

ПРИМЕНЕНИЕ АЛГОРИТМОВ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ПРОГНОЗИРОВАНИЯ АКАДЕМИЧЕСКОЙ УСПЕВАЕМОСТИ КУРСАНТОВ

Мороговский А.А., Степанец Е.В.

Учреждение образования «Белорусский государственный университет информатики и радиоэлектроники», г. Минск, Республика Беларусь, alexworld121@mail.ru, e.stepanets@bsuir.by

Abstract. The modern development of military education requires transitioning from reactive to proactive management of academic performance through early identification of cadets at risk of academic difficulties. This is particularly important in military institutions, where cadets face heavy workloads and limited time for remediation, while traditional assessment systems often reveal problems too late for effective intervention. The substantial volume of digital data generated during training—electronic grade books, test results, and activity logs—can serve as a source of features for predicting future performance. Machine learning methods, particularly ensemble algorithms such as gradient boosting over decision trees (e.g., XGBoost), are highly effective for this task due to their ability to handle nonlinear dependencies and heterogeneous features while maintaining interpretability for educational decision-making.

Современный этап развития военного образования характеризуется расширением применения цифровых технологий для управления образовательным процессом. Одним из перспективных направлений является использование методов машинного обучения для анализа учебной деятельности и раннего выявления курсантов, испытывающих затруднения в освоении учебных дисциплин [1]. Актуальность данного направления обусловлена необходимостью перехода от преимущественно реактивного управления успеваемостью, при котором проблемы фиксируются после получения неудовлетворительных результатов контроля, к проактивному управлению, основанному на прогнозировании возможных негативных результатов и своевременной корректировке образовательной траектории.

Особую значимость данная задача приобретает в условиях военного вуза. Учебная деятельность курсантов осуществляется в условиях жестко регламентированного распорядка дня, значительной учебной и служебной нагрузки и ограниченного времени, которое может быть использовано для самостоятельного восполнения пробелов в знаниях. Вследствие этого недостаточное освоение базовых положений математических, физических и технических дисциплин на начальных этапах образования способно последовательно приводить к накоплению пробелов и снижению результатов последующего обучения. Наиболее критичной данная проблема является для дисциплин, содержание которых имеет выраженную последовательную структуру: усвоение последующих разделов непосредственно зависит от освоения предшествующего материала.

Традиционная система контроля преимущественно ориентирована на фиксацию уже достигнутого результата. Итоговые и промежуточные формы контроля позволяют определить наличие академической задолженности, однако информация о ней зачастую становится доступной преподавателю в момент, когда для полноценной коррекции образовательной траектории остается ограниченный временной ресурс. Дополнительным фактором является невозможность постоянного детального анализа индивидуальной динамики каждого обучаемого преподавателем при работе с учебной группой численностью 25-30 человек. Опыт преподавателя позволяет выявлять наиболее очевидные признаки от-

ставания, однако субъективная оценка не может полностью заменить количественный анализ накопленной информации об учебной деятельности.

В связи с этим возникает противоречие между значительным объемом цифровых данных, формируемых в процессе обучения, и ограниченными возможностями их оперативной интерпретации человеком. Электронные журналы, системы дистанционного обучения, результаты тестирования, количество попыток выполнения заданий и их временные характеристики формируют цифровой след образовательной деятельности. При соответствующей организации хранения и обработки эти данные могут рассматриваться не только как средство фиксации результатов, но и как источник признаков для прогнозирования будущей успеваемости [2].

В научной литературе прогнозирование академической успеваемости рассматривается как одно из устойчивых направлений применения образовательной аналитики и машинного обучения. Систематические обзоры показывают, что для решения подобных задач применяются методы опорных векторов, нейронные сети и ансамблевые алгоритмы. При этом ансамблевые методы демонстрируют высокую эффективность на задачах прогнозирования с разнородными признаками [3]. Современный обзор исследований 2026 года также отмечает доминирование деревьев решений и нейросетевых моделей в задачах прогнозирования успеваемости, одновременно подчеркивая значение интерпретируемости моделей при использовании результатов прогнозирования в образовательных решениях [1].

Перспективным вариантом для рассматриваемой задачи является применение алгоритмов поэтапного улучшения над решающими деревьями, в частности XGBoost или аналогичных методов. Преимущество данного подхода заключается в способности обрабатывать нелинейные зависимости между признаками, взаимодействии различных факторов и неоднородные массивы учебных данных. В отличие от сложных нейросетевых архитектур, результаты работы моделей на основе деревьев могут быть дополнительно интерпретированы с использованием методов оценки значимости признаков и средств объяснимого машинного обучения. Прогноз модели должен использоваться не как автоматическое основание для принятия кадрового или дисциплинарного

решения, а как аналитический инструмент поддержки деятельности преподавателя и командования [4].

Алгоритм XGBoost представляет собой масштабируемую реализацию поэтапного улучшения, ориентированную на построение ансамбля деревьев решений. Авторами метода показана его эффективность при обработке больших и разреженных массивов данных, а также возможность масштабирования вычислений [5]. Для задачи прогнозирования академической успеваемости важна не только точность классификации, но и возможность получить оценку вероятности принадлежности обучаемого к группе повышенного риска.

Предлагаемая информационно-аналитическая система может функционировать по принципу цифрового «навигатора» академической успеваемости. На первом этапе формируется исторический массив данных за несколько учебных лет. В качестве исходных признаков целесообразно рассматривать результаты входного контроля, текущие оценки по дисциплинам, результаты модульных и тематических проверок, посещаемость, факт и своевременность выполнения заданий, количество попыток выполнения контрольных мероприятий, а также показатели работы в электронной образовательной среде. При этом персональные данные должны обрабатываться в соответствии с установленными требованиями информационной безопасности и действующего законодательства [1; 3].

Особое значение имеет не только абсолютное значение оценки, но и динамика изменения учебных показателей. Например, устойчивое снижение результатов тестирования в течение нескольких последовательных контрольных периодов может являться более информативным признаком, чем единичная неудовлетворительная оценка. Аналогично увеличение времени решения типовых задач при неизменной сложности заданий может свидетельствовать о снижении степени освоения учебного материала. Для технических дисциплин, включая радиофизику, схемотехнику и радиоэлектронику, целесообразно исследовать взаимосвязь между скоростью выполнения типовых заданий, количеством ошибок и последующими результатами контрольных мероприятий. Следовательно, одним из направлений исследования может стать выделение так называемых маркеров риска. В качестве таких признаков могут рассматриваться снижение средней результативности выполнения заданий за заданный временной интервал, увеличение количества пересдач, систематическое невыполнение небольших учебных заданий, снижение активности в электронной образовательной среде, ухудшение результатов входного или промежуточного контроля. Конкретные пороговые значения, например, снижение скорости решения задач на 30 % за две недели, не должны задаваться априорно. Они должны определяться экспериментально на основании статистического анализа реальных данных и проверяться на независимой выборке [2].

Принципиальная схема функционирования предлагаемой системы может быть представлена

следующим образом: сбор и предварительная обработка учебных данных → формирование набора признаков → обучение прогнозной модели на исторических данных → валидация модели → расчет индивидуальной вероятности академического риска → визуализация результатов → педагогическая коррекция → повторный анализ динамики. Подобный подход соответствует современной концепции обучаемой аналитики, в рамках которой выделяются описательная, прогнозная, диагностическая и предписывающая аналитика. В документации различных аналитических компаний также предусмотрена архитектура, в которой модели используют индикаторы и целевые показатели, формируют прогнозы, уведомления и действия по результатам аналитики [2].

На этапе практического внедрения целесообразно сформировать несколько уровней риска. Например, низкий уровень может соответствовать устойчивой положительной динамике и отсутствию признаков академических затруднений; средний уровень - появлению отдельных негативных тенденций, требующих педагогического внимания; высокий уровень - совокупности признаков, статистически связанных с высокой вероятностью неудовлетворительного результата. Цветовая индикация «зеленая - желтая - красная» может использоваться исключительно как средство визуализации и не должна интерпретироваться как окончательная оценка способностей курсанта.

Ключевым элементом предлагаемой системы является автоматизация не только прогнозирования, но и последующей педагогической реакции. Если модель выявляет повышенный риск по конкретной дисциплине, информационная система может сформировать индивидуальное задание для самостоятельной подготовки. Например, при выявлении недостаточного освоения раздела по радиоэлектронике курсанту могут назначаться дополнительные задачи в виртуальной лаборатории. Сложность заданий при этом может изменяться в зависимости от результатов их выполнения: при успешном освоении базового уровня система переводит обучаемого к более сложным задачам, а при сохранении ошибок возвращает его к необходимому элементу подготовки. Такой подход позволяет использовать машинное обучение не только для констатации риска, но и для формирования адаптивной образовательной траектории [1; 4].

При этом использование машинного обучения требует решения задачи интерпретируемости. Высокая точность модели сама по себе недостаточна, если преподаватель не понимает причины формирования высокого значения риска. Поэтому целесообразно применять методы объяснимого искусственного интеллекта, позволяющие определить вклад отдельных признаков в прогноз. Современные исследования в области прогнозирования академической успеваемости подчеркивают необходимость развития ХАИ-подходов для превращения прогнозной модели в практически применимый инструмент педагогического управления [4].

Для оценки эффективности предлагаемого подхода необходимо провести экспериментальное исследование. Исторические данные могут быть разделены на обучающую, валидационную и тестовую выборки с учетом временной структуры образовательного процесса. При этом принципиально важно исключить попадание в признаки информации, которая становится известной только после наступления прогнозируемого события. В противном случае возникает утечка данных, приводящая к завышению оценки качества модели. Ожидаемым результатом внедрения системы является не автоматическое принятие решения об отчислении или переводе курсанта, а увеличение временного интервала между моментом выявления негативной тенденции и итоговым контролем [1;4]. Если традиционная система фиксирует проблему преимущественно по результатам контрольного мероприятия, прогнозная система потенциально способна выявлять изменение учебной динамики за несколько недель до экзамена. Именно увеличение времени на педагогическую коррекцию представляет основной практический эффект предлагаемой технологии.

При этом количественные показатели, например, снижение количества отчислений на 15-20 %, на первоначальном этапе должны рассматриваться исключительно как целевой ориентир экспериментального исследования, а не как доказанный эффект. Для подтверждения такого результата потребуются сопоставить контрольную и экспериментальную группы, определить исходный уровень академической неуспеваемости и исключить влияние сторонних факторов. Научно обоснованным результатом будет считаться только тот эффект, который подтвержден статистически на независимой выборке.

Следует учитывать и ограничения предлагаемого подхода. Машинная модель обучается на исторических данных, поэтому ошибки и систематические искажения исходной образовательной практики способны переноситься в прогноз. Кроме того, корреляционная связь между определенным учебным поведением и академическим результатом не означает наличия причинно-следственной зависимости [3; 4]. По этой причине прогноз не должен использоваться как самостоятельное основание для дисциплинарного воздействия. Его назначение заключается в своевременном обнаружении признаков, требующих дополнительного педагогического анализа.

В результате применение алгоритмов машинного обучения для прогнозирования академической успеваемости курсантов может рассматриваться как перспективное направление цифровой трансформации управления образовательным процессом в военном вузе. Использование поэтапного улучшения, анализа цифрового следа и методов объяснимого искусственного интеллекта позволяет сформировать инструмент поддержки решений, сохраняющий ведущую роль преподавателя и командования.

Главным практическим результатом такого подхода является не замена преподавателя алгоритмом, а расширение возможностей педагогического наблюдения. При наличии учебной группы значи-

тельной численности цифровая модель способна выполнять непрерывный количественный анализ тех показателей, которые невозможно полноценно оценивать вручную каждый день. Это создает предпосылки для своевременной индивидуальной коррекции обучения, рационального распределения педагогических ресурсов и повышения устойчивости образовательного процесса. Для вуза такая система может стать элементом единого цифрового контура управления подготовкой специалистов, в котором данные текущей учебной деятельности используются не только для отчетности, но и для прогнозирования и предупреждения негативных результатов.

Таким образом, ключевым инструментом смены реактивного подхода на проактивную стратегию, к контролю успеваемости, выступает интеграция технологий машинного обучения, в частности алгоритма XGBoost, направленного на заблаговременное выявление академических рисков. Цифровые следы образовательного процесса обладают высоким предиктивным потенциалом, но их анализ классическими методами затруднен из-за дефицита времени и высокой интенсивности потоков данных. Разрабатываемая информационно-аналитическая платформа позиционируется не как система автопринятия решений, а как механизм педагогической поддержки, расширяющий временные возможности для корректировки учебного процесса и его индивидуализации. Результативность данного метода нуждается в эмпирической проверке на независимых данных, а прогнозы не должны вытеснять профессиональную интуицию педагога. Предлагаемая методика расширяет инструментарий аналитических систем в военных учебных заведениях, обеспечивая сбалансированное сочетание цифровых технологий, педагогической обоснованности и этического контроля над применением инфокоммуникационных ресурсов.

Литература

1. Application of machine learning in higher education to predict students' performance, learning engagement and self-efficacy: a systematic literature review // *Asian Education and Development Studies*. 2025. Vol. 14, No. 2. P. 205–240. DOI: 10.1108/AEDS-08-2024-0166.
2. Moodle Documentation. Analytics. Официальная документация Moodle.
3. A systematic review of the literature on machine learning application of determining the attributes influencing academic performance // *Decision Analytics Journal*. 2023. Vol. 7. 100204. DOI: 10.1016/j.dajour.2023.100204.
4. Predicting student performance: A comprehensive review of machine learning, deep learning, and explainable AI approaches // *Computers and Education: Artificial Intelligence*. 2026. Vol. 10. 100548. DOI: 10.1016/j.caeai.2026.100548.
5. Chen T., Guestrin C. XGBoost: A Scalable Tree Boosting System // *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2016. P. 785–794. DOI: