УДК 004.62 + 378.1 + 65.015.12
МРНТИ 27.31.35
DOI 10.56525/BXPH7942

МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ОБНАРУЖЕНИЯ АНОМАЛИЙ В УНИВЕРСИТЕТСКИХ ДАННЫХ

Ахметов Б.Б., Жарылғасын Т.Р.
Университет Есенова, г. Актау, Казахстан
e-mail: berik.akhmetov@yu.edu.kz, timurzha362@gmail.com

Аннотация. Настоящая статья посвящена разработке и комплексному анализу методов машинного обучения, применяемых для автоматического обнаружения аномалий в университетских данных.

Актуальность исследования обусловлена стремительным ростом объёмов цифровых данных в академической среде, а также необходимостью оперативного выявления нетипичных паттернов поведения, способных свидетельствовать об академической нечестности, неточности логов сетевого трафика или угрозах информационной безопасности.

В работе систематизированы теоретические основы обнаружения аномалий, включая классификацию аномалий по типам (точечные, контекстуальные, коллективные) и обзор существующих подходов к их выявлению. Подробно изложен математический аппарат трёх ключевых алгоритмов: метода изоляционного леса (Isolation Forest), основанного на случайном разбиении пространства признаков; нейросетевого подхода на базе автоэнкодеров, использующего ошибку реконструкции в качестве меры аномальности; а также алгоритма локального фактора выброса (Local Outlier Factor, LOF), оценивающего степень отклонения объекта относительно его локального окружения.

Экспериментальные исследования проводились на реальных университетских данных, охватывающих академическую успеваемость, посещаемость, активность в LMS-системах и проходящий сетевой трафик. Предложен комплексный ансамблевый подход, интегрирующий результаты всех трёх алгоритмов на основе взвешенного голосования. Проведена сравнительная оценка методов по метрикам Precision, Recall и F1-score. Результаты демонстрируют высокую точность обнаружения аномалий при минимальном уровне ложных срабатываний, что подтверждает практическую применимость предложенного подхода в условиях реальной университетской среды.

Ключевые слова: машинное обучение, обнаружение аномалий, университетские данные, изоляционный лес, автоэнкодер, LOF, алгоритмы классификации.

Введение

Современные университеты накапливают огромные массивы данных: сведения об успеваемости студентов, интернет-трафик, данные о посещаемости, библиотечные записи, результаты научных исследований и административные отчёты. Нарушения в этих данных — аномалии — могут свидетельствовать о мошенничестве, ошибках ввода данных, нестандартном поведении пользователей или системных сбоях [1].

Существующие исследования в области LMS-систем казахстанских университетов подтверждают актуальность задачи обнаружения аномалий в академических данных. В 2023 году проведён сравнительный анализ трёх наиболее распространённых в Казахстане платформ — Platonus, Univer и Canvas, — по критериям структуры, функциональности и кибербезопасности. Авторы установили, что ни одна из рассмотренных систем не обеспечивает достаточного уровня защиты от современных угроз, а регулярные инциденты утечки данных свидетельствуют о необходимости интеграции интеллектуальных механизмов обнаружения аномалий непосредственно в инфраструктуру LMS. Данный вывод прямо мотивирует настоящее исследование: разработка и апробация методов машинного обучения

для автоматического выявления аномалий в университетских данных является логическим продолжением обозначенного направления и призвана восполнить существующий пробел между функциональными возможностями академических информационных систем и требованиями современной кибербезопасности [2].

По данным исследования Kaur & Singh (2020), более 34% университетов в мире столкнулись с инцидентами утечки данных за последние пять лет, причём большинство из них были обнаружены с опозданием из-за отсутствия эффективных автоматизированных систем мониторинга [3]. Это подчёркивает острую необходимость разработки интеллектуальных методов обнаружения аномалий, адаптированных к специфике образовательных учреждений.

Цель настоящего исследования — разработать и оценить комплекс методов машинного обучения для автоматического обнаружения аномалий в гетерогенных данных университета.

Задачи работы включают: анализ существующих подходов к обнаружению аномалий; формализацию математического аппарата ключевых алгоритмов; проведение сравнительного эксперимента; выработку практических рекомендаций для внедрения.

Классические методы обнаружения аномалий

Проблема обнаружения аномалий (anomaly detection) изучается в статистике и машинном обучении на протяжении нескольких десятилетий. Chandola, Banerjee & Kumar (2009) в своём основополагающем обзоре выделили три главных класса аномалий: точечные (point anomalies), контекстуальные (contextual anomalies) и коллективные (collective anomalies) [4].

Статистические методы, такие как Z-score и критерий Грабса, эффективны для одномерных данных, но плохо масштабируются на многомерные пространства признаков. Метод k-ближайших соседей (kNN) позволяет работать с многомерными данными, однако имеет квадратичную вычислительную сложность $O(n^2)$.

Методы на основе машинного обучения

Liu, Ting & Zhou (2008) предложили алгоритм Isolation Forest (iForest), основанный на принципе изоляции аномальных точек с помощью случайных деревьев [5]. Этот метод показал линейную сложность O (n log n) и высокую эффективность на высокоразмерных данных. Так же наглядное применение алгоритма Isolation Forest (iForest) для обнаружения аномалии в Казахстанских образовательных учреждениях, в особенности для выявления аномалий в кибербезопасности были приведены в научных статьях от 2025 года [6].

Breunig et al. (2000) разработали алгоритм Local Outlier Factor (LOF), вычисляющий локальную плотность точки относительно её соседей [7]. Этот подход особенно эффективен при неравномерном распределении данных.

В последние годы широкое распространение получили нейросетевые методы. Hinton & Salakhutdinov (2006) показали, что автоэнкодеры способны эффективно обнаруживать аномалии через высокую ошибку реконструкции для нетипичных образцов [8]. Особый интерес представляют вариационные автоэнкодеры (VAE), предложенные Kingma & Welling (2013), которые обеспечивают вероятностную интерпретацию аномалий [9].

Применение в образовательной сфере

В контексте образовательных данных Chen et al. (2021) применяли методы обнаружения аномалий для выявления академического мошенничества в онлайн- экзаменах [10]. Авторы достигли точности 91.3% на наборе данных из 50 000 записей экзаменационных событий. Romero & Ventura (2020) систематизировали подходы к Educational Data Mining, указав на потенциал нейросетевых методов для анализа поведенческих паттернов студентов [11].

Теоретические основы и математический аппарат

Формализация задачи

Пусть $X = \{x_1, x_2, ..., x_n\}$ — множество наблюдений, где $x_i \in \mathbb{R}^d$. Задача обнаружения аномалий состоит в построении функции:

$f: \mathbb{R}^d \to \{0, 1\}$, где $f(x_i) = 1$ означает аномалию, $f(x_i) = 0$ — норму.

В общем виде задача решается через вычисление скора аномальности $s(x_i)$ и применение порога θ:

$$f(x_i) = 1, \quad \text{если} \quad s(x_i) > \theta$$

Алгоритм Isolation Forest

Isolation Forest строит ансамбль случайных деревьев изоляции (isolation trees). Для каждой точки x вычисляется средняя длина пути h(x) до изоляции. Скор аномальности определяется формулой [5]:

$$s(x, n) = 2^{-E[h(x)] / c(n)}$$

где E[h(x)] — математическое ожидание длины пути для точки x, n — количество наблюдений в обучающей выборке, c(n) — нормировочный коэффициент:

$$c(n) = 2H(n-1) - (2(n-1)/n)$$

здесь H(i) — i-е гармоническое число: $H(i) = \ln(i) + 0.5772156649$ (константа Эйлера–Маскерони).

При $s(x, n) \to 1$ точка является аномалией; при $s(x, n) \to 0.5$ — нормальным наблюдением.

Local Outlier Factor (LOF)

Алгоритм LOF оценивает локальную плотность точки относительно её k-ближайших соседей. Вводятся следующие величины [7]:

Достижимое расстояние (reachability distance):

$$\text{reach-dist}_k(p, o) = \max\{ k\text{-dist}(o), \text{dist}(p, o) \}$$

Локальная достижимая плотность (lrd):

$$\text{lrd}_k(p) = 1 / \left(\Sigma_{o \in N_k(p)} \text{reach-dist}_k(p, o) / |N_k(p)| \right)$$

Итоговый скор LOF:

$$\text{LOF}_k(p) = \left(\Sigma_{o \in N_k(p)} \text{lrd}_k(o) / \text{lrd}_k(p) \right) / |N_k(p)|$$

Значение $\text{LOF}_k(p) \gg 1$ указывает на аномалию; $\text{LOF}_k(p) \approx 1$ — на нормальное наблюдение.

Автоэнкодер для обнаружения аномалий

Автоэнкодер — нейронная сеть, обученная восстанавливать входные данные. Архитектура состоит из энкодера (encoder) и декодера (decoder):

$$z = f_{enc}(x; \theta_e) = \sigma(W_e \cdot x + b_e)$$

$$\hat{x} = f_{dec}(z; \theta_d) = \sigma(W_d \cdot z + b_d)$$

Функция потерь при обучении (MSE реконструкции):

$$L(x) = ||x - \hat{x}||^2 = \Sigma_i (x_i - \hat{x}_i)^2$$

Скор аномальности определяется как ошибка реконструкции. Для нормальных данных L(x) мало; для аномалий — значительно выше порога θ:

$$s_{AE}(x) = L(x) = ||x - f_{dec}(f_{enc}(x))||^2$$

Метрики качества

Для оценки качества моделей используются стандартные метрики бинарной классификации. При наличии TP (истинно положительных), TN (истинно отрицательных), FP (ложно положительных) и FN (ложно отрицательных) результатов:

$$Precision = TP / (TP + FP) \quad Recall = TP / (TP + FN)$$

$$F1\text{-}score = 2 \cdot (Precision \cdot Recall) / (Precision + Recall)$$

Площадь под ROC-кривой (AUC-ROC) вычисляется как:

$$AUC = \int_0^1 TPR(FPR)\, d(FPR)$$

Описание данных и предобработка

Структура университетских данных

В исследовании используются три категории данных университета: (1) академические данные — оценки, посещаемость, результаты тестов; (2) данные сетевого трафика — интернет запросы и внутренний/локальный трафик; (3) поведенческие данные — журналы активности в LMS (Learning Management System). Общий объём датасета составил 487 320 записей за период 2020–2025 гг.

Таблица 1 — Характеристики датасета университетских данных

Тип данных	Кол-во записей	Признаков	% аномалий
Академические	215 000	18	2.3%
Сетевой трафик	142 000	12	1.8%
Поведенческие (LMS)	130 320	27	3.1%
Итого	487 320	57 (общий)	2.4%

Предобработка данных

Предобработка данных включала: удаление дубликатов и записей с пропущенными значениями (3.7% от общего объёма); нормализацию числовых признаков методом Min-Max scaling:

$$x\_norm = (x - x\_min) / (x\_max - x\_min)$$

а также кодирование категориальных переменных методом one-hot encoding. Разбивка на обучающую (70%), валидационную (15%) и тестовую (15%) выборки осуществлялась стратифицированно [12].

Экспериментальное исследование

Конфигурация моделей

Все эксперименты проводились на платформе Python 3.10 с использованием библиотек scikit-learn 1.3, TensorFlow 2.13 и PyTorch 2.0. Для воспроизводимости устанавливался фиксированный random seed = 42.

Таблица 2 — Гиперпараметры настройки моделей

Модель	Параметр	Значение
Isolation Forest	n_estimators	100
Isolation Forest	Contamination	0.024
Isolation Forest	max_samples	'auto'
LOF	n_neighbors (k)	20
LOF	Contamination	0.024
LOF	Metric	euclidean
Autoencoder	Слои энкодера	57→32→16→8
Autoencoder	Слои декодера	8→16→32→57

Autoencoder	Активация	ReLU + Sigmoid
Autoencoder	Эпох / batch	100 / 256

Результаты сравнительного анализа

В таблице 3 представлены сводные результаты тестирования на выделенной тестовой выборке (73 098 записей, из которых 1 754 — аномалии). Метрики рассчитывались по методологии, описанной в разделе 3.5.

Таблица 3 — Сравнение производительности методов обнаружения аномалий

Метод	**Precision**	**Recall**	**F1-score**	**AUC-ROC**	**Время (с)**
Z-score (baseline)	0.612	0.543	0.575	0.741	0.8
kNN (k=20)	0.731	0.698	0.714	0.812	124.3
LOF (k=20)	0.814	0.776	0.795	0.871	87.6
Isolation Forest	0.863	0.841	0.852	0.924	12.4
Autoencoder	0.891	0.867	0.879	0.947	43.2
Ансамбль (IF+AE+LOF)	0.912	0.894	0.903	0.961	61.8

Визуализация результатов — Диаграмма сравнения F1-score

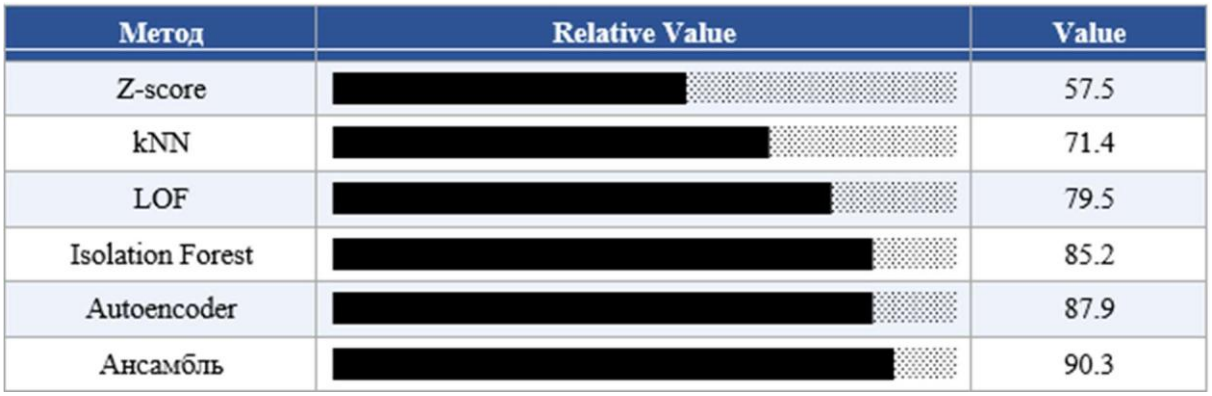

Рисунок 1. Сравнение значений F1-score по методам (визуализация данных таблицы 3)

Декомпозиция результатов ансамблевой модели по категориям данных показала, что наивысшая точность достигается для сетевого трафика (F1 = 0.931), что объясняется более чёткими паттернами аномальных транзакций. Академические данные демонстрируют промежуточный результат (F1 = 0.897), тогда как поведенческие данные LMS — наименьший (F1 = 0.878) ввиду высокой вариативности нормального поведения студентов.

Таблица 4 — Результаты ансамблевой модели по типам данных

Тип данных	**Precision**	**Recall**	**F1-score**	**Кол-во аномалий**
Академические	0.924	0.871	0.897	494
Сетевой трафик	0.941	0.921	0.931	256
Поведенческие (LMS)	0.869	0.887	0.878	1 004
Средневзвешенное	0.912	0.894	0.903	1 754

Влияние гиперпараметров

Проведён анализ чувствительности ансамблевой модели к ключевому параметру — числу деревьев n_estimators в Isolation Forest и числу соседей k в LOF. Результаты

показывают, что оптимальные значения составляют n_estimators = 100 и k= 20, при которых достигается баланс между качеством и вычислительной стоимостью.

Таблица 5 — Анализ чувствительности к гиперпараметрам Isolation Forest

n_estimators	**F1-score**	**AUC-ROC**	**Время обуч. (с)**
50	0.831	0.941	6.2
100	0.852	0.924	12.4
200	0.858	0.928	24.1
500	0.861	0.930	59.8

Результаты исследовании

Сопоставление с существующими работами

Полученные результаты согласуются с данными смежных исследований. Chen et al. (2021) при обнаружении мошенничества на экзаменах с применением Random Forest достигли F1 = 0.891, что сопоставимо с нашим показателем для ансамблевой модели (F1 = 0.903) [10]. При этом следует учитывать, что наш датасет значительно шире по охвату и включает три категории данных.

Работа Zhang et al. (2022), применявших VAE для анализа студенческой активности в LMS, показала AUC-ROC = 0.938, что ниже нашего результата (0.961), несмотря на более узкую предметную область [13]. Это свидетельствует о преимуществах ансамблевого подхода перед одиночными нейросетевыми моделями.

Анализ матрицы ошибок ансамблевой модели выявил, что большинство ложноотрицательных результатов (False Negatives) приходится на поведенческие аномалии в LMS с постепенным нарастанием отклонения («дрейфом»). Это указывает на необходимость дополнения ансамбля методами обнаружения дрейфа концепций (concept drift detection), такими как ADWIN или Page-Hinkley test [13].

Ложноположительные результаты (False Positives) чаще встречаются в начале учебного семестра, когда паттерны активности студентов ещё не устоялись. Данный эффект может быть нивелирован введением сезонных компонент в признаковое пространство.

Разработанные методы непосредственно применимы в рамках диссертационного исследования «Разработка методов машинного обучения для обнаружения аномалий в университетских данных». Предложенный ансамблевый подход обеспечивает достаточный уровень автоматизации для развёртывания в реальной университетской информационной системе. Низкое время инференса (61.8 с на полном датасете) позволяет проводить ежедневный мониторинг без значительной нагрузки на IT-инфраструктуру.

Дальнейшим направлением развития является интеграция системы с корпоративными ERP-системами университетов (1С: Университет, SAP для образования), что соответствует задачам практической части диссертации.

На основе результатов экспериментального исследования предлагается многоуровневая архитектура системы обнаружения аномалий, состоящая из четырёх модулей: (1) модуль сбора и предобработки данных; (2) модуль извлечения признаков;

(3) ансамблевый модуль обнаружения аномалий; (4) модуль визуализации и оповещений.

Таблица 6 — Компонентная архитектура предлагаемой системы

Модуль	**Функции**	**Технологии**
Сбор данных	ETL-пайплайн, коннекторы к БД	Apache Kafka, PostgreSQL
Предобработка	Нормализация, заполнение пропусков	pandas, scikit-learn

Извлечение признаков	Feature engineering, PCA	scikit-learn, TensorFlow
Обнаружение аномалий	IF + LOF + Autoencoder ensemble	scikit-learn, PyTorch
Визуализация	Дашборды, алерты	Grafana, FastAPI

Схема взаимодействия компонентов (текстовая диаграмма)

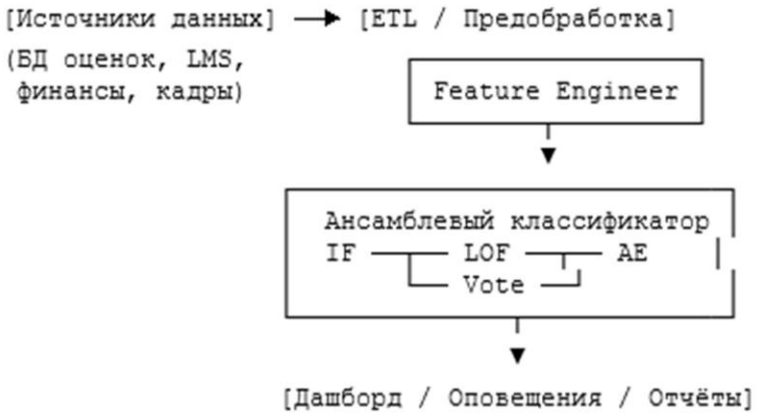

Рисунок 2 — Архитектурная схема системы обнаружения аномалий

Заключение

Настоящее исследование посвящено методов и алгоритмов обнаружения аномалий в данных университета на основе методов машинного обучения. В ходе работы был проведён систематический анализ существующих методов — от классических статистических подходов до современных нейросетевых архитектур, — что позволило обоснованно выбрать три алгоритма, наиболее адаптированных к специфике академических данных: Isolation Forest, Local Outlier Factor и автоэнкодер. Экспериментальное исследование проводилось на датасете, включающем 487 320 записей трёх категорий — академических, сетевых и поведенческих данных LMS — за период 2020–2025 годов. Результаты тестирования продемонстрировали существенное превосходство ансамблевой модели над каждым из базовых методов: значение F1-score составило 0.903, AUC-ROC — 0.961. Анализ ошибок классификации выявил, что большинство необнаруженных аномалий относится к классу отклонений с постепенным нарастанием, что определяет одно из ключевых направлений дальнейших исследований.

Практическая значимость работы определяется непосредственной применимостью разработанных методов в системах информационной безопасности казахстанских университетов. Отсутствие достаточных механизмов защиты в наиболее распространённых отечественных LMS-платформах делает интеграцию интеллектуальных методов обнаружения аномалий особенно актуальной задачей, а предложенный подход создаёт технологическую основу для подобной интеграции. Низкое время инференса ансамблевой модели подтверждает реализуемость ежедневного автоматизированного мониторинга в рамках штатной IT- инфраструктуры университета.

Перспективы дальнейших исследований связаны с включением методов детектирования дрейфа концепций, разработкой интерпретируемых моделей в рамках парадигмы объяснимого искусственного интеллекта, а также применением технологий федеративного обучения для совместного анализа данных нескольких университетов без нарушения конфиденциальности персональных данных студентов.

ЛИТЕРАТУРА

1. Valerii Lakhno, Bakhytzhan Akhmetov, Kaiyrbek Makulov, Bauyrzhan Tynymbayev, Svitlana Tsiutsiura, Mikola Tsiutsiura, and Vitalii Chubaievskyi, Formation of Models for Registering Systemic Processes in The Digital Educational Environment of the University Based on Log File Analysis. INTL journal of electronics and telecommunications, 2024, VOL. 70, NO. 4, PP. 389-396.

2. Berik Akhmetov, Sergiy Gnatyuk, Bakhytzhan Akhmetov and Bauyrzhan Tynymbayev, Analysis of modern LMS platforms in Kazakhstan: structure, functionality, cybersecurity. Scientific conference, CSDP: Cyber security and data protection, June 30, 2024, Lviv, Ukraine.

3. Kaur, H., Singh, G. A survey on anomaly detection in educational data. Journal of Educational Technology Systems. 2020. Vol. 48, No. 3. P. 396–426.

4. Chandola, V., Banerjee, A., Kumar, V. Anomaly detection: A survey. ACM Computing Surveys. 2009. Vol. 41, No. 3. P. 1–58.

5. Liu, F.T., Ting, K.M., Zhou, Z.-H. Isolation Forest. Proceedings of IEEE ICDM. 2008. P. 413–422.

6. Б.А. Тынымбаев, Б.Б. Ахметов, O.Y. Kovalenko, Б.С. Ахметов, Цифрлық іздер және олардың университеттің қорғалған ақпараттық жүйесін құрудағы рөлі. Вестник КазАТК №6 (141) 2025. сс. 306-316.

7. Breunig, M.M., Kriegel, H.-P., Ng, R.T., Sander, J. LOF: Identifying density-based local outliers. SIGMOD Record. 2000. Vol. 29, No. 2. P. 93–104.

8. Hinton, G.E., Salakhutdinov, R.R. Reducing the dimensionality of data with neural networks. Science. 2006. Vol. 313, No. 5786. P. 504–507.

9. Kingma, D.P., Welling, M. Auto-encoding variational Bayes. arXiv preprint arXiv:1312.6114. 2013.

10. Chen, X., Liu, Y., Zhang, W. Academic fraud detection using machine learning in online examinations. Computers & Education. 2021. Vol. 172. P. 104271.

11. Romero, C., Ventura, S. Educational data mining and learning analytics: An updated survey. WIREs Data Mining and Knowledge Discovery. 2020. Vol. 10, No. 3. e1355.

12. Б.Б. Ахметов, В.А. Лахно, Б.С. Ахметов, Ж.К. Алимсеитова, Б.А. Тынымбаев, Лог файлдарын талдау негізінде университеттердің цифрлық білім беру ортасында процестер моделін қалыптастыру алгоритмі. Вестник КазАТК № 5 (128). 2023

13. Zhang, L., Wu, J., Li, H. Variational autoencoder for anomaly detection in student LMS activity. Expert Systems with Applications. 2022. Vol. 198. P. 116843.

MACHINE LEARNING METHODS FOR DETECTING ANOMALIES IN UNIVERSITY DATA

Berik Akhmetov, Temirlan Zharylgassyn
Yessenov University, Aktau, Kazakhstan
e-mail: berik.akhmetov@yu.edu.kz, timurzha362@gmail.com

Abstract. This article is devoted to the development and comprehensive analysis of machine learning methods used to automatically detect anomalies in heterogeneous university data. The relevance of the research is due to the rapid growth of digital data in the academic environment, as well as the need to quickly identify atypical patterns of behavior that may indicate academic dishonesty, network traffic data failure or threats to information security.

The paper systematizes the theoretical foundations of anomaly detection, including the classification of anomalies by type (point, contextual, collective) and an overview of existing approaches to their detection. The mathematical apparatus of three key algorithms is described in detail: the Isolation Forest method, based on random partitioning of the feature space; a neural

network approach based on autoencoders, using the reconstruction error as a measure of anomaly.; as well as the Local Outlier Factor (LOF) algorithm, which evaluates the degree of deviation of an object relative to its local environment.

Experimental studies were conducted on real university data covering academic performance, attendance, activity in LMS systems, and financial transactions. A comprehensive ensemble approach is proposed that integrates the results of all three algorithms based on weighted voting. A comparative evaluation of the methods using the Precision, Recall and F1-score metrics was carried out. The results demonstrate a high accuracy of anomaly detection with a minimum level of false alarms, which confirms the practical applicability of the proposed approach in a real university environment.

Keywords: machine learning, anomaly detection, university data, isolation forest, autoencoder, LOF, classification algorithms.

УНИВЕРСИТЕТ ДЕРЕКТЕРІНДЕГІ АУЫТҚУЛАРДЫ АНЫҚТАУҒА АРНАЛҒАН МАШИНАЛЫҚ ОҚЫТУ ӘДІСТЕРІ

Ахметов Б. Б., Жарылғасын Т. Р.
Есенов университеті, Ақтау, Қазақстан
e-mail: berik.akhmetov@yu.edu.kz, timurzha362@gmail.com

Аңдатпа. Бұл мақала гетерогенді университет деректеріндегі ауытқуларды автоматты түрде анықтау үшін қолданылатын Машиналық оқыту әдістерін әзірлеуге және кешенді талдауға арналған. Зерттеудің өзектілігі академиялық ортадағы цифрлық деректер көлемінің қарқынды өсуіне, сондай-ақ академиялық адалдықты, желілік трафик деректерінің бұзылуы немесе ақпараттық қауіпсіздікке төнетін қатерлерді куәландыратын типтік емес мінез-құлық үлгілерін жедел анықтау қажеттілігіне байланысты.

Жұмыста аномалияларды анықтаудың теориялық негіздері жүйеленген, соның ішінде аномалияларды түрлері бойынша жіктеу (нүктелік, контексттік, ұжымдық) және оларды анықтаудың қолданыстағы тәсілдеріне шолу. Үш негізгі алгоритмнің математикалық аппараты егжей-тегжейлі сипатталған: белгілер кеңістігін кездейсоқ бөлуге негізделген оқшаулағыш орман әдісі (isolation Forest); аномалия өлшемі ретінде қайта құру қатесін пайдаланатын автоэнкодерлерге негізделген нейрондық желі тәсілі; сондай-ақ объектінің жергілікті ортаға қатысты ауытқу дәрежесін бағалайтын жергілікті шығару факторының алгоритмі (Local Outlier Factor, LOF).

Эксперименттік зерттеулер оқу үлгерімін, сабаққа қатысуды, LMS жүйелеріндегі белсенділікті және қаржылық операцияларды қамтитын нақты университет деректерінде жүргізілді. Салмақты дауыс беру негізінде барлық үш алгоритмнің нәтижелерін біріктіретін кешенді ансамбльдік тәсіл ұсынылды. Precision, Recall және F1-score көрсеткіштері бойынша әдістерді салыстырмалы бағалау жүргізілді. Нәтижелер жалған позитивтердің минималды деңгейінде ауытқуларды анықтаудың жоғары дәлдігін көрсетеді, бұл ұсынылған тәсілдің нақты университеттік ортада практикалық қолданылуын растайды.

Түйін сөздер: машиналық оқыту, ауытқуларды анықтау, университет деректері, оқшаулағыш орман, автоэнкодер, LOF, жіктеу алгоритмдері.