

DOI: 10.32517/0234-0453-2025-40-4-6-17



АНАЛИТИКА И НЕЙРОСЕТЕВАЯ ГЕНЕРАЦИЯ ЦИФРОВОГО СЛЕДА ДЛЯ ПОСТРОЕНИЯ ПЕРСОНАЛИЗИРОВАННЫХ ОБРАЗОВАТЕЛЬНЫХ ТРАЕКТОРИЙ

А. Г. Леонов^{1,2,3,4} ✉, К. А. Машченко^{1,2,4}, Н. С. Мартынов¹, А. В. Шляхов¹, Т. Г. Хан¹

¹ Федеральный научный центр Научно-исследовательский институт системных исследований Национального исследовательского центра «Курчатовский институт», г. Москва, Россия

² Московский государственный университет имени М. В. Ломоносова, г. Москва, Россия

³ Московский педагогический государственный университет, г. Москва, Россия

⁴ Государственный университет управления, г. Москва, Россия

✉ dr.l@vip.niisi.ru

Аннотация

В процессе работы преподавателя накапливается большое количество данных о прохождении курсов обучающимися. Эти данные необходимо обрабатывать для усовершенствования материалов и проверки педагогических гипотез. Однако с увеличением количества информации это становится практически нереализуемо. В статье рассматриваются подходы к анализу цифрового следа студентов и его нейросетевой генерации для построения интеллектуальных персонализированных траекторий обучения. Исследуются существующие подходы к адаптивному обучению и предлагается новая методика, основанная на анализе временных рядов и машинном обучении. Приводятся результаты применения цифровой образовательной платформы «Мирера» для сбора и обработки данных о студенческой активности, а также рассматриваются возможности прогнозирования учебных успехов студентов на основе их цифрового следа. Предлагается метод категоризации обучающихся на основе данных об их образовательных характеристиках и поведении, собранных в ходе обучения. Используется генерация синтетических данных для улучшения точности предсказательных моделей и построения долгосрочных предсказаний. Проведенный вычислительный эксперимент подтверждает эффективность предложенного подхода в сравнении с традиционными методами анализа образовательных данных. Полученные результаты могут быть использованы для автоматизации адаптивного обучения и повышения качества образовательного процесса.

Ключевые слова: цифровая образовательная платформа «Мирера», цифровой след, образовательная аналитика, сигнатуры, адаптивные траектории, образовательные технологии.

Для цитирования:

Леонов А. Г., Машченко К. А., Мартынов Н. С., Шляхов А. В., Хан Т. Г. Аналитика и нейросетевая генерация цифрового следа для построения персонализированных образовательных траекторий. *Информатика и образование*. 2025;40(4):6–17. DOI: 10.32517/0234-0453-2025-40-4-6-17.

ANALYTICS AND NEURAL NETWORK GENERATION OF DIGITAL FOOTPRINT FOR CONSTRUCTING PERSONALISED LEARNING TRAJECTORIES

A. G. Leonov^{1,2,3,4} ✉, K. A. Mashchenko^{1,2,4}, N. S. Martynov¹, A. V. Shlyakhov¹, T. G. Khan¹

¹ Scientific Research Institute for Systems Analysis of the National Research Centre "Kurchatov Institute", Moscow, Russia

² Lomonosov Moscow State University, Moscow, Russia

³ Moscow Pedagogical State University, Moscow, Russia

⁴ State University of Management, Moscow, Russia

✉ dr.l@vip.niisi.ru

Abstract

In a teacher's work, a large amount of data accumulates about students taking courses. The data must be processed to improve materials and test pedagogical hypotheses. But with an increase in the amount of information, this becomes almost impossible. The

article discusses approaches to the analysis of students' digital footprint and its neural network generation for constructing intelligent personalized learning trajectories. Existing approaches to adaptive learning and propose a new methodology based on time series analysis and machine learning are investigated. Application results of digital learning platform Mirera for collecting and processing student activity data and the ability to predict students' academic success based on their digital footprint are presented. A method for categorizing learners based on their learning characteristics and behavior data collected during training is proposed. Synthetic data generation is used to improve the accuracy of prediction models and build long-term predictions. The conducted computational experiment confirms the effectiveness of the proposed approach compared to traditional methods of educational data analysis. The results obtained can be used to automate adaptive learning and improve the quality of the educational process.

Keywords: digital educational platform Mirera, digital footprint, educational analytics, signatures, adaptive trajectories, educational technologies

For citation:

Leonov A. G., Mashchenko K. A., Martynov N. S., Shlyakhov A. V., Khan T. G. Analytics and neural network generation of digital footprint for constructing personalised learning trajectories. *Informatics and Education*. 2025;40(4):6–17. (In Russian.) DOI: 10.32517/0234-0453-2025-40-4-6-17.

1. Введение

Сегодня персонализированные системы обучения применяются практически на всех ступенях образования: в начальной, основной, средней школе [1, 2], в высших учебных заведениях [3]. Адаптивные методы используются не только в обучении, но и в тестировании, например, тест GMAT показывает высокую точность в оценке когнитивных способностей испытуемого [4].

Существующие персонализированные системы обучения можно разделить на группы в зависимости от используемой в них методики.

Первую группу составляют системы, основывающиеся на методах персонализации. Так, система ALEKS [5] адаптируется под обучаемого, используя начальный тест оценки знаний в качестве отправной точки образовательного процесса. Далее, при достижении необходимого количества успешно выполненных заданий, обучаемому предлагается промежуточный тест оценки знаний по теме. Менее детерминированной является система Knewton [6] — с каждым следующим ответом обучаемого система адаптируется к его уровню знаний, включая выявление недостаточно освоенных тем, что влечет повторение студентом уже пройденного материала для его закрепления.

Ко второй группе можно отнести системы, использующие специализированные технологии, направленные на изучение и закрепление материала. SCHOLAR — одна из первых адаптивных обучающих систем, использующая экспертную систему на основе правил для принятия решения о том, как корректно оценивать ответы обучаемого [7]. Более сложные технологии, такие как машинное обучение, использует, например, система DreamBox [8] — для принятия решения о том, какое следующее задание должно быть предоставлено студенту, система опирается не только на корректность ответов, но и на скорость реакции тестируемого, а также на другие подобные характеристики процесса решения задач.

К третьей группе относятся методы адаптивного обучения, строящие анализ поведения студента, исходя из сохраненного цифрового следа обучаемого.

К таким методам можно отнести подходы **дата-майнинга** [9]. Дата-майнинг — это процесс извлечения полезной информации и закономерностей из обширной базы данных.

В широком смысле методы дата-майнинга можно разделить на три крупные последовательные группы: 1) предобработку, 2) анализ и 3) постобработку. Методы анализа включают в себя хорошо знакомые классификацию и кластеризацию, а также менее часто ассоциируемые с дата-майнингом регрессию и корреляционный анализ.

В целом дата-майнинг уже достаточно давно находит свое применение в сфере образования [10]. Методы дата-майнинга могут быть полезны преподавателю для выявления обучающихся, нуждающихся в дополнительной помощи, для предсказания их итоговой оценки за курс, для поиска типичных и атипичных закономерностей поведения студентов, для идентификации наиболее распространенных ошибок и т. д.

В работе [11], исходя из накопленного за 15 недель цифрового следа обучаемого, студенты распределялись по 16 категориям. Эта кластеризация осуществлялась соответственно тесту Фельдера—Силверман, в котором обучаемый оценивается по четырем параметрам: 1) визуальный или вербальный, 2) действующий или мыслящий, 3) чувствующий или интуитивный, 4) последовательный или целостный. Оказывается, алгоритм машинного обучения k-средних показывает высокую метрику точности (англ. *accuracy*) — 89,7 % — в определении категории студента. В то же время алгоритм сокращения размерностей показал чуть меньшую точность, которая составила 85,2 %.

В статье [12] показано, как использование цифрового следа обучаемого помогло улучшить завершаемость образовательных курсов — 86 % студентов смогли освоить курс в полной мере в отведенное время. Для проверки эффективности персонализированного подхода в данной работе использовались непараметрические методы математической статистики. Для сравнения тестовой и контрольной групп авторы использовали тест Манна—Уитни.

Еще одним механизмом для имплементации и автоматизации адаптивного обучения являются **блокчейн-технологии**. Так, был проведен эксперимент с использованием тестовой сети блокчейна Ethereum [13].

Использование блокчейн-технологий для организации адаптивного обучения дает ряд преимуществ:

- частично автоматизирует сбор данных для регистрации учебных событий;
- повышает надежность и прослеживаемость данных учебного процесса;
- позволяет объективно оценить стиль обучения, поскольку поведение обучающихся регистрируется автоматически.

Однако существует ряд ограничений, связанных с применением цифрового следа в адаптивном образовании:

- **этические вопросы:** кто является истинным владельцем цифрового следа — студент или учебное заведение; что произойдет с данной информацией по завершении обучения; как хранить чувствительные для студента данные;
- **интеграционные сложности,** в частности, отсутствие единого стандарта регистрации цифрового следа, что осложняет обмен данными между различными системами и учебными заведениями;
- остается также открытым **вопрос о добавлении внешних данных в сформированный образовательной системой цифровой след студента**, например, данных о самообучении или обучении вне учебного заведения.

В рамках представленного в настоящей статье исследования рассмотрена реализация адаптивной методики на базе цифровой образовательной платформы (ЦОП) «Мирера». Это отечественная платформа, оптимизированная для российской системы высшего и среднего образования. ЦОП «Мирера» разрабатывалась как такой инструмент для преподавателя, который снимает значительную часть нагрузки путем автоматизации большинства рутинных образовательных процессов. В ЦОП «Мирера» реализована автоматическая проверка широкого класса заданий из различных предметных областей, при этом количество возможных типов заданий с автоматической проверкой постоянно расширяется. ЦОП «Мирера» ориентирована на разработку и проведение курсов, в которых сочетаются классическое очное обучение и дистанционные технологии. Использование такой платформы улучшает весь образовательный процесс, включая освоение материалов студентами, и обеспечивает эффективную организацию взаимодействия «ученик—учитель». Система использует методы искусственного интеллекта, что приводит к значительному росту результативности обучения.

Во время прохождения курсов на платформе «Мирера» студенты решают набор заданий внутри структурных единиц курса, называемых **контекстами**.

При этом сохраняется детальный цифровой след обучения, который отражает в частности:

- статистику посещаемости;
- активность;
- набранные баллы;

- количество и наполнение попыток решений задач;
- время, потраченное на решение задач;
- результаты проверки на плагиат;
- скорость набора на клавиатуре;
- другие информативные аналитические характеристики обучения.

Для преподавателя аналитика, построенная на цифровом следе студента, важна для персонального адаптивного воздействия на образовательный процесс и изменения содержания курса для конкретной группы или отдельных студентов [14]. Это также помогает своевременно выявлять как отстающих студентов — чтобы педагог имел возможность дополнительно помочь потенциальным двоечникам, так и студентов высоких достижений — для направления последних на более глубокое освоение предмета.

При массовом образовательном процессе итоговой статистики обучения недостаточно, чтобы быстро и эффективно оценить, кто из слушателей недостаточно хорошо освоил материал, плохо и(или) несамостоятельно выполняет задания (использует заимствования), а кто, наоборот, слишком быстро выполняет задания и ему нужно дать дополнительные интересные задачи, чтобы он смог двигаться вперед в изучении материала, не скучал и не отвлекал других обучающихся. Сведения о накопленных баллах, посещаемости и плагиате не дают возможность сформировать представление об индивидуальном «почерке» в решениях студента. Такими сходными характеристиками также могут обладать различные студенты, что позволяет отнести почерк решения заданий к групповому атрибуту. Таким образом, *возникает необходимость в разработке детальной аналитики, охватывающей цифровой след студентов на протяжении всего времени их обучения* [15].

2. Методология и стратегия исследования

2.1. Категориальная декомпозиция студентов

В рамках рассматриваемого исследования студенты были разделены на три категории, условно названные «отличники», «хорошисты», «двоечники»:

- **«отличники»** — это студенты, которые способны самостоятельно осваивать материал и которым, как правило, не требуется помощь преподавателя. Они посещают все занятия и всегда первыми выполняют задания. Чтобы полностью раскрыть потенциал таких студентов, необходимо большое количество сложных дополнительных заданий. В их работах практически нулевые показатели плагиата, поскольку работы этих студентов сами являются источниками заимствования для остальных студентов;
- **«хорошисты»** — это студенты, которые также осваивают материал, имеют высокий уровень посещаемости, стараются выполнять задания самостоятельно, однако иногда просят помощи у «отличников», чтобы разобраться

в каком-то материале, — это приводит к наличию заимствований в решенных заданиях. Такие студенты образуют большую группу и являются наиболее интересным объектом для проведения исследований и анализа;

- **«двоечники»** — это студенты, которые не смогли освоить курс за время обучения. В процессе прохождения курса разрыв уровня с остальными обучаемыми достигает такого размера, что «двоечники» не видят смысла в посещении занятий, чем еще больше усугубляют свое отставание. На контрольных такие студенты списывают, сдавая задания с первой попытки ближе к окончанию работы, когда другие студенты уже успешно закончили выполнять свои задания.

Из градации типов студентов следует, что «отличникам» практически не требуется вмешательство со стороны преподавателя, так как они успешно усваивают материал, в то время как работа с «двоечниками» требует значительных усилий со стороны преподавателя, в том числе на дополнительные индивидуальные занятия. Но основное внимание преподавателя должно быть сосредоточено на «хорошистах». Можно поставить педагогическую цель помочь «хорошистам» перейти в высшую категорию, при этом не допустить понижения категории для успевающих студентов.

Тогда **важной становится задача построения итоговой прогностической модели**: уметь предсказать тот самый момент, когда слушатель начинает переходить в категорию неуспевающего студента, чтобы преподаватель мог уделить временно отстающему особое внимание, а если такая проблема становится массовой, то изменить наполнение курса, например, проведя декомпозицию сложных заданий. Для выявления такого момента можно воспользоваться множеством различных критериев. Базовым критерием является снижение в динамике курса количества баллов, получаемых студентом. Другие поведенческие критерии — падение посещаемости и увеличение процента плагиата. Важным критерием является количество попыток, затраченных на выполнение задания.

Таким образом, **предварительный портрет неуспевающего студента выглядит так**: либо наблюдается низкое количество баллов на протяжении длительного промежутка времени, либо фиксируется такое поведение «двоечника», когда он сдает все задачи с первой попытки, но с высоким уровнем заимствования.

2.2. Формирование временного ряда обучения

Поскольку ЦОП «Мирера» хранит детальные данные о попытках решения задач студентами, можно проанализировать временной ряд попыток, что позволяет, например, обнаружить закономерность о сдаче «двоечниками» заданий ближе к концу контеста. Представление попыток решений и результатов контестов в виде многомерного временного ряда открывает возможности для построения поведенческой аналитики. Чистые итоговые метрики, без совместных взаимосвязей и привязки ко времени,

показывают весьма ограниченную картину, содержащую значительное количество ложных результатов и временами приводящую к ошибочным выводам. Помимо этого, обучившись на поведенческой аналитике и «почерке» студентов, прогнозирующая модель может детектировать изменение этого «почерка», что является основным сигналом того, что студент в скором времени может ухудшить свои результаты.

Использование аппарата временных рядов позволяет не только предсказать финальную категорию студента на основе небольшого количества начальных данных, но и прогнозировать дальнейшую траекторию его обучения (какие баллы студент будет набирать в будущем, с каким уровнем плагиата и посещаемости и т. д.). В дальнейшем можно рассчитать различные метрики и критерии, основанные на траекторных данных. Результат работы данной модели своевременно подскажет преподавателю о необходимости оказания целенаправленной помощи конкретному студенту, определив отстающего, например, по поведенческому «почерку».

2.3. Генерация синтетических данных для обучающей выборки

Полученный в образовательном процессе цифровой след в виде временного ряда используется для построения индивидуальной адаптивной траектории студента. Множество собранных цифровых следов студентов после предварительной обработки используется в качестве обучающего набора данных для глубоких нейронных сетей с помощью генерации дополнительных синтетических студенческих данных, похожих по распределению на исходные. Такой подход решает общеизвестную проблему недостатка входных данных для обучения моделей, основанных на нейросетевых технологиях. Кроме того, шкалирование временного ряда зачастую может быть неоднородным. Описанные выше поведенческие характеристики имеют временные точки, связанные с датой контеста. Также собираются данные для формирования временного ряда, основанного на информации об упорядоченных по времени попытках решения заданий (количество времени, потраченное студентом на попытку решения, ее статус, полученные за нее баллы, мера плагиата, процент совпадения с эталонным решением и т. д.). Такой ряд имеет *временное шкалирование*, которое основано на времени отправки попытки решения и не совпадает с упомянутым ранее *контестным шкалированием*. Следовательно, возникает необходимость с помощью аппроксимации выравнивать временную шкалу различных временных рядов. Для генерации недостающих данных используется математическая теория сигнатур путей [16–18]. Авторы использовали генеративные модели, основанные на сигнатурных методах, которые можно применить для создания дополнительных образовательных траекторий студентов точно такого же распределения, что и настоящие данные, и которые в дальнейшем используются для нейронных сетей и других глубоких моделей машинного обучения с описанной выше прогнозирующей

целью. На основании прогнозируемой моделью траектории обучения принимается решение о текущем статусе освоения курса студентом и о построении его индивидуальной адаптивной траектории обучения. Отличительной особенностью сигнатур является то, что они способны улавливать взаимосвязи между различными размерностями временного ряда. Например, появляется возможность обнаружить и учесть такую зависимость, когда при падении посещаемости растет мера плагиата и падает доля набранных за констест баллов.

2.4. Практическая значимость исследования

Исследование о формировании временных рядов по студенческим решениям открывает множество возможностей для улучшения образовательного процесса. Рассмотрим три ключевых аспекта:

- **Сравнение групп для проверки педагогических экспериментов.** Используя временные ряды, можно эффективно сравнивать результаты различных групп студентов, что позволяет проверять гипотезы педагогических экспериментов. Это дает возможность не только оценить эффективность изменений, но и адаптировать подходы на основе полученных данных.
- **Автоматическое выделение лучших студентов.** Анализ временных рядов позволяет отслеживать успехи обучаемых на протяжении временных периодов и автоматически фиксировать лучших студентов. При этом преподаватель может визуализировать студенческие траектории, что помогает оценивать достижения обучающихся. Это также способствует выявлению студентов, которые значительно улучшили свои результаты по сравнению с началом обучения. Такой подход позволяет более эффективно управлять образовательным процессом.
- **Прогнозирование успехов студентов и создание индивидуальных адаптивных траекторий курса.** Используя данные о предыдущих успехах студента, можно прогнозировать его будущие достижения и строить индивидуальные адаптивные траектории обучения. Это позволяет создать рекомендации, которые помогают обучаемым выбирать наиболее подходящие курсы и учебные материалы.

Указанные аспекты свидетельствуют о высоком педагогическом потенциале использования временных рядов в образовательной аналитике.

3. Проведение вычислительного эксперимента

3.1. Построение цифровой аналитики для выдвижения педагогических гипотез и сравнения экспериментальных групп

Основу данных, рассматриваемых в качестве информативных студенческих характеристик, составляют:

- процент набранных студентом баллов по каждому констесту;
- посещаемость студента;
- процент плагиата в решениях студента;
- количество попыток на решение каждой из задач.

В рамках исследования использовались данные двухлетнего курса «Работа на ЭВМ и программирование» обучения двух групп на механико-математическом факультете МГУ имени М. В. Ломоносова. Временной ряд составлялся по данным, сгруппированным по констестам полного курса, упорядоченным по временной оси конкретного констеста. Формирование цифрового следа студента с помощью временных рядов позволяет проверять различные педагогические гипотезы.

Так, на рисунке 1 проиллюстрирована цифровая аналитика для учебной группы студентов 2021 года поступления. Построен график количества попыток решения задач в констесте, которые отправлялись студентами в систему автоматической проверки. На графике каждая точка — это среднее количество попыток студента за конкретный констест по всем задачам. По горизонтальной оси расположены отсортированные по времени констесты, по вертикальной оси — усредненное по всем задачам констеста количество попыток студента, потраченное на эти задачи.

На графике можно отчетливо наблюдать четыре длительных непрерывных временных интервала, что соответствует факту семестрового деления двухлетнего курса, а также один короткий в январе 2023 года, когда проводился экзамен по пройденному материалу. В середине первого семестра обучения можно отметить небольшой рост количества попыток решения, что объясняется проведением промежуточных контрольных испытаний. Также заметен всплеск студенческой активности ближе к концу первого семестра, когда подавляющее большинство неуспевающих студентов пытаются закрыть зачетные мероприятия, при этом растет число попыток сдачи задач. Аналогичная ситуация наблюдается и во втором семестре, однако в третьем семестре нет такого лавинообразного роста, так как студенты к этому моменту уже сформировали определенного уровня предметные компетенции, несмотря на рост сложности отдельных заданий. В четвертом семестре отмечается падение среднего количества попыток при изучении таких тем, как «Программирование на Python», «Библиотека OpenGL», где задания подобраны таким образом, что не требуют большого количества попыток.

На рисунке 2 представлен график плагиата (заимствований) в работах студента при прохождении им курса.

Временная ось совпадает с отсортированной по констестам временной осью предыдущего графика, а на вертикальной оси отображена мера (процент) плагиата студента, усредненная по всем задачам внутри констеста. К концу каждого семестра усредненная мера плагиата падает, что объясняется спецификой

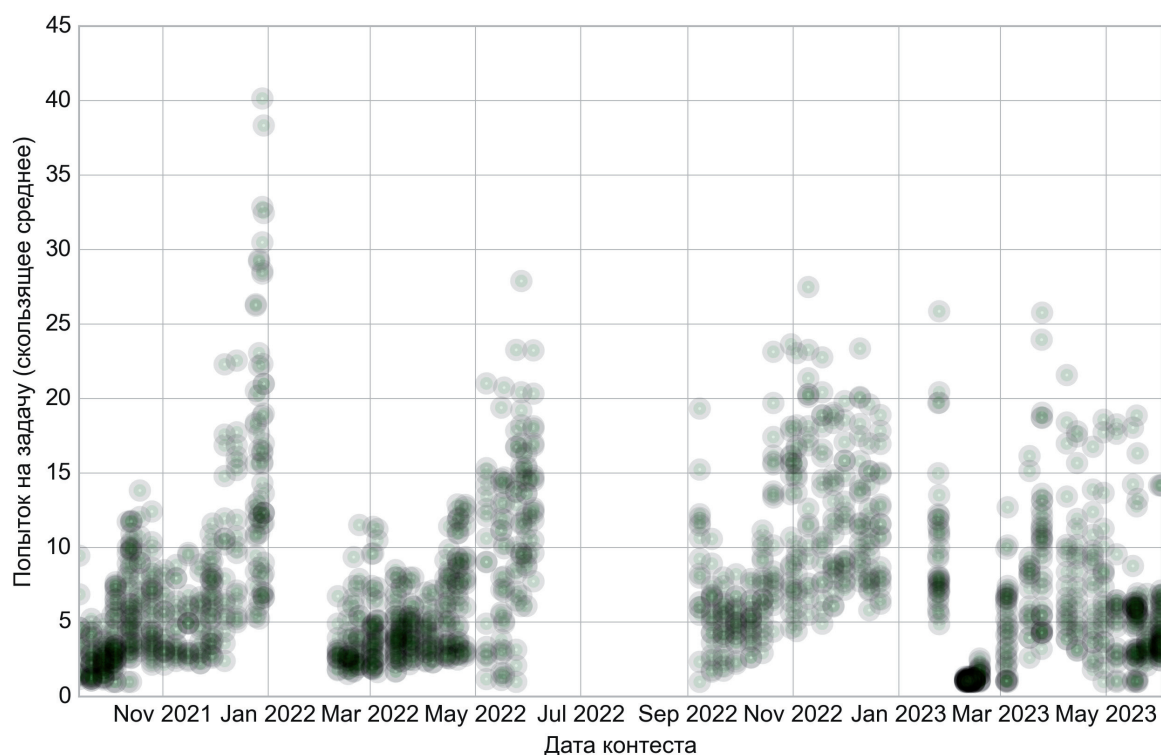


Рис. 1. Среднее количество потраченных на задачу попыток студентов, начавших обучение в 2021 году.
Каждая точка на графике соответствует конкретному студенту в группе

Fig. 1. Average number of attempts spent on a task by students who started their studies in 2021.
Each point on the graph corresponds to a specific student in the group

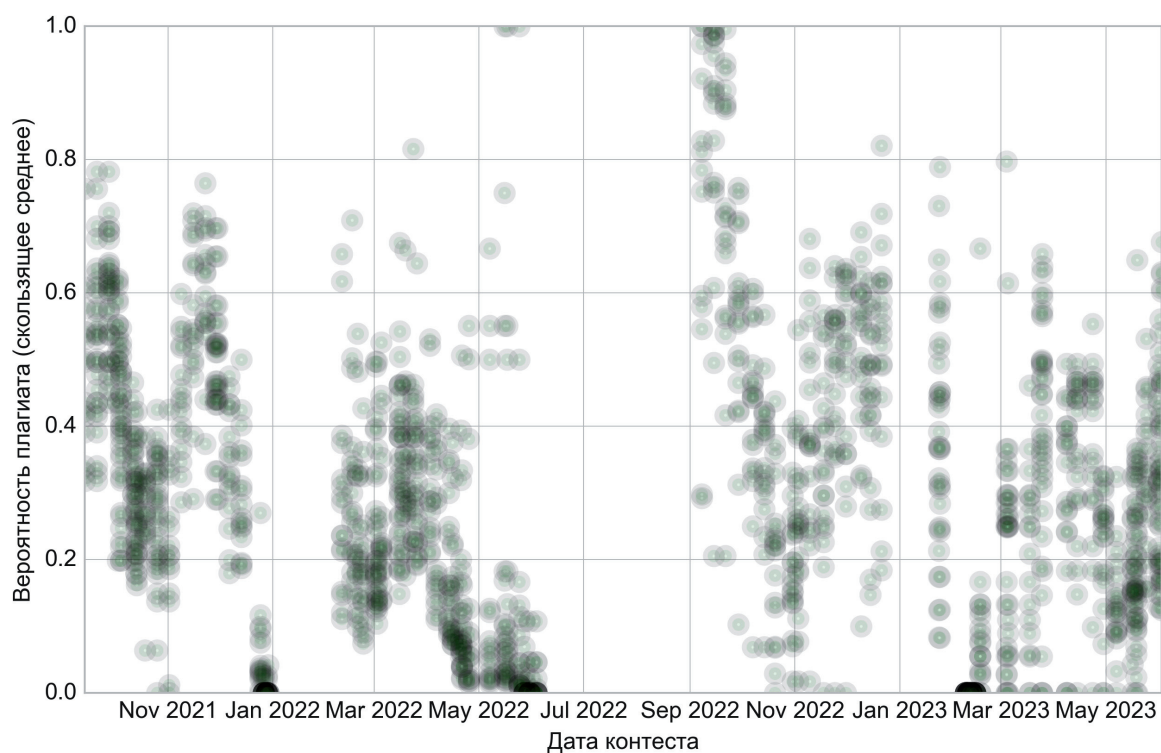


Рис. 2. Средний плагиат по всем задачам за конкурс у студентов, начавших учебу в 2021 году.
Каждая точка на графике соответствует конкретному студенту в группе

Fig. 2. Average plagiarism across all tasks during the contest for students who started their course in 2021.
Each point on the graph corresponds to a specific student in the group

ведения цифрового курса: к концу семестра «отличники» уже получили свои отметки, им блокируется доступ в систему, и отстающим студентам становится просто не у кого списывать. Раздача индивидуальных заданий на финальных контрольных мероприятиях также приводит к падению показателя плагиата.

Визуализация цифрового следа количественных характеристик обучаемых вдоль временной оси позволяет практикующему педагогу выдвигать различные гипотезы в процессе ведения курса. Такие гипотезы требуют не только количественной проверки, но и качественного визуального анализа временных рядов, чтобы понять, как, когда и какие изменения происходят при изучении курса.

Так, гипотеза проведенного эксперимента состояла в том, что рост качества обучения студентов произошел в результате внедрения новаций в ЦОП «Мирера» [19, 20]. Рассматривались студенты 2021 и 2022 годов поступления. Они проходили один и тот же курс, но в 2022 году были внедрены указанные улучшения и корректировки по итогам обучения с 2021-го по 2022 год. В среднем за месяц проводится четыре семинара и одна домашняя работа. На рисунке 3 построен сравнительный аналитический график доли баллов, набранных студентами разных годов поступления, сглаженный скользящим средним с окном в $4 + 1 = 5$ контестов.

По горизонтальной оси отложены контесты, которые для наглядности совмещены по временной

оси по группам в разные годы поступления. Сплошной линией на графике отображена средняя доля набранных баллов по контесту студентами группы, а пунктирные линии заключают между собой область от квантиля 20 % до квантиля 80 % набранных баллов внутри одной группы студентов. Студентам, решившим все задачи, как правило, предлагаются сложные задания, дающие дополнительные баллы. Также за определенные достижения педагог может премировать студентов дополнительными баллами. Таким образом, суммарные баллы за контест потенциально могут превысить зачетный уровень, т. е. нормированный результат может быть больше единицы. Фактически в данном эксперименте сравниваются диапазоны условных «хорошистов».

На графике, представленном на рисунке 3, явно видно, особенно в третьем семестре (когда студентам предлагались сложные задания), что у группы студентов, поступивших в 2022 году, квантиль 20 % и квантиль 80 % стали значительно ближе к среднему значению по группе, т. е. по итогу обучения сформировался больший костяк уверенных «хорошистов» (получивших в результате обучения по курсу оценку «отлично»), набирающих без наличия видимого разброса большое количество баллов. Можно сделать вывод, что чем меньше разброс квантилей для «хорошистов», тем более качественное и равномерное обучение предлагается по курсу и тем успешнее результаты применения цифровой образовательной платформы.



Рис. 3. Доля баллов, набранных по контестам, в среднем у студентов группы, среднее, а также квантили 20 % и 80 % при сравнении результатов студентов 2021 и 2022 годов поступления на механико-математическом факультете МГУ имени М.В. Ломоносова

Fig. 3. Average part of points scored in contests by students in the group, along with the average and the 20 % and 80 % quantiles when comparing the results of the students of the 2021 and 2022 admission years at the Faculty of Mechanics and Mathematics, Lomonosov Moscow State University

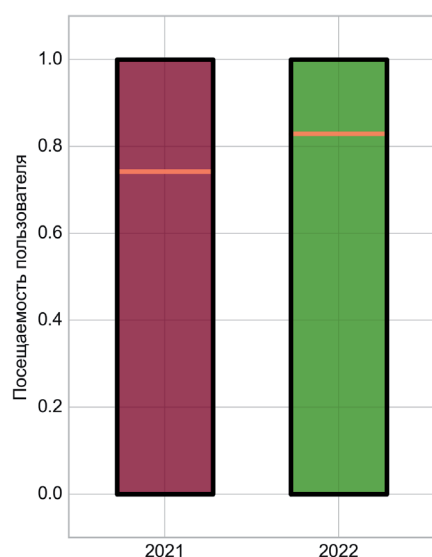


Рис. 4. Средняя посещаемость по всем контекстам и всем студентам в группах, среднее, а также квантили 20 % и 80 % при сравнении результатов студентов 2021 и 2022 годов поступления механико-математического факультета МГУ имени М.В.Ломоносова

Fig. 4. Average attendance in contests by students in the group, along with the average and the 20 % and 80 % quantiles when comparing the results from the 2021 and 2022 admission years at the Faculty of Mechanics and Mathematics, Lomonosov Moscow State University

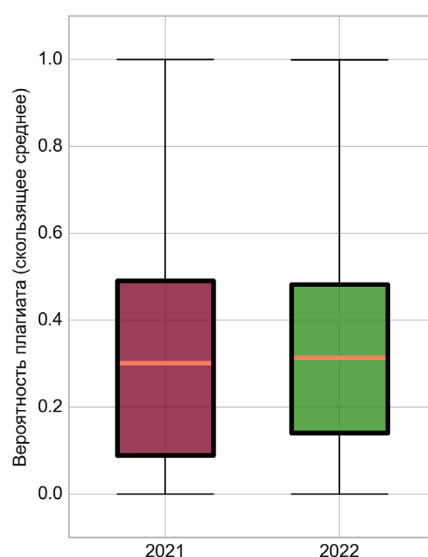


Рис. 5. Средняя мера плагиата по всем контекстам и всем студентам в группах, среднее, а также квантили 20 % и 80 % при сравнении результатов студентов 2021 и 2022 годов поступления механико-математического факультета МГУ имени М.В.Ломоносова

Fig. 5. Average plagiarism measure in contests by students in the group, along with the average and the 20 % and 80 % quantiles when comparing the results from the 2021 and 2022 admission years at the Faculty of Mechanics and Mathematics, Lomonosov Moscow State University

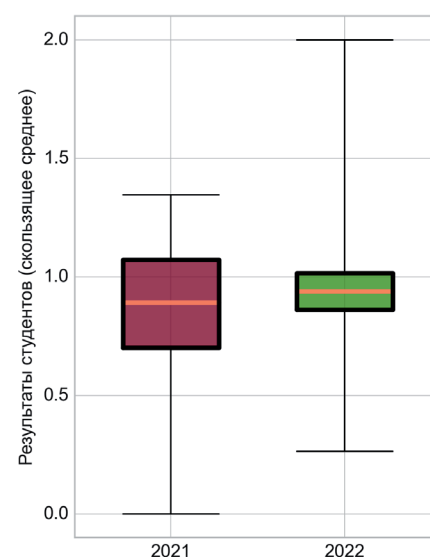


Рис. 6. Средняя доля баллов по всем контекстам и всем студентам в группах, среднее, а также квантили 20 % и 80 % при сравнении результатов студентов 2021 и 2022 годов поступления механико-математического факультета МГУ имени М.В.Ломоносова

Fig. 6. Average part of points scored in contests by students in the group, along with the average and the 20 % and 80 % quantiles when comparing the results from the 2021 and 2022 admission years at the Faculty of Mechanics and Mathematics, Lomonosov Moscow State University

Для проведения итогового сравнения групп полезно рассматривать различные статистические характеристики временного ряда процесса обучения, выраженные в виде количественных величин. Например, можно найти среднее значение по траектории студента, усреднив величины по всей временной оси, т. е. по всем контекстам курса. Для наглядной визуализации статистических характеристик удобно использовать график типа boxplot, на котором горизонтальная линия внутри прямоугольной области — это среднее значение, верхняя и нижняя границы прямоугольника — это квантили 20 % и квантили 80 %, а оставшиеся горизонтальные линии соответствуют минимальному и максимальному значениям. Проведено сравнение результатов двух групп разных годов обучения с помощью построенных подобных графиков для учитываемых признаков, а именно изображенного на рисунке 4 графика посещаемости, графика меры плагиата, проиллюстрированного на рисунке 5 и графика доли набранных баллов, который можно видеть на рисунке 6.

Аналитическая картина плагиата не сильно менялась в рамках эксперимента, однако стоит отметить, что среднее значение в 2022 году стало выше, чем в 2021 году. Также область между квантилями стала более узкой, что означает уменьшение отклонения от среднего значения. Из небольшого сдвига

верхнего квантиля к среднему значению следует, что основная масса студентов стала меньше списывать. Можно заметить, что средняя доля набранных баллов увеличилась. К тому же квантили заметно приблизились к среднему значению, что говорит о более успешном усвоении курса группой в целом. То есть обучение на курсе можно охарактеризовать как более однородное и качественное для группы студентов, поступивших в 2022 году.

Внедрение описанного метода сравнения результатов групп в образовательную платформу позволяет проверять гипотезы различных педагогических экспериментов.

3.2. Анализ процесса студенческого обучения и выделение наиболее успешных студентов

В продолжение описанного выше эксперимента рассмотрена возможность выявления подгруппы наиболее успевающих студентов с помощью визуализации их индивидуальных траекторий обучения. В группе студентов 2021 года поступления была отслежена последовательная сумма посещаемости каждого студента, в процессе вычисления которой происходит прибавление к текущей накопленной сумме единицы, если студент посетил занятие, и нуля, если не посетил (рис. 7).

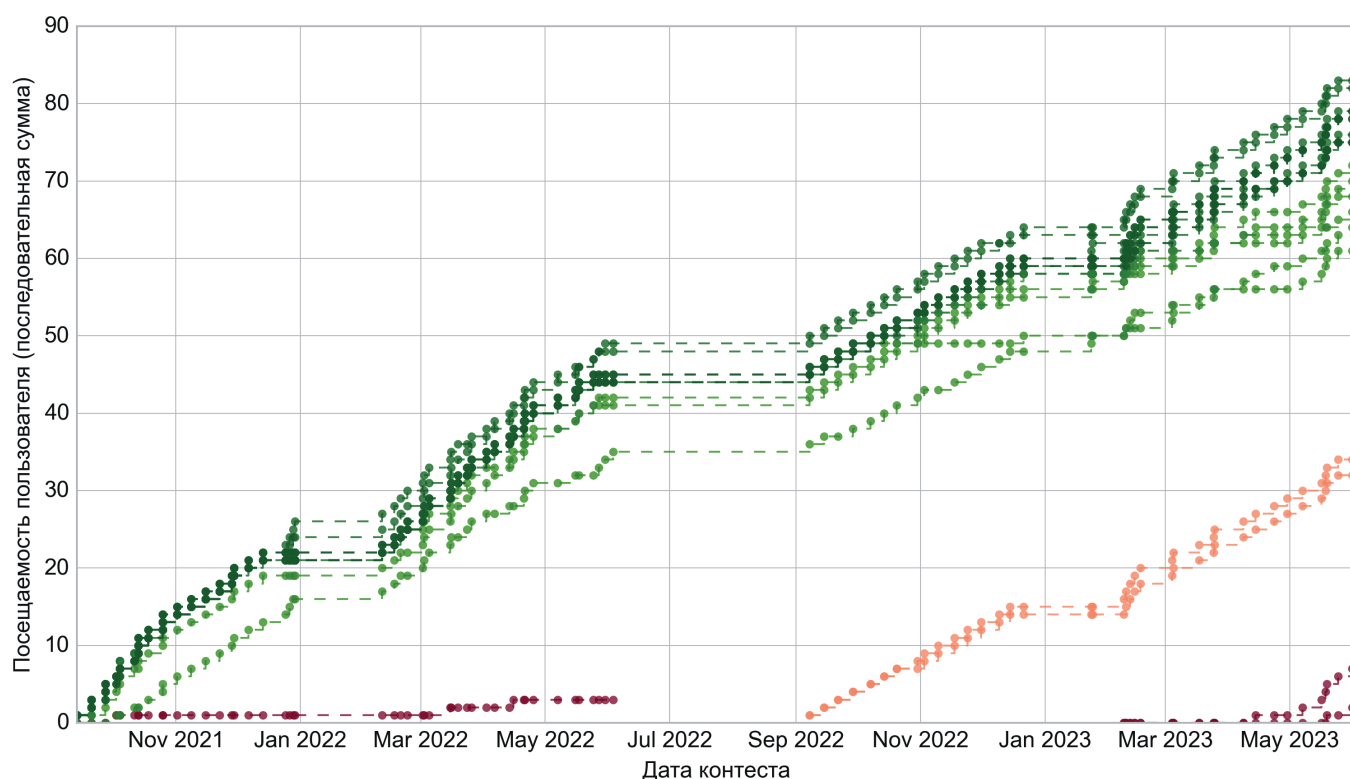


Рис. 7. Последовательная сумма посещаемости по контекстам студентов, начавших учиться в 2021 году.
Каждая точка на графике соответствует посещаемости конкретного студента

Fig. 7. Cumulative attendance sum over time for contests among students started learning in 2021.
Each point on the graph represents the attendance of a specific student in the group

На рисунке 7 каждый скачок вверх на единицу означает, что студент посетил контекст. Если на продолжительном по времени промежутке значение остается постоянным для данного студента и таких постоянных промежутков для него становится много, если падает усредненный наклон графика посещаемости, то можно сделать вывод, что приближается момент возможной «потери» студента в случае отсутствия дополнительного вмешательства педагога. В конечном итоге легко выделяются студенты, посещающие большинство контекстов, а также можно отметить студентов, которые посещают мало занятий, — на графике они всегда будут расположены внизу.

3.3. Прогнозирование студенческих успехов для построения адаптивной траектории курса

Итоговой целью работы модели прогнозирования студенческих результатов является предсказание будущей образовательной траектории студента (т. е. каждой характеристики цифрового следа), которое даст возможность не только предвидеть, например, станет ли некий неуспевающий студент через месяц «двоечником», но и автоматически создавать для студентов персональные адаптивные траектории будущего обучения. Визуальное подтверждение повышения уровня предметной компетенции студента с переходом в другую категорию может быть предоставлено системой автоматической оценки

достижений студента с помощью вычислительной модели, которая и выявит успешность студента. А в дальнейшем предварительно обученная на сгенерированных сигнатурными методами данных нейросетевая модель способна также предложить помощь в предсказании успешности студента по выстроенной траектории обучения.

Можно зафиксировать месячное скользящее окно пяти связанных по времени контекстов, по которым строится предсказание. Обучающая выборка собирается следующим образом: рассматриваются пять последовательных по времени контекстов, по которым строится временной ряд образовательной траектории студента, затем для рассматриваемого временного ряда вычисляются сигнатуры вдоль пути, а значение, следующее непосредственно за подвыборкой, принимается за корректное значение ответа модели. После сбора обучающей выборки производится запуск регрессионной нейросетевой модели, которая прогнозирует следующее значение временного ряда, используя только извлеченный из него набор сигнатур.

На графике, представленном на рисунке 8, проиллюстрированы результаты работы обученной прогностической модели на тестовой выборке. Вертикальной прерывистой линией отмечен момент, с которого эта модель начинает свою независимую работу по предсказанию следующего значения. При сдвиге окна каждый раз в качестве нового

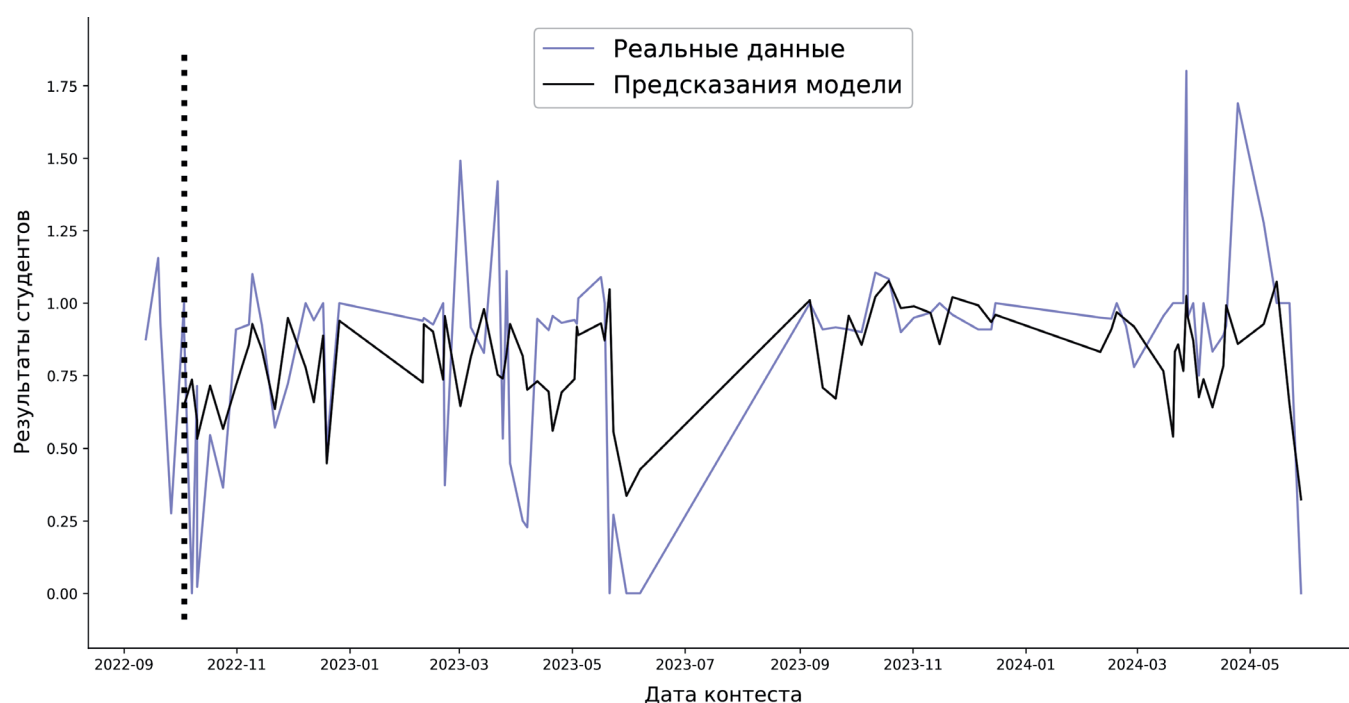


Рис. 8. Пример работы модели предсказания траектории студента с помощью сигнатур

Fig. 8. An example of the work of the model for predicting the trajectory of a student using signatures

значения добавляется истинное значение. Такой подход хорошо соотносится с реальным педагогическим применением модели, так как она в режиме реального времени имеет доступ к актуальным данным студента на текущую дату. На основании этих данных прогностическая модель служит рекомендательной системой, дающей подсказку преподавателю по проведению курса. Однако при использовании модели для других задач анализа траектории студента может быть полезной практикой добавлять в качестве нового значения вновь предсказанное. При использовании модели в таком режиме можно получить аппроксимацию дальнейшего поведения студента, хотя ошибка прогнозирования будет иметь накопительный характер.

4. Заключение

В работе предложена методика анализа цифрового следа студентов и механизмы ее использования для построения персонализированных траекторий обучения. В исследовании рассмотрены три категории студентов, а также основные показатели их образовательного поведения, влияющие на прогнозирование успеваемости. Проведенные вычислительные эксперименты показали, что анализ временных рядов и применение машинного обучения позволяют эффективно прогнозировать динамику обучения студентов и выявлять потенциальные риски снижения успеваемости.

Реализация предложенного подхода в рамках цифровой образовательной платформы «Мирера» демонстрирует его значимость для автоматизированного управления образовательным процессом.

Результаты исследования подтвердили, что применение цифрового следа и нейросетевых методов анализа способствуют повышению точности прогнозирования академической успеваемости студентов, что открывает дополнительные возможности для своевременной корректировки образовательных стратегий и повышения эффективности обучения.

Финансирование

Работа выполнена в рамках темы государственного задания НИЦ «Курчатовский институт» - НИИСИ по теме № FNEF-2024-0001 (1023032100070-3-1.2.1).

Funding

The work was carried out within the framework of the state assignment NRC “Kurchatov Institute” - SRISA on the topic № FNEF-2024-0001 (1023032100070-3-1.2.1).

Список источников / References

1. Moltudal S. H., Krumsvik R. J., Høydal K. L. Adaptive learning technology in primary education: Implications for professional teacher knowledge and classroom management. *Frontiers in Education*. 2022;7:1–18. EDN: OVOKQG. DOI: 10.3389/feduc.2022.830536.
2. Tosheva S., Martinovska C. Adaptive e-learning system in secondary education. *International Journal of Emerging Technologies in Learning (iJET)*. 2012;7(S1):36–41. DOI: 10.3991/ijet.v7iS1.1913.
3. Muñoz J. L., Ojeda F. M., Jurado D. L., Peña P. F., Carranza C. P., Berrios H. Q., Molina S. U., Farfan A. R., Arias-González J. L., Vasquez-Pauca M. J. Systematic review of adaptive learning technology for learning in higher education. *Eurasian Journal of Educational Research*. 2022;98:221–233. DOI: 10.14689/ejer.2022.98.014.
4. Oh I.-S., Schmidt F. L., Shaffer J. A., Le H. The Graduate Management Admission Test (GMAT) is even more valid than we thought: A new development in meta-analysis and its implications for the validity of the GMAT. *Academy of*

Management Learning & Education. 2008;7(4):567–569. DOI: 10.5465/AMLE.2008.35882196.

5. Cosyn E., Uzun H. B., Doble C., Matayoshi J. A practical perspective on knowledge space theory: ALEKS and its data. *Journal of Mathematical Psychology*. 2021;101:102512. EDN: FYNMIL. DOI: 10.13140/RG.2.2.20082.53448.

6. Conklin T. A. Knewton (An adaptive learning platform available at <https://www.knewton.com/>). *Academy of Management Learning & Education*. 2016;15(3):635–647. DOI: 10.5465/amle.2016.0206.

7. Carbonell J. R. AI in CAI: An artificial-intelligence approach to computer-assisted instruction. *IEEE Transactions on Man-Machine Systems*. 1970;11(4):190–202. DOI: 10.1109/TMMS.1970.299942.

8. Gjermeni F., Prodani F. AI and student engagement: A comparative analysis. *Interdisciplinary Journal of Research and Development*. 2024;11(3):194–208. DOI: 10.56345/ijrdv11n326.

9. Shahiri A. M., Husain W., Rashid N. A. A review on predicting student's performance using data mining techniques. *Procedia Computer Science*. 2015;72:414–422. DOI: 10.1016/j.procs.2015.12.157.

10. Romero C., Ventura S. Educational data mining: A review of the state of the art. *IEEE Transactions on Systems, Man, and Cybernetics*. 2010;40(6):601–618. DOI: 10.1109/TSMCC.2010.2053532.

11. Almazroei E., Alsharif E. F. Automated learning style classification using autoencoders and Felder—Silverman model from students' digital footprints. *Machine Intelligent Research*. 2024;18(02):1–23.

12. Горбатов С. В., Краснова Е. А. Цифровой след как механизм индивидуализации образовательной траектории студента (на примере курса «Цифровые технологии самообразования»). *Перспективы науки и образования*. 2022;58(4):193–208. EDN: SPPWXY. DOI: 10.32744/pse.2022.4.12.

[Gorbatov S. V., Krasnova E. A. A digital footprint as a mechanism of individualizing a student's educational trajectory (on the case of the “Digital technologies of self-education” course). *Perspectives of Science and Education*. 2022;58(4):193–208. (In Russian.) DOI: 10.32744/pse.2022.4.12.]

13. Зими́на Д. В., Му́ромцев Д. И. Автоматизация разработки адаптивного электронного обучения с использованием технологии блокчейн. *Вестник Астраханского государственного технического университета. Серия: Управление, вычислительная техника и информатика*. 2022;(2):119–128. EDN: WVSUAE. DOI: 10.24143/2072-9502-2022-2-119-128.

[Zimina D. V., Muromtsev D. I. Automation of developing adaptive e-learning using blockchain technology. *Vestnik of Astrakhan State Technical University. Series: Management, Computer Science and Informatics*. 2022;(2):119–128. (In Russian.) EDN: WVSUAE. DOI: 10.24143/2072-9502-2022-2-119-128.]

14. Багдасарян Л. Ш., Ардеев А. Х., Куликова Т. А., Поддубная Н. А. Применение цифрового следа образовательных интеллектуальных технологий. *Вестник Северо-Кавказского федерального университета*. 2024;(2):150–157. EDN: PXGKJF. DOI: 10.37493/2307-907X.2024.2.18.

[Bagdasaryan L. Sh., Ardeev A. H., Kulikova T. A., Poddubnaya N. A. Application of a digital footprint of educational intelligent technologies. *Newsletter of North-Caucasus Federal University*. 2024;(2):150–157. (In Russian.) EDN: PXGKJF. DOI: 10.37493/2307-907X.2024.2.18.]

15. Azcona D., Hsiao I.-H., Smeaton A. F. Detecting students-at-risk in computer programming classes with learning analytics from students' digital footprints. *User Modeling and User-Adapted Interaction*. 2019;29(2):759–788. EDN: HNAIFM. DOI: 10.1007/s11257-019-09234-7.

16. Chen K.-T. Integration of paths, geometric invariants and a generalized Baker—Hausdorff formula. *Annals of Mathematics*. 1957;65(1):163–178. DOI: 10.2307/1969671.

17. Chen K.-T. Integration of paths — a faithful representation of paths by non-commutative formal power series. *Transactions of the American Mathematical Society*. 1958;89:395–407. DOI: 10.1090/S0002-9947-1958-0106258-0.

18. Youngs R., Fisher E. Learning from the past and predicting the future. *The Journal of Laryngology & Otolaryngology*. 2015;129(6):519. DOI: 10.1017/S0022215115001383.

19. Васильев И. А., Караваева А. С., Леонов А. Г., Мащенко К. А., Шляхов А. В. О нововведениях в цифровой образовательной платформе Мирера. *Труды научно-исследовательского института системных исследований Российской академии наук*. 2023;13(4):127–134. EDN: MPQKZN. DOI: 10.25682/NIISI.2023.4.0018.

[Vasilyev I. A., Karavaeva A. S., Leonov A. G., Mashchenko K. A., Shlyakhov A. V. About innovations in the digital educational platform of Mirera. *Proceedings of the Research Institute for System Research of the Russian Academy of Sciences*. 2023;13(4):127–134. (In Russian.) EDN: MPQKZN. DOI: 10.25682/NIISI.2023.4.0018.]

20. Леонов А. Г., Мартынов Н. С., Мащенко К. А., Холькина А. А., Шляхов А. В. Автоматизация проверки семантической составляющей текстовых ответов обучающихся в цифровой образовательной платформе. *Программные продукты и системы*. 2024;37(3):440–452. EDN: СТТАJN. DOI: 10.15827/0236-235X.142.440-452.

[Leonov A. G., Martynov N. S., Mashchenko K. A., Kholkina A. A., Shlyakhov A. V. Automation of semantic analysis for textual responses of students in a digital educational platform. *Software & Systems*. 2024;37(3):440–452. (In Russian.) EDN: СТТАJN. DOI: 10.15827/0236-235X.142.440-452.]

Информация об авторах

Леонов Александр Георгиевич, доктор пед. наук, канд. физ.-мат. наук, доцент, зав. кафедрой дополнительного профессионального образования, Федеральный научный центр Научно-исследовательский институт системных исследований Национального исследовательского центра «Курчатовский институт», г. Москва, Россия; ведущий научный сотрудник лаборатории вычислительных методов, кафедра вычислительной математики, отделение математики, механико-математический факультет, Московский государственный университет имени М. В. Ломоносова, г. Москва, Россия; профессор кафедры математики и информатики в начальной школе, факультет начального образования, Институт детства, Московский педагогический государственный университет, г. Москва, Россия; руководитель магистерской программы «Цифровые технологии в государственном и корпоративном управлении», доцент кафедры математических методов в экономике и управлении, Институт информационных систем, Государственный университет управления, г. Москва, Россия; *ORCID*: <https://orcid.org/0000-0001-9622-1526>; *e-mail*: dr.l@vip.niisi.ru

Мащенко Кирилл Алексеевич, младший научный сотрудник отдела учебной информатики, Федеральный научный центр Научно-исследовательский институт системных исследований Национального исследовательского центра «Курчатовский институт», г. Москва, Россия; младший научный сотрудник лаборатории вычислительных методов, кафедра вычислительной математики, отделение математики, механико-математический факультет, Московский государственный университет имени М. В. Ломоносова, г. Москва, Россия; преподаватель магистерской программы «Цифровые технологии в государственном и корпоративном управлении», кафедра математических методов в экономике и управлении, Институт информационных систем, Государственный университет управления, г. Москва, Россия; *ORCID*: <https://orcid.org/0000-0003-0355-6699>; *e-mail*: kirill.mashchenko@niisi.ru

Мартынов Николай Сергеевич, младший научный сотрудник отдела учебной информатики, Федеральный научный центр Научно-исследовательский институт системных

исследований Национального исследовательского центра «Курчатовский институт», г. Москва, Россия; *ORCID*: <https://orcid.org/0009-0001-3071-2485>; *e-mail*: nikolai.martynov@math.msu.ru

Шляхов Артем Вячеславович, младший научный сотрудник отдела учебной информатики, Федеральный научный центр Научно-исследовательский институт системных исследований Национального исследовательского центра «Курчатовский институт», г. Москва, Россия; *ORCID*: <https://orcid.org/0009-0005-7108-6707>; *e-mail*: shlyakhov@vip.niisi.ru

Хан Татьяна Григорьевна, аспирант, Федеральный научный центр Научно-исследовательский институт системных исследований Национального исследовательского центра «Курчатовский институт», г. Москва, Россия; *ORCID*: <https://orcid.org/0009-0002-7497-5460>; *e-mail*: tatyana.khan@bk.ru

Information about the authors

Alexander G. Leonov, Doctor of Sciences (Education), Candidate of Sciences (Physics and Mathematics), Docent, Head of the Department of Continuing Professional Education, Scientific Research Institute for Systems Analysis of the National Research Centre “Kurchatov Institute”, Moscow, Russia; Leading Researcher at the Laboratory of Computational Methods, The Department of Computational Mathematics, Branch of Mathematics, Faculty of Mechanics and Mathematics, Lomonosov Moscow State University, Moscow, Russia; Professor at the Department of Mathematics and Informatics in Elementary School, Faculty of Elementary Education, Institute of Childhood, Moscow Pedagogical State University, Moscow, Russia; Master’s Program Supervisor of the “Digital Technologies in Public and Corporate Governance”, Associate Professor at the Department of Mathematical Methods in Economics and Management, Institute of Information Systems, State University of Management, Moscow,

Russia; *ORCID*: <https://orcid.org/0000-0001-9622-1526>; *e-mail*: dr.l@vip.niisi.ru

Kirill A. Mashchenko, Junior Researcher at the Department of Educational Informatics, Scientific Research Institute for Systems Analysis of the National Research Centre “Kurchatov Institute”, Moscow, Russia; Junior Researcher at the Laboratory of Computational Methods, The Department of Computational Mathematics, Branch of Mathematics, Faculty of Mechanics and Mathematics, Lomonosov Moscow State University, Moscow, Russia; Master’s Program Instructor of the “Digital Technologies in Public and Corporate Governance”, The Department of Mathematical Methods in Economics and Management, Institute of Information Systems, State University of Management, Moscow, Russia; *ORCID*: <https://orcid.org/0000-0003-0355-6699>; *e-mail*: kirill010399@vip.niisi.ru

Nikolay S. Martynov, Junior Researcher at the Department of Educational Informatics, Scientific Research Institute for Systems Analysis of the National Research Centre “Kurchatov Institute”, Moscow, Russia; *ORCID*: <https://orcid.org/0009-0001-3071-2485>; *e-mail*: nikolai.martynov@math.msu.ru

Artem V. Shlyakhov, Junior Researcher at the Department of Educational Informatics, Scientific Research Institute for Systems Analysis of the National Research Centre “Kurchatov Institute”, Moscow, Russia; *ORCID*: <https://orcid.org/0009-0005-7108-6707>; *e-mail*: shlyakhov@vip.niisi.ru

Tatyana G. Khan, a postgraduate student, Scientific Research Institute for Systems Analysis of the National Research Centre “Kurchatov Institute”, Moscow, Russia; *ORCID*: <https://orcid.org/0009-0002-7497-5460>; *e-mail*: tatyana.khan@bk.ru

Поступила в редакцию / Received: 19.03.25.

Поступила после рецензирования / Revised: 19.05.25.

Принята к печати / Accepted: 20.05.25.