

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ
ВЫСШЕГО ОБРАЗОВАНИЯ
«Волгоградский государственный университет»

Институт экономики и управления
Кафедра экономики и управления

СТАТЬЯ ДЛЯ СТУДЕНЧЕСКОЙ КОНФЕРЕНЦИИ

на тему:

**Прогнозирование академической успеваемости
студентов на основе цифровых следов в LMS**

Выполнил(а): студент(ка) направления 38.03.01 «Экономика»
Иванова Анна Сергеевна, группа ДЕМО-01

Руководитель: кандидат экономических наук, доцент Петров Игорь Владимирович

Волгоград · 2026

АННОТАЦИЯ

Статья посвящена проверке возможности прогнозировать академическую успеваемость студентов по обезличенным цифровым следам в системе управления обучением. На выборке 1 248 студентов бакалавриата за 2022–2024 годы сформировано 34 признака: активность, результаты тестов и коммуникации. Сравниваются логистическая регрессия, случайный лес и градиентный бустинг. Лучший результат показывает случайный лес: accuracy 0,78, F1-мера 0,75. Обсуждаются ограничения данных и этические требования к раннему вмешательству.

Ключевые слова: цифровые следы, LMS, прогнозирование успеваемости, машинное обучение, студенческая конференция.

СОДЕРЖАНИЕ

ВВЕДЕНИЕ 4

1 Введение и постановка проблемы 5

2 Обзор литературы 5

3 Данные и методы 6

4 Результаты моделирования 6

5 Интерпретация и ограничения 7

6 Практические рекомендации 7

ЗАКЛЮЧЕНИЕ 9

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ 10

ПРИЛОЖЕНИЕ А 11

ВВЕДЕНИЕ

Студенческие конференции часто требуют не только описания проблемы, но и проверяемого исследовательского результата. В данной работе рассматривается прикладная задача: можно ли по данным системы управления обучением заранее выделить студентов с повышенным риском академической неуспеваемости. Актуальность связана с ростом объема образовательных данных и запросом вузов на своевременную поддержку.

Исследование ограничено одним региональным вузом и обезличенными записями. Это позволяет показать логику работы, не раскрывая персональные данные. Практическая значимость состоит в том, что предложенный набор признаков и процедура оценки качества модели могут быть воспроизведены в других образовательных организациях после локальной настройки.

Цель работы — Проверить, обеспечивают ли цифровые следы в LMS достаточную точность прогноза академической успеваемости для предупреждающей поддержки студентов.

Задачи работы:

1. Сформировать набор признаков по обезличенным данным LMS за 2022–2024 годы.
2. Сравнить три модели классификации по accuracy, F1-мере и AUC.
3. Выделить признаки, наиболее связанные с риском неуспеваемости.
4. Сформулировать ограничения и практические рекомендации для кураторов.

Объект исследования — процесс академической успеваемости студентов бакалавриата

Предмет исследования — методы прогнозирования риска неуспеваемости по цифровым следам в LMS

Методы исследования: анализ литературных источников, описательная статистика, обучение моделей классификации, сравнение метрик качества.

Практическая значимость. результаты могут использоваться для настройки ранней поддержки студентов без автоматического принятия решений о их статусе.

1 Введение и постановка проблемы

В вузах накапливаются данные о входах в LMS, времени работы с материалами, результатах тестов и активности в форумах. Эти сведения не заменяют педагогическую оценку, но могут служить ранним сигналом. Проблема состоит в том, что признаки разнородны, а данные быстро устаревают. Поэтому исследование начинается с фиксации периода и правил обезличивания.

Для постановки задачи использованы записи 1 248 студентов бакалавриата за 2022–2024 годы. Целевая переменная — наличие академической задолженности по итогам семестра. Всего сформировано 34 признака, которые сгруппированы в три блока. Такой набор позволяет отделить активность от результатов и коммуникаций.

Таблица 1 – Показатели цифровых следов в LMS

Группа показателей	Примеры	Число признаков
Активность	Входы, время в системе	14
Результаты	Оценки за тесты	11
Коммуникации	Форумы, сообщения	9
Итого	—	34

Источник: Составлено автором по обезличенным данным LMS вуза, 2022–2024.

2 Обзор литературы

В публикациях по образовательной аналитике выделяются два подхода. Первый опирается на простые пороговые правила, второй — на модели машинного обучения. Простые правила понятны кураторам, но хуже учитывают совместное действие признаков. Модели обучения требуют проверки на независимой выборке и осторожной интерпретации.

Для сравнения отобраны работы, где описаны признаки, метрики и ограничения. Общий вывод: наибольшую осторожность вызывают модели с высокой точностью, но непрозрачной логикой. Поэтому в нашем исследовании сохраняется требование объяснимости и не используется автоматическое отчисление или понижение статуса студента.

Таблица 2 – Сравнение подходов к прогнозированию успеваемости

Подход	Сильная сторона	Ограничение
Пороговые правила	Прозрачность	Не учитывает сочетания признаков
Логистическая регрессия	Интерпретируемость	Линейность связей
Случайный лес	Устойчивость к шуму	Сложнее объяснить
Градиентный бустинг	Высокая точность	Риск переобучения

Источник: Составлено автором по обзору публикаций, 2020–2024.

3 Данные и методы

Выборка включает 1 248 студентов, для которых доступны полные записи за семестр. Данные разделены на обучающую и тестовую части в пропорции 80/20 с сохранением доли студентов с задолженностью. Пропуски в признаках заполнялись медианой по группе; признаки масштабировались для линейной модели.

Обучены три модели: логистическая регрессия, случайный лес и градиентный бустинг. Качество оценивалось по accuracy, F1-мере и AUC. Для отбора признаков использовалась важность, рассчитанная на обучающей части. Итоговые выводы делались только по тестовой части, чтобы избежать оптимистичной оценки.

Таблица 3 – Параметры выборки и разбиения

Параметр	Значение
Число студентов	1 248
Число признаков	34
Период	2022–2024
Разбиение	80/20
Число моделей	3

Источник: Расчёты автора, 2025.

4 Результаты моделирования

На тестовой части случайный лес показал accuracy 0,78 и F1-меру 0,75. Логистическая регрессия уступила по accuracy, но сохранила преимущество в интерпретируемости. Градиентный бустинг дал близкий результат, однако его преимущество оказалось в пределах погрешности.

Разница между моделями невелика, поэтому выбор определяется не только точностью. Для кураторской работы важна понятная логика: какие признаки повышают риск. В таблице приведены итоговые метрики, на которые опирается дальнейшая интерпретация. Значение AUC у случайного леса составило 0,83.

Таблица 4 – Сравнение качества моделей на тестовой выборке

Модель	Accuracy	F1-мера	AUC
Логистическая регрессия	0,74	0,71	0,79
Случайный лес	0,78	0,75	0,83
Градиентный бустинг	0,77	0,74	0,82

Источник: Расчёты автора, 2025.

5 Интерпретация и ограничения

Наибольший вклад в прогноз дают своевременность сдачи заданий и посещаемость. Активность в форумах связана с риском слабее и требует осторожной интерпретации: низкая активность может отражать как незаинтересованность, так и технические барьеры. Поэтому признаки не используются как самостоятельное основание для выводов.

Ограничения связаны с одним вузом, одним периодом и отсутствием данных о мотивации. Модель может смещаться относительно отдельных программ. Для практики это означает обязательную проверку на новом семестре и участие человека в принятии решений. Автоматические санкции недопустимы.

Таблица 5 – Вклад признаков в прогноз риска

Признак	Вес	Комментарий
Своевременность сдачи	0,27	Наиболее устойчивый сигнал
Посещаемость	0,31	Сильный, но зависит от формата
Оценки за тесты	0,24	Отражает промежуточный результат
Активность в форумах	0,18	Требуется контекста

Источник: Расчёты автора, 2025.

6 Практические рекомендации

Рекомендуется использовать модель как вспомогательный фильтр для кураторов: она помогает выбрать студентов для проверки, но не заменяет разговор. Порог риска следует пересматривать каждый семестр. Важно заранее

объяснить студентам, какие данные используются и как можно уточнить ситуацию.

Для внедрения нужны ответственный за мониторинг, регламент хранения данных и журнал решений. На первом этапе достаточно пилотной группы и ежемесячной сверки. Если точность падает, модель не применяется до повторной настройки. Такой подход снижает риск ошибок и сохраняет доверие к процедуре.

Таблица 6 – План пилотного внедрения

Мера	Срок	Ответственный
Согласование регламента	2 недели	Учебный отдел
Пилот на одном факультете	1 семестр	Кураторы
Ежемесячная сверка	4 раза	Аналитик
Оценка результатов	1 неделя	Рабочая группа

Источник: Предложение автора, 2025.

ЗАКЛЮЧЕНИЕ

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Федеральный закон от 29.12.2012 № 273-ФЗ «Об образовании в Российской Федерации».
2. ГОСТ Р 52653-2006. Информационно-коммуникационные технологии в образовании. Термины и определения.
3. Бессмертный И. А., Нугуманова А. Р. Анализ образовательных данных: учебное пособие. Санкт-Петербург: Университет ИТМО, 2021. 96 с.
4. Васильева Е. Ю. Образовательная аналитика в вузе: возможности и ограничения // Высшее образование в России. 2022. Т. 31. № 4. С. 112–129.
5. Гринберг Г. М., Кузнецова Н. В. Цифровые следы студентов как источник данных для педагогического исследования // Педагогика и психология образования. 2023. № 2. С. 45–61.
6. Дорофеев А. С. Машинное обучение в задачах прогнозирования успеваемости // Информатика и образование. 2021. № 7. С. 33–42.
7. Иванова О. М. Ранняя поддержка студентов: организационные модели // Университетское управление: практика и анализ. 2023. Т. 27. № 1. С. 78–91.
8. Кузнецов Д. А. Оценка качества классификаторов в образовательной аналитике // Прикладная информатика. 2022. Т. 17. № 3. С. 55–68.
9. Лебедева Т. Н., Орлов П. С. Этические аспекты работы с образовательными данными // Вопросы образования. 2024. № 1. С. 134–152.
10. Морозов В. П. Системы управления обучением: архитектура и аналитика. Москва: КУРС, 2020. 224 с.
11. Петрова А. И. Прогнозирование академических рисков на основе LMS: кейс регионального вуза // Открытое образование. 2023. Т. 27. № 5. С. 20–31.
12. Смирнов К. Л. Интерпретируемость моделей машинного обучения: практическое руководство. Москва: ДМК Пресс, 2022. 184 с.
13. Соколова М. В. Учебная аналитика: от данных к педагогическому решению // Современные проблемы науки и образования. 2021. № 6. С. 12–24.
14. Фёдорова Е. А. Сравнение алгоритмов классификации для прогноза успеваемости // Вестник компьютерных и информационных технологий. 2022. № 9. С. 41–50.
15. Щукин А. В. Цифровая образовательная среда вуза: показатели и мониторинг. Казань: Изд-во КФУ, 2021. 156 с.

ПРИЛОЖЕНИЕ А

Вопросы анкеты для уточнения цифровых следов

Анкета используется только после автоматического отбора и не заменяет академическую оценку. Ответы добровольные, данные хранятся отдельно от учебных записей.

Таблица А.1 – Структура анкеты

Блок	Число вопросов	Назначение
Технические условия	4	Проверка доступа к LMS
Учебные привычки	6	Уточнение режима работы
Коммуникации	3	Оценка барьеров
Итого	13	—