

ПОДСИСТЕМА ДЛЯ АНАЛИТИКИ ПРОГРЕССА ОБУЧЕНИЯ С ПРЕДИКТИВНЫМИ МОДЕЛЯМИ ОТСЕВА НЕУСПЕВАЮЩИХ СТУДЕНТОВ

Вальц М.Э. (Университет ИТМО)

Научный руководитель – Кулинич Я.В.

(Факультет программной инженерии и компьютерной техники, Университет ИТМО)

Введение. Рынок образовательных технологий (EdTech), достигнувший объема около \$185 млрд в глобальном масштабе, характеризуется смещением фокуса с экстенсивного роста на эффективность и удержание клиентов [1]. Одной из главных проблем онлайн-образования остаются высокие показатели отсева и академической неуспеваемости, обусловленные, в частности, изолированным характером обучения и отсутствием своевременной поддержки [2]. В ответ на эти вызовы все большее распространение получают методы машинного обучения и предиктивной аналитики для идентификации студентов группы риска. Однако, несмотря на успехи в области предиктивной аналитики, остаются проблемы репродуцируемости, обобщаемости и интерпретируемости существующих моделей.

В то же время цифровые образовательные платформы функционируют в условиях неоднородной нагрузки, характеризующейся резкими и трудно прогнозируемыми всплесками активности (например, в периоды сдачи заданий, прохождения тестирований или завершения учебных модулей). Это создает дополнительные требования к архитектуре аналитических подсистем, включая масштабируемость, устойчивость к пиковым нагрузкам и корректную обработку больших объемов событий цифрового следа обучающихся в реальном времени или с минимальной задержкой.

Таким образом, актуальность настоящей работы обусловлена необходимостью создания комплексной аналитической подсистемы, интегрируемой в платформу дистанционного обучения, которая позволит не только прогнозировать риски отсева студентов на основе их цифрового следа, но и обеспечивать интерпретируемость результатов и устойчивость работы в условиях неоднородной и пиковых нагрузок для принятия своевременных педагогических решений.

Основная часть. В данной работе была получена формула для раннего прогнозирования неуспевающих студентов на основе их активности на образовательной платформе. Данная формула позволяет не только обнаружить потенциально неуспевающего студента на ранних этапах, но и продемонстрировать, на основе каких данных было сделано это предсказание.

Также для формулы на основе различных исследований и интернет источников был выбран набор необходимых для предсказания метрик, которые собираются в LMS на основе цифрового следа обучающихся.

Для хранения и анализа метрик аналитиками в рамках работы была реализована система DWH (Data Warehouse) для хранения и подготовки данных для бизнес аналитики. Так как нагрузка на цифровые образовательные системы является неоднородной, были проведены нагрузочные и стресс тесты, которые продемонстрировали возможность системы выдерживать неоднородную высокую нагрузку с непредсказуемыми всплесками.

Поток работы системы состоит из следующих этапов:

1. Собранные метрики сохраняются в самой образовательной системе.
2. На основе полученных метрик для обучающихся по упомянутой выше формуле вычисляется их принадлежность к группе потенциально неуспевающих студентов.

3. Полученные на 1-ом этапе метрики в изначальном виде попадают в систему DWH.
4. Происходит обработка полученных метрик: удаление невалидных метрик, нормализация, трансформация.
5. На основе обработанных данных строятся витрины данных для аналитиков системы.

Выводы. Выводы. В данной работе разработана подсистема аналитики образовательной платформы, предназначенная для раннего выявления студентов группы риска на основе их цифрового следа. Полученная формула прогнозирования позволяет не только обнаруживать потенциальную академическую неуспеваемость на ранних этапах обучения, но и обеспечивать интерпретируемость результатов за счет явной связи предсказания с используемыми метриками активности.

Для поддержки аналитических процессов была спроектирована и реализована система хранения и подготовки данных (DWH), обеспечивающая сбор, очистку, нормализацию и трансформацию образовательных метрик с последующим формированием витрин данных для аналитиков. Проведенные нагрузочные и стресс-тесты подтвердили устойчивость архитектуры к неоднородной и пиковым нагрузкам, характерным для цифровых образовательных систем.

Разработанное решение демонстрирует возможность интеграции предиктивной аналитики в образовательную платформу без нарушения ее основной функциональности и создает основу для принятия своевременных педагогических решений, направленных на снижение академической неуспеваемости и повышение эффективности обучения.

Список использованных источников:

1. Фокин А. Эффективность вместо гонки за ростом: как изменился рынок EdTech в 2025-м // РБК Тренды: [Электронный ресурс] – 2025. – URL: <https://trends.rbc.ru/trends/education/6952218e9a79473ffc6f6b06> (дата обращения: 18.02.2026)
2. Минобрнауки назвало количество отчисленных из вузов студентов в 2025 году // РИА Новости: [Электронный ресурс] - 2025. - URL: <https://ria.ru/20251106/rossiya-2053262266.html> (дата обращения: 18.02.2026)

Вальц М.Э. (автор) Подпись

Кулинич Я.В. (научный руководитель) Подпись