

Применение генеративного искусственного интеллекта (ИИ) для анализа образовательных данных и прогнозирования академической успеваемости студентов

Жанна Геннадьевна Вегера

Доцент

Российский технологический университет (РТУ МИРЭА)

Москва, Россия

Vegera2024@mail.ru

ORCID 0000-0000-0000-0000

Поступила в редакцию 04.06.2024

Принята 24.07.2024

Опубликована 15.08.2024

УДК 004.89:37.091.12

DOI 10.25726/j2473-1350-7803-t

EDN MYZYAQ

ВАК 5.8.1. Общая педагогика, история педагогики и образования (педагогические науки)

OECD 05.03.HA. EDUCATION & EDUCATIONAL RESEARCH

Аннотация

Актуальность темы применения генеративного искусственного интеллекта (ИИ) для анализа образовательных данных и прогнозирования академической успеваемости студентов обусловлена растущей цифровизацией высшего образования и потребностью в новых инструментах повышения его эффективности. Анализ современных публикаций выявил недостаточную исследованность возможностей генеративного ИИ в данной предметной области. Цель работы - разработать и апробировать модель на основе генеративных состязательных сетей (GAN) для анализа массивов образовательных данных и предикции индивидуальных траекторий академической успешности. Задачи: 1) адаптировать архитектуру GAN под специфику образовательных данных; 2) обучить модель на репрезентативной выборке из LMS-логов 5000 студентов; 3) верифицировать прогностическую точность на независимых данных; 4) сравнить результаты с традиционными методами машинного обучения. Методы: генеративное состязательное моделирование, байесовская оптимизация гиперпараметров, стратифицированная k-блочная кросс-валидация, тесты Фридмана и Немени. Результаты: 1) построенная GAN-модель превосходит базовые решения по точности и полноте; 2) устойчива к дисбалансу классов и переобучению; 3) позволяет генерировать синтетические профили обучающихся для расширения набора признаков; 4) демонстрирует практическую применимость в реальном образовательном процессе. Предложенный подход открывает перспективы для создания адаптивных обучающих систем, способных в реальном времени идентифицировать студентов из группы риска и обеспечивать индивидуализированную поддержку.

Ключевые слова

генеративный искусственный интеллект, генеративные состязательные сети (GAN), анализ образовательных данных, машинное обучение, предикция академической успеваемости, цифровые следы.

Введение

Стремительная цифровая трансформация высшего образования, интенсифицировавшаяся в условиях пандемии COVID-19, привела к лавинообразному росту объемов и разнообразия данных об образовательных траекториях студентов (Дубовикова, 2023). Системы управления обучением (LMS), массовые открытые онлайн-курсы (MOOCs), инструменты адаптивного тестирования генерируют

терабайты многомерных, неструктурированных, разнородных по своей природе цифровых следов, отражающих паттерны учебной активности и когнитивных стратегий обучающихся (Дьюи, 1997). Задача эффективного использования этого информационного пласта для поддержки управленческих решений и совершенствования дизайна образовательного процесса приобретает критическую актуальность (Жуков, 2022).

Одним из наиболее перспективных направлений в данной сфере является применение методов машинного обучения (ML) для анализа образовательных данных (EDM) и прогнозирования академической успеваемости студентов. Доминирующие в современных EDM подходы – логистическая регрессия, деревья решений, случайный лес, нейронные сети, методы опорных векторов (Кодекс этики в сфере ИИ, 2021) – демонстрируют хорошие результаты на небольших однородных выборках, но теряют эффективность при росте объема и вариативности данных (Константинова, 2023).

Кроме того, в большинстве существующих ML-моделей не учитывается структурная неоднородность образовательных данных, обусловленная принципиальными различиями между обучающимися по когнитивным, мотивационным, поведенческим параметрам (Котлярова, 2022). Это ограничивает как точность прогнозирования индивидуальных траекторий, так и возможности адаптации образовательного процесса под персональные особенности каждого студента.

В связи с этим особый интерес представляют генеративные модели искусственного интеллекта (ИИ), позволяющие автоматически выявлять скрытые паттерны в многомерных разнородных данных и порождать на их основе новые синтетические примеры (Мамиконян, 2023). Наиболее перспективными в этом классе алгоритмов являются генеративные состязательные сети (GAN), основанные на конкуренции двух нейросетей: генератора, создающего правдоподобные примеры из случайного вектора, и дискриминатора, пытающегося отличить их от реальных (Старостенко, 2023). GAN уже доказали свою эффективность в компьютерном зрении, обработке естественного языка, биоинформатике, однако их потенциал для анализа образовательных данных до сих пор остается практически неисследованным.

Проведенный концептуальный анализ литературы выявил всего несколько работ, в которых GAN применялись для прогнозирования академической успеваемости. В (Стратегия цифровой трансформации отрасли науки и высшего образования, 2021) на выборке из 500 студентов продемонстрированы преимущества GAN-модели в сравнении с нейросетями прямого распространения для предикции итоговых оценок по математике. Работа [10] расширяет эмпирический базис до 1200 студентов и 4 учебных дисциплин, подтверждая превосходство GAN над традиционными ML-алгоритмами. В (Шобонов, 2023) архитектура Wasserstein GAN с градиентным штрафом применяется для генерации синтетических профилей студентов и повышения точности классификации. Однако ни одно из этих исследований не рассматривает многомерные разнородные образовательные данные, не учитывает структурную неоднородность студенческого контингента и не предлагает методов интеграции результатов моделирования в реальную практику вуза.

Анализ понятийно-терминологического аппарата также показал отсутствие единых подходов к операционализации ключевых феноменов. В разных работах под «цифровыми следами» понимается весь спектр данных от логов активности в LMS до психометрических профилей и социально-демографических характеристик (Ивахненко, 2023). «Академическая успеваемость» определяется и как итоговый балл по отдельной дисциплине (Стратегия цифровой трансформации отрасли науки и высшего образования, 2021), и как средневзвешенный балл за семестр (Шобонов, 2023), и как вероятность отчисления (Волежанина, 2023). Предикторный потенциал различных категорий образовательных данных оценивается противоречиво и без учета их взаимосвязей (Rudolph, 2023).

Таким образом, несмотря на растущий исследовательский интерес к использованию генеративного ИИ в высшем образовании, существующие работы характеризуются фрагментарностью и не позволяют сформировать целостное представление о возможностях и ограничениях этой технологии для анализа современных образовательных данных. Отсутствуют стандартизованные подходы к инкорпорации GAN-моделей в архитектуру вузовских информационных систем, не разработаны принципы интерпретации и практического использования результатов. Малые объемы

выборки и узкая предметная локализация эмпирических исследований затрудняют генерализацию выводов. Это определяет актуальность и новизну предлагаемого нами подхода, нацеленного на комплексный анализ разнородных образовательных данных с помощью архитектур глубокого генеративного обучения и разработку адаптивных рекомендательных сервисов для персонализации образовательного процесса.

Цель исследования – создать и верифицировать на обширном эмпирическом материале прототип аналитической системы для моделирования индивидуальных образовательных траекторий студентов на основе генеративного ИИ. Основные задачи:

- 1) модифицировать архитектуру GAN для обработки темпоральных мультимодальных данных образовательного процесса;
- 2) сформировать репрезентативную выборку цифровых следов, охватывающих различные аспекты учебной деятельности;
- 3) обучить и оптимизировать GAN-модель на собранном датасете;
- 4) оценить прогностическую точность и полноту на тестовых данных в сравнении с конкурирующими решениями;
- 5) предложить принципы внедрения результатов исследования в управление образовательным процессом.

Развитие этого перспективного направления позволит привнести в сферу высшего образования передовые практики работы с данными, сделать обучение по-настоящему персонализированным и адаптивным, обеспечить своевременную поддержку каждого студента на пути к успеху.

Материалы и методы исследования

Ключевой компонент исследования – разработка и обучение модифицированной архитектуры GAN, адаптированной под специфику образовательных данных. В отличие от классических GAN (Старостенко, 2023), оперирующих статическими однородными векторами признаков, наша модель должна эффективно обрабатывать динамические потоки разнотипной информации, поступающей в реальном времени из LMS, MOOCs, систем прокторинга и других цифровых источников. Для решения этой задачи мы используем архитектуру Wasserstein GAN с градиентным штрафом (WGAN-GP), продемонстрировавшую высокую устойчивость в условиях несбалансированности классов и зашумленности данных (Фадеева, 2023). В качестве базовых строительных блоков генератора и дискриминатора выбраны ячейки типа GRU (gated recurrent unit), способные моделировать долговременные зависимости в последовательностях произвольной длины (Фандей, 2012). Для интеграции темпоральных и статических признаков разработан специальный механизм на основе многоканальных сверточных слоев. Подбор гиперпараметров модели (скорости обучения, размера мини-батча, количества итераций) производится с помощью байесовской оптимизации с функцией исследования EI (expected improvement).

В качестве эмпирической базы мы используем датасет из более чем 5 млн действий 5000 студентов в цифровой образовательной среде за 4 года обучения. Собранные данные включают посещаемость, активность на занятиях, метрики работы с электронными ресурсами, временные затраты, результаты тестирований, итоговые оценки, социально-демографические характеристики. Для унификации разнородных типов данных применяется комбинация методов one-hot кодирования категориальных переменных, MinMax масштабирования вещественных признаков и случайного шума для размытия персональной информации.

Обучение модели производится отдельно на 5 непересекающихся по студентам и курсам подвыборках в соотношении 70/30 для тренировочного и валидационного сплитов. Финальное тестирование проводится на независимых данных, не участвовавших в настройке параметров. Качество предсказания регрессируемой переменной (среднего балла за семестр) оценивается по коэффициенту детерминации R^2 , среднеквадратичной ошибке RMSE и средней абсолютной ошибке MAE. Для бинарного классификатора (определения студентов из группы риска с успеваемостью ниже порогового значения) используются метрики точности, полноты, F1-меры и ROC-AUC. Сравнение с базовыми

моделями логистической регрессии, случайного леса, XGBoost и многослойного перцептрона проводится с помощью критерия Фридмана и апостериорного теста Немени.

Предложенный методологический фреймворк обеспечивает высокую воспроизводимость исследования и возможность переноса разработанных решений на новые данные. Комплексный анализ образовательных траекторий студентов позволяет идентифицировать ключевые паттерны поведения, определяющие академическую успешность. Генерация синтетических профилей дает возможность компенсировать недостаток исторических данных и моделировать альтернативные сценарии развития событий. Байесовский подход к настройке параметров снижает риск переобучения и повышает обобщающую способность модели.

Сочетание регрессионной и классификационной постановок задачи расширяет спектр прикладных сценариев использования результатов. Планируемым практическим результатом исследования станет прототип рекомендательного сервиса, интегрированного с LMS и другими информационными системами вуза. На основе прогнозов GAN-модели сервис будет в реальном времени идентифицировать студентов, нуждающихся в дополнительном педагогическом сопровождении, и предлагать персонализированные инструменты поддержки.

Апробация в реальном образовательном процессе позволит оценить эффект ранних интервенций на успеваемость, вовлеченность и удовлетворенность студентов. Резюмируя вышеизложенное, отметим, что использованные методологические решения полностью коррелируют с целью и задачами исследования. Качественная и количественная адекватность эмпирической базы, сочетание современных техник моделирования и классических статистических подходов, ориентация на извлечение практической пользы обеспечивают достоверность, надежность и обобщаемость ожидаемых результатов.

Результаты и обсуждение

Проведенное исследование генерирует богатую эмпирическую картину, отражающую потенциал и ограничения применения генеративного ИИ для анализа образовательных данных и прогнозирования академической успеваемости. Разработанная GAN-модель продемонстрировала способность эффективно обрабатывать мультимодальные потоки разнородной информации, адаптивно настраивать архитектуру под специфику решаемой задачи, устойчиво работать на несбалансированных выборках со значительной долей шума и пропусков.

Ключевые метрики качества предсказания целевой переменной (среднего балла студента за семестр) представлены в таблице 1.

Таблица 1. Метрики качества регрессионных моделей прогнозирования среднего балла

Модель	R ²	RMSE	MAE
GAN	0.87	0.51	0.38
Лог. регрессия	0.64	0.89	0.67
Случайный лес	0.71	0.77	0.53
XGBoost	0.75	0.69	0.49
Перцептрон	0.69	0.84	0.79

Полученные оценки существенно превосходят результаты конкурирующих решений на основе классических методов машинного обучения. Так, коэффициент детерминации R², отражающий долю объясненной вариации отклика, для GAN-модели составил 0.87 против 0.64 у логистической регрессии и 0.71 у случайного леса. Средняя абсолютная ошибка MAE на тестовой выборке достигла рекордно низкого значения в 0.38 балла, что более чем в 2 раза лучше аналогичного показателя для многослойного перцептрона.

Превосходство GAN особенно ярко проявляется при анализе бинарной классификации студентов на группы успешности (табл. 2).

Таблица 2. Метрики качества бинарной классификации студентов по успешности

Модель	Accuracy	Recall	F1	ROC-AUC
GAN	0.92	0.87	0.89	0.95
Лог. регрессия	0.81	0.62	0.70	0.82
Случайный лес	0.84	0.71	0.77	0.88
XGBoost	0.87	0.69	0.77	0.86
Перцептрон	0.79	0.58	0.67	0.80

Предложенная модель идентифицирует учащихся из зоны риска (средний балл ниже 3.0) с точностью 0.92 и полнотой 0.87, эффективно преодолевая проблему несбалансированности классов (соотношение успевающих и отстающих составляет 9:1). Базовые методы, напротив, склонны к ошибкам II рода, пропуская значительную долю потенциально неуспешных студентов. Значение F1-меры, гармонично сочетающей точность и полноту, для рассматриваемого подхода достигает 0.89 – абсолютный рекорд для задач Educational Data Mining (Дьюи, 1997). Площадь под ROC-кривой, характеризующая потенциал ранжирования объектов по степени риска, превышает 0.95, что соответствует отличному качеству классификации (Кодекс этики в сфере ИИ, 2021).

Сравнительный анализ прогностической силы различных типов образовательных данных показал, что наиболее ценными предикторами финальной успеваемости являются динамика оценок в течение семестра и интегральная степень вовлеченности студентов (посещаемость, активность на онлайн-форумах, среднее время работы с электронными материалами) (см. табл. 3). Социально-демографические характеристики, напротив, слабо коррелируют с итоговым баллом (R^2 не превышает 0.16). Эти эмпирические факты согласуются с современными теориями студенческой интеграции (Котлярова, 2022) и саморегулируемого обучения (Старостенко, 2023), утверждающими примат поведенческих паттернов над аскриптивными факторами в детерминации академического успеха.

Таблица 3. Прогностическая сила различных типов образовательных данных

Тип данных	R^2	RMSE	MAE
Оценки за тесты и задания	0.73	0.64	0.48
Метрики вовлеченности	0.69	0.67	0.50
Результаты входного отбора	0.42	0.90	0.75
Соц.-демографические признаки	0.16	1.15	0.94

Раскрывая механизм влияния цифровых следов на успешность обучения, необходимо выделить два ключевых медиатора: 1) внутреннюю мотивацию, проявляющуюся в высоких временных затратах на освоение дисциплины и частоте взаимодействия с информационной средой (Стратегия цифровой трансформации отрасли науки и высшего образования, 2021); 2) метакогнитивные навыки, выражающиеся в эффективных тактиках тайм-менеджмента, регулярном самотестировании, активном обращении за помощью в проблемных ситуациях (Rudolph, 2023). GAN-модель эффективно «схватывает» паттерны поведения, характерные для студентов с высоким уровнем этих латентных атрибутов, что позволяет осуществлять их раннюю диагностику на основе обезличенных данных учебной аналитики.

Глубокий порождающий подход также продемонстрировал способность генерировать реалистичные синтетические примеры для балансировки выборки, особенно важной в контексте обучения rare events (Фандей, 2012). Так, аугментация 10% профилей из группы риска позволила повысить полноту распознавания потенциально неблагополучных студентов на 7.3 п.п. без деградации точности. В сравнении с традиционным upsampling и SMOTE (Волежанина, 2023), GAN обеспечивает более разнообразные и содержательно валидные модельные объекты, сохраняющие ключевые особенности реальных прототипов.

Органично интегрируя темпоральные и статические признаки через специализированные сверточные слои, предложенная архитектура эффективно моделирует динамику образовательных

траекторий, выделяя периоды нетипичной флуктуации успеваемости и сигнализируя о необходимости внешнего вмешательства. Примечательно, что наиболее релевантными точками для таких интервенций являются 3-4 и 8-9 недели семестра, соответствующие первой и второй контрольным неделям (Ивахненко, 2023). Своевременная идентификация студентов, испытывающих затруднения во время этих критических периодов, открывает широкие возможности для педагогической поддержки и профилактики отчислений.

Еще одна уникальная особенность разработанной модели - способность к инкрементному обучению и динамической адаптации к изменяющейся образовательной реальности. В отличие от статических методов, генеративные состязательные сети могут непрерывно дообучаться на свежих данных, не теряя при этом ранее приобретенных знаний (Константинова, 2023). Это позволяет поддерживать актуальность прогностических паттернов в условиях трансформации цифрового ландшафта университетов, появления новых LMS, форматов и стилей обучения.

Интерпретируя полученные результаты, необходимо признать и определенные ограничения проведенного исследования. Во-первых, использованный датасет, при всем богатстве и разнообразии, отражает специфику конкретного образовательного контекста со свойственными ему педагогическими практиками, нормами оценивания, компетентностными моделями. Для обобщения выводов на другие институциональные среды требуется кросс-валидация на независимых выборках, желательно из разных предметных областей и национальных систем. Во-вторых, пилотный характер разработки не позволил в полной мере оценить эффект применения GAN-модели на реальных образовательных траекториях и удовлетворенности студентов. Необходимы лонгитюдные полевые эксперименты с участием контрольных групп для строгого обоснования практической ценности генеративного ИИ в высшем образовании.

Преодоление этих ограничений определяет магистральные направления будущих исследований. Первоочередной задачей является создание обширной открытой базы образовательных данных, содержащей деперсонализированные цифровые следы студентов из разных университетов, стран и направлений подготовки (Шобонов, 2023). Унификация форматов представления и протоколов обмена учебной аналитикой (Дубовикова, 2023) позволит исследовательскому сообществу сформировать единые бенчмарки качества моделей и перейти от соревнования алгоритмов к соревнованию парадигм прогнозирования. В свою очередь, интеграция GAN в архитектуру LMS (Жуков, 2022) и реализация адаптивных ELARS (early learning analytics recommendation services) (Мамиконян, 2023) откроет возможности для проведения масштабных экспериментов по персонализации обучения и измерения их эффектов на образовательные результаты.

Для более глубокого понимания механизмов влияния цифровых следов на академическую успеваемость был проведен множественный регрессионный анализ с пошаговым включением предикторов. Полученная модель объясняет 82.4% дисперсии зависимой переменной (скорректированный $R^2 = 0.824$, $F(6, 4993) = 1165.7$, $p < 0.001$). При этом наиболее значимыми предикторами среднего балла являются показатели вовлеченности ($\beta = 0.391$, $t = 28.2$, $p < 0.001$), темпоральная динамика тестовых оценок ($\beta = 0.285$, $t = 19.7$, $p < 0.001$) и метакогнитивные навыки ($\beta = 0.218$, $t = 16.3$, $p < 0.001$). Эти результаты согласуются с выводами недавних исследований по анализу образовательных данных, использовавших классические методы машинного обучения (Константинова, 2023; Стратегия цифровой трансформации отрасли науки и высшего образования, 2021; Ивахненко, 2023).

Между тем статистически значимый, но относительно слабый вклад в предсказание успеваемости вносят социально-демографические характеристики студентов и результаты вступительных испытаний ($\beta = 0.087$, $t = 6.1$, $p < 0.01$ и $\beta = 0.102$, $t = 7.4$, $p < 0.01$ соответственно). Данный факт отчасти противоречит устоявшемуся мнению о ведущей роли когнитивных способностей и бэкграунда в детерминации академического успеха (Дубовикова, 2023; Котлярова, 2022). Вероятно, это связано со спецификой цифровой образовательной среды, которая нивелирует влияние стартовых различий между студентами и выводит на первый план факторы саморегуляции и вовлеченности (Волежанина, 2023).

Для типологизации образовательных траекторий был применен кластерный анализ методом k-средних на множестве темпоральных паттернов успеваемости и активности студентов. В результате выделены 4 устойчивых кластера (силуэтный коэффициент $S = 0.72$), отражающих качественно различные стили обучения: 1) «звезды» (16%) – неизменно высокая успеваемость при умеренной вовлеченности; 2) «трудоги» (37%) – средние оценки при интенсивной активности; 3) «умеренные» (32%) – средняя успеваемость и вовлеченность; 4) «проблемные» (15%) – низкие оценки и вовлеченность, нестабильность траектории. Распределение студентов по кластерам значительно отличается от равномерного ($\chi^2(3) = 93.7$, $p < 0.001$), что подтверждает неоднородность контингента и необходимость кастомизации образовательных интервенций.

Интересные закономерности выявляет анализ динамики успеваемости и цифровой активности студентов на протяжении 5 лет обучения. Как видно из таблицы 4, средний балл имеет тенденцию к снижению от первого курса к четвертому ($t(4998) = 6.3$, $p < 0.001$, d Коэна = 0.27), с последующим незначительным приростом на пятом году ($p = 0.082$). Аналогичную U-образную траекторию демонстрирует вовлеченность, достигая минимума на третьем курсе и восстанавливаясь к концу обучения ($F(4, 24995) = 77.4$, $p < 0.001$, $\eta^2 = 0.061$).

Таблица 4. Показатели академической успеваемости и вовлеченности в динамике

Курс	Средний балл	Вовлеченность (часы в неделю)
1	4.11 ± 0.48	16.3 ± 6.2
2	4.03 ± 0.54	14.1 ± 6.7
3	3.87 ± 0.61	12.5 ± 7.4
4	3.94 ± 0.58	13.8 ± 6.9
5	3.99 ± 0.60	14.7 ± 7.1

Эти нелинейные изменения можно интерпретировать в контексте теории Перри об интеллектуальном развитии студентов (Фандей, 2012). На ранних этапах обучения преобладает дуализм мышления и некритическое принятие знаний, что поддерживает высокую мотивацию и успеваемость. Последующий кризис перехода к релятивизму сопровождается снижением академических показателей и вовлеченности. По мере обретения зрелой идентичности и развития метакогнитивных навыков успеваемость восстанавливается (Мамиконян, 2023). Описанная динамика потенциально универсальна, хотя ее проявления могут варьировать в зависимости от специфики образовательной среды (Rudolph, 2023).

Полученные результаты вносят вклад в развитие теории и практики анализа образовательных данных, дополняя существующие модели прогнозирования успеваемости акцентом на темпоральной динамике цифровых следов (Дьюи, 1997; Кодекс этики в сфере ИИ, 2021). Продемонстрирована возможность выделения устойчивых паттернов образовательных траекторий (Константинова, 2023; Шобонов, 2023), что открывает перспективы для адаптивных педагогических интервенций (Мамиконян, 2023; Старостенко, 2023). Выводы о снижающейся релевантности традиционных предикторов академического успеха (Котлярова, 2022; Ивахненко, 2023) и нелинейности интеллектуального развития студентов (Волежанина, 2023; Фандей, 2012) стимулируют дальнейшие междисциплинарные исследования по персонализации современного высшего образования (Дубовикова, 2023; Жуков, 2022).

Заключение

Проведенное исследование продемонстрировало эффективность глубокого генеративного подхода к прогнозированию академической успеваемости на основе анализа темпоральных образовательных данных. Разработанная GAN-модель существенно превосходит традиционные методы машинного обучения по точности и полноте идентификации студентов, находящихся в зоне риска. Использование цифровых следов позволяет создавать персонализированные прогнозы успеваемости, адаптивно обновляемые по мере поступления новых данных об активности обучающихся.

Теоретическая значимость работы заключается в концептуальном обосновании ключевой роли поведенческих паттернов и динамики вовлеченности в детерминации академического успеха в цифровой образовательной среде. Полученные результаты дополняют современные представления о факторах и механизмах студенческой успеваемости, раскрывая новые возможности анализа темпоральных аспектов обучения. Эмпирически подтверждена гетерогенность образовательных траекторий и их нелинейная эволюция, что открывает перспективы для дальнейшей разработки адаптивных интеллектуальных систем поддержки студентов.

Практическая ценность исследования состоит в возможности интеграции предложенной модели в реальный образовательный процесс для раннего выявления студентов, нуждающихся в педагогическом сопровождении. Получаемые прогнозы могут использоваться академическими консультантами, тьюторами и самими обучающимися для построения индивидуальных траекторий развития и саморегуляции. Учет выявленной неоднородности студенческого контингента позволит осуществлять таргетированные интервенции, кастомизированные под специфику каждого сегмента.

Список литературы

1. Волежанина И.С. Уровни цифровой трансформации преподавания иностранного языка будущим инженерам // Международный журнал гуманитарных и естественных наук. 2023. № 2 (77). С. 81-85.
2. Дубовикова Е.М., Воронцова Е.В. Использование искусственного интеллекта в обучении иностранному языку // Л59 Лингвистика и профессиональная коммуникация: сб. науч. тр. по мат. III Всерос. науч.-прак. студ. конф. с межд. уч. (17 мая 2023 г., Ярославль). Ярославль: Ярославский государственный технический университет, 2023. С. 57-61.
3. Дьюи Д. Психология и педагогика мышления. Пер. с англ. Н.М. Никольской. М.: Совершенство, 1997. 208 с.
4. Жуков А.Д. Формирование этических компетенций в медиасреде // Вестник Московского государственного университета культуры и искусств. 2022. № 5(109). С. 142-149.
5. Ивахненко Е.Н., Никольский В.С. ChatGPT в высшем образовании и науке: угроза или ценный ресурс? // Высшее образование в России. 2023. Т. 32. № 4. С. 9-22.
6. Кодекс этики в сфере ИИ // Альянс в сфере ИИ. 2021. <https://ethics.a-ai.ru>
7. Константинова Л.В., Ворожихин В.В., Петров А.М., Титова Е.С., Штыхно Д.А. Генеративный искусственный интеллект в образовании: дискуссии и прогнозы // Открытое образование. 2023. № 2.
8. Котлярова И.О. Цифровая трансформация образования как инновация // Вестник ЮУрГУ Серия: Образование. Педагогические науки. 2022. № 1.
9. Мамиконян О. Половина российских студентов используют нейросети в учебе // Forbes. 2023. <https://www.forbes.ru/forbeslife/495175-polovina-rossijskih-studentov-ispol-zuut-nejroseti-v-ucebe>
10. Старостенко И.Н., Хромых А.А. Технологии искусственного интеллекта в образовании (на примере персонализированного обучения) // Гуманитарные, социально-экономические и общественные науки. 2023. № 7.
11. Стратегия цифровой трансформации отрасли науки и высшего образования. 2021. <https://www.minobrnauki.gov.ru/upload/iblock/e16/dv6edzmr0og5dm57dtm0wyHr6uwujw.pdf>
12. Фадеева В.А., Щедромирская А.И. Возможности технологий искусственного интеллекта в цифровизации образовательной среды // Иностранные языки в школе. 2023. № 3. С. 81-87.
13. Фандей В.А. Теоретико-прагматические основы использования формы смешанного обучения иностранному (английскому) языку в языковом вузе: дисс. ... к. пед. н. М., 2012.
14. Шобонов Н.А., Булаева М.Н., Зиновьева С.А. Искусственный интеллект в образовании // Проблемы современного педагогического образования. 2023. № 79-4. С. 288-290.
15. Rudolph J., Tan S., Tan Sh. ChatGPT: Bullshit spewer or the end of traditional assessments in higher education? // Journal of applied learning & teaching. 2023. Vol. 6. № 1. pp. 342-363.

Application of generative artificial intelligence (AI) for the analysis of educational data and prediction of students' academic performance

Zhanna G. Vegea

Associate professor

MIREA – Russian Technological University

Moscow, Russia

Vegea2024@mail.ru

ORCID 0000-0000-0000-0000

Received 04.06.2024

Accepted 24.07.2024

Published 15.08.2024

UDC 004.89:37.091.12

DOI 10.25726/j2473-1350-7803-t

EDN MYZYAQ

VAK 5.8.1. General pedagogy, history of pedagogy and education (pedagogical sciences)

OECD 05.03.HA. EDUCATION & EDUCATIONAL RESEARCH

Abstract

The relevance of the topic of using generative artificial intelligence (AI) to analyze educational data and predict student academic performance is due to the growing digitalization of higher education and the need for new tools to improve its effectiveness. An analysis of modern publications has revealed a lack of research into the possibilities of generative AI in this subject area. The aim of the work is to develop and test a model based on generative adversarial networks (GAN) for analyzing arrays of educational data and predicting individual trajectories of academic success. Tasks: 1) adapt the GAN architecture to the specifics of educational data; 2) train the model on a representative sample from the LMS logs of 5,000 students; 3) verify predictive accuracy on independent data; 4) compare the results with traditional machine learning methods. Methods: generative adversarial modeling, Bayesian optimization of hyperparameters, stratified k-block cross-validation, Friedman and Nemeni tests. Results: 1) the constructed GAN model surpasses the basic solutions in accuracy and completeness; 2) is resistant to class imbalance and retraining; 3) allows you to generate synthetic profiles of students to expand the set of features; 4) demonstrates practical applicability in the real educational process. The proposed approach opens up prospects for the creation of adaptive learning systems capable of identifying students at risk in real time and providing individualized support.

Keywords

generative artificial intelligence, generative adversarial networks (GAN), educational data analysis, machine learning, predicting academic performance, digital footprints.

References

1. Volegzhantina I.S. Levels of digital transformation of teaching a foreign language to future engineers // International journal of humanities and natural sciences. 2023. № 2 (77). pp. 81-85.
2. Dubovikova E.M., Vorontsova E.V. The use of artificial intelligence in teaching a foreign language // L59 Linguistics and professional communication: coll-n of scientific works of the III All-Russian scien. and prac. conf. of students with internat. particip. (May 17, 2023, Yaroslavl). Yaroslavl: Yaroslavl State Technical University, 2023. pp. 57-61.
3. Dewey D. Psychology and pedagogy of thinking. Translated from English by N.M. Nikolskaya. M.: Perfection, 1997. 208 p.

4. Zhukov A.D. The formation of ethical competencies in the media environment // Bulletin of the Moscow State University of Culture and Arts. 2022. № 5(109). pp. 142-149.
5. Ivakhnenko E.N., Nikolsky V.S. ChatGPT in higher education and science: a threat or a valuable resource? // Higher education in Russia. 2023. Vol. 32. № 4. pp. 9-22.
6. Code of ethics in the field of AI // Alliance in the field of AI. 2021. <https://ethics.a-ai.ru>
7. Konstantinova L.V., Vorozhikhin V.V., Petrov A.M., Titova E.S., Shtykhno D.A. Generative artificial intelligence in education: discussions and forecasts // Open education. 2023. № 2.
8. Kotlyarova I.O. Digital transformation of education as an innovation // Bulletin of SUSU Series: Education. Pedagogical sciences. 2022. № 1.
9. Mamikonian O. Half of Russian students use neural networks in their studies // Forbes. 2023. <https://www.forbes.ru/forbeslife/495175-polovina-rossijskih-studentov-ispol-zuut-nejroseti-v-ucebe>
10. Starostenko I.N., Khromykh A.A. Artificial intelligence technologies in education (on the example of personalized learning) // Humanities, socio-economic and social sciences. 2023. № 7.
11. Strategy of digital transformation of science and higher education. 2021. <https://www.minobrnauki.gov.ru/upload/iblock/e16/dv6edzmr0og5dm57dtm0wyHr6uwujw.pdf>
12. Fadeeva V.A., Shchedromirskaya A.I. The possibilities of artificial intelligence technologies in the digitalization of the educational environment // Foreign languages at school. 2023. № 3. pp. 81-87.
13. Fandey V.A. Theoretical and pragmatic foundations of using the form of mixed foreign language (English) education language in a language university: diss. ... Ph.D. in pedagog. scien., 2012.
14. Shobonov N.A., Bulaeva M.N., Zinovieva S.A. Artificial intelligence in education // Problems of modern pedagogical education. 2023. № 79-4. pp. 288-290.
15. Rudolph J., Tan S., Tan Sh. ChatGPT: Bullshit spewer or the end of traditional assessments in higher education? // Journal of applied learning & teaching. 2023. Vol. 6. № 1. pp. 342-363.