

<https://doi.org/10.25559/SITITO.020.202402.371-378>
УДК 378+004.6

Оригинальная статья

Применение данных цифровой образовательной истории для прогнозирования риска неуспешности студента

Т. А. Кустицкая^{1*}, Р. В. Есин¹, А. А. Кытманов², Т. В. Зыкова¹

¹ ФГАОУ ВО «Сибирский федеральный университет», г. Красноярск, Российская Федерация

Адрес: 660041, Российская Федерация, г. Красноярск, пр. Свободный, д. 79

² ФГБОУ ВО «МИРЭА – Российский технологический университет», г. Москва, Российская федерация

Адрес: 119452, Российская федерация, г. Москва, пр. Вернадского, д. 78

* tkustitskaya@sfu-kras.ru

Аннотация

В работе рассмотрено несколько постановок задачи прогнозирования успешности обучения на основе образовательных данных. Предложен подход к решению этих задач с использованием цифровой образовательной истории обучающихся - многомерных структурированных динамически изменяющихся данных об их образовательной деятельности и текущих результатах обучения. Сформулированы принципы проектирования базы образовательных данных, позволяющие формировать цифровую образовательную историю студента таким образом, чтобы максимально детально охарактеризовать его образовательный путь, как с точки зрения успеваемости, так и с точки зрения динамики образовательной активности и изменения траектории обучения. В ходе эмпирического исследования на основе образовательных данных студентов Сибирского федерального университета решались два типа задач прогнозирования успешности обучения для студентов: задача прогнозирования риска появления академических задолженностей и задача прогнозирования риска отчисления. Для этих типов задач была исследована предиктивная значимость показателей цифровой образовательной истории и эффект от использования этих показателей в предиктивных моделях с точки зрения качества прогнозирования. Подтверждено, что включение показателей цифровой образовательной истории в состав предикторов моделей позволяет значительно повысить точность прогнозирования отчислений. Выявлены наборы наиболее значимых предикторов таких прогнозных моделей. В состав этих наборов, в частности, вошли данные о предыдущих отчислениях студентов из вуза и причинах этих отчислений, академических отпусках и переводах, о динамике изменения оценок студентов в ходе пересдач. Включение данных цифровой образовательной истории в прогнозные, аналитические или описательные модели может быть полезно и для решения других задач учебной аналитики.

Ключевые слова: учебная аналитика, база образовательных данных, цифровая образовательная история, прогнозирование успешности обучения

Финансирование: Исследование выполнено за счет гранта Российского научного фонда (проект № 22-28-00413 «Цифровая образовательная история и ее применение в системах поддержки успешности обучения»).

Конфликт интересов: авторы заявляют об отсутствии конфликта интересов.

Для цитирования: Применение данных цифровой образовательной истории для прогнозирования риска неуспешности студента / Т. А. Кустицкая [и др.] // Современные информационные технологии и ИТ-образование. 2024. Т. 20, № 2. С. 371-378. <https://doi.org/10.25559/SITITO.020.202402.371-378>

© Кустицкая Т. А., Есин Р. В., Кытманов А. А., Зыкова Т. В., 2024



Контент доступен под лицензией Creative Commons Attribution 4.0 License.
The content is available under Creative Commons Attribution 4.0 License.



Application of Digital Educational History Data to Predict the Risk of Student Failure

T. A. Kustitskaya^{a*}, R. V. Esin^a, A. A. Kytmanov^b, T. V. Zyкова^a

^a Siberian Federal University, Krasnoyarsk, Russian Federation

Address: 79 Svobodny Pr., Krasnoyarsk 660041, Russian Federation

^b MIREA – Russian Technological University, Moscow, Russian Federation

Address: 78 Vernadsky Ave., Moscow 119454, Russian Federation

* tkustitskaya@sfu-kras.ru

Abstract

The paper considers several variants of tasks of predicting learning success based on education data. An approach to solving these problems using digital educational history of students – multidimensional structured dynamically changing data on their educational activities and current learning outcomes – is proposed. Principles for designing an education database are formulated, allowing to form a digital educational history of the student in such a way that characterizes his educational path in maximum detail in terms of academic performance, as well as in terms of the dynamics of educational activity and changes in the trajectory of learning. In an empirical study based on educational data of students at the Siberian Federal University we solved two types of tasks: predicting the risk of academic debts and predicting student dropouts. For these types of tasks, the feature importances for digital educational history indicators, as well as the effect of using these indicators in predictive models on performance metrics were investigated. It was confirmed that including digital educational history in predictive model significantly increases the quality of predicting dropouts. We identified sets of the most important features for such predictive models. These sets included data on previous dropouts of students from the university and the reasons for these dropouts, academic leaves and transfers, and the dynamics of changes in students' grades during retakes.

Keywords: learning analytics, education database, digital educational history, student success prediction

Funding: The study was carried out at the expense of a grant from the Russian Scientific Foundation (project 22-28-00413 “Digital Educational History and Its Application in the Systems of Support of Learning Success”).

Conflict of interests: The authors declares no conflict of interest.

For citation: Kustitskaya T.A., Esin R.V., Kytmanov A.A., Zyкова T.V. Application of Digital Educational History Data to Predict the Risk of Student Failure. *Modern Information Technologies and IT-Education*. 2024;20(2):371-378. <https://doi.org/10.25559/SITITO.020.202402.371-378>



Введение

В настоящее время наиболее очевидным и распространенным направлением использования образовательных данных является мониторинг учебного процесса и управление административными процессами внутри университета. Однако все большую популярность приобретают различные инструменты учебной аналитики, направленные на оптимизацию учебного процесса на основе анализа данных о студентах и процессе обучения [1, 2].

Как вузы, так и образовательные онлайн платформы используют панели учебной аналитики, чтобы предоставить учащимся краткую и наглядно представленную информацию о прогрессе и успехе их обучения, а преподавателям — отчеты о процессе обучения. Наиболее известным и исторически одним из первых примеров такого использования является система Course Signal [3], которая визуализирует прогнозируемые результаты обучения студента в виде светофора. Прогноз основан на оценках по учебной дисциплине, времени, затраченному учащимся на задание, и общих показателях успеваемости в прошлом. Используемая несколькими вузами США система My Learning Analytics [4] состоит из нескольких информационных панелей, предназначенных для помощи учащимся в планировании своего обучения в рамках учебных курсов и оценке своей успеваемости по сравнению с одноклассниками.

Для повышения эффективности инструменты мониторинга часто сопровождаются рекомендациями и методической помощью [5, 6]. Кроме того, рекомендательные системы разрабатываются не только для поддержки учебного процесса по той или иной дисциплине, но и для решения других, более специфических задач. Например, авторы в [7] предлагают рекомендательную систему для выбора места стажировки.

Системы раннего предупреждения, системы взаимодействия со студентами и системы поддержки обучения, включая мониторинг, рекомендательные сервисы и меры поддержки студентов из групп риска, и другие инструменты набирают популярность благодаря более широкому использованию вузами онлайн-обучения, которое, однако, приводит к уменьшению личного контакта и поддержки со стороны преподавателей. Такие системы могут быть реализованы на различных уровнях: например, в исследованиях [6], [8, 9] решения по образовательной аналитике выполняются на уровне отдельных курсов, тогда как в работах¹ [10-13] можно найти описание общеуниверситетских систем поддержки обучения.

Во всех этих решениях образовательные данные являются важным компонентом соответствующих систем. Полнота, непротиворечивость и разнонаправленность этих данных во многом определяют эффективность инструментов, построенных на их основе. В работе [14] нами были предложены принципы проектирования университетской базы образовательных данных, основанные на учете предшествующей образовательной истории обучающегося, а также описана архитектура такой базы. Мы полагаем, что за счет применения этих принципов при построения базы образовательных данных можно

повысить эффективность инструментов учебной аналитики, разработанных на ее основе.

Целью данной работы является изучение влияния некоторых из предложенных в [14] принципов проектирования базы образовательных данных на качество прогнозирования риска неуспешности студента вуза.

Цифровая образовательная история и ее учет при проектировании базы образовательных данных

В [15: 575] было введено понятие *цифровой образовательной истории* обучающегося, которая представляет собой «многомерные структурированные динамически изменяющиеся данные о его образовательной деятельности и текущих результатах обучения». К цифровой образовательной истории студента могут быть отнесены различные данные об академических достижениях студента до поступления в вуз (оценки аттестата о среднем образовании, результаты конкурсов и олимпиад, оценки по вступительным экзаменам и баллы ЕГЭ), итоговые оценки студента в предыдущих семестрах и информация о пересдачах, текущая успеваемость и данные об образовательной активности в процессе обучения. Концепция цифровой образовательной истории лежит в основе следующих принципов проектирования базы образовательных данных университета: принципа студентоориентированности и принципа непрерывности данных.

Студент является ключевым участником образовательного процесса – любая образовательная программа разрабатывается и любая образовательная деятельность проводится для студента, большинство инструментов анализа обучения направлены непосредственно или косвенно на повышение успеваемости студента. Поэтому, чтобы отразить центральную роль студента в образовательной базе данных, мы предлагаем следовать принципу *студентоориентированности*, который, в частности, предполагает использование идентификатора студента в качестве ключевого поля, связывающего данные студента со всеми остальными таблицами базы данных и хранение всех данных об образовательной деятельности студента, собранных информационными системами вуза.

Принцип *непрерывности* данных предполагает сохранение всех динамически изменяющихся данных о студентах с соответствующими временными метками. Он гарантирует, что никакие данные не будут перезаписаны и, следовательно, не останутся недоступными для целей анализа образовательной истории. Сохранение промежуточных значений характеристик, связанных с обучением студентов, является чрезвычайно важным, так как они могут быть важными свидетельствами мотивации, применяемой студентом в тот или иной момент времени стратегии обучения и ее успешности.

Например, если вместе с итоговыми оценками по курсу мы регистрируем информацию о датах и результатах пересдач экзаменов и зачетов, мы получаем информацию о трудностях в обучении, которые испытывал студент, и о динамике улучшения его результатов.

¹ Sclater N., Peasgood A., Mullan J. Learning Analytics in Higher Education: A review of UK and international practice. Jisc, 2016. 39 p. [Электронный ресурс]. URL: <https://www.jisc.ac.uk/sites/default/files/learning-analytics-in-he-v3.pdf> (дата обращения: 11.02.2024).



Прогнозирование риска неуспешности обучающихся

Задача прогнозирования успешности (неуспешности) обучения является одной из важных задач для администрации вузов (на уровне вуза в целом, института или факультета) как с точки зрения повышения сохранности контингента обучающихся, так и с точки зрения повышения конкурентноспособности вуза. Постановка задачи может варьироваться в зависимости от управленческой команды и может быть сформулирована как прогнозирование удовлетворенности студента обучением, вычисление вероятности успешного завершения студентом некоторого учебного курса, вероятности отчисления студента в текущем семестре или вероятность успешного завершения обучения студента по образовательной программе [16, 17].

В рамках данной работы будем говорить только о двух возможных постановках задачи - о задаче прогнозирования возникновения академических задолженностей и о задаче прогнозировании отчисления студента.

Первая задача (*задача 1*) является достаточно специфичной для отечественного образования, так как в отличие от систем образования многих других стран, система российского высшего образования допускает ситуацию, когда студент может в течение года ликвидировать задолженности по дисциплинам, не будучи отчисленным. Поэтому задача прогнозирования появления задолженностей в текущую сессию, во-первых, приобретает самостоятельную значимость, а во-вторых, может заметно отличаться от задачи прогнозирования успешности обучения по отдельным дисциплинам, а также от задачи прогнозирования отчисления с образовательной программы прежде всего по набору значимых предикторов неуспешности.

Что касается задачи прогнозирования отчисления студентов (*задача 2*), то в литературе описано значительное количество исследований по построению определенных моделей для прогнозирования отсева первокурсников [18, 19], что обусловлено высокой востребованностью таких моделей, поскольку отчисления студентов происходит в основном в начале обучения. Исследования общих закономерностей отчисления не так широко распространены. Тем не менее, авторы исследований [20, 21] отмечают, что универсальная модель для прогнозирования отсева студентов на протяжении всей учебы может быть полезна для определения основных причин отчисления и неудовлетворенности обучением с течением времени. При решении задачи прогнозирования отсева студентов методами машинного обучения большинство рассматриваемых признаков являются статичными и не отражают динамику активности студента. Обычно в качестве предикторов используются такие данные, как успеваемость и социально-демографические данные [22-25].

Основываясь на существующих исследованиях по прогнозированию отсева учащихся, была выдвинута гипотеза о том, что добавление признаков, отражающих динамику успеваемости учащихся во времени, а также данных о перемещении

студентов (перевод, академический отпуск, отчисление и восстановление на образовательную программу), будет способствовать значительному улучшению качества моделей прогнозирования, используемых для решения *задачи 1* и *задачи 2*.

Для проверки гипотезы мы провели эмпирическое исследование по решению задач 1 и 2 для студентов Сибирского федерального университета (СФУ). Данные о 46535 студентах были извлечены из баз данных информационной системы СФУ (использовались данные за 5 предшествующих лет). Информация об учебном профиле студентов (учебная группа и институт, направление подготовки, форма обучения), итоговых оценках по дисциплинам учебного плана, приказах об изменении статуса и перемещениях студентов были получены из информационных систем вуза².

Для решения обеих задач было решено использовать одинаковый набор моделей прогнозирования (бустинговые классификаторы Catboost и Xgboost), а также одинаковые обучающие и тестовые наборы данных с тем, чтобы изолированно оценить вклад показателей цифровой образовательной истории в прогнозирование.

Первоначальный набор предикторов (без использования показателей образовательной истории) сформирован на основе характеристик студентов, которые обычно собираются в вузах и используются для прогнозирования успеваемости и отсева обучающихся:

- демографические и социальные характеристики: возраст, гражданство, тип финансирования обучения;
 - общая характеристика учебного процесса: форма обучения, факультет, год и семестр обучения, уровень подготовки, специальность, обучение по индивидуальному учебному плану;
 - показатели успеваемости за 3 предыдущих семестра: средние баллы зачетной книжки, количество экзаменов и зачетов.
- Расширенный набор предикторов (с использованием данных образовательной истории) дополняет первоначальный набор следующими характеристиками:
- история движения контингента: количество академических отпусков и успешных восстановлений после них, число переводов и отчислений, и их причины;
 - история оценок: средние оценки до и после периода пересдач (индивидуальных и комиссионных) за 3 предшествующих семестра, количество пересдач.

Рассмотрим более подробно результаты решения задач 1 и 2 на первоначальном и расширенном наборах предикторов.

Задача 1. В этой задаче откликом моделей (то есть прогнозируемой переменной) является появление академических задолженностей в экзаменационную сессию (0 – долгов нет, 1 – не сдан хотя бы один экзамен или зачет по дисциплинам семестра).

Модели Catboost и Xgboost были обучены и протестированы на нескольких разбиениях данных на обучающие и тестовые выборки. Усредненные по этим выборкам показатели качества прогнозирования представлены в табл. 1.

² Институт космических и информационных технологий СФУ [Электронный ресурс]. URL: <https://dec.sfu-kras.ru> (дата обращения: 11.02.2024); Siberian Federal University "1C: Document Management" Electronic Service [Электронный ресурс]. URL: <https://d.sfu-kras.ru> (дата обращения: 11.02.2024).



Таблица 1. Значения метрик качества классификации при решении задачи прогнозирования задолженностей по дисциплинам текущего семестра в сессию

Table 1. Values of classification quality metrics when solving the problem of forecasting debts in disciplines of the current semester in the session

метрики	Классификатор CatBoost		Классификатор Xgboost	
	без использования образовательной истории	с использованием образовательной историй	без использования образовательной истории	с использованием образовательной истории
recall	0.737	0.751	0,732	0.743
precision	0.769	0.795	0,782	0.788
F1-score	0.752	0.773	0,756	0.765
ROC-AUC	0.887	0.900	0.888	0.897

Источник: здесь и далее в статье все рисунки и таблицы составлены авторами.

Source: Hereinafter in the article, all tables and figures are compiled by the authors.

Качество прогнозирования для обеих моделей на расширенном наборе предикторов оказалось несколько выше (на 1-3 % в зависимости от метрики качества), однако в целом точность прогноза не очень велика и ее небольшое улучшение за счет данных предыдущей образовательной истории не кажется значимым. Отметим, однако, что такие характеристики как *общее количество отчислений студента* в предыдущие периоды обучения и *количество отчислений по причине невыполнения учебного плана*, а также *доля дисциплин семестра, по которым у студента возникли задолженности в предыдущую сессию*, вошли в список семи наиболее значимыми предикторов для модели Xgboost согласно метрике *feature importance*.

Задача 2. Для данной задачи откликом модели выступает отчисление студента в текущем семестре (0 – студент продолжает обучение, 1 – студент отчислен). Процедура обучения и валидации моделей происходила аналогично задаче 1. Качество моделей, построенных на основе стандартного набора предикторов, оказалось неудовлетворительным (см. табл. 2), особенно с точки зрения выявления учащихся из группы риска (метрика *recall* была равна 0.615 и 0.595, соответственно). Классификаторы, переобученные на расширенном наборе данных, показали гораздо лучшее качество – значение всех метрик качества довольно сильно выросло. Самый значительный рост показала метрика *recall* (до 0,958 и 0,950 соответственно).

Таблица 2. Значения метрик качества классификации при решении задачи прогнозирования отчисления студента в текущем семестре

Table 2. Values of classification quality metrics when solving the problem of predicting student expulsion in the current semester

метрики	Классификатор CatBoost		Классификатор Xgboost	
	без использования образовательной истории	с использованием образовательной историй	без использования образовательной истории	с использованием образовательной истории
recall	0.615	0.958	0.593	0.950
precision	0.737	0.913	0.723	0.908
F1-score	0.671	0.935	0.652	0.928
ROC-AUC	0.967	0.999	0.962	0.999

Добавленные предикторы обладают высокой прогностической силой. На рис. 1 показаны десять наиболее важных предикторы (с точки зрения метрики *feature importance*) как для модели Catboost, так и для Xgboost. Темно-зеленые столбцы диаграммы соответствуют переменным, включенным как в исходную, так и в расширенную модели, тогда как светло-зеленые столбцы – только в расширенные версии. Заметим, что у модели Xgboost десять наиболее значимых предикторов – это показатели образовательной истории.

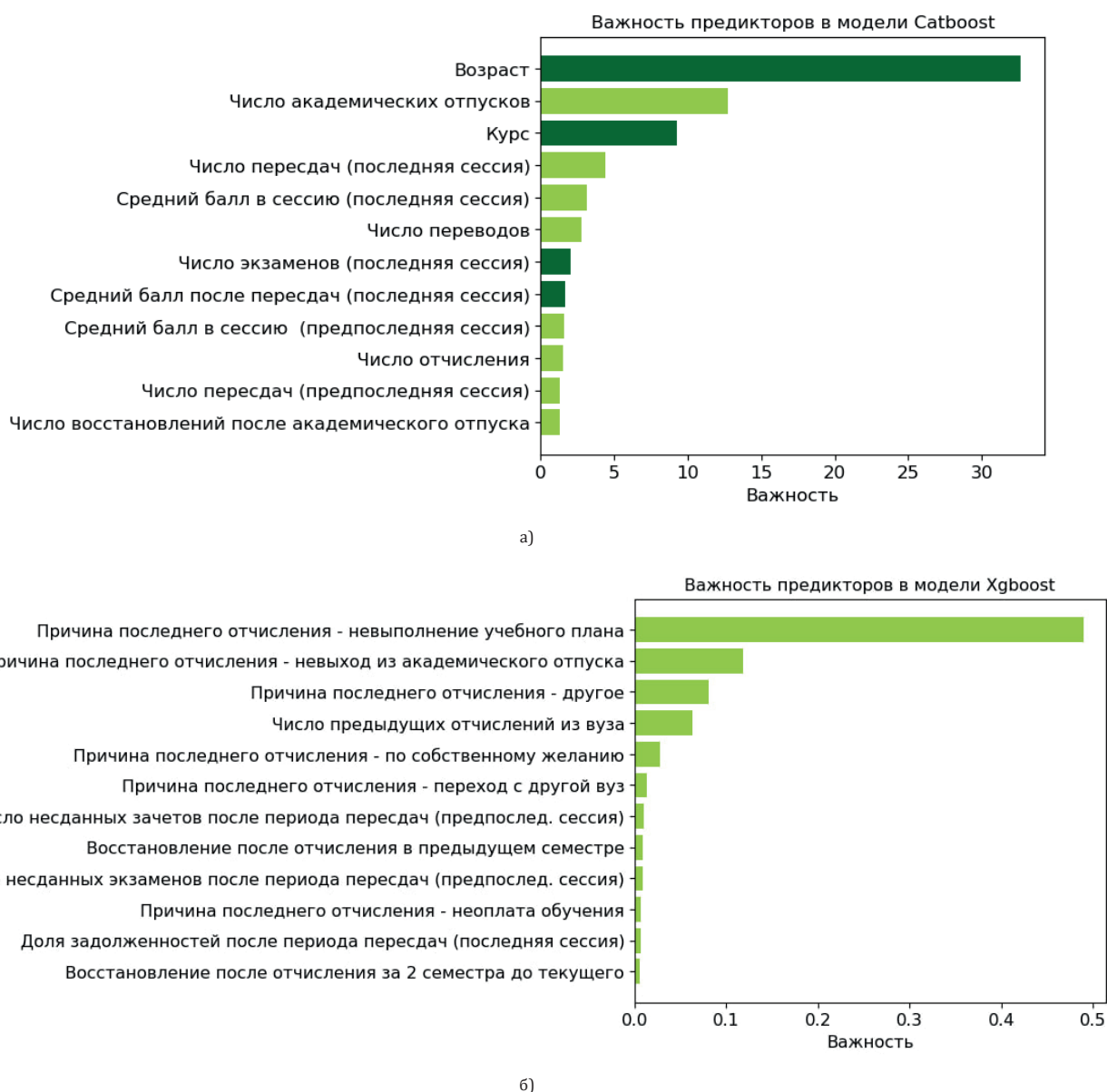
Новые предикторы включают информацию о факторах риска отчисления (поскольку часто перемещения (академический отпуск, перевод на другую специальность либо на индивидуальный план, отчисления), а также пересдачи экзаменов и зачетов указывают на трудности с учебой), поэтому расширенные модели лучше выявляют учащихся из групп риска.

Отметим, что включение данных цифровой образовательной истории в модели прогнозирования значительно улучшило

точность моделей в задаче прогнозирования отчислений, но при этом совсем незначительно улучшило точность в задаче прогнозирования академических задолженностей. Скорее всего, это связано с тем, что признаки появления задолженностей в текущем семестре нужно искать не в истории обучения в предыдущие периоды, а в текущем образовательном поведении студента (его текущей успеваемости по дисциплинам данного семестра, активности в электронной среде обучения и т.п.). В будущих исследованиях мы планируем совместить описанные прогнозные модели, работающие на уровне образовательной программы, с моделями прогнозирования успешности обучения на уровне отдельных дисциплин [8], построенными по показателям текущих показателей успешности обучения в семестре. Источником данных о текущих показателях активности и успеваемости студентов станет, прежде всего, система электронного обучения «Е-Курсы»³.

³ Система электронного обучения СФУ [Электронный ресурс]. URL: <https://e.sfu-kras.ru> (дата обращения: 11.02.2024).





Р и с. 1. Наиболее важные предикторы а) для модели Catboost, б) для модели Xgboost
 Fig. 1. The most important predictors а) for the Catboost model, б) for the Xgboost model

Представленные в работе прогнозные модели были разработаны для системы поддержки успешности обучения СФУ, которая в настоящее время находится на стадии опытной эксплуатации [25]. Следует отметить, что в ходе работы над реализацией принципа непрерывности для данных, используемых прогнозными моделями, мы столкнулись со следующей проблемой – используемая СФУ архитектура базы данных не позволяет хранить временные показатели всех необходимых данных о студентах (например, информация о том, что студент был отчислен, не сохраняется в базе данных об учающихся, если затем этот студент был восстановлен, либо поступил в вуз заново). Проблема была решена ретроспективно путем восстановления исторических данных из резервных копий

баз образовательных данных и университетской базы приказов. Переход к предложенной нами в [14] архитектуре базы образовательных данных, разработанной с учетом принципа непрерывности, потребует переработки всех внутривузовских систем сбора и хранения данных.

Заключение

Задача прогнозирования риска неуспешности обучения наряду с многими другими задачами учебной аналитики может эффективно решаться в вузе только если используемые для их решения данные всесторонне характеризуют образовательный путь студента. Предложенный в работе подход к решению



задачи прогнозирования с использованием данных цифровой образовательной истории позволяет учесть не только текущую успеваемость и образовательное поведение обучающегося, но и информацию о динамике его обучения на предыдущих этапах, об испытанных им ранее затруднениях.

На примере прогнозирования риска отчисления для студентов

Сибирского федерального университета было показано, что включение в состав предикторов моделей показателей цифровой образовательной истории значительно повышает точность прогнозирования, что позволит в дальнейшем эффективно использовать эти модели в университетской системе поддержки успешности обучения.

References

- [1] Kew S.N., Tasir Z. Learning Analytics in Online Learning Environment: A Systematic Review on the Focuses and the Types of Student-Related Analytics Data. *Technology, Knowledge and Learning*. 2022;27:405-427. <https://doi.org/10.1007/s10758-021-09541-2>
- [2] Wong B.T., Li K.C. A review of learning analytics intervention in higher education (2011-2018). *Journal of Computers in Education*. 2020;7(1):7-28. <https://doi.org/10.1007/s40692-019-00143-7>
- [3] Arnold K.E., Pistilli M.D. Course Signals at Purdue: Using learning analytics to increase student success. In: Proceedings of the 2nd International Conference on Learning Analytics and Knowledge (LAK'12), Vancouver, BC, Canada. New York, NY, USA: Association for Computing Machinery; 2012. p. 267-270. <https://doi.org/10.1145/2330601.2330666>
- [4] Teasley S.D., Kay M., Elkins S., Hammond J. User-Centered Design for a Student-Facing Dashboard Grounded in Learning Theory. In: Sahin M., Ifenthaler D. (eds) Visualizations and Dashboards for Learning Analytics. *Advances in Analytics for Learning and Teaching*. Cham: Springer; 2021. p. 191-212. https://doi.org/10.1007/978-3-030-81222-5_9
- [5] Gonçalves A.L., Carlos L.M., da Silva J.B., Alves G.R. Personalized Student Assessment based on Learning Analytics and Recommender Systems. In: Proceedings of the 3rd International Conference of the Portuguese Society for Engineering Education (CISPEE). Aveiro, Portugal: IEEE Press; 2018. p. 1-7. <https://doi.org/10.1109/CISPEE.2018.8593493>
- [6] Yilmaz F.G.K., Yilmaz R. Learning analytics as a metacognitive tool to influence learner transactional distance and motivation in online learning environments. *Innovations in Education and Teaching International*. 2020;58(5):575-585. <https://doi.org/10.1080/14703297.2020.1794928>
- [7] Permana A.A.J., Pradnyana G.A. Recommendation Systems for internship place using artificial intelligence based on competence. *Journal of Physics: Conference Series*. 2019;1165(1):012007. <https://doi.org/10.1088/1742-6596/1165/1/012007>
- [8] Esin R.V., Kustitskaya T.A., Noskov M.V. *Prognostirovaniye uspeshnosti obucheniya po discipline na osnove universalnykh pokazatelyej tzirovogo sleda* [Predicting academic performance in a course by universal features of LMS Moodle digital footprint]. *Informatica i obrazovaniye = Informatics and Education*. 2023;38(3):31-41. (In Russ., abstract in Eng.) <https://doi.org/10.32517/0234-0453-2023-38-3-31-41>
- [9] Ozerova G.P., Pavlenko G.F. Prediction of student performance in blended learning utilizing learning analytics data. *Science for Education Today*. 2019;9(6):73-87. <https://doi.org/10.15293/2658-6762.1906.05>
- [10] Arthars N., Dollinger M., Vigentini L., Liu D.Y.T., Kondo E., King D.M. Empowering Teachers to Personalize Learning Support. In: Ifenthaler D., Mah D.K., Yau J.Y.K. (eds.) Utilizing Learning Analytics to Support Study Success. Cham: Springer; 2019. p. 223-248. https://doi.org/10.1007/978-3-319-64792-0_13
- [11] Ortigosa A., Carro R.M., Bravo-Agapito J., Lizcano D., Alcolea J. J. Blanco Ó. From lab to production: Lessons learnt and real-life challenges of an early student-dropout prevention system. *IEEE Transactions on Learning Technologies*. 2019;12(2):264-277. <https://doi.org/10.1109/TLT.2019.2911608>
- [12] Mussida P., Lanzi P.L. A computational tool for engineer dropout prediction. In: 2022 IEEE Global Engineering Education Conference (EDUCON). Tunis, Tunisia: IEEE Press; 2022. p. 1571-1576. <https://doi.org/10.1109/EDUCON52537.2022.9766632>
- [13] Salas-Martínez Á., Ramírez-Martinell A. Learning Analytics in Higher Education: A Decade in Systematic Literature Review. *Programming and Computer Software*. 2024;50:690-700. <https://doi.org/10.1134/S0361768824700701>
- [14] Kustitskaya T.A., Esin R.V., Kytmanov A.A., Zyкова T.V. Designing an Education Database in a Higher Education Institution for the Data-Driven Management of the Educational Process. *Education Sciences*. 2023;13(9):947. <https://doi.org/10.3390/educsci13090947>
- [15] Esin R.V., Zyкова T.V., Kustitskaya T.A., Kytmanov A.A. *Tzifrovaya obrazovatel'naya istoriya kak sostavlyaychaya tzifrovogo obrazovatel'nogo profilya obuchayuchego v usloviyah transformatsii obrazovaniya*. [Digital educational history as a component of the digital student's profile in the context of education transformation]. *Perspektivy Nauki i Obrazovaniya = The prospects of science and education*. 2022;(5):566-584. (In Russ., abstract in Eng.) <https://doi.org/10.32744/pse.2022.5.34>
- [16] York T.T., Gibson C., Rankin S. Defining and Measuring Academic Success. *Practical Assessment, Research, and Evaluation*. 2015;20(1):5. <https://doi.org/10.7275/hz5x-tx03>
- [17] Hommel M., Egetenmeier A., Maier U. Supporting Stakeholders with Learning Analytics to Increase Study Success. In: Ifenthaler D., Mah D.K., Yau J.K. (Eds.) Utilizing Learning Analytics to Support Study Success. Cham: Springer; 2019. p. 37-60. https://doi.org/10.1007/978-3-319-64792-0_3
- [18] Naseem M., Chaudhary K., Sharma B. Predicting Freshmen Attrition in Computing Science using Data Mining. *Education and Information Technologies*. 2022;27(7):9587-9617. <https://doi.org/10.1007/s10639-022-11018-3>
- [19] Del Bonifro F., Gabbriellini M., Lisanti G., Zingaro S.P. Student Dropout Prediction. In: Bittencourt I., Cukurova M., Muldner K., Luckin R., Millán E. (eds.) Artificial Intelligence in Education. AIED 2020. *Lecture Notes in Computer Science*. Vol. 12163. Cham: Springer; 2020. p. 129-140. https://doi.org/10.1007/978-3-030-52237-7_11



- [20] Song Z., Sung S.-H., Park D.-M., Park B.-K. All-Year Dropout Prediction Modeling and Analysis for University Students. *Applied Sciences*. 2023;13(2):1143. <https://doi.org/10.3390/app13021143>
- [21] Oqaidi K., Aouhassi S., Mansouri K. Towards a Students' Dropout Prediction Model in Higher Education Institutions Using Machine Learning Algorithms. *International Journal of Emerging Technologies in Learning*. 2022;17(18):103-117. <https://doi.org/10.3991/ijet.v17i18.25567>
- [22] Sulak S.A., Koklu N. Predicting Student Dropout Using Machine Learning Algorithms. *Intelligent Methods In Engineering Sciences*. 2024;3(3):91-98. <https://doi.org/10.58190/imiens.2024.103>
- [23] Alnasyan B., Basher M., Allassafi M. The power of Deep Learning techniques for predicting student performance in Virtual Learning Environments: A systematic literature review. *Computers and Education: Artificial Intelligence*. 2024;6:100231. <https://doi.org/10.1016/j.caeai.2024.100231>
- [24] Kabathova J., Drlik M. Towards Predicting Student's Dropout in University Courses Using Different Machine Learning Techniques. *Applied Sciences*. 2021;11(7):3130. <https://doi.org/10.3390/app11073130>
- [25] Kustitskaya T.A., Noskov M.V., Vainshtein Y.V. Predicting learning success: research problems and challenges. *Nauka i shkola*. 2023;(4):71-83. (In Russ., abstract in Eng.) <https://doi.org/10.31862/1819-463X-2023-4-71-83>

Поступила 17.02.2024; одобрена после рецензирования 28.04.2024; принята к публикации 11.06.2024.

Submitted 17.02.2024; approved after reviewing 28.04.2024; accepted for publication 11.06.2024.

Об авторах:

Кустицкая Татьяна Алексеевна, доцент кафедры прикладной математики и анализа данных Института космических и информационных технологий, ФГАОУ ВО «Сибирский федеральный университет» (660041, Российская Федерация, г. Красноярск, пр. Свободный, д. 79), кандидат физико-математических наук, **ORCID: <https://orcid.org/0000-0001-9854-1259>**, tkustitskaya@sfu-kras.ru

Есин Роман Витальевич, доцент кафедры прикладной математики и анализа данных Института космических и информационных технологий, ФГАОУ ВО «Сибирский федеральный университет» (660041, Российская Федерация, г. Красноярск, пр. Свободный, д. 79), кандидат педагогических наук, **ORCID: <https://orcid.org/0000-0002-9682-4690>**, resin@sfu-kras.ru

Кытманов Алексей Александрович, заведующий кафедрой высшей математики Института перспективных технологий и индустриального программирования, ФГБОУ ВО «МИРЭА – Российский технологический университет» (119452, Российская Федерация, г. Москва, пр. Вернадского, д. 78), доктор физико-математических наук, доцент, **ORCID: <https://orcid.org/0000-0003-3325-099X>**, aakyt@gmail.com

Зыкова Татьяна Викторовна, доцент кафедры прикладной математики и компьютерной безопасности Института космических и информационных технологий, ФГАОУ ВО «Сибирский федеральный университет» (660041, Российская Федерация, г. Красноярск, пр. Свободный, д. 79), кандидат физико-математических наук, доцент, **ORCID: <https://orcid.org/0000-0002-7332-2372>**, zykovatv@mail.ru

Все авторы прочитали и одобрили окончательный вариант рукописи.

About the authors:

Tatyana A. Kustitskaya, Associate Professor of the Chair of Applied Mathematics and Computer Security, School of Space and Information Technology, Siberian Federal University (79 Svobodny Pr., Krasnoyarsk 660041, Russian Federation), Cand. Sci. (Phys.-Math.), **ORCID: <https://orcid.org/0000-0001-9854-1259>**, tkustitskaya@sfu-kras.ru

Roman V. Esin, Associate Professor of the Chair of Applied Mathematics and Computer Security, School of Space and Information Technology, Siberian Federal University (79 Svobodny Pr., Krasnoyarsk 660041, Russian Federation), Cand. Sci. (Ped.), **ORCID: <https://orcid.org/0000-0002-9682-4690>**, resin@sfu-kras.ru

Alexey A. Kytmanov, Head of the Chair of Higher Mathematics, Institute for Advanced Technologies and Industrial Programming, MIREA – Russian Technological University (78 Vernadsky Ave., Moscow 119454, Russian Federation), Dr. Sci. (Phys.-Math.), Associate Professor, **ORCID: <https://orcid.org/0000-0003-3325-099X>**, aakyt@gmail.com

Tatyana V. Zyкова, Associate Professor of the Chair of Applied Mathematics and Computer Security, School of Space and Information Technology, Siberian Federal University (79 Svobodny Pr., Krasnoyarsk 660041, Russian Federation), Cand. Sci. (Phys.-Math.), Associate Professor, **ORCID: <https://orcid.org/0000-0002-7332-2372>**, zykovatv@mail.ru

All authors have read and approved the final manuscript.

