

DOI: 10.32517/0234-0453-2024-39-3-16-28

# ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ ОБРАЗОВАТЕЛЬНЫХ РЕЗУЛЬТАТОВ СТУДЕНТОВ УНИВЕРСИТЕТА С ИСПОЛЬЗОВАНИЕМ LMS MOODLE

Е. В. Карманова<sup>1</sup> ✉, Я. М. Захарова<sup>1</sup>, А. В. Киселев<sup>1</sup><sup>1</sup> *Магнитогорский государственный технический университет им. Г. И. Носова, г. Магнитогорск, Челябинская область, Россия*✉ [ev.karmanova@magtu.ru](mailto:ev.karmanova@magtu.ru)

## Аннотация

Интеллектуальный анализ образовательных данных представляет собой технологию обработки больших объемов информации, собранной с помощью образовательных порталов, для оценки качества учебного контента и прогнозирования образовательных результатов. При использовании данной технологии применяются различные методы анализа данных: поиск аномалий и ассоциаций, классификация, кластерный, регрессионный и факторный анализ. Каждый из этих методов использует специальные инструменты, такие как электронные таблицы, высокоуровневые языки программирования (Python или R), аналитические low-code платформы (Loginom, Orange) или специализированные модули аналитики, внедренные в образовательные порталы.

В статье раскрываются возможности модуля «Аналитика», встроенного в LMS Moodle, для отслеживания и оптимизации образовательного процесса. Авторы дают рекомендации по настройке и использованию этого модуля в учебном процессе. Модуль «Аналитика» позволяет без навыков программирования создавать прогнозы успешности освоения отдельного учебного курса, не требует предобработки данных для анализа и дает возможность быстро принимать меры по совершенствованию учебного процесса на основе результатов прогноза. Однако модуль имеет и ряд недостатков (ограниченность признаков и моделей прогнозирования, высокие требования к вычислительным мощностям для запуска обучения моделей и построения прогнозов), поэтому администраторы образовательных порталов часто отключают его. В связи с этим был проведен эксперимент по обучению модели классификации для предсказания экзаменационных оценок студентов на основе деятельности обучающихся на образовательном портале МГТУ им. Г. И. Носова.

В ходе эксперимента были взяты данные об активности студентов на образовательном портале университета за один учебный год, построены модели машинного обучения для предсказания экзаменационной оценки по отдельному студенту и предмету. Результаты эксперимента показали, что точность предсказания оценки студента на промежуточной аттестации (по метрике F1) составляет 74 %. Это подтверждает эффективность использования учебной аналитики на основе данных пользователей образовательных платформ.

**Ключевые слова:** образовательный портал, учебная аналитика, модель машинного обучения, прогнозирование образовательных результатов, LMS Moodle.

## Для цитирования:

Карманова Е. В., Захарова Я. М., Киселев А. В. Интеллектуальный анализ образовательных результатов студентов университета с использованием LMS Moodle. *Информатика и образование*. 2024;39(3):16–28. DOI: 10.32517/0234-0453-2024-39-3-16-28.

## EDUCATIONAL DATA MINING IN A UNIVERSITY USING LMS MOODLE

E. V. Karmanova<sup>1</sup> ✉, Ya. M. Zakharova<sup>1</sup>, A. V. Kiselev<sup>1</sup><sup>1</sup> *Nosov Magnitogorsk State Technical University, Magnitogorsk, Chelyabinsk Region, Russia*✉ [ev.karmanova@magtu.ru](mailto:ev.karmanova@magtu.ru)

## Abstract

Educational data mining is a technology for processing big data collected through educational portals and used for assessing the quality of educational content and predicting educational outcomes. Various data analysis methods are applied in the implementation of this technology, including anomaly detection, association rule mining, classification, cluster analysis, regression analysis, and factor analysis. Each of these methods uses special tools, such as spreadsheets, high-level programming languages (Python or R), low-code platforms (Loginom, Orange), or specialized analytics modules integrated into educational portals.

The article reveals the features of the “Analytics” module embedded in LMS Moodle for tracking and optimizing the educational process. The authors provide recommendations for configuring and using this module in the educational process. The module “Analytics” allows to create forecasts of the success rate of mastering a separate training course without programming skills, does not require data preprocessing for analysis, and allows to quickly take measures to improve the educational process based on the results of the forecast. However, the module has a number of disadvantages (limited attributes and forecasting models, high requirements for computing

power to run model training and build forecasts), so often administrators of educational portals disable this module. In this regard, an experiment was conducted to train a classification model for predicting students' examination scores based on students' activities on the educational portal of Nosov Magnitogorsk State Technical University.

In the course of the experiment, data on students' activity on the university's educational portal for one academic year were taken, and machine learning models were built to predict examination scores for an individual student and subject. The results of the experiment showed F1 74 % in predicting the score that the students would receive on the intermediate assessment. This confirms the effectiveness of using educational data mining based on user data from educational platforms.

**Keywords:** educational portal, educational data mining, machine learning model, forecasting educational results, LMS Moodle.

**For citation:**

Karmanova E. V., Zakharova Ya. M., Kiselev A. V. Educational data mining in a university using LMS Moodle. *Informatics and Education*. 2024;39(3):16–28. (In Russian.) DOI: 10.32517/0234-0453-2024-39-3-16-28.

## 1. Введение

Современные высшие образовательные учреждения в качестве средств обучения и коммуникации используют образовательные порталы, размещая на них электронные курсы по отдельным дисциплинам. Как правило, электронные курсы содержат все необходимые элементы обучения: теоретические, практические и контрольно-измерительные материалы. Даже при обучении в очной форме, предполагающей обязательное присутствие студентов на лекционных и практических занятиях в стенах учебного заведения, наличие электронных курсов на образовательных порталах очень важно.

Портальные технологии (технологии, обеспечивающие работу крупного сетевого телекоммуникационного узла) позволяют не только размещать различного рода учебный контент, но и сохранять каждое действие студентов в виде логов<sup>1</sup>, а значит, фиксировать образовательные результаты. На сегодняшний день при использовании таких технологий в вузах накапливается огромный массив данных (большие данные), умелая обработка и правильный анализ которых позволят не только получать информацию о качестве предоставляемого контента и организации учебного процесса, но и прогнозировать образовательные результаты.

Для качественного анализа больших данных сегодня используются методы и инструменты интеллектуального анализа данных (*англ.* Data Mining). В области образования эта технология еще не нашла широкого применения, но уже получила свое название — интеллектуальный анализ образовательных данных (ИАОД) (*англ.* Educational Data Mining) [1, 2], или учебная аналитика [3].

Одним из примеров использования методов интеллектуального анализа данных является построение поведенческого паттерна. Так, в торговле этот паттерн формируется на основе поведения покупателей, зафиксированного в истории их покупок, а целью анализа такого паттерна будет прогноз: какой продукт и в какое время приобретет отдельный покупатель. Для банков поведенческий паттерн — это действия клиентов, установленные на основе

проводимых ими транзакций, а цель анализа — определение того, является ли текущая транзакция мошеннической.

В рамках ИАОД можно получать поведенческий паттерн обучающегося на основе его деятельности на электронной образовательной платформе [4] и предсказывать, какую оценку в результате он получит по отдельной дисциплине. Реализация таких прогнозов с использованием методов машинного обучения позволяет выявлять важные факторы, влияющие на результаты процесса обучения. В частности, с помощью этих прогнозов можно доказать, что информирование студентов о целях учебных курсов в рамках онлайн-обучения имеет важное значение, поскольку именно осознание целей обучения стимулирует познавательную деятельность обучающихся и, как следствие, влияет на результаты [5]. В то же время сами студенты, видя предсказание системы, будут на ранних этапах обучения понимать, насколько верно они организуют свою самостоятельную учебную деятельность.

В целом с помощью ИАОД можно определить перечень учебных предметов, трудных для освоения, узнать, какие учебные темы вызывают повышенный интерес у обучающихся, проверить, как быстро выпускники с высокими оценками находят работу и есть ли зависимость между их образовательными результатами и заработной платой [6, 7].

Несмотря на описанные преимущества, использование учебной аналитики все еще мало распространено: существует проблема в управлении учебным процессом на базе портальных технологий, поскольку преподаватели не владеют специализированными инструментами и методами анализа моделей поведения студентов на основе больших данных. В связи с этим в рамках нашего исследования была поставлена **цель продемонстрировать возможности использования современных инструментов ИАОД и описать опыт их применения.**

## 2. Методы и инструменты интеллектуального анализа образовательных данных

Основным назначением ИАОД является обнаружение скрытых закономерностей, полезных знаний об образовательном процессе и его участниках для поддержки принятия решений. В статье Е. В. Ширинкиной «Методы интеллектуального анализа

<sup>1</sup> Логи (*англ.* log — протокол, журнал, судовой журнал) — записи в базе данных, которые содержат информацию о действиях пользователей в системе (времени входа и выхода из системы, просмотрах курсов и учебных материалов, выполнении заданий, обсуждениях на форумах и т. д.).

данных и образовательной аналитики» [8] подробно рассматриваются основные методы образовательной аналитики:

- *поиск аномалий* — определение «нетипичных» студентов (не успевающих за темпом программы или, наоборот, значительно опережающих ее) либо выявление подозрительного поведения студентов на образовательной платформе (слишком высокая активность) и др.;
- *поиск ассоциаций* — выявление взаимосвязи между поведенческими паттернами студентов и трудностями, которые они испытывают при обучении, между полом студента и его успехами в определенных предметах, определение зависимости успешного выполнения заданий студентами от погодных условий или времени выполнения задания;
- *классификация* — выявление студентов, которые находятся в зоне риска отчисления, либо определение качества учебного материала в электронных курсах;
- *кластерный анализ* — объединение учебного материала или слушателей в группы (кластеры) по схожим паттернам;
- *регрессионный анализ* — оценка статистической значимости отдельных показателей образовательного процесса и характеристик студентов;
- *факторный анализ* — определение того, какие черты поведения студентов приводят к неуспеваемости, отказу от обучения и др.

Для реализации описанных методов ИАОД необходим удобный и понятный для преподавателя инструментарий. Перечислим основные виды инструментов для реализации ИАОД и их характеристики:

- *электронные таблицы* позволяют визуализировать данные, вести описательную статистику по табличным данным (рассчитывать среднее, медианное, максимальные, минимальные значения и т. д.), однако не предназначены для моделей машинного или глубокого обучения, которые могли бы предсказать риск отчисления студентов, их аномальное поведение и т. д.;
- *языки программирования (как правило, Python или R)* имеют уже готовые библиотеки и модули для реализации ИАОД. Основная сложность их применения заключается в том, что пользователи должны владеть возможностями данного инструмента;
- *аналитические low-code платформы* имеют готовые наборы компонентов и моделей машинного обучения для реализации ИАОД (к примеру, Loginom). Преимуществом использования данных платформ является возможность выполнения анализа данных без знания основ программирования.

Поскольку современные образовательные учреждения используют при организации учебного

процесса системы управления обучением (Moodle, iSpring Learn, Open edX и др.), именно наличие в этих системах функционала по реализации ИАОД является наиболее эффективным решением, которое позволило бы не только преподавателям, но и руководству учреждений гибко, качественно и своевременно получать необходимую информацию об учебном процессе.

В работе Г. В. Садыковой и О. И. Северьянова «Дистанционное обучение в эпоху пандемии: опыт российских преподавателей» [9] отмечается, что одной из самых распространенных платформ для организации электронного обучения в России является Moodle. А. И. Шамсутдинова и Г. Ф. Хасанова в статье «Образовательная аналитика результатов повышения квалификации преподавателей в LMS Moodle» [10] подчеркивают эффективность Moodle, особенно для университетов, реализующих программы повышения квалификации профессорско-преподавательского состава. Е. К. Якунина и М. В. Романова в работе «Преимущества и возможности применения СДО Moodle в общеобразовательных учебных заведениях» [11] акцентируют внимание на том, что популярность системы Moodle обусловлена ее открытым исходным кодом, бесплатностью, а также разнообразием встроенных модулей и инструментов для создания интерактивных образовательных материалов.

Система управления обучением (*англ.* Learning Management System, LMS) Moodle (*англ.* Modular Object-Oriented Dynamic Learning Environment — модульная объектно-ориентированная динамическая обучающая среда) хранит журнал событий (логи), который можно использовать для реализации методов интеллектуального анализа образовательных данных (ИАОД). Например, Е. Н. Мащенко и Е. В. Ченгарь в статье «Анализ данных журналов событий системы дистанционного обучения для задач учебной аналитики» [12] применяют визуализацию данных из журналов событий Moodle для исследования активности обучающихся в электронном курсе. А. Б. Степанов в работе «Данные цифровой образовательной среды как инструмент успешного взаимодействия субъектов электронного образования» [13] акцентирует внимание на аналитическом подходе к обработке данных в Moodle, выделяя ключевые типы данных и факторы, влияющие на успеваемость обучающихся, такие как частота взаимодействий с платформой, продолжительность работы с заданиями и время посещения страниц курса. Р. В. Есин Т. А. Кустицкая, М. В. Носков в статье «Прогнозирование успешности обучения по дисциплине на основе универсальных показателей цифрового следа LMS Moodle» [14] описывают способы использования обобщенных показателей цифрового следа обучающегося, полученных из журналов событий Moodle, для раннего прогнозирования успешности промежуточной аттестации. Е. Terbusheva, X. Piotrowska, S. Kalmykova в работе «Analytics of the digital behavior of Russian first-year university students: Case study» [15] раскрывают аналитический потенциал встроенных и внешних инструментов Moodle, демонстрируя, что

даже самые простые аналитические инструменты дают полезные результаты. В исследовании В. И. Токтаровой и О. Г. Поповой «Анализ образовательных данных взаимосвязи успешности обучения и поведения студентов в цифровой образовательной среде вуза» [16] показана линейная зависимость между поведением обучающегося в электронном курсе, размещенном на платформе Moodle, и его итоговой оценкой. Анализ зависимостей проводился в рамках одного курса по программированию. В то же время R. R. Estacio и R. C. Raga «Analyzing students online learning behavior in blended courses using Moodle» [17] показывают: корреляция между активностью обучающихся в онлайн-среде и их успеваемостью может сильно варьироваться, что связано, среди прочего, с дизайном электронных курсов.

Таким образом, многие исследователи отмечают высокий аналитический потенциал системы Moodle благодаря накоплению в ней данных об активности обучающихся. В связи с этим с версии 3.4, выпущенной в ноябре 2017 года, Moodle включает в себя аналитику обучения с открытым исходным кодом, выходящую за рамки описательной аналитики. Модели Moodle позволяют использовать методы машинного обучения для предсказания успеваемости обучающегося и выдачи рекомендаций студентам и преподавателям по совершенствованию качества учебного процесса. Для работы с аналитическими моделями в Moodle представлен отдельный модуль «Аналитика».

Несмотря на то, что Moodle является самой популярной LMS в России, научных трудов, посвященных ее аналитическим возможностям, крайне мало. В основном это статьи небольшого объема, которые не раскрывают потенциал платформы полностью, не дают подробных рекомендаций по настройке и использованию модуля «Аналитика». Необходимо отметить, что серьезной проблемой научной литературы является быстрое устаревание источников. Инструменты аналитики LMS Moodle постоянно совершенствуются и обновляются, поэтому статьи, написанные более двух лет назад, сегодня не могут считаться актуальными и достоверными.

### 3. Возможности модуля «Аналитика» LMS Moodle для интеллектуального анализа образовательных результатов

Инструменты, собранные в модуле «Аналитика», позволяют оценивать текущие результаты образовательного процесса, находить скрытые закономерности и связи в большом объеме образовательных данных, делать прогнозы на основании имеющихся данных, уведомлять участников образовательного процесса о возможных проблемах и рисках.

Аналитика обучения Moodle поддерживает два вида моделей:

- **прогностические модели**, которые обучаются на основе исторических данных об-

разовательного портала (активности всех пользователей) с использованием методов машинного обучения. Эти модели предназначены для обнаружения скрытых взаимосвязей в образовательных данных. Примером является встроенная модель «Студенты с риском отчисления», которая позволяет предсказать, кто из студентов недостаточно активно работает на курсе и рискует его не закончить. Кроме того, можно самостоятельно создавать различные модели для прогнозирования академических достижений обучающихся, выбирая для них соответствующие показатели;

- **статистические модели**, которые не требуют предварительного обучения и выводят уведомления о предстоящих событиях, основанные на метаданных отдельных электронных курсов. Метаданными могут являться сведения о количестве записанных студентов и преподавателей на курс, сроках выполнения заданий и др., событиями — запуск курса, окончание курса и др.

В таблице 1 представлена краткая информация об встроенных моделях аналитики LMS Moodle.

В рамках нашего исследования особый интерес представляет модель «Студенты с риском отчисления», которая на основе изученных паттернов поведения обучающихся определяет студентов, рискующих не завершить курс. В ней используется модель вовлеченности обучающихся Ч. Гаррисона «сообщество исследований» (англ. Community of Inquiry), которая состоит из трех компонентов [18, 19]:

- **познавательная деятельность** (англ. Cognitive Presence) — реализация условий для активизации познавательной деятельности студентов через эффективное взаимодействие «студент — контент»;
- **педагогическое сопровождение** (англ. Teacher Presence) — создание эффекта кругло-суточного присутствия преподавателя в электронной образовательной среде и возможность взаимодействия «студент — преподаватель»;
- **социальная составляющая** (англ. Social Presence) — создание условий для формирования учебного сообщества и возможностей проявления личностных качеств в процессе учебной коммуникации «студент — студент».

Стоит отметить, что у модели «Студенты с риском отчисления» в Moodle имеются некоторые ограничения:

1. Модель требует присутствия в Moodle определенного количества данных, с помощью которых можно делать прогнозы. Курс должен содержать элементы Moodle, посредством которых можно оценить познавательную деятельность и социальную составляющую.
2. Модель предполагает, что курсы имеют фиксированные даты начала и окончания обучения, и не предназначена для использования на курсах со скользящей регистрацией.



Таблица 1 / Table 1

**Встроенные модели аналитики LMS Moodle****The built-in LMS Moodle analytics models**

№ п/п	Название	Вид	Функциональное назначение	Технология реализации
1	«Ближайшие учебные мероприятия»	Статистическая	Позволяет отправлять пользователям уведомления о ближайших учебных мероприятиях	Проверяет элементы курса с предстоящими сроками выполнения и выводит данные на страницу календаря пользователя
2	«Курсы, которые могут не начаться»	Статистическая	Данная модель выводит отчет о том, какие курсы с предстоящей датой могут не начаться, так как на них не назначен преподаватель или не зачислены студенты	Основывает прогнозы на следующих предположениях: <ul style="list-style-type: none"> <li>• если нет обучающихся, то курс может не начаться;</li> <li>• если нет преподавателя, то курс может не начаться</li> </ul>
3	«Студенты, которые давно не заходили в курс»	Статистическая	Определяет студентов, не заходивших в электронный курс, на который они зачислены, в течение длительного времени (по умолчанию — один месяц)	Основывается на том, какие курсы доступны для студента
4	«Студенты, которые еще не заходили в курс»	Статистическая	Определяет студентов, которые никогда не заходили в электронный курс, на который они зачислены	Основывается на том, какие электронные курсы доступны для студента
5	«Студенты с риском отчисления»	Прогностическая	Позволяет определить студентов, недостаточно активно работающих на курсе и рискующих его не закончить	Создает прогнозы с помощью технологии машинного обучения

3. Курс должен иметь определенную длительность, чтобы его можно было разбить на временные интервалы.

4. Курсы не будут включены в обучение или прогнозы, если даты начала и окончания обучения отличаются друг от друга более чем на один год.

Помимо использования моделей, встроенных по умолчанию, модуль «Аналитика» позволяет создавать прогностические модели самостоятельно. Получение точных прогнозов требует правильной настройки модуля «Аналитика». Доступ к настройкам имеет пользователь с правами администратора. В прогностических моделях Moodle выделяют четыре основных компонента, которые необходимо задать при создании новой модели:

**1. Функции:**

- выявление студентов, которые рискуют не достичь компетенций, обеспечиваемых курсом;
- выявление студентов, которые рискуют не выполнить условия завершения курса;
- выявление студентов с риском отчисления;
- выявление студентов, которые рискуют не набрать минимальную оценку для завершения курса.

**2. Критерии анализа:**

- предоставление доступа к курсу после даты окончания;

- наличие доступных курсов;
- возможность записи любого действия в курсе;
- активность;
- возможность отслеживания завершения и т. д.

**3. Интервалы анализа:**

- без ограничений по времени;
- через месяц после начала;
- через одну неделю после начала;
- последние три дня;
- последний месяц;
- прошлая неделя;
- последний квартал и т. д.

**4. Обработчик прогнозов:**

- механизм машинного обучения PHP реализует логистическую регрессию с использованием PHP-ML (предоставлено Moodle);
- механизм машинного обучения Python реализует нейронную сеть прямого распространения одного скрытого слоя с использованием TensorFlow.

На рисунке 1 представлены основные этапы работы с модулем «Аналитика», которые мы рассмотрим более подробно.

На этапе постановки цели необходимо ответить на следующие вопросы:

1) Для чего внедряется аналитика?



Рис. 1. Основные этапы работы с модулем «Аналитика»

Fig. 1. The main stages of working with the "Analytics" module

- 2) Какие показатели необходимо получить?
- 3) Как это поможет оптимизировать процесс обучения?
- 4) Что делать, если результат весьма вероятен или маловероятен?
- 5) Кто и каким способом должен быть уведомлен о результатах аналитики?

На этапе сбора и организации данных важно определить качественные показатели (какие элементы мы будем использовать для анализа) и количественные показатели (сколько элементов предполагается проанализировать). Указываются интервалы анализа, выбираются обработчик прогнозов и электронные курсы, на которых будет проводиться обучение.

На этапе обучения модели запускается процесс ее обучения на всех исторических данных, которые накопились за время работы образовательного портала, с учетом выбранных электронных курсов. В случае, если данных недостаточно для обучения, система выводит соответствующее предупреждение и останавливает процесс.

На этапе анализа и интерпретации данных пользователь, имеющий соответствующие права, просматривает созданный моделью прогноз и принимает решение о необходимости рассылки уведомлений с результатами прогноза заинтересованным лицам, о корректировке элементов обучения и т. д.

#### 4. Эксперимент по использованию модуля «Аналитика» LMS Moodle на базе образовательного портала вуза

В рамках исследования была проведена апробация использования модуля «Аналитика» на базе образовательного портала Магнитогорского государственного технического университета им. Г. И. Носова (МГТУ им. Г. И. Носова). В ходе эксперимента были выявлены следующие ограничения:

- производительность образовательного портала резко снизилась, ресурсы портала стали загружаться медленнее;
- не выполнялись требования модуля «Аналитика», поскольку большинство электронных курсов не имели настроек с фиксированными датами начала и окончания разделов.

В связи с этим была развернута тестовая система на базе LMS Moodle и проведен эксперимент на искусственно сгенерированных данных. Для обучения прогностической модели было создано четыре курса длительностью в одну неделю. На этих курсах имитировалась деятельность десяти студентов, из которых:

- четыре студента полностью выполняли все задания курса;
- три студента не выполнили одно или несколько из контрольных заданий курсов;
- у троих студентов была нулевая активность.

Соответственно, успешно закончили курсы четыре студента из десяти.

Далее для анализа работоспособности прогностической модели «Студенты с риском отчисления» был создан электронный курс «Интеллектуальный анализ данных», состоящий из пяти учебных тем. Каждая тема содержала контрольное задание, вводная тема включала в себя тест из десяти вопросов. В курс было добавлено десять студентов и один преподаватель. На протяжении двух недель имитировалась деятельность студентов и преподавателя. Двое из десяти студентов выполнили все задания курса, остальные только посещали разделы курса. После запуска прогноза система представила список студентов с риском отчисления, который отображен на рисунке 2.

В результате модель предсказала, что девять из десяти студентов недостаточно активно работают и рискуют не закончить курс. Студент 2, которого модель не включила в список студентов с риском отчисления, проявлял активность на курсе, хотя выполнил еще не все задания. При этом Студент 1 и Студент 3 также регулярно заходили в электронный курс и уже выполнили все условия для его успешного завершения. Таким образом, можно утверждать, что прогноз не полностью соответствует действительности. На рисунке 3 представлен отчет по деятельности Студента 1.

Кроме того, предоставляется отчет со значениями по каждому показателю, на основе которого был сделан прогноз. На рисунке 4 представлен такой отчет по Студенту 1.

На наш взгляд, причиной полученных результатов прогноза могли стать следующие обстоятельства:

- модель обучалась на искусственных данных, и, возможно, паттерны поведения всех студентов были схожими между собой;
- при обучении модели не хватило данных (небольшое количество электронных курсов, студентов, элементов курса, на которых обучалась модель);
- модель рассматривает как важный признак высокую активность студентов по освоению курса.

Таким образом, LMS Moodle предоставляет набор встроенных инструментов для реализации учебной аналитики. Можно выделить следующие преимущества использования модуля «Аналитика» Moodle:

- не требует установки дополнительных программных средств;
- не требует навыков программирования для создания прогнозов на основе методов машинного обучения;
- не требует дополнительных манипуляций с данными (не нужно выгружать, чистить, нормировать данные);
- позволяет быстро принимать меры на основе результатов прогноза (например, сразу отправить сообщение студенту).

К недостаткам модуля «Аналитика» стоит отнести следующее:

- большинство встроенных моделей являются статистическими и не предполагают прогнозов;

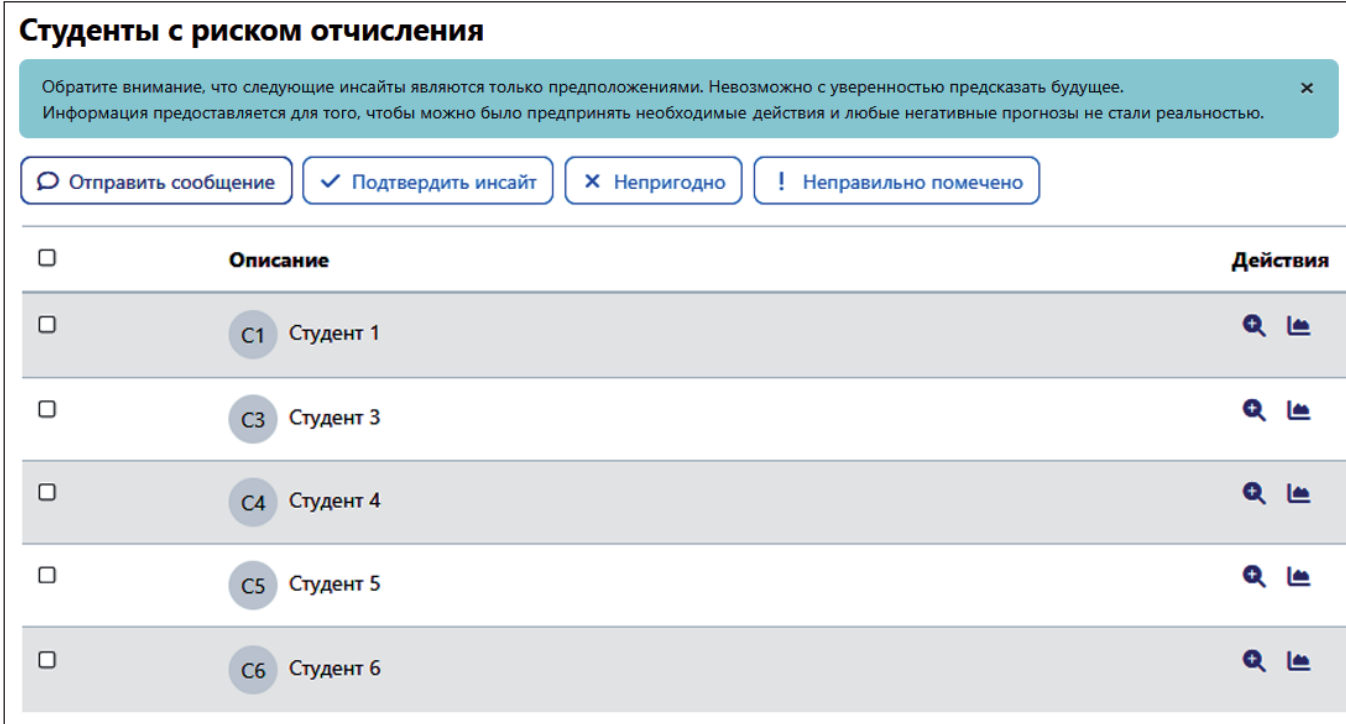


Рис. 2. Результат прогноза модели «Студенты с риском отчисления»  
Fig. 2. The result of the “Students at risk of expulsion” forecast

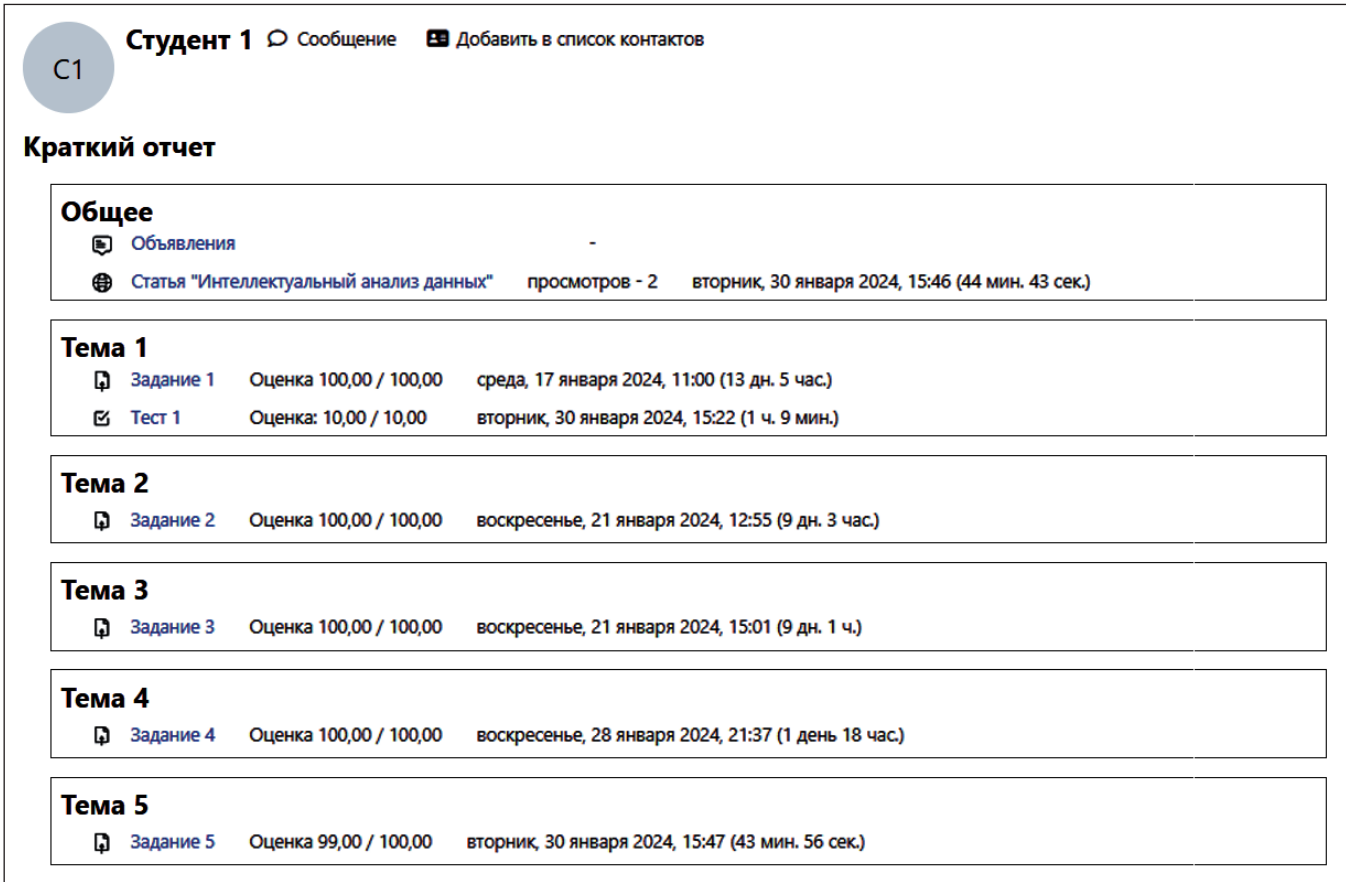


Рис. 3. Краткий отчет по Студенту 1  
Fig. 3. The brief report for Student 1

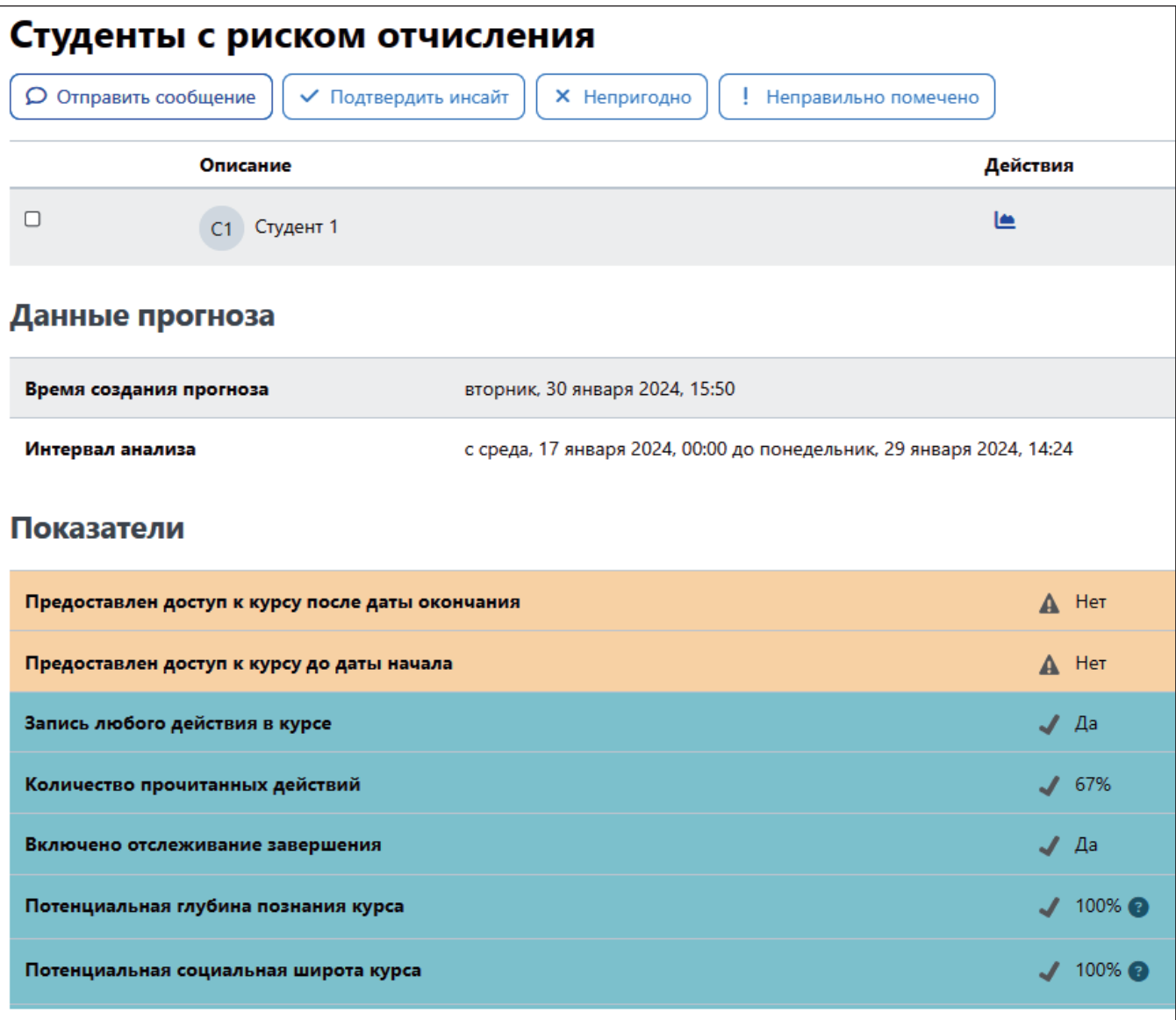


Рис. 4. Значения показателей для Студента 1  
Fig. 4. Indicator values for Student 1

- модели, использующие технологии машинного обучения, не работают для курсов с плавающей датой;
- признаки, на которых может обучаться модель, ограничены;
- прогнозы не всегда точны (особенно если у модели было мало данных для обучения).

### 5. Методы машинного обучения для предсказания результатов учебной деятельности студентов на основе логов LMS Moodle

Описанные ранее ограничения модуля «Аналитика» не позволили провести эксперимент по определению возможности прогнозирования образовательных результатов, основываясь на паттернах поведения студентов при работе с электронными курсами на базе образовательного портала МГТУ им. Г. И. Носова. В связи с этим было решено провести ИАОД, выгрузив отдельно все логи пользователей из базы данных LMS Moodle, и построить модель классификации — предсказания оценок студентов.

Данные были импортированы из Moodle MySQL в базу данных ClickHouse<sup>1</sup> [20]. Общий объем записей составил более 100 млн различных событий. Каждый раз, когда студент совершает какое-либо действие на образовательном портале, происходит генерация определенного события. Например, событие «Просмотр курса» фиксируется, когда пользователь зашел на

<sup>1</sup>

ClickHouse — столбцовая система управления базами данных для онлайн-обработки аналитических запросов. Подробнее см.: <https://clickhouse.com/docs/ru>



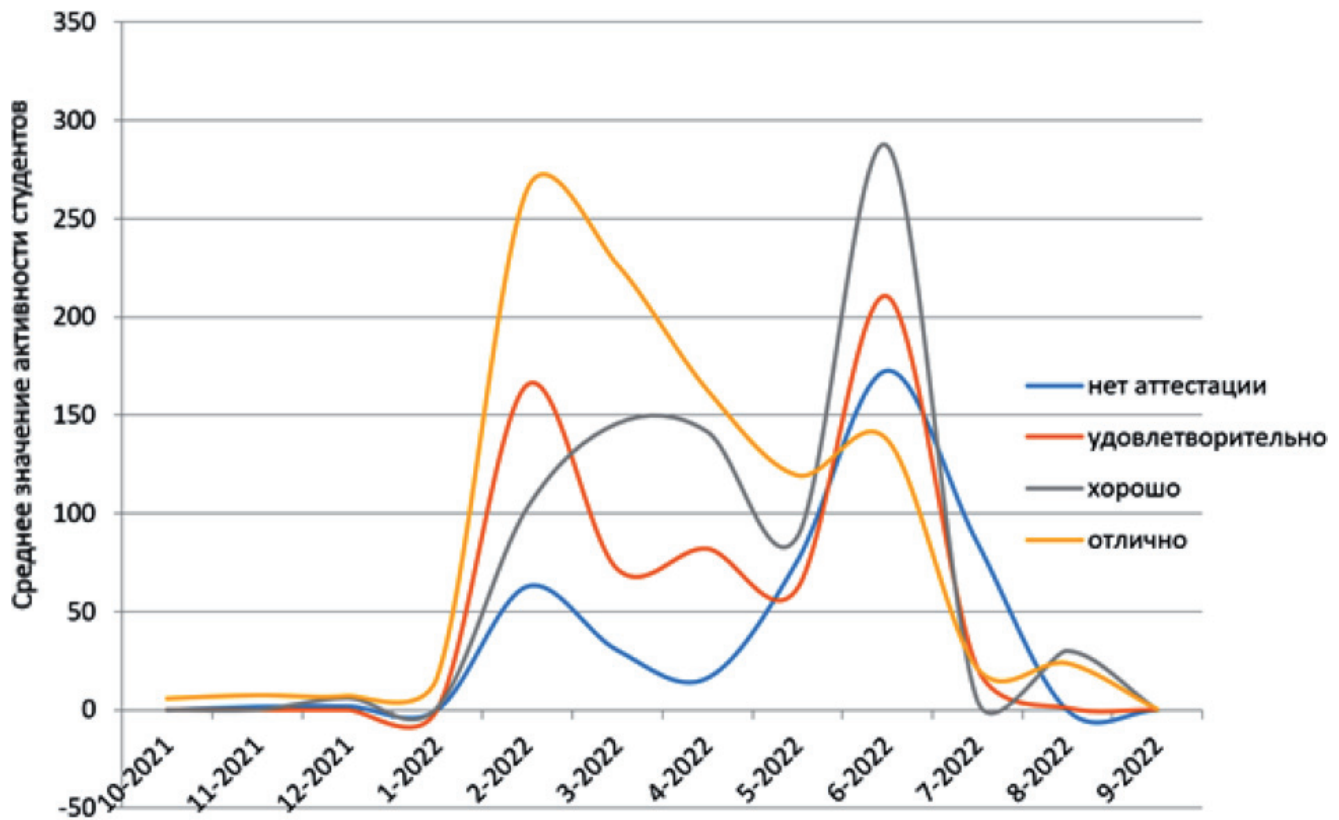


Рис. 5. Распределение активности по виду аттестации

Fig. 5. Distribution of activity by type of certification

курс; событие «Оценка студента» сохраняется, когда преподаватель оценил выложенное студентом задание. Насчитывается более 200 различных типов событий.

Данные по активности студентов были взяты за период с октября 2021 года по октябрь 2022 года. Для визуализации использовался инструмент Yandex DataLens<sup>1</sup>, который берет информацию напрямую из базы данных ClickHouse.

В ходе исследования была сформулирована гипотеза: *чем больше событий генерирует пользователь, тем выше вероятность получить более высокую оценку в рамках промежуточной аттестации* (данное предположение было выдвинуто в результате применения модели предсказания студентов с риском отчисления с помощью модуля «Аналитика» Moodle). Уточним, что были взяты данные для дисциплин, заканчивающихся формой промежуточного контроля «экзамен». В импортированном наборе данных содержалась информация о 6316 студентах и 893 электронных курсах.

Для проверки гипотезы была проведена предобработка данных: агрегированы логи студентов с образовательного портала и добавлены результаты промежуточной аттестации из информационной системы университета. Визуализация распределения

активностей студентов по результатам промежуточной аттестации показала (рис. 5), что площадь под графиком функции у отличников больше, чем у хорошистов, и т. д. Кроме того, наблюдаются периоды высокой активности в начале и в конце семестра: студенты генерируют больше событий, что подтверждают их активные действия перед сессией и во время начала учебного семестра.

Затем все активности студентов были сгруппированы по неделям, поскольку занятия по отдельным дисциплинам повторяются в среднем один или два раза в неделю и не имеет смысла обрабатывать логи по каждому дню. Кроме того, некоторые студенты выделяют время на учебу в будние дни, а другие могут учиться только по выходным, поэтому неделя — оптимальный временной интервал, за который можно делать срез и анализировать промежуточные результаты.

Чтобы сохранять нарастающий итог и понимать темп прироста событий по каждому из них, нами были введены дополнительные критерии с добавлением `_avg` в имени, которые отражают среднее значение событий конкретного типа по определенному студенту и отдельной учебной дисциплине за эту неделю, а также предыдущие события этой недели. Таким образом, был сформирован набор данных об активности студентов образовательного портала, содержащий 20 признаков и 136 795 записей. Срез этого набора представлен в таблице 2.

<sup>1</sup> Yandex DataLens — сервис анализа и визуализации данных от Yandex Cloud. Подробнее см.: <https://datalens.yandex.cloud/>

Таблица 2 / Table 2

## Срез набора данных с информацией об активности студентов по неделям

## Dataset slice with information on student activity by week

courseid	userid	num_week	s_all	s_all_avg	...	name_formopril	leveled	num_sem	kurs
71 262	34 527	6	9	9	...	1	1	2	2
71 262	34 527	7	0	4,5	...	1	1	2	2
71 262	34 527	8	0	3	...	1	1	2	2
...	...	...	...	...	...	...	...	...	...

**Представим описание признаков полученного набора данных:**

- `courseid` — уникальный идентификатор курса, дисциплины;
- `userid` — уникальный идентификатор студента (не используется в обучении);
- `num_week` — номер недели в году;
- `s_all` — количество всех событий на текущий момент;
- `s_all_avg` — среднее количество всех событий в неделю;
- `s_course_viewed` — количество просмотров курса;
- `s_course_viewed_avg` — среднее количество просмотров курса в неделю;
- `s_q_attempt_viewed` — количество просмотров теста;
- `s_q_attempt_viewed_avg` — среднее количество просмотров теста в неделю;
- `s_a_course_module_viewed` — количество просмотров модуля в курсе;
- `s_a_course_module_viewed_avg` — среднее количество просмотров модуля в курсе в неделю;
- `s_a_submission_status_viewed` — количество отправленных заданий на проверку;
- `s_a_submission_status_viewed_avg` — среднее количество ответов;
- `namer_level` — оценка за дисциплину;
- `depart` — номер кафедры;
- `name_osno` — основа обучения (имеет два значения: бюджет или контракт);
- `name_formopril` — форма обучения;

- `leveled` — уровень образования (имеет два значения: бакалавриат, магистратура);
- `num_sem` — номер семестра;
- `kurs` — номер курса учебной группы.

В рамках исследования нами были применены методы машинного обучения для прогнозирования оценок студентов на экзамене по отдельной учебной дисциплине начиная со второй недели от начала обучения. Для реализации задачи прогнозирования были обучены четыре модели классификации [21]:

- *k*-ближайших соседей (англ. *k*-nearest Neighbors Classifier);
- деревья решений (англ. Decision Tree Classifier);
- случайный лес (англ. Random Forest Classifier);
- CatBoost (англ. CatBoost Classifier).

Полученные модели были сравнены по доле правильных ответов (метрика *Accuracy*), по среднему гармоническому значению точности и полноты (метрика *F1*), а также по времени обучения (табл. 3).

**В результате анализа полученных моделей было выявлено, что модель на основе случайного леса дает лучший прогноз.** В таблице 4 представлен классификационный отчет по каждому предсказываемому классу (в нашем случае класс — это итоговая оценка студента). Помимо метрики *F1*, данный отчет содержит следующие показатели:

- точность модели (метрика *Precision*) показывает, насколько точно модель предсказывает положительные примеры для каждого класса;
- полнота модели (метрика *Recall*) показывает, насколько хорошо модель находит все положительные примеры для каждого класса;

Таблица 3 / Table 3

## Сравнение моделей машинного обучения по предсказанию экзаменационных оценок у студентов

## Comparison of machine learning models for predicting student examination scores

Модель	Метрика <i>Accuracy</i>	Метрика <i>F1</i>	Время обучения (с.)
<i>k</i> -nearest Neighbors Classifier	0,66	0,63	47
Decision Tree Classifier	0,73	0,71	5
Random Forest Classifier	0,76	0,74	123
CatBoost Classifier	0,69	0,66	571

Таблица 4 / Table 4

## Метрики качества модели на основе случайного леса

## Quality metrics of random forest classifier

Класс (оценка)	Метрика <i>Precision</i>	Метрика <i>Recall</i>	Метрика <i>F1</i>	Показатель <i>Support</i>
2	0,76	0,66	0,70	14 555
3	0,74	0,72	0,73	23 220
4	0,72	0,69	0,70	43 079
5	0,79	0,86	0,82	55 941
Macro avg	0,75	0,73	0,74	136 795

- показатель *Support* отображает количество образцов (в нашем случае студентов) каждого класса в наборе данных.

Как видно из таблицы 4, модель случайного леса лучше всего предсказывает оценку «отлично» (точность — 79 %), хуже всего — оценку «хорошо» (точность — 72 %).

С помощью модели случайного леса было также определено, какие признаки набора данных являются наиболее важными для прогнозирования, насколько каждый признак влияет на итоговую оценку. На рисунке 6 представлен график ранжирования при-

знаков по степени влияния для модели случайного леса. Как видно из рисунка, ведущими признаками для классификационной модели являются:

- учебная дисциплина;
- среднее количество всех событий в неделю;
- среднее количество просмотров и т. д.

Полученные модели классификации студентов позволяют сделать вывод о возможности прогнозирования успеваемости студентов на основе их активности начиная со второй недели обучения. Модель случайного леса показала большую метрику качества *F1* в сравнении с другими заявленными моделями.

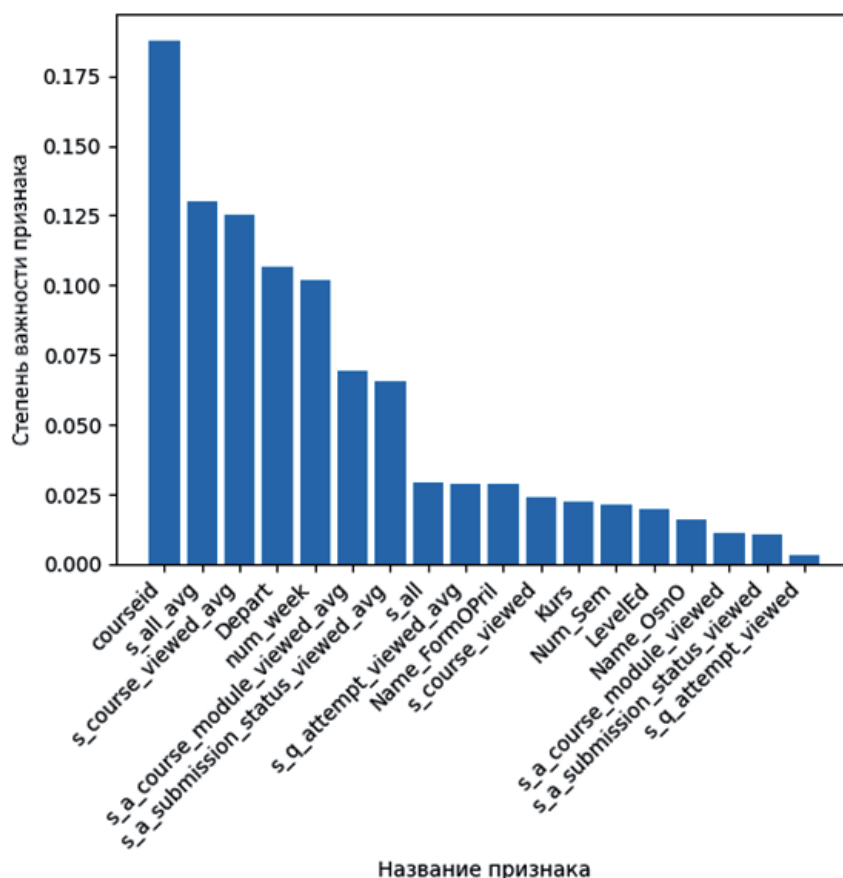


Рис. 6. Ранжирование признаков по степени влияния для модели «Случайный лес»

Fig. 6. Ranking of features by degree of influence for the Random Forest Classifier

Возможно, модели глубокого обучения позволят увеличить метрику качества модели и тем самым способность к прогнозированию, что является перспективным направлением дальнейшего исследования.

## 6. Заключение

Современные платформы управления образовательным процессом и накопившиеся большие данные о деятельности обучающихся и ее результатах позволили активно развиваться новому направлению в образовании — интеллектуальному анализу образовательных данных, или учебной аналитике. Данная технология позволяет не только понять факторы, влияющие на обучение, но и спрогнозировать его результаты, а значит, вовремя оптимизировать или скорректировать образовательный процесс.

Реализация интеллектуального анализа данных предполагает использование разнообразных инструментов, технологий и методов машинного обучения. В рамках исследования были изучены возможности модуля «Аналитика» LMS Moodle, который входит в ядро образовательной платформы и позволяет выявлять студентов с риском отчисления, а также находить проблемы формирования отдельных компетенций в рамках учебного плана. Эксперимент показал, что модуль выполняет заявленные функциональные требования, однако требует больших вычислительных мощностей сервера для обучения моделей и вычисления прогнозов, а также тщательной настройки параметров отдельных электронных курсов.

В связи с обнаруженными ограничениями модуля «Аналитика» параллельно был проведен эксперимент по созданию моделей машинного обучения на базе набора данных об активности студентов на образовательном портале МГТУ им. Г. И. Носова. В результате обучения модель случайного леса позволила предсказывать с точностью 76 %, какую оценку на экзамене получит студент в рамках промежуточной аттестации, начиная со второй недели семестра.

Результаты эксперимента показали эффективность интеллектуального анализа образовательных данных, который позволил выявлять слабых студентов на ранних этапах обучения и тем самым вовремя корректировать образовательный процесс.

## Список источников / References

1. Тербушева Е. А. Методика обучения интеллектуальному анализу образовательных данных студентов педагогического вуза. *Открытое образование*. 2019;23(3):14–24. EDN: TJCZTW. DOI: 10.21686/1818-4243-2019-3-14-24. [Terbusheva E. A. Methods of teaching educational data mining for pedagogical students. *Open Education*. 2019;23(3):14–24. (In Russian.) EDN: TJCZTW. DOI: 10.21686/1818-4243-2019-3-14-24.]
2. Баранова Е. В., Верещагина Н. О., Швецов Г. В. Цифровые инструменты для анализа учебной деятельности студентов. *Известия Российского государственного педагогического университета им. А. И. Герцена*. 2020;(198):56–65. EDN: AXDDPU. DOI: 10.33910/1992-6464-2020-198-56-65. [Baranova E. V., Vereshchagina N. O., Shvetsov G. V. Digital tools for learning activities analysis. *Izvestia: Herzen University*

*Journal of Humanities & Sciences*. 2020;(198):56–65. (In Russian.) EDN: AXDDPU. DOI: 10.33910/1992-6464-2020-198-56-65.]

3. Вилкова К. А., Захарова У. С. Учебная аналитика в традиционном образовании: ее роль и результаты. *Университетское управление: практика и анализ*. 2020;24(3):59–76. EDN: LZRMLY. DOI: 10.15826/umpa.2020.03.026.

[Vilkova K. A., Zakharova U. S. Learning analytics in conventional education: Its role and outcomes. *Journal University Management: Practice and Analysis*. 2020;24(3):59–76. (In Russian.) EDN: LZRMLY. DOI: 10.15826/umpa.2020.03.026.]

4. Nammakhunt A., Porouhan P., Premchaiswadi W. Creating and collecting e-learning event logs to analyze learning behavior of students through process mining. *International Journal of Information and Education Technology*. 2023;13(2):211–222. DOI: 10.18178/ijiet.2023.13.2.1798.

5. Sun W., Hong J.-C., Dong Y., Huang Y., Fu Q. Self-directed learning predicts online learning engagement in higher education mediated by perceived value of knowing learning goals. *Asia-Pacific Education Researcher*. 2023;32:307–316. DOI: 10.1007/s40299-022-00653-6.

6. Toradmal M. B., Mehta M., Mehendale S. Machine learning approaches for educational data mining. *Inventive Systems and Control. Lecture Notes in Networks and Systems*. Singapore, Springer Nature Singapore; 2023;672:737–748. DOI: 10.1007/978-981-99-1624-5\_55.

7. Arora S., Agarwal M., Kawatra R. Prediction of educationist's performance using regression model. *Proc. 7th Int. Conf. on Computing for Sustainable Global Development (INDIACom-2020)*. New Delhi, India, Institute of Electrical and Electronics Engineers Inc.; 2020;88–93. DOI: 10.23919/INDIACom49435.2020.9083708.

8. Ширинкина Е. В. Методы интеллектуального анализа данных и образовательной аналитики. *Современное образование*. 2022;(1):51–67. EDN: DINNOW. DOI: 10.25136/2409-8736.2022.1.37582.

[Shirinkina E. V. Methods of data mining and educational analytics. *Modern Education*. 2022;(1):51–67. (In Russian.) EDN: DINNOW. DOI: 10.25136/2409-8736.2022.1.37582.]

9. Садыкова Г. В., Северьянов О. И. Дистанционное обучение в эпоху пандемии: опыт российских преподавателей. *Информатика и образование*. 2023;38(1):72–82. EDN: SOCMEP. DOI: 10.32517/0234-0453-2023-38-1-72-82.

[Sadykova G. V., Severyanov O. I. Distance learning during the pandemic era: Russian instructors' experience. *Informatics and Education*. 2023;38(1):72–82. (In Russian.) EDN: SOCMEP. DOI: 10.32517/0234-0453-2023-38-1-72-82.]

10. Шамсутдинова А. И., Хасанова Г. Ф. Образовательная аналитика результатов повышения квалификации преподавателей в LMS MOODLE. *Управление устойчивым развитием*. 2021;(6(37)):110–115. EDN: VJLLCS.

[Shamsutdinova A. I., Khasanova G. F. Learning analytics of the results of faculty further education in LMS MOODLE. *Managing Sustainable Development*. 2021;(6(37)):110–115. (In Russian.) EDN: VJLLCS.]

11. Якунина Е. К., Романова М. В. Преимущества и возможности применения СДО MOODLE в общеобразовательных учебных заведениях. *Ab ovo... (С самого начала...)*. Сборник научных трудов. Магнитогорск: Магнитогорский государственный технический университет им. Г. И. Носова; 2019;4:86–87. EDN: JUYTTH.

[Yakunina E. K., Romanova M. V. Advantages and possibilities of using LMS MOODLE in general education institutions. *Ab ovo... (S Samogo Nachala...)*. Collection of Scientific Articles. Magnitogorsk, Nosov Magnitogorsk State Technical University; 2019;4:86–87. (In Russian.) EDN: JUYTTH.]

12. Мащенко Е. Н., Ченгарь О. В. Анализ данных журналов событий системы дистанционного обучения для задач



учебной аналитики. *Естественные и технические науки*. 2023;(5(180)):34–40. EDN: EQZSEO.

[Maschenko E. N., Chengar O. V. Analysis of data logs of events of the system of distance learning for tasks of learning analytics. *Natural and Technical Sciences*. 2023;(5(180)):34–40. (In Russian.) EDN: EQZSEO.]

13. Стенанов А. Б. Данные цифровой образовательной среды как инструмент успешного взаимодействия субъектов электронного образования. *EdCrunch Томск. Материалы международной конференции по новым образовательным технологиям*. Томск: Издательский дом Томского государственного университета; 2019;140–145. EDN: FLQKZT.

[Stepanov A. B. Digital educational environment data as a tool for successful interaction of e-learning entities. *EdCrunch Tomsk. Proc. Int. Conf. on New Educational Technologies*. Tomsk, Publishing House of Tomsk State University; 2019;140–145. (In Russian.) EDN: FLQKZT.]

14. Есин Р. В., Кустицкая Т. А., Носков М. В. Прогнозирование успешности обучения по дисциплине на основе универсальных показателей цифрового следа LMS Moodle. *Информатика и образование*. 2023;38(3):31–41. EDN: RCEOVE. DOI: 10.32517/0234-0453-2023-38-3-31-41.

[Esin R. V., Kustitskaya T. A., Noskov M. V. Predicting academic performance in a course by universal features of LMS Moodle digital footprint. *Informatics and Education*. 2023;38(3):31–41. (In Russian.) EDN: RCEOVE. DOI: 10.32517/0234-0453-2023-38-3-31-41.]

15. Terbusheva E., Piotrowska X., Kalmykova S. Analytics of the digital behavior of Russian first-year university students: Case study. *CEUR Workshop Proceedings. Proc. 15th Int. Conf. “New Educational Strategies in Modern Information Space”*. Aachen, Germany, M. Jeusfeld c/o Redaktion Sun SITE, Informatik V, RWTH Aachen; 2020:28–39. EDN: BYGVFO. Available at: [https://ceur-ws.org/Vol-2630/paper\\_4.pdf](https://ceur-ws.org/Vol-2630/paper_4.pdf)

16. Токтарова В. И., Попова О. Г. Анализ образовательных данных взаимосвязи успешности обучения и поведения студентов в цифровой образовательной среде вуза. *Информатика и образование*. 2022;37(4):54–63. EDN: ESVZJD. DOI: 10.32517/0234-0453-2022-37-4-54-63.

[Toktarova V. I., Popova O. G. An analysis of educational data on the correlation between learning success and students' behavior in the university digital educational environment. *Informatics and Education*. 2022;37(4):54–63. (In Russian.) EDN: ESVZJD. DOI: 10.32517/0234-0453-2022-37-4-54-63.]

17. Estacio R. R., Raga R. C. Analyzing students online learning behavior in blended courses using Moodle. *Asian Association of Open Universities Journal*. 2017;12(1):52–68. DOI: 10.1108/AAOUJ-01-2017-0016.

18. Garrison D. R., Anderson T., Archer W. Critical inquiry in a text-based environment: Computer conferencing in higher education. *The Internet and Higher Education*. 1999;2(2–3):87–105. DOI: 10.1016/S1096-7516(00)00016-6.

19. Глотова А. В. Организация групповой работы студентов в условиях онлайн-обучения в системе высшей школы. *Преподаватель XXI век*. 2021;(1–1):122–132. EDN: KDWKWE. DOI: 10.31862/2073-9613-2021-1-122-132.

[Glотова А. В. The issues of online group learning in higher education. *Prepodavatel XXI Vek*. 2021;(1–1):122–132. (In Russian.) EDN: KDWKWE. DOI: 10.31862/2073-9613-2021-1-122-132.]

20. Карманова Е. В., Киселев А. В. Возможности ClickHouse для обработки данных образовательного портала вуза. *Актуальные проблемы современной науки, техники и образования. Тезисы докладов 81-й междуна-*

*родной научно-технической конференции*. Магнитогорск: Магнитогорский государственный технический университет им. Г. И. Носова; 2023;1:447. EDN: YNNONO.

[Karmanova E. V., Kiselev A. V. The capabilities of ClickHouse for data processing of the educational portal of the university. *Current Problems of Modern Science, Technology and Education. Theses of the 81st Int. Scientific and Technical Conf.* Magnitogorsk, Nosov Magnitogorsk State Technical University; 2023;1:447. (In Russian.) EDN: YNNONO.]

21. Токтарова В. И., Пашкова Ю. А. Предиктивная аналитика в цифровом образовании: анализ и оценка успешности обучения студентов. *Сибирский педагогический журнал*. 2022;(1):97–106. EDN: NSPMDR. DOI: 10.15293/1813-4718.2201.09.

[Toktarova V. I., Pashkova Yu. A. Predictive analytics in digital education: Analysis and evaluation of students' learning success. *Siberian Pedagogical Journal*. 2022;(1):97–106. (In Russian.) EDN: NSPMDR. DOI: 10.15293/1813-4718.2201.09.]

#### Информация об авторах

**Карманова Екатерина Владимировна**, канд. пед. наук, доцент, доцент кафедры бизнес-информатики и информационных технологий, Институт энергетики и автоматизированных систем, Магнитогорский государственный технический университет им. Г. И. Носова, г. Магнитогорск, Челябинская область, Россия; *ORCID*: <https://orcid.org/0000-0003-1807-5388>; *e-mail*: [ev.karmanova@magtu.ru](mailto:ev.karmanova@magtu.ru)

**Захарова Яна Михайловна**, ассистент кафедры менеджмента и государственного управления, Институт экономики и управления, Магнитогорский государственный технический университет им. Г. И. Носова, г. Магнитогорск, Челябинская область, Россия; *ORCID*: <https://orcid.org/0000-0003-0498-3051>; *e-mail*: [yana\\_zakharova\\_2000@bk.ru](mailto:yana_zakharova_2000@bk.ru)

**Киселев Андрей Викторович**, магистрант направления «Прикладная информатика», профиль «Технологии Data Science», кафедра бизнес-информатики и информационных технологий, Институт энергетики и автоматизированных систем, Магнитогорский государственный технический университет им. Г. И. Носова, г. Магнитогорск, Челябинская область, Россия; *ORCID*: <https://orcid.org/0009-0003-2867-6758>; *e-mail*: [kiselev.andrew.168@gmail.com](mailto:kiselev.andrew.168@gmail.com)

#### Information about the authors

**Ekaterina V. Karmanova**, Candidate of Sciences (Education), Docent, Associate Professor at the Business Computing and IT Department, Institute of Power Engineering and Automated Systems, Nosov Magnitogorsk State Technical University, Magnitogorsk, Chelyabinsk Region, Russia; *ORCID*: <https://orcid.org/0000-0003-1807-5388>; *e-mail*: [ev.karmanova@magtu.ru](mailto:ev.karmanova@magtu.ru)

**Yana M. Zakharova**, Assistant at the Management and Public Administration Department, Institute of Economics and Management, Nosov Magnitogorsk State Technical University, Magnitogorsk, Chelyabinsk Region, Russia; *ORCID*: <https://orcid.org/0000-0003-0498-3051>; *e-mail*: [yana\\_zakharova\\_2000@bk.ru](mailto:yana_zakharova_2000@bk.ru)

**Andrei V. Kiselev**, a master's student of the educational program “Applied Informatics”, profile “Data Science”, The Business Computing and IT Department, Institute of Power Engineering and Automated Systems, Nosov Magnitogorsk State Technical University, Magnitogorsk, Chelyabinsk Region, Russia; *ORCID*: <https://orcid.org/0009-0003-2867-6758>; *e-mail*: [kiselev.andrew.168@gmail.com](mailto:kiselev.andrew.168@gmail.com)

**Поступила в редакцию / Received:** 05.02.24.

**Поступила после рецензирования / Revised:** 29.04.24.

**Принята к печати / Accepted:** 07.05.24.