЕ КАЧЕСТВА МОДЕЛИ ПРОГНОЗИРОВАНИЯ ОБРАЗОВАТЕЛЬНЫХ РЕЗУЛЬТАТОВ СТУДЕНТОВ УНИВЕРСИТЕТОВ

Р. Б. Куприянов¹ ✉, Д. Ю. Звонарев¹

¹ *Московский городской педагогический университет, г. Москва, Россия*

✉ kupriyanovrb@mgpu.ru

Аннотация

Прогнозирование образовательных успехов обучающихся является одной из актуальных задач интеллектуального анализа образовательных данных. В данной работе рассмотрено два исследовательских вопроса: повышение качества модели прогнозирования образовательных результатов обучающихся университета и внедрение разработанной модели в реальный образовательный процесс университета. Исследуемые модели прогнозирования основаны на алгоритмах градиентного бустинга над решающими деревьями и линейной регрессии. По результатам проведенного исследования было выявлено, что данные по использованию электронной и университетской библиотек позволяют повысить качество прогнозирования образовательных результатов обучающихся, а также подтверждается факт, что мониторинг образовательных результатов учащихся в динамике оказывается информативнее при принятии управленческих решений в образовательном процессе, чем ориентир на итоговые результаты по успеваемости. Изученные в настоящей работе модели прогнозирования образовательных результатов обучающихся могут быть использованы в образовательных организациях высшего образования для своевременного выявления обучающихся с высоким риском отчисления, предоставления обратной связи студентам и преподавателям в части образовательных успехов обучающихся и управления образовательным процессом.

Ключевые слова: цифровая трансформация образовательного процесса, прогнозирование, интеллектуальный анализ образовательных данных.

Для цитирования:

Куприянов Р. Б., Звонарев Д. Ю. Повышение качества модели прогнозирования образовательных результатов студентов университетов. *Информатика и образование*. 2021;36(9):40–46. DOI: 10.32517/0234-0453-2021-36-9-40-46

IMPROVING THE QUALITY OF THE UNIVERSITY STUDENTS' ACADEMIC PERFORMANCE PREDICTION MODEL

R. B. Kupriyanov¹ ✉, D. Yu. Zvonarev¹

¹ *Moscow City University, Moscow, Russia*

✉ kupriyanovrb@mgpu.ru

Abstract

Predicting the educational success of students is one of the actual tasks of the intellectual analysis of educational data. In this article, two research issues are considered: improving the quality of the university students' academic performance prediction model and implementation the developed model into the real university educational process. The models predicting academic performance are based on XGBoost algorithm and the linear regression algorithm. According to the results of the study, it was revealed that data on the use of electronic and university libraries make it possible to improve the quality of predicting the students' academic performance, and also confirm the fact that monitoring the students' academic performance in dynamics is more informative in making managerial decisions in the educational process than the absolute values of the academic performance results. The models for predicting the students' academic performance studied in this work can be used in educational institutions of higher education for the timely identification of at-risk students, providing feedback to students and teachers regarding the educational success of students and managing the educational process.

Keywords: digital transformation of educational process, predicting model, educational data mining.

For citation:

Kupriyanov R. B., Zvonarev D. Yu. Improving the quality of the university students' academic performance prediction model. *Informatics and Education*. 2021;36(9):40–46. DOI: 10.32517/0234-0453-2021-36-9-40-46 (In Russian.)

1. Введение

Развитие цифровых технологий и средств коммуникации позволяет учащимся получить неограниченный доступ к образовательным материалам и способствует их самореализации. Одной из реакций системы высшего образования на комплекс явлений, связанных с резким ростом доступности знаний и обобщаемых понятием «интернет-революция в образовании», становится переход от статичных образовательных программ к гибким индивидуальным образовательным траекториям, формирующимся вокруг ядра образовательной программы и сочетающим в себе вариативность профессионального образования с реальными потребностями компаний из различных секторов экономики. Так, в работе [1] Дж. Беккер, рассматривая проблему индивидуальных образовательных траекторий в контексте реализации университетами системы свободных искусств и наук, определяет свободу выбора обучающимся как основополагающий элемент данной системы и указывает на необходимость отсутствия единого для всех обучающихся учебного плана и универсального пути к получению высшего образования.

Внедрение индивидуальных образовательных траекторий ведет к существенному усложнению образовательной программы, сказывается на трудозатратах на ее реализацию и администрирование.

Качество образовательной программы, как комплексная характеристика, требует одновременного анализа множества аспектов ее реализации, а рост сложности персональных образовательных треков существенно усложняет возможность экспертного прогнозирования результатов обучения. Перспективным направлением повышения уровня подготовки специалистов в системе высшего образования в современных условиях является разработка новых подходов к реализации педагогического мониторинга и прогнозирования образовательных успехов обучающихся с использованием средств современных информационных технологий [2].

Одним из методов, используемых для достижения данной цели, является **интеллектуальный анализ образовательных данных** [3], который позволяет решать следующие задачи:

- выявление студентов, входящих в группу риска отчисления;
- своевременный анализ большого объема данных;
- отслеживание отклонений от нормального хода образовательного процесса;
- поиск проблемных ситуаций в учебном процессе и анализ причин их возникновения;
- обработка и систематизация накопленных данных с целью трансформации образовательного процесса.

Следует отметить, что анализ и оценка образовательных результатов в силу специфики данного объекта исследования всегда подвержены суще-

ственному воздействию человеческого фактора — как со стороны обучающихся, так и со стороны педагогов. Использование методов искусственного интеллекта и интеллектуального анализа данных нивелирует вероятность субъективной интерпретации данных [4]. Экстраполируя этот вывод на проблему в целом, можно предположить, что применение технологии больших данных позволит обеспечить достоверность и объективность полученных результатов.

Целью настоящего исследования является изучение следующих вопросов:

1. Можно ли повысить качество ранее разработанной нами модели прогнозирования образовательных результатов обучающихся [5] за счет добавления данных о пользовании электронными и университетской библиотеками?
2. Является ли мониторинг образовательных результатов учащихся в динамике информативнее для принятия управленческих решений в образовательном процессе, чем ориентир на итоговые результаты по успеваемости?

2. Обзор литературы

Можно выделить ряд работ, посвященных проблематике исследования в области прогнозирования образовательных результатов обучающихся.

Так, проблему прогнозирования предметной обучаемости изучали Я. В. Луценко и В. Е. Коржаков [6]. Они предложили интеллектуальную технологию синтеза типовых моделей детерминации уровней предметной обученности студентов. В данном исследовании анализируются данные о социальном статусе студента и выполняется их сопоставление с учебными достижениями, где среди наиболее значимых обобщенных признаков присутствуют «социальное положение матери», «социальное положение отца», «количество детей в семье», «место рождения».

В статье [7] описано, как исследователи используют психологические факторы для прогнозирования образовательных достижений студентов, осуществляя их проверку на данных, полученных по результатам экзаменов.

Другое исследование предсказывало успеваемость обучающихся с помощью систем онлайн-обучения 145 студентов, используя данные онлайн-учебных мероприятий и форумов [8].

В исследовании [9] анализируется поведение учащихся в процессе обучения. В этом исследовании авторы применяют метод отслеживания взгляда для определения поведения учащегося при чтении. Эти данные помогли создать индивидуальную обратную связь между особенностями виртуальной реальности и обучением для улучшения навыков обучения студентов.

Существуют и другие подходы, направленные на совершенствование моделей прогнозирования. В работе [10] рассматривается применение нейронных

сетей, SVM* и алгоритмов дерева решений к трем академическим наборам данных. Согласно исследованию [11], в моделях прогнозирования также часто используются ансамблевые методы, которые объединяют несколько простых алгоритмов. Такой подход приводит к повышению точности по сравнению с простыми алгоритмами. Для восстановления пропорций между классами в несбалансированных наборах данных ученые обычно используют процедуры фильтрации примеров мажоритарного класса, кроме того, дублируются примеры миноритарного класса [12].

Обзор литературы приводит нас к выводу, что большинство исследователей применяют расширенные методы интеллектуального анализа данных, включающих как образовательные, так и личные данные учащихся. Решение проблемы объективной оценки успеваемости студентов и раннего выявления студентов из группы риска лежит в рамках многокритериального подхода к анализу педагогических процессов и явлений.

3. Методология

3.1. Модель данных и выборки

Выборка данных исследования составила 3695 студентов 2016–2018 годов поступления.

Из базы данных Московского городского педагогического университета для каждого студента были извлечены следующие метаданные:

- институт, в котором обучается студент;
- год, в котором студент поступил в университет;
- учебные дисциплины;
- баллы ЕГЭ;
- успеваемость по каждой учебной дисциплине за все периоды обучения;
- пол;
- город, в котором зарегистрирован студент;
- город, в котором проживает студент;
- участие в общественной деятельности;
- активность в использовании цифровой образовательной среды;
- использование книг из электронной или университетской библиотеки.

В качестве общественной деятельности рассматривались работа студента в качестве вожакого в летнем лагере, волонтерская деятельность, а также участие в университетских мероприятиях.

Для проверки первого исследовательского вопроса была использована вся выборка данных.

Для проверки второго исследовательского вопроса была выделена подвыборка данных с общей совокупностью анализируемого числа обучающихся в 1040 студентов.

3.2. Подготовка данных

В целях использования данных в модели прогнозирования образовательных результатов обучающихся была проведена предварительная обработка данных, включающая:

- общую оценку количества данных по каждому показателю;
- оценку количества данных по каждому показателю по годам поступления студентов;
- разложение поля участия в общественной деятельности в разные колонки;
- добавление колонки участия хотя бы в одном мероприятии;
- разложение поля с успеваемостью по разным колонкам таким образом, чтобы каждая оценка была в отдельной колонке;
- объединение данных с оценками по предметам с остальными данными студента.

Для проверки первого исследовательского вопроса были подготовлены два набора данных:

- первый набор данных (*набор данных 1*) был эквивалентен набору данных из нашего предыдущего исследования [5] и не содержал данные об использовании книг из электронной или университетской библиотеки;
- второй набор данных (*набор данных 2*) был эквивалентен набору данных из нашего предыдущего исследования [5], но содержал данные об использовании книг из электронной или университетской библиотеки.

3.3. Алгоритмы прогнозирования

Согласно нашему предшествующему исследованию [5], модель прогнозирования образовательных результатов обучающихся состояла из двух последовательных блоков:

- 1) прогнозирование образовательных успехов обучающихся;
- 2) кластеризация обучающихся на основе прогнозных значений их образовательных успехов.

Для изучения первого исследовательского вопроса были применены три модели прогнозирования.

Первая модель была основана на линейной регрессии. Для построения данной модели был использован набор данных 1.

Вторая и третья модели были основаны на алгоритме градиентного бустинга над решающими деревьями (XGBoost) [13–15]. При этом вторая модель была обучена на наборе данных 1, а третья модель — на наборе данных 2.

Все модели использовались для прогнозирования образовательных успехов студентов на один семестр вперед.

Для обучения каждой модели исходные выборки данных были разделены на обучающее и тестовое множества в соотношении 70:30. Для предотвращения переобучения моделей [16] и правильности оценки точности предсказания использовалась кросс-валидация с контрольной подвыборкой объемом в 20 %.

* SVM (англ. Support Vector Machine) — метод опорных векторов — один из наиболее популярных методов обучения, который применяется для решения задач классификации и регрессии.

Для кластеризация обучающихся на основе прогнозных значений их образовательных успехов был применен алгоритм k-средних [17, 18]. Согласно нашему предыдущему исследованию [19], кластеризация образовательных успехов обучающихся проводилась на четыре кластера.

3.4. Апробация модели прогнозирования

Для изучения второго исследовательского вопроса была проведена апробация модели прогнозирования, основанной на алгоритме градиентного бустинга над решающими деревьями, в четырех независимых подразделениях Московского городского педагогического университета.

Из исследуемой совокупности была выделена **выборка, сформированная из трех категорий студентов:**

1. Студенты с низкими текущими образовательными результатами:

$$\frac{\sum_{i=1}^m S_{ij}}{m} < X_{cr},$$

где:

S_{ij} — результат изучения j -й дисциплины i -м обучающимся;

m — количество изучаемых в текущем семестре дисциплин;

X_{cr} — значение верхней границы кластера группы риска.

2. Студенты с низкими прогнозируемыми результатами:

$$X_{pr_i} < X_{cr},$$

где X_{pr_i} — прогнозируемое значение образовательного результата i -го студента на следующей семестр.

3. Обычно успевающие студенты с низкими образовательными результатами по итогам последней сессии.

Предполагается, что у данных студентов средний балл по последней сессии будет ниже прогнозируемого:

$$\frac{\sum_{i=1}^m S_{ij}}{m} - X_{pr_i} < 0.$$

Полученные в результате выборки сведения уточнялись по причине неуспеваемости. Такими причинами становились сами преподаватели и особенности их преподавания, сложный для изучения предмет

и особенность его реализации, а также некоторые другие причины. По сути данный подход является альтернативой экспертным методам анализа, ориентированным на поиск «узких» мест образовательного процесса. В традиционном подходе для получения идентичных результатов потребовалось бы провести предварительный сбор экспертных данных с их последующей обработкой, при этом локализация исследуемой выборки оказалась бы затрудненной вследствие наличия невалидных данных.

4. Результаты

4.1. Оценка качества моделей прогнозирования

Оценка качества моделей прогнозирования позволяет ответить на первый исследовательский вопрос. Для оценки качества применяемых моделей прогнозирования были использованы общепринятые метрики оценки точности прогнозных моделей [20]: точность (Precision), полнота (Recall), F1-мера (F1 measure). Результаты оценки качества применяемых моделей прогнозирования представлены в таблице 1.

Данные в таблице 1 показывают заметный прирост качества модели прогнозирования на наборе данных 2, что подтверждает целесообразность применения данных о пользовании книгами из электронной или университетской библиотеки в задачах прогнозирования образовательных результатов обучающихся.

4.2. Результат апробации модели прогнозирования

По выделенным критериям (см. п. 3.4) были получены результаты, представленные в таблице 2.

Следует отметить, что студенты, имеющие критически низкие образовательные результаты, не всегда попадают в группу риска отчисления, тогда как студенты, демонстрирующие достаточно резкое снижение образовательных результатов, сопровождающееся переходом в другой кластер, имеют более высокие шансы быть отчисленными.

Дальнейший анализ осуществлялся путем сравнения среднего арифметического выборки (определенного по таблице 3) со средним по каждой дисциплине сессии и средним баллом проблемных студентов.

Формируемая структура текущей успеваемости в студенческой группе позволяет выделить направ-

Таблица 1 / Table 1

Оценка качества моделей прогнозирования Assessment of the quality of forecasting models

№ п/п	Модель	Точность (Precision)	Полнота (Recall)	F1-мера (F1 measure)
1	Линейная регрессия	0.662	0.627	0.640
2	XGBoost (набор данных 1)	0.698	0.677	0.673
3	XGBoost (набор данных 2)	0.710	0.716	0.711

Таблица 2 / Table 2

Результаты предварительного анализа успеваемости студентов

Results of the preliminary analysis of student performance

Критерии	Институт 1	Институт 2	Институт 3	Институт 4
$\frac{\sum_{i=1}^m S_{ij}}{m} < X_{cr}$	6,2 %	3,3 %	0,9 %	3,5 %
$X_{pr_i} < X_{cr}$	8,4 %	3,3 %	1,4 %	6,3 %
$\frac{\sum_{i=1}^m S_{ij}}{m} - X_{pr_i} < 0$	7,5 %	2,0 %	0,9 %	4,8 %

Таблица 3 / Table 3

Структура данных для детализированного анализа образовательных успехов обучающихся

Data structure for detailed analysis of students' educational success

	Наименование дисциплины					Среднее по семестру	Прогнозируемое
	S_1	...	S_j	...	S_m		
ФИО студента	L_1					$\frac{\sum_{i=1}^m S_{ij}}$	X_{pr_i}
	...						
	L_i						
	...						
	L_n						
Средний балл по дисциплине	$\frac{\sum_{i=1}^n S_{j1}}{n}$						
Средний балл по группе	$\frac{\sum_{i=1}^m \sum_{j=1}^n S_{ij}}{m}$						

ления для персонализированной работы со студентами.

Также представляется возможным выделить дисциплины, вызывающие у обучающихся проблемы с освоением. Далее причины неуспеваемости по отдельным дисциплинам могут быть уточнены по иным признакам: технологии преподавания, методические материалы и т. д. Отдельно следует обратить внимание на повторяемость низких образовательных результатов по дисциплине.

Помимо непосредственно анализа результатов причин низкой успеваемости был проведен опрос студентов, отобранных по критериям низкой успеваемости (см. табл. 3). В качестве дополнительных факторов, которые могут повлиять на улучшение образовательных результатов, студенты выделили следующие организационные факторы:

- изменение времени начала занятий;
- оптимизация расписания занятий с целью уменьшения количества учебных дней;

- реализация части практических и семинарских занятий с использованием технологий электронного обучения или дистанционных образовательных технологий.

Проведенная оценка уровня отчислений обучающихся в контрольных группах, обусловленных неудовлетворительными образовательными результатами, зафиксирована на уровне следующих значений:

- институт 1 — 1,4 %;
- институт 2 — 1,5 %;
- институт 3 — 0 %;
- институт 4 — 0,9 %.

Также были проведены интервью с обычно успевающими студентами с низкими образовательными результатами по итогам последней сессии. Было установлено, что значительное число респондентов рассматривают вариант прекращения своего образования. В связи с этим был проведен отдельный анализ причин отчисления и структуры контин-

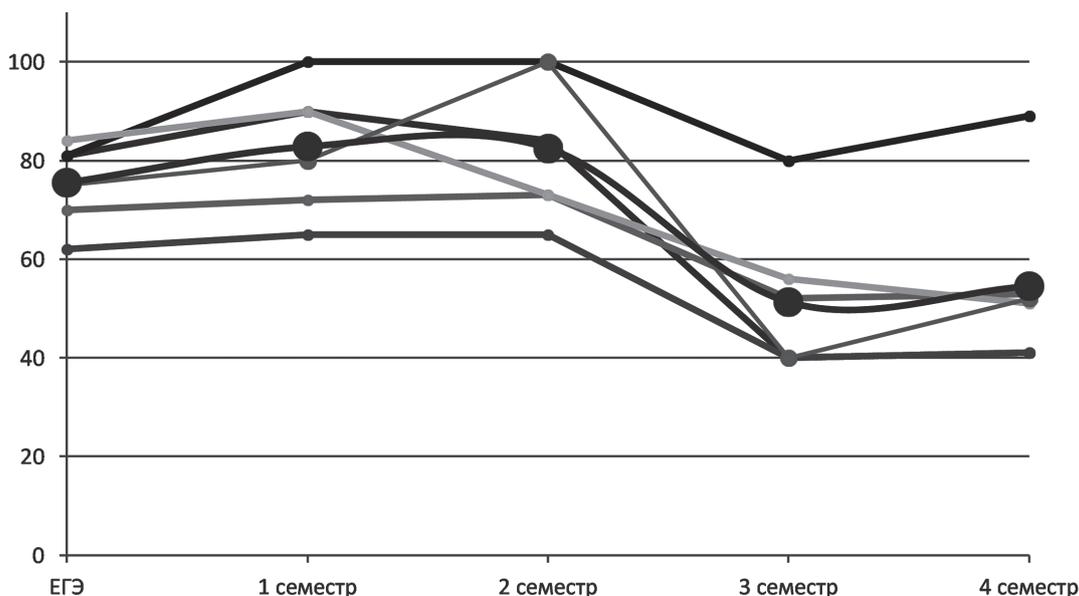


Рис. Кривые образовательных результатов, типичные для студентов, планирующих прекратить обучение (линией с крупными маркерами показана обобщенная кривая)

Fig. Learning outcome curves typical for students planning to drop out (the line with large markers shows the generalized curve)

гента обучающихся, отчисленных из университета по собственному желанию (объем исследуемой совокупности составил 46 обучающихся, данная выборка сформирована из исходной совокупности в 3695 обучающихся). В качестве причин нежелания продолжать обучение в вузе обучающимися были указаны следующие:

- ошибка в выборе будущей профессии (57 %);
- необходимость сократить время, затрачиваемое на дорогу в университет (12 %);
- сложность совмещения учебы и работы (9 %);
- иные причины (22 %).

Характерным маркером потери интереса к профессии оказалось снижение образовательных результатов по ядерным курсам образовательной программы, что особенно характерно для студентов «высоких» кластеров. Типичная траектория образовательных результатов представлена на рисунке.

Как правило, студенты демонстрируют стабильные или растущие образовательные результаты, после чего характерно резкое снижение этих результатов, часто сопровождающееся переходом в зону низких и крайне низких образовательных результатов.

5. Заключение

Обобщение полученных результатов исследования показывает, что интеллектуальный анализ данных об образовательных успехах обучающихся и последующая интерпретация результатов такого анализа могут быть успешно применены для решения различных управленческих, педагогических и образовательных задач: администрацией университета, преподавателями и студентами.

Преподаватель, читающий дисциплину, получает не обобщенный результат своей деятельности,

а стратифицирует его по выделенным адресным группам обучающихся, соизмеряет результат своего труда с аналогичной деятельностью других преподавателей. Он получает эффективный инструмент для самооценки и может искать новые педагогические подходы с учетом структуры обучаемых групп.

В свою очередь студенты могут объективно оценивать себя, не утрачивая при этом уверенности в своих силах, так как понимают, что о возможных предстоящих затруднениях знает также университет. Данный факт способствует предотвращению возникновения неприятных последствий как для студента, так и для университета.

Наконец, администрация получает математически выверенный инструмент для оценки эффективности деятельности всего университета в целом и каждого сотрудника в частности.

Огромным преимуществом системы интеллектуального анализа данных является ее самообучаемость. Последовательные итерации совершенствования модели обеспечивают ее развитие совместно с процессами реализации образовательной деятельности. В итоге выработанный алгоритм сначала построения модели, а затем ее практического применения может быть направлен на поиск оптимальных путей удовлетворения запросов участников образовательного процесса, делая этот процесс комфортным и в то же время повышая его эффективность и результативность.

Список источников / References

1. Беккер Дж. Образование по системе свободных искусств и наук: ответ на вызовы XXI в. *Вопросы образования*. 2015;(4):33–61. DOI: 10.17323/1814-9545-2015-4-33-61
 Becker J. Liberal arts and sciences education: Responding to the challenges of the XXIst century. *Educational Studies*. 2015;(4):33–61. DOI: 10.17323/1814-9545-2015-4-33-61

2. Matzavela V., Alepis E. Decision tree learning through a predictive model for student academic performance in intelligent m-learning environments. *Computers & Education*. 2021;2(6):100035. DOI: 10.1016/j.caeai.2021.100035

3. Rodrigues M. W., Zárate L. E., I sotani S., Educational Data Mining: A review of evaluation process in the e-learning. *Telematics and Informatics*. 2018;35(6):1701–1717. DOI: 10.1016/j.tele. 2018.04.015

4. Прошкина Е. Н., Балашова И. Ю. Анализ прогнозирования успеваемости студентов на основе радиальной базисной нейросети. *Технические науки: традиции и инновации. Материалы III Международной научной конференции*. Казань: Молодой ученый; 2018:24–27. Режим доступа: <https://moluch.ru/conf/tech/archive/287/13683/>

Proshkina E. N., Balashova I. Yu. Analysis of the academic success prediction based on the radial basis neural network. *Engineering Sciences: Tradition and Innovation. Proc. 3rd Int. Scientific Conf.* Kazan, Molodoy uchenyj; 2018:24–27. Available at: <https://moluch.ru/conf/tech/archive/287/13683/>

5. Куприянов Р. Б., Звонарев Д. Ю. Разработка модели прогнозирования образовательных результатов обучающихся для университетов. *Искусственный интеллект и принятие решений*. 2021;(2):11–20. DOI 10.14357/20718594210202

Kupriyanov R. B., Zvonarev D. Yu. Developing of the student's educational success prediction model for universities. *Artificial Intelligence and Decision Making*. 2021;(2):11–20. DOI 10.14357/20718594210202

6. Луценко Я. В., Коржаков В. Е. Прогнозирование уровня предметной обученности студентов путем СК-анализа данных об их социальном статусе. *Вестник Адыгейского государственного университета. Серия: Педагогика и психология*. 2007;(3):53–61. Режим доступа: https://www.elibrary.ru/download/elibrary_11933816_22068154.pdf

Lutsenko Ya. V., Korzhakov V. E. Predicting the level of students' subject proficiency by using the social cognitive method to analyze their social status data. *The Bulletin of the Aduyge State University. Series: Pedagogy and Psychology*. 2007;(3):53–61. Available at: https://www.elibrary.ru/download/elibrary_11933816_22068154.pdf

7. Halde R. R., Deshpande A., Mahajan A. Psychology assisted prediction of academic performance using machine learning. Psychology assisted prediction of academic performance using machine learning. *2016 IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT)*. P. 431–435. DOI: 10.1109/RTEICT.2016.7807857

8. Pardo A., Han F., Ellis R. A. Combining university student self-regulated learning indicators and engagement with online learning events to predict academic performance. *IEEE Transactions on Learning Technologies*. 2016;10(1):82–92. DOI: 10.1109/TLT.2016.2639508

9. Ramkumar Rajendran, Anurag Kumar, Carter K. E., Levin D. T., Biswas G. Predicting learning by analyzing eye-gaze data of reading behavior. *EDM*. 2018.

10. Czubala G., Mihai A., Crivei L. S PRAR: A novel relational association rule mining classification model applied for academic performance prediction. *Procedia Computer Science*. 2019;(159):20–29. DOI: 10.1016/j.procs.2019.09.156

11. Ajibade S.-S. M., Nor Bahiah Hj. A., Siti M. Hj. Sh. A data mining approach to predict academic performance of students using ensemble techniques. *Proc. Int. Conf. on Intelligent Systems Design and Applications (ISDA)*. 2018:749–760. DOI: 10.1007/978-3-030-16657-1_70

12. Mudasir Ashraf, Majid Zaman, Muheet Ahmed. An intelligent prediction system for educational data min-

ing based on ensemble and filtering approaches. *Procedia Computer Science*. 2020;(167):1471–1483. DOI: 10.1016/j.procs.2020.03.358

13. Breiman L. Random forests. *Machine Learning*. 2001;(45):5–32. Available at: https://www.cise.ufl.edu/~anand/fa11/Breiman_Random_Forests.pdf

14. Friedman J. H. Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*. 2001;29: 1189–1232. Available at: <https://www.jstor.org/stable/2699986?origin=JSTOR-pdf>

15. Friedman J. H. Stochastic gradient boosting. *Computational Statistics & Data Analysis*. 2002;38:367–378. DOI: 10.1016/S0167-9473(01)00065-2

16. Helmrich I. R., van Klaveren D., Steyerberg E. W. Research Note: Prognostic model research: overfitting, validation and application. *Journal of Physiotherapy*. 2019;65(4):243–245. DOI: 10.1016/j.jphys.2019.08.009

17. Lloyd S. P. Least squares quantization in PCM. *IEEE Transactions on Information Theory*. 1982;28(2):129–137. DOI: 10.1109/TIT.1982.1056489

18. MacQueen J. Some methods for classification and analysis of multivariate observations. *Proc. 5th Berkeley Symp. on Math. Statistics and Probability*. 1967:281–297. Available at: <http://www.cs.cmu.edu/~bhiksha/courses/mlsp.fall2010/class14/macqueen.pdf>

19. Куприянов Р. Б., Семенов А. Л. Анализ динамики образовательных результатов студентов крупного педагогического университета. *Вестник Московского городского университета. Серия: Информатика и информатизация образования*. 2018;(1 (43)):66–71. Режим доступа: <https://www.mgpu.ru/wp-content/uploads/2018/06/Vestnik-MGPU-Informatika-1-2018.pdf>

Kupriyanov R. B., Semenov A. L. The analysis of dynamics of educational results of students of a large teacher training university. *Vestnik of Moscow City University. Series: Informatics and Informatization of Education*. 2018;(1 (43)):66–71. Available at: <https://www.mgpu.ru/wp-content/uploads/2018/06/Vestnik-MGPU-Informatika-1-2018.pdf>

20. Luque A., Carrasco A., Martín A., de las Heras A. The impact of class imbalance in classification performance metrics based on the binary confusion matrix. *Pattern Recognition*. 2019;91:216–231. DOI: 10.1016/J.PATCOG.2019.02.023

Информация об авторах

Куприянов Роман Борисович, зам. начальника управления информационных технологий, Московский городской педагогический университет, г. Москва, Россия; ORCID: <http://orcid.org/0000-0001-5153-4334>; e-mail: kupriyanovrb@mgpu.ru

Звонарев Дмитрий Юрьевич, канд. тех. наук, доцент, директор многофункционального студенческого центра, Московский городской педагогический университет, г. Москва, Россия; ORCID: <http://orcid.org/0000-0002-6520-0746>; e-mail: zvonarevd@mgpu.ru

Information about the authors

Roman B. Kupriyanov, Deputy Head of the IT Department, Moscow City University, Moscow, Russia; ORCID: <http://orcid.org/0000-0001-5153-4334>; e-mail: kupriyanovrb@mgpu.ru

Dmitry Yu. Zvonarev, Candidate of Sciences (Engineering), Docent, Head of the Multifunctional Student Center, Moscow City University, Moscow, Russia; ORCID: <http://orcid.org/0000-0002-6520-0746>; e-mail: zvonarevd@mgpu.ru

Поступила в редакцию / Received: 13.10.2021.

Поступила после рецензирования / Revised: 08.11.2021.

Принята к печати / Accepted: 09.11.2021.