

ПРОГНОЗИРОВАНИЕ УСПЕШНОСТИ ОБУЧЕНИЯ: ПРОБЛЕМЫ И ЗАДАЧИ¹

Т. А. Кустицкая, М. В. Носков, Ю. В. Вайнштейн

Аннотация. Статья посвящена проблемам прогнозирования успешности обучения. Целью работы является обсуждение актуальных задач и возможных сложностей, касающихся создания сервисов прогнозирования успешности обучения в цифровой среде образовательной организации. Среди многообразия задач прогнозирования, возникающих в учебной аналитике, авторами были выделены и подробно рассмотрены два основных направления: прогнозирование отсева обучающихся и прогнозирование академической успеваемости по дисциплинам учебного плана. В статье рассмотрены примеры создания и использования моделей прогнозирования в учебном процессе организациями среднего общего и высшего образования. Несмотря на большое число исследований в данном направлении, существуют лишь единичные примеры успешно внедренных на региональном уровне или хотя бы на уровне отдельной образовательной организации систем прогнозирования. Основными препятствиями к построению хорошо масштабируемой системы поддержки успешности обучения на основе прогнозных моделей, по мнению авторов, являются сложности с унификацией образовательных данных, отсутствие политики использования личных данных в учебной аналитике, недостаточная проработанность механизмов обратной связи и мероприятий коррекции поведения обучающихся. Решение каждой из проблем, поставленных в статье, является отдельной серьезной научной задачей. Обозначены перспективы использования результатов исследования.

Ключевые слова: прогнозирование успешности обучения, учебная аналитика, система поддержки успешности обучения, отсев обучающихся, цифровая среда вуза.

Для цитирования: Кустицкая Т. А., Носков М. В., Вайнштейн Ю. В. Прогнозирование успешности обучения: проблемы и задачи // Наука и школа. 2023. № 4. С. 71–83. DOI: 10.31862/1819-463X-2023-4-71-83.

¹ Исследование проводилось за счет средств гранта в форме субсидии на реализацию программы стратегического академического лидерства «Приоритет 2030».



PREDICTING LEARNING SUCCESS:
RESEARCH PROBLEMS AND CHALLENGES

T. A. Kustitskaya, M. V. Noskov, Y. V. Vainshtein

Abstract. *The article is devoted to the problems of learning success prediction. The aim of the work is to discuss current tasks and possible difficulties related to the development of services for predicting learning success in the digital environment of an educational institution. Among the variety of forecasting tasks arising in educational analytics, two main directions were identified and examined in detail: prediction of student dropout and prediction of academic performance for courses of the curriculum. The article discusses examples of creating and using predictive models in the educational process by secondary and higher education organizations. It is noted that despite the large number of studies in this problem field, there are only few examples of successfully implemented regional or at least organizational-level forecasting systems. The authors believe that the main obstacles to building a well-scalable system for supporting learning success based on predictive models are difficulties with data unification, lack of policy of using personal data in learning analytics, lack of feedback mechanisms and activities for correcting learning behavior. Solving each of these problems is a separate serious scientific task. The prospects for using the results of the research are indicated.*

Keywords: *learning success prediction, learning analytics, learning success support system, student dropout, digital educational environment.*

Cite as: Kustitskaya T. A., Noskov M. V., Vainshtein Y. V. Predicting learning success: research problems and challenges. *Nauka i shkola*. 2023, No. 4, pp. 71–83. DOI: 10.31862/1819-463X-2023-4-71-83.

Введение

С развитием информатизации образования в информационных системах вузов стало накапливаться большое количество образовательных данных, а управление образованием постепенно стало осуществляться на их основе. Вопрос о максимально эффективном использовании этих данных стал отдельной исследовательской задачей. Это привело к возникновению нового научного направления, соединяющего образовательные данные, математические методы и информационные технологии. Это научное направление получило название «анализ образовательных данных» или учебная аналитика. К сегодняшнему дню сформировалось несколько ветвей учебной аналитики, одной из которых является прогнозирование результатов обуче-

ния. Одной из главных задач прогнозирования результатов обучения является прогноз отсева студентов, так как значительные потери контингента приводят к финансовым и репутационным потерям вуза и зачастую отрицательно влияет на судьбу каждого отчисленного студента. Официальной первопричиной отсева, за редким исключением, является академическая задолженность студента. Поэтому прогнозирование успешности обучения по дисциплинам учебного плана может стать важным инструментом для профилактики отчисления студентов. На формирование академической задолженности и на отсев студентов влияют многие причины, учебного, демографического и социально-экономического характера. Эти же данные являются основой для раннего прогнозирования факта возникновения академической за-

долженности и отсева студентов с помощью методов искусственного интеллекта или статистического анализа данных. Под термином «раннее прогнозирование» мы будем понимать, что прогнозирование осуществлено тогда, когда в случае негативного прогноза еще есть достаточное время для корректирующего воздействия и исправления ситуации. Наличие данных, методов их обработки, прогноза и механизмов коррекции поведения обучающихся говорит о том, что можно создать сервис прогнозирования в рамках цифровой среды вуза [1].

В Сибирском федеральном университете (СФУ) в настоящее время ведется работа по созданию сервиса раннего прогнозирования успешности обучения для решения двух задач:

Задача А. Прогнозирование отсева студентов

Задача Б. Прогнозирование успешности обучения студентов по дисциплинам учебного плана.

В ходе решения этих задач возникли определенные сложности, с которыми будут сталкиваться любые научно-исследовательские коллективы, работающие над их решениями. Целью настоящей статьи является обсуждение актуальных задач и возможных проблем, касающихся создания сервисов прогнозирования успешности обучения студентов в цифровой среде вузов.

Можно ли создать сервис прогнозирования в цифровой среде вуза?

Казалось бы, что при всех возможных достоинствах сервисов прогнозирования успешности обучения в вузах, при той пользе, которую они могут принести руководству вуза в части экономики, руководству подразделений для повышения качества обучения и отдельным студентам для успешного обучения, библиографические базы должны быть наполнены статьями по их успешному созданию

и исследованию. Однако внимательное изучение литературы показывает почти полное отсутствие прямого указания на университеты, в цифровой среде которых присутствуют такие сервисы. Это отмечается, например, в [2]. В исследовании [3] утверждается, что имеются только две успешные попытки создать сервисы прогнозирования отчисления обучающихся: *Wisconsin Dropout Early Warning System (DEWS)*, разработанная Департаментом народного просвещения штата Висконсин, и *Course Signals*, разработанная в университете Пердью (Purdue University, штат Индиана, США). К этим системам можно добавить, по нашему мнению, успешные разработки системы раннего прогнозирования в Миланском политехническом университете (Politecnico di Milano) [4] и в университете Вупперталя земли Северный Рейн-Вестфалия (University of Wuppertal). Остановимся подробнее на этих системах.

Система *DEWS* была создана как сервис информационной системы штата Висконсин для образования, благодаря инвестициям в ее разработку в виде миллионов долларов федеральных и государственных средств. Эта система года была создана весной 2012 г., и она решает задачу А. *DEWS* предоставляет прогнозы более чем для 225 000 учащихся из Висконсина более чем из 1000 школ штата два раза в год. Система охватывает учащихся младших классов, и точность ее прогноза достигает 75% [5]. Отметим, что она создана и функционирует в рамках государственной поддержки, сбор данных согласован и унифицирован. Набор предикторов системы включает в себя:

- дни пропущенных занятий;
- дни удаления из-за плохого поведения;
- школы и округа, которые посещались ранее;
- оценки по чтению и математике;
- демографические данные: раса, пол и т. п.

Перейдем к университетским системам раннего прогнозирования. Из [6] следует, что исследователям из университета Вупперталь удалось создать действующую систему раннего прогнозирования (*EDS*), решающую задачу *A* и использующую данные о студентах, которые собираются и хранятся в соответствии с Законом о статистике высшего образования (*HStatG*) Германии. Это персональные данные, данные о предыдущем обучении, данные об условиях зачисления, академические данные, состоящие из количества экзаменов, средней оценки за семестр и результатов сдачи экзаменов (по семестрам). Для классификации студентов с высоким риском отчисления используется ансамбль методов: модель линейной регрессии, нейросети, алгоритм дерева решений и алгоритм *AdaBoost* [7]. Система протестирована в государственном университете на 23 000 студентов и в частном университете на 6500 студентов. В статье утверждается, что «результаты показывают, что 74% студентов государственного университета и 72% частного университета правильно идентифицированы в конце первого семестра с использованием обширных демографических данных и данных об успеваемости; кроме того, точность *EDS* повышается по мере того, как в конце каждого семестра становятся доступными новые данные об успеваемости студентов: после четвертого семестра *EDS* правильно предсказывает 80% отсева студентов государственных университетов и 83% частных университетов прикладных наук». Очевидным достоинством системы *EDS* является ее переносимость, ввиду стандартизированных данных.

Система, созданная в Миланском политехническом университете (*Politecnico di Milano*), также нацелена на решение задачи *A* [8]. Ее особенностью является то, что она предназначена для прогнозирования отсева первокурсников. Поэтому данные для прогнозирования исполь-

зуется существенно меньше, а именно: демографические данные, информация о финансовом положении, средние данные по сессии первого семестра. К сожалению, в работе [8] не представлено оценки результативности системы, отмечено лишь, что в качестве мер по предупреждению отсева по результатам прогнозирования применяется стратегия, которая «состоит в предоставлении учащимся репетиторства и дополнительной помощи в учебе для преодоления возникающих трудностей» [8, с. 1575].

На решение задачи *B* нацелена система *Course Signals* университета *Пердью*. Главная цель создания системы в том, чтобы использовать бизнес-аналитику для повышения успеваемости студентов на уровне дисциплины. Таким образом, это позволяет повысить показатели сохранности контингента и эффективности выпуска. Данные, на которых построено прогнозирование успешности обучения, извлекаются из системы управления обучением (в России наибольшее распространение получила система *LMS Moodle*), системы управления учебным процессом (в России широко используется АИС «Планы») и электронного журнала преподавателя. Для выделения группы риска используется индикатор типа «светофор», показывающий, насколько подвержен риску каждый учащийся. После обработки данных каждый учащийся получает цветовой сигнал степени риска: красный указывает на высокую вероятность неуспеха, желтый – на потенциальные проблемы, а зеленый – на высокую вероятность успеха. Как утверждается в [9], «потенциальные проблемы выявляются уже на второй неделе семестра. Это позволяет своевременно принять решение о воздействии, например, опубликовать сигнал на странице студента, отправить ему электронное письмо, сообщение или организовать встречу. Положительная оценка в виде зеленого сигнала светофора может сопрово-

ждаться позитивным сообщением» [9, с. 8]. Из текста в [9] следует, что помимо успеваемости система *Course Signals* оценивает и вовлеченность студентов по данным из LMS и общения с преподавателями.

Как отмечают разработчики, в результате работы системы по некоторым предметам успеваемость увеличилась на 12–14%; улучшились взаимоотношения между студентами и преподавателями; по мнению преподавателей, у учащихся повысилась вовлеченность в предмет после воздействия сигналов светофора и улучшилось отношение к учебной деятельности; студенты в рамках анкетирования отметили, что «уменьшилось их чувство “просто числа”». Также в работе [9] представлен опыт создания сервисов раннего прогнозирования других университетов США и Великобритании.

Заметим, что из рассмотренных систем только о первых двух можно уверенно говорить как о вполне завершенных проектах, остальные, мы полагаем, возможно рассматривать как пилотные проекты.

Работы по созданию систем раннего прогнозирования идут во многих странах мира – в Перу [10], Испании [3], Словакии [11] и других. И хотя значительная часть статей в области учебной аналитики посвящена поиску методик обработки данных, дающих наиболее точное решение задач прогнозирования на заданных множествах данных, общая направленность работ на создание полнофункциональных систем не вызывает сомнений.

Рассмотренные примеры систем и сервисов подтверждают возможность создания систем раннего прогнозирования. Такая система позволяет точно выявлять и воздействовать на студентов с высоким риском неуспешности обучения и решать задачи сохранности контингента. Этот вывод подтверждается и первыми результатами работы сервиса прогнозирования СФУ.

Таким образом, по нашему мнению, *одной из главных задач учебной аналитики в России является создание прогностической системы сохранности контингента вуза в условиях российского образования, причем целесообразно говорить о создании переносимой системы (сервиса, платформы), которая может быть внедрена в любом вузе страны.* Поэтому данные, используемые в сервисе, должны опираться на достаточно разумную легкость сбора и хранения, а главное, они должны быть унифицированы для системы высшего образования, как, например, в рассмотренном примере системы в университете Вупперталь, которая использует данные о студентах, «которые по закону обязаны хранить и регулярно обновлять все немецкие университеты; таким образом подразумевается, что ее (систему) можно легко внедрить в каждом университете страны» [6].

Какие данные нужны?

Качество прогноза зависит от качества данных об обучаемых, процессе обучения и от выбранного набора предикторов. Анализ литературы позывает, что возможные предикторы можно сгруппировать в три категории:

- результаты предыдущего обучения: оценки из аттестата, баллы при поступлении, результаты предыдущих сессий, то есть данные цифровой образовательной истории [12];
- демографические данные учащихся: пол, возраст, раса / этническая принадлежность, социально-экономический статус (образование и род занятий родителей, место жительства, доход семьи и т. д.), то есть данные цифрового профиля обучающегося;
- показатели текущего обучения: количество входов в LMS, количество выполненных заданий или тестов, количество «кликов» в системе, время выполнения задания и т. д.

Для решения задачи А такие данные, как «пол» и «средний балл за сессию», присутствуют практически во всех статьях. Примерно в половине всех работ в качестве предикторов используются «национальность», «финансовая обеспеченность», «специальность обучения», «образование родителей» и «условия проживания». Другие предикторы встречаются значительно реже или специфично для страны, в которой ведется исследование – например, «каста» в Индии. Отметим, что в работах, связанных с задачей А, текущая успеваемость практически не используется. Поэтому результаты прогнозирования обновляются не чаще двух раз в году (после каждой сессии). Достаточно ли это для эффективного управления контингентом? На наш взгляд, решение задачи Б на основе данных о текущем образовательном процессе может внести существенный вклад в уточнение прогноза отсева обучаемых, получаемого моделью А.

Для решения задачи Б, как правило, используются многочисленные данные из LMS и журналов преподавателей. В некоторых случаях пытаются привлечь такие демографические данные, как «пол», «национальность» и направление подготовки. Главной проблемой здесь является непрерывность и корректность поступающих данных, а также их единообразие. Так, в Мадридском открытом университете (UDIMA) электронные курсы «имеют общую структуру: все курсы состоят из 10 дидактических блоков и нескольких мероприятий, включающих от 3 до 6 форумов или глоссариев, от 2 до 6 анкет (с 2 попытками) и от 2 до 6 заданий» [3]. Все ли данные об образовательном поведении из LMS необходимо использовать для эффективного прогнозирования успешности обучения, также является большим вопросом. Так, например, такой предиктор, как время, затраченное на выполнение действий в MOODLE,

вызывает дискуссию в литературе. Например, в [3, с. 2] указывается, что «время не оказалось значимой переменной для прогнозирования успеваемости», а в [13, 14] этот параметр успешно используется для исследования.

До сих пор мы говорили о данных, сбор которых автоматически ведется в информационных системах. Но зачастую в исследованиях используются данные, которые можно собрать через опросы, требующие, как правило, добровольного согласия студентов, или данные, относящиеся к конфиденциальной информации. К таким данным, относится, например, информация об образовании родителей [15], инвалидность [16], финансовое состояние родителей [17] и т. д. Нужны ли такого рода данные для построения систем раннего прогнозирования успешности обучения (СРП) в России? Во всяком случае, не всегда мы можем напрямую переносить зарубежные результаты исследований на отечественные разработки, прежде всего, из-за различий в организации обучения и законодательства в области защиты данных.

Таким образом, *проблема, которую необходимо решить научному сообществу, заключается в определении состава предикторов, на основе которых может работать высокоточная СРП, и необходимых для этих предикторов данных. Результатом решения этой проблемы должны стать рекомендации по сбору данных для университетов, планирующих внедрение СРП.*

Завершая тему о данных, заметим, что «вопрос этики учебной аналитики, при том, что данная тема является одной из самых обсуждаемых среди зарубежных исследователей» [18], по-прежнему не вызывает большого интереса в России. Поэтому *задача выработки общей политики по использованию личных данных студента в научных исследованиях является одной из основных в российской учебной аналитике.*

Как использовать данные?

В системах раннего прогнозирования успешности обучения образовательные данные используются для прогнозирования результата обучения. В случае задачи *А* этот успешный результат означает получения диплома, в случае задачи *Б* – сдачу экзамена или зачета. Но «предсказать, сдаст или нет учащийся, будет сложнее, чем выявить учащихся из групп риска и отличников» [19, с. 6]. Иначе говоря, используемые для решения задач *А* и *Б* методы машинного обучения выдают прогноз успешности завершения обучения (в вузе или по данной дисциплине), но при каком именно прогнозе следует студента отнести к группе риска, необходимо решать из практических соображений.

Рассмотрим этот вопрос подробнее на примере методов классификации, выдающих вероятностный прогноз, которые оценивают риск отсева числом от 0 до 1. Для принятия решения об отнесении студента к группе риска СРП требуется пороговое значение, выше которого, на основании результатов прогноза, потенциальные неуспевающие определяются как подверженные риску (отсева или неаттестации по дисциплине). Чем ниже выбранный порог, тем выше процент выявленных системой неуспевающих. Но в то же время процент идентифицированных в качестве неуспевающих снижается: многие студенты, которые не будут отчислены, рассматриваются как обучающиеся из группы риска, то есть алгоритм прогнозирования «перестраховывается». Определившись с желаемым балансом, пороговое значение следует подбирать, добиваясь баланса на валидационной выборке (путем оптимизации различных функционалов качества на этой выборке). Затем прогноз распространяется на тестовую выборку студентов и оценивается точность. Получить прогноз с такой же точностью на новых данных возможно

только в случае однородности данных. В стабильных обществах для каждого университета подобрать метод прогнозирования, для которого задача *А* будет иметь удовлетворительное решение, достаточно долго.

Иное дело задача *Б*. Здесь мы сразу имеем дело с относительно небольшим количеством студентов. Мы можем это количество увеличить, если курс читается достаточно долгое время с одинаковым дидактическим сопровождением, например, одним и тем же преподавателем, с одним и тем же электронным курсом и т. д., как это описано в [20]. Если же принять во внимание разнообразие и количество дисциплин в вузе, то поиск единого метода прогнозирования ко всем дисциплинам становится сложной задачей. Многочисленные попытки найти такую единую методику относительно близких по направлению дисциплин, например для курсов информационного цикла [21], дают, как правило, невысокую точность. Наиболее интересный и масштабный эксперимент был проведен в Университете Овьедо (Испания) [19]. В этом университете учился 29 602 студента с поддержкой 5112 электронных курсов. После фильтрации по установленным исследователями критериям относительно однообразности курсов ими было отобрано 699 курсов. Применяя разнообразные алгоритмы, авторы получили точность прогноза до 93% к середине семестра.

Таким образом, можно сделать вывод, что создать полноценную систему, решающую задачу *Б*, опираясь только на данные электронных курсов, скорее всего, не удастся. Возможно, что подход к решению этой задачи, в смысле выделения кластера студентов высокого риска неуспешности обучения, лежит в построении прогноза на тех данных, которые характеризуют личностные качества студента. Как пример такого подхода можно рассматривать методики, предложенные в [22; 23], которые

нашли подтверждение в эксперименте. Очевидным недостатком этих работ является интуитивность в выборе предикторов, характеризующих личностные характеристики студентов. Поэтому задача о том, *какие из предикторов или их сочетаний и с какой достоверностью характеризуют личностные черты студента, влияющие на успешность обучения*, является одной из самых трудных и интересных задач учебной аналитики, по своим результатам далеко уходящих за пределы прогнозирования успешности обучения.

Что делать со студентами с высоким риском неуспешности?

Анализируя методы и модели работы с данными, представленные в литературе, можно заметить, что наибольшее распространение получили нейронные сети, логистическая регрессия, модели на основе решающих деревьев, байесовские сети и другие алгоритмы машинного обучения, их модификации и ансамбли. Результаты их работы позволяют делать выводы о принадлежности или непринадлежности студента к группе высокого риска. Однако затем возникает следующий вопрос: что делать с этими студентами дальше?

Большая часть статей говорит о том, что прогнозирование успешности позволит «определить учеников, которым требуется помощь, и поможет преподавателям проложить путь повышения успеваемости ученика» [24, с. 4]. Однако, кроме предупреждений о попадании в группу риска и бесед со студентами и их родителями, иных реальных стратегий, как правило, не предлагается. В этом смысле интересна статья [25], где предлагается совмещение учебной аналитики с адаптивными обучающими ресурсами. Однако, кроме вышеприведенных данных по университету в Пердью, мы не нашли информации о том, как изменяется ситуация даже при про-

стом предупреждении студентов. А ведь здесь все далеко не просто. Например, в [26, с. 3] отмечено, что «существуют значительные различия в том, как учащиеся реагируют на прогноз (имеется в виду попадание в группу риска): они были любопытны и мотивированы, утешены и настроены скептически, растеряны и напуганы, а некоторые учащиеся не интересовались этим или сомневались в точности предсказаний». С другой стороны, как отмечено в [27, с. 1], «обратная связь, полученная на основе данных... без надлежащего учета характера учащихся, может помешать эффективному вмешательству». Эти соображения приводят к задаче *исследований реакции обучаемых из групп рисков на прогноз неудачного обучения и на внешние воздействия, с учетом типов воздействия*.

По существу, об этой задаче говорится в [27, с. 5]: «в больших классах, где индивидуальная обратная связь невозможна, а общая обратная связь не очень информативна, оптимальным путем является выделение различных профилей обучения и акцент на создании обратной связи и вмешательстве, специфичных для этих профилей».

Как показано в [20], простое предупреждение о попадании в группу высокого риска зачастую приводит к изменению отношения студентов к обучению. Это приводит и к изменению статистических закономерностей в данных. Таким образом, в моделях классификации обучаемых по риску неуспешности необходим учет обратных связей. Иначе говоря, система должна иметь возможность динамически обновлять и обучать модель при изменении основных атрибутов, определяющих результаты студентов. Поэтому интересно было бы *построить и обосновать модель классификации обучаемых по риску неуспешности обучения при непрерывной обратной связи (то есть при непрерывном воздействии на обучаемого)*. Пока таких

моделей не существует, но без их наличия мы всегда будем полагаться на интуицию преподавателей, кураторов или других участников учебно-воспитательного процесса и надеяться на то, что их действия, по крайней мере, не нанесут вреда студентам.

Заключение

В настоящей работе были рассмотрены актуальные задачи и возможные проблемы, касающиеся создания сервисов прогнозирования успешности обучения студентов в цифровой среде вузов. Мы полагаем, что комплексное решение выявленных проблем позволит создать эффективную систему поддержки успешности обучения. Разумеется, рассмотренными в статье задачами прогнозирования успешности обучения

задачи учебной аналитики далеко не исчерпываются. Интерес к анализу образовательных данных в последние годы значительно возрос, что выражается, в частности, в появлении большого числа научных статей. Анализируя существующие исследования, мы затронули только те задачи, которые опираются на данные цифровой среды образовательной организации. Но пандемия новой короновирусной инфекции привела еще к одному явлению – применению социальных сетей в качестве инструмента обучения и организации учебного процесса. Мы считаем, что, помимо использования классических источников образовательных данных, следует обратить внимание на этот новый источник, как имеющий высокий потенциал с точки зрения исследования образовательного поведения обучающихся.

СПИСОК ЛИТЕРАТУРЫ

1. Активная информационная система вуза в информационно-образовательной среде / Г. М. Цибельский, М. В. Носков, Р. А. Барышев, М. В. Сомова // Педагогика. 2017. № 3. С. 28–32.
2. Learning Analytics: Analysis of Methods for Online Assessment / V. Renò, E. Stella, C. Patruno [et al.] // Appl. Sci. 2022. № 12. 9296.
3. *Bravo-Agapito J., Romero S. J., Pamplona S.* Early prediction of undergraduate Student's academic performance in completely online learning: A five-year study // Computers in Human Behavior. Feb. 2021. Vol. 115. Art. 106595.
4. *Mussida P., Lanzi P. L.* A computational tool for engineer dropout prediction // 2022 IEEE Global Engineering Education Conference (EDUCON). 2022, March. P. 1571–1576. DOI: <https://doi.org/10.1109/EDUCON52537.2022.9766632>.
5. *Knowles J. E.* Of Needles and Haystacks: Building an Accurate Statewide Dropout Early Warning System in Wisconsin // Journal of Educational Data Mining. 2015. Vol. 7, Iss. 3. P. 18–67.
6. Early Detection of Students at Risk – Predicting Student Dropouts Using Administrative Student Data and Machine Learning Methods / K. Schneider, J. Berens, S. Oster, J. Burghoff // Beiträge zur Jahrestagung des Vereins für Socialpolitik 2018: Digitale Wirtschaft. 2018. Session: Education II. No. D20-V1. ZBW – Leibniz-Informationszentrum Wirtschaft, Kiel, Hamburg.
7. *Schapire E., Freund Y.* A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting // Journal of Computer and System Science. 1997. Vol. 55. P. 119–139.
8. *Mussida P., Lanzi P. L.* A computational tool for engineer dropout prediction // IEEE Global Engineering Education Conference, EDUCON. 2022, March. P. 1571–1576.
9. *Sclater N., Peasgood A., Mullan J.* Learning Analytics in Higher Education. A review of UK and international practice. Full report. Published under the CC BY 4.0 license creativecommons.org/licenses/by/4.0/. 41 p.

10. A proposal of Machine Learning model to improve learning process and reduce dropout rate at technical training institutes / L.A.V. Apaza, J.A.R. Huamani, J.O.A. Bernedo, A.G.Z. Chauca // 16th Iberian Conference on Information Systems and Technologies. CISTI 2021, Chaves, 23–26 June 2021. Art. 170402.
11. *Kabathova J., Drlik M.* Towards Predicting Student's Dropout in University Courses Using Different Machine Learning Techniques // *Appl. Sci.* 2021. Vol. 11. 3130.
12. Цифровая образовательная история как составляющая цифрового профиля обучающегося в условиях трансформации образования / Р. В. Есин, Т. В. Зыкова, Т. А. Кустицкая, А. А. Кытманов // *Перспективы науки и образования.* 2022. № 5 (59). С. 566–584.
13. Model for Predicting Academic Performance Through Artificial Intelligence / J. Silva, L. Romero, D. Solano [et al.] // Singh V., Asari V., Kumar S., Patel R. (eds.) *Computational Methods and Data Engineering, Advances in Intelligent Systems and Computing.* 2021. Vol. 1227. Springer, Singapore.
14. Predicting Student Academic Performance by Means of Associative Classification / L. Cagliero, L. Canale, L. Farinetti [et al.] // *Appl. Sci.* 2021. Vol. 11. 1420.
15. Student Dataset from Tecnologico de Monterrey in Mexico to Predict Dropout in Higher Education / J. Alvarado-Uribe, P. Mejía-Almada, A. L. Masetto Herrera [et al.] // *Data.* 2022. Vol. 7, Iss. 9. 119. DOI: <https://doi.org/10.3390/data7090119>.
16. *Radovanović S., Delibašić B., Suknović M.* Predicting Dropout in Online Learning Environments // *Computer Science and Information Systems.* 2021. Vol. 18, No. 3. P. 957–978.
17. *Singh H. P., Alhulail H. N.* Predicting Student-Teachers Dropout Risk and Early Identification: A Four-Step Logistic Regression Approach // *IEEE Access.* 2022. Vol. 10. P. 6470–6482.
18. *Кустицкая Т. А., Носков М. В.* Развитие учебной аналитики в России // Информатизация образования и методика электронного обучения: цифровые технологии в образовании: материалы V Междунар. науч. конф.: в 2 ч. Красноярск, 21–24 сентября 2021. Ч. 1. С. 273–278.
19. *Riestra-González M., Paule-Ruiz M. del P., Ortin F.* Massive LMS log data analysis for the early prediction of course-agnostic student performance // *Computers & Education.* 2020. Vol. 163, Iss. 1. Art. 104108. DOI: <https://doi.org/10.1016/j.compedu.2020.104108>.
20. *Kustitskaya T. A., Kytmanov A. A., Noskov M. V.* Early student-at-risk detection by current learning performance and learning behavior indicators // *Cybernetics and Information Technologies.* 2022. Vol. 22, No. 1. P. 117–133.
21. Model for Predicting Academic Performance Through Artificial Intelligence / J. Silva, L. Romero, D. Solano [et al.] // Singh V., Asari V., Kumar S., Patel R. (eds.) *Computational Methods and Data Engineering, Advances in Intelligent Systems and Computing Series, Vol. 1227.* Springer, Singapore, 2021. P. 519–526.
22. *Носков М. В., Сомова М. В., Федотова И. М.* Управление успешностью обучения студента на основе марковской модели // *Информатика и образование.* 2018. № 10 (299). С. 4–11.
23. Прогностическая модель оценки успешности предметного обучения в условиях цифровизации образования / М. В. Носков, Ю. В. Вайнштейн, М. В. Сомова, И. М. Федотова // *Вестн. Рос. Ун-та дружбы народов. Сер.: Информатизация образования.* 2023. Т. 20, № 1. С. 7–19.
24. Data Mining in Educational Systems for Effective Student Mentoring / M. Mehra, D. R. Kalbande, S. Mankar, S. Mutsaddi // *International Conference on Advances in Computing, Communication and Control (ICAC3).* Mumbai, India, 2019. P. 1–5.
25. *Токтарова В. И., Попова О. Г.* Анализ образовательных данных взаимосвязи успешности обучения и поведения студентов в цифровой образовательной среде вуза // *Информатика и образование.* 2022. Т. 37, № 4. С. 54–63.
26. *Joseph-Richard P., Jaffrey A., Uhomoihi J.* Predictive Learning Analytics and the Creation of Emotionally Adaptive Learning Environments in Higher Education Institutions: A Study of Students' Affect Responses // *International Journal of Information and Learning Technology.* 2021. Vol. 38, Iss. 2. P. 243–257.
27. *Tempelaar D., Rienties B., Nguyen Q.* Dispositional Learning Analytics for supporting individualized learning // *Frontiers in Education.* 2021. Vol. 6. Art. 703773.

REFERENCES

1. Tsibulskiy G. M., Noskov M. V., Baryshev R. A., Somova M. V. Aktivnaya informatsionnaya sistema vuza v informatsionno-obrazovatelnoy srede. *Pedagogika*. 2017, No. 3, pp. 28–32.
2. Renò V., Stella E., Patrino C. et al. Learning Analytics: Analysis of Methods for Online Assessment. *Appl. Sci.* 2022, No. 12, 9296.
3. Bravo-Agapito J., Romero S. J., Pamplona S. Early prediction of undergraduate Student's academic performance in completely online learning: A five-year study. *Computers in Human Behavior*. Feb. 2021, Vol. 115, Art. 106595.
4. Mussida P., Lanzi P. L. A computational tool for engineer dropout prediction. In: 2022 IEEE Global Engineering Education Conference (EDUCON). 2022, March, pp. 1571–1576. DOI: <https://doi.org/10.1109/EDUCON52537.2022.9766632>.
5. Knowles J. E. Of Needles and Haystacks: Building an Accurate Statewide Dropout Early Warning System in Wisconsin. *Journal of Educational Data Mining*. 2015, Vol. 7, Iss. 3, pp. 18–67.
6. Schneider K., Berens J., Oster S., Burghoff J. Early Detection of Students at Risk – Predicting Student Dropouts Using Administrative Student Data and Machine Learning Methods. In: Beiträge zur Jahrestagung des Vereins für Socialpolitik 2018: Digitale Wirtschaft. 2018. Session: Education II. No. D20-V1. ZBW – Leibniz-Informationszentrum Wirtschaft, Kiel, Hamburg.
7. Schapire E., Freund Y. A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. *Journal of Computer and System Science*. 1997, Vol. 55, pp. 119–139.
8. Mussida P., Lanzi P. L. A computational tool for engineer dropout prediction. In: IEEE Global Engineering Education Conference, EDUCON. 2022, March. Pp. 1571–1576.
9. Sclater N., Peasgood A., Mullan J. *Learning Analytics in Higher Education. A review of UK and international practice. Full report*. Published under the CC BY 4.0 license creativecommons.org/licenses/by/4.0/. 41 p.
10. Apaza L.A.V., Huamani J.A.R., Bernedo J.O.A., Chauca A.G.Z. A proposal of Machine Learning model to improve learning process and reduce dropout rate at technical training institutes. In: 16th Iberian Conference on Information Systems and Technologies. CISTI 2021, Chaves, 23–26 June 2021. Art. 170402.
11. Kabathova J., Drlik M. Towards Predicting Student's Dropout in University Courses Using Different Machine Learning Techniques. *Appl. Sci.* 2021. Vol. 11. 3130.
12. Esin R. V., Zykova T. V., Kustitskaya T. A., Kytmanov A. A. Tsfirovaya obrazovatel'naya istoriya kak sostavlyayushchaya tsifrovogo profilya obuchayushchegosya v usloviyakh transformatsii obrazovaniya. *Perspektivy nauki i obrazovaniya*. 2022, No. 5 (59), pp. 566–584.
13. Silva J., Romero L., Solano D. et al. Model for Predicting Academic Performance Through Artificial Intelligence. In: Singh V., Asari V., Kumar S., Patel R. (eds.) *Computational Methods and Data Engineering. Advances in Intelligent Systems and Computing*. 2021. Vol. 1227. Springer, Singapore.
14. Cagliero L., Canale L., Farinetti L. et al. Predicting Student Academic Performance by Means of Associative Classification. *Appl. Sci.* 2021. Vol. 11. 1420.
15. Alvarado-Urbe J., Mejía-Almada P., Masetto Herrera A.L. [et al.] Student Dataset from Tecnológico de Monterrey in Mexico to Predict Dropout in Higher Education. *Data*. 2022. Vol. 7, Iss. 9. 119. DOI: <https://doi.org/10.3390/data7090119>.
16. Radovanović S., Delibašić B., Suknović M. Predicting Dropout in Online Learning Environments. *Computer Science and Information Systems*. 2021, Vol. 18, No. 3, pp. 957–978.
17. Singh H. P., Alhulail H. N. Predicting Student-Teachers Dropout Risk and Early Identification: A Four-Step Logistic Regression Approach. *IEEE Access*. 2022, Vol. 10, pp. 6470–6482.

18. Kustitskaya T. A., Noskov M. V. Razvitie uchebnoy analitiki v Rossii. In: Informatizatsiya obrazovaniya i metodika elektronnoogo obucheniya: tsifrovye tekhnologii v obrazovanii. *Proceedings of the V International scientific conference*. In 2 vols. Krasnoyarsk, 21–24 Sept. 2021. Vol. 1, pp. 273–278.
19. Riestra-González M., Paule-Ruíz M. del P., Ortin F. Massive LMS log data analysis for the early prediction of course-agnostic student performance. *Computers & Education*. 2020, Vol. 163, Iss. 1. Art. 104108. DOI: <https://doi.org/10.1016/j.compedu.2020.104108>.
20. Kustitskaya T. A., Kytmanov A. A., Noskov M. V. Early student-at-risk detection by current learning performance and learning behavior indicators. *Cybernetics and Information Technologies*. 2022, Vol. 22, No. 1, pp. 117–133.
21. Silva J., Romero L., Solano D. et al. Model for Predicting Academic Performance Through Artificial Intelligence. In: Singh V., Asari V., Kumar S., Patel R. (eds.) *Computational Methods and Data Engineering. Advances in Intelligent Systems and Computing Series*, Vol. 1227. Springer, Singapore, 2021. Pp. 519–526.
22. Noskov M. V., Somova M. V., Fedotova I. M. Upravlenie uspešnostyu obucheniya studenta na osnove markovskoy modeli. *Informatika i obrazovanie*. 2018, No. 10 (299), pp. 4–11.
23. Noskov M. V., Vaynshteyn Yu. V., Somova M. V., Fedotova I. M. Prognosticheskaya model otsenki uspešnosti predmetnogo obucheniya v usloviyakh tsifrovizatsii obrazovaniya. *Vestn. Ros. Un-ta družby narodov. Ser.: Informatizatsiya obrazovaniya*. 2023, Vol. 20, No. 1, pp. 7–19.
24. Mehra M., Kalbande D. R., Mankar S., Mutsaddi S. Data Mining in Educational Systems for Effective Student Mentoring. In: International Conference on Advances in Computing, Communication and Control (ICAC3). Mumbai, India, 2019. Pp. 1–5.
25. Toktarova V. I., Popova O. G. Analiz obrazovatelnykh dannykh vzaimosvyazi uspešnosti obucheniya i povedeniya studentov v tsifrovoy obrazovatelnoy srede vuza. *Informatika i obrazovanie*. 2022, Vol. 37, No. 4, pp. 54–63.
26. Joseph-Richard P., Jaffrey A., Uhomoihi J. Predictive Learning Analytics and the Creation of Emotionally Adaptive Learning Environments in Higher Education Institutions: A Study of Students' Affect Responses. *International Journal of Information and Learning Technology*. 2021, Vol. 38, Iss. 2, pp. 243–257.
27. Tempelaar D., Rienties B., Nguyen Q. Dispositional Learning Analytics for supporting individualized learning. *Frontiers in Education*. 2021, Vol. 6, Art. 703773.

Кустицкая Татьяна Алексеевна, кандидат физико-математических наук, доцент кафедры прикладной математики и анализа данных, Сибирский федеральный университет
e-mail: tkustitskaya@sfu-kras.ru

Kustitskaya Tatiana A., PhD in Physics and Mathematics, Assistant Professor, Applied Mathematics and Data Analysis Department, Siberian Federal University
e-mail: tkustitskaya@sfu-kras.ru

Носков Михаил Валерианович, доктор физико-математических наук, профессор, профессор кафедры прикладной математики и анализа данных, Сибирский федеральный университет
e-mail: mnoskov@sfu-kras.ru

Noskov Mikhail V., ScD in Physics and Mathematics, Professor, Full Professor, Professor, Applied Mathematics and Data Analysis Department, Siberian Federal University
e-mail: mnoskov@sfu-kras.ru

Вайнштейн Юлия Владимировна, доктор педагогических наук, доцент, профессор кафедры прикладной математики и анализа данных, Сибирский федеральный университет
e-mail: yweinstein@sfu-kras.ru

Vainshtein Yulia V., ScD in Education, Associate Professor, Professor, Applied Mathematics and Data Analysis Department, Siberian Federal University
e-mail: yweinstein@sfu-kras.ru

Статья поступила в редакцию 04.05.2023
The article was received on 04.05.2023