УДК 004.89:336.717

5.2.2 Математические, статистические и инструментальные методы в экономике

СРАВНИТЕЛЬНЫЙ АНАЛИЗ СТАТИСТИЧЕСКИХ И МАШИННО-ОБУЧАЮЩИХ МЕТОДОВ ВЫЯВЛЕНИЯ ЭКОНОМИЧЕСКИ ЗНАЧИМЫХ АНОМАЛИЙ В ДАННЫХ КОРПОРАТИВНЫХ ИНФОРМАЦИОННЫХ СИСТЕМ

Курдаев О.И.
магистрант 2 курса, направление «Информационные системы и технологии»

Алашеев В.В.
канд. техн. наук, доцент кафедры компьютерных технологий и систем
ФГБОУ ВО «Кубанский государственный аграрный университет имени И.Т. Трубилина», Краснодар, Российская Федерация

Сравниваются статистический MAD-критерий, Isolation Forest, XGBoost, автокодировщик последовательностей и каскады MAD+IF и MAD+IF+SeqAE на данных корпоративных информационных систем. Эксперименты проведены на наборе ULB Credit Card Fraud (0,17 % мошеннических операций) и на сессионном наборе, собранном агрегацией журнала платёжной КИС (4,3 % аномалий). Оценка выполнена по F1, PR-AUC и функция ущерба с разными весами ошибок. При наличии разметки лучший результат даёт XGBoost (F1 = 0,81 и 0,77); среди методов без учителя на транзакциях предпочтителен Isolation Forest без SMOTE (F1 = 0,17 против 0,12 с балансировкой). Каскад MAD+IF после подбора порогов на валидации сокращает объём записей, передаваемых на углублённый анализ, на 58 % относительно полного потока; добавление SeqAE качество не повышает, но увеличивает время обработки. Полученные выводы могут использоваться при построении систем внутреннего контроля и финансового мониторинга

Ключевые слова: ЭКОНОМИЧЕСКИЕ АНОМАЛИИ, СТАТИСТИЧЕСКИЕ МЕТОДЫ, МАШИННОЕ ОБУЧЕНИЕ, КОРПОРАТИВНЫЕ ИНФОРМАЦИОННЫЕ СИСТЕМЫ, ОБНАРУЖЕНИЕ МОШЕННИЧЕСТВА, ИЗОЛЯЦИОННЫЙ ЛЕС, ГРАДИЕНТНЫЙ БУСТИНГ, НЕСБАЛАНСИРОВАННЫЕ ДАННЫЕ, ВНУТРЕННИЙ АУДИТ, ФИНАНСОВЫЙ МОНИТОРИНГ

http://dx.doi.org/10.21515/1990-4665-220-055

UDC 004.89:336.717

5.2.2 Mathematical, statistical and instrumental methods in economics

COMPARATIVE ANALYSIS OF STATISTICAL AND MACHINE LEARNING METHODS FOR IDENTIFYING ECONOMICALLY SIGNIFICANT ANOMALIES IN CORPORATE INFORMATION SYSTEM DATA

Kurdaev O.I.
2nd year Master's student, direction "Information systems and technologies"

Alasheev V.V.
Candidate of Technical Sciences, Associate Professor of the Department of Computer Technologies and Systems
Kuban State Agrarian University named after I.T. Trublin, Krasnodar, Russian Federation

Statistical MAD, Isolation Forest, XGBoost, a sequence autoencoder (SeqAE), and two- and three-stage MAD+IF and MAD+IF+SeqAE cascades are compared on transaction and session data resembling corporate IS logs. Experiments are conducted on the ULB Credit Card Fraud dataset (0.17% fraud) and on a session dataset aggregated from payment CIS transactions (4.3% anomalies). With labeled data, XGBoost achieves F1 = 0.81 and 0.77; among unsupervised methods, Isolation Forest without SMOTE performs best on transactions (F1 = 0.17 vs. 0.12 with SMOTE). After validation tuning, the MAD+IF cascade reduces the data volume for in-depth analysis by about 58% relative to scoring the full stream; adding SeqAE does not improve F1 but increases latency. The findings can be used in internal control and financial monitoring systems

Keywords: ECONOMIC ANOMALIES, STATISTICAL METHODS, MACHINE LEARNING, CORPORATE INFORMATION SYSTEMS, FRAUD DETECTION, ISOLATION FOREST, GRADIENT BOOSTING, IMBALANCED DATA, INTERNAL AUDIT, FINANCIAL MONITORING

Введение

Корпоративные информационные системы – ERP, платёжные модули, системы электронного документооборота – ежедневно формируют большие журналы операций. Суммы проводок, время транзакций, последовательность действий пользователя несут прямую экономическую нагрузку: ошибка или злоупотребление быстро превращаются в финансовый ущерб, штраф регулятора или репутационные потери [10]. Классические средства защиты по сигнатурам хорошо отрабатывают известные угрозы, но хуже справляются с ситуацией, когда сотрудник или внешний злоумышленник пользуется штатным доступом и действует «внутри регламента» [1, 4].

Практика показывает три типичных затруднения. Доля подозрительных событий в реальных потоках редко превышает 0,2 %, поэтому доля верных ответов как метрика почти бесполезна. Отдельная операция может выглядеть нормальной, тогда как аномальной оказывается цепочка действий – значит, нужен учёт временной структуры. Нормальное поведение пользователей со временем меняется, и модель, обученная на старых данных, теряет точность [1]. На платёжных потоках это подтверждают и отраслевые исследования: при смене поведения клиентов устойчивее оказываются не одиночные классификаторы, а ансамблевые схемы с учётом прошлых концепций [13]. В такой постановке разумно оценивать алгоритмы по F1, PR-AUC и функции ущерба, где штраф за пропуск мошенничества и за ложную тревогу задаётся по-разному [12].

Цель статьи – сравнить статистические и машинно-обучающие подходы к выявлению экономически значимых отклонений в данных КИС в рамках специальности 5.2.2 (п. 9, 10, 13 паспорта). Мы формализуем задачу через минимизацию ожидаемого ущерба, описываем единый протокол эксперимента, получаем численные оценки на двух наборах данных и

проверяем, насколько каскады MAD+IF и MAD+IF+SeqAE сокращают поток кандидатов и как это соотносится с метриками качества и временем обработки.

В отличие от разрозненных публикаций по отдельным алгоритмам, здесь все методы прогоняются в одинаковых условиях разбиения выборки и подбора порога. Отдельно измеряется, насколько предварительная фильтрация MAD и Isolation Forest сокращает поток, поступающий на более тяжёлые модели – это напрямую связано с нагрузкой на службу внутреннего контроля.

Обзор литературы

Montgomery [1] обобщает подходы к поиску выбросов и подчёркивает, что универсального алгоритма нет – выбор зависит от структуры данных и того, что именно считается ошибкой классификации. На первом этапе обработки потока часто достаточно простой статистики: Z-критерия, межквартильного размаха или модифицированного Z-критерия на базе MAD. Yaro et al. [2] для непрерывных данных показывают снижение ложных срабатываний при робастной оценке масштаба, но их схема не работает с зависимостями между соседними событиями.

Isolation Forest оценивает аномальность по глубине изоляции точки в случайном дереве и не требует разметки. Prajesha, Veni [3] комбинируют его с робастным масштабированием и методом главных компонент; Al-Shehari et al. [4] применяют его к внутренним угрозам при доле атак около 0,05 %; при этом синтетическое уравнивание классов через SMOTE, наоборот, ухудшает результат – в литературе это обсуждают как «парадокс изоляции» [5]. Там, где разметка есть, Bhattacharyya et al. [10] на транзакционных данных получают высокое качество ансамблевых моделей, в том числе градиентного

бустинга; Paldino et al. [13] показывают, что при дрейфе поведения клиентов ансамблевые схемы устойчивее одиночных классификаторов на потоках платёжных транзакций. По данным [13], на выборках порядка 30–50 млн транзакций ансамбль с критерием разнообразия моделей лучше учитывает смену поведения клиентов (concept drift), сохраняя знание о прошлых схемах мошенничества, чем одиночный классификатор без такой «памяти» – это согласуется с задачей периодического переобучения моделей внутреннего контроля.

Последовательные аномалии – когда «нормальные» по отдельности шаги в сумме образуют подозрительный сценарий – чаще ищут автокодировщиками, в том числе LSTM-AE [6, 7]. Elkan [12] отдельно показывает, что при разной цене пропуска мошенничества и ложной тревоги оптимизировать нужно взвешенную функцию потерь, а не долю верных ответов. Обзор Phua et al. [9] фиксирует устойчивый интерес экономических приложений к выявлению мошенничества; в отечественной практике мониторинг опирается и на требования ФЗ-115 [11].

Сопоставить статистику, методы, основанные на изоляции, бустинг и автокодировщик на одних и тех же журналах КИС с учётом ущерба в опубликованных работах нам не удалось – этот пробел и закрывает экспериментальная часть статьи.

Методология

Экономическая постановка задачи

Рассмотрим поток событий в журнале КИС. Событию i сопоставляется вектор признаков $x_i \in R^d$ (сумма, время, тип операции, частота действий пользователя и т.п.) и метка $y_i \in \{0; 1\}$, где единица означает операцию,

требующую реакции службы контроля – мошенничество, нарушение регламента, несанкционированное действие. Классификатор минимизирует не абстрактную ошибку, а ожидаемый ущерб:

$$Cost = C_{FN} \cdot FN + C_{FP} \cdot FP \quad (1)$$

где FN и FP – числа пропусков и ложных тревог; C_{FN} – оценка ущерба от пропущенного инцидента, C_{FP} – затраты на проверку «холостого» срабатывания. В расчётах положили C_{FP} = 1, а C_{FN} варьировали: 10, 100 и 1000 – так моделируются разные соотношения «цена пропуска / цена проверки».

Наборы данных и протокол эксперимента

Сравнение выполнено на двух открытых наборах, которые мы трактуем как упрощённые модели потоков КИС. Сводные характеристики – в таблице 1.

Таблица 1 – Характеристика эталонных наборов данных

Набор	**Экономическая интерпретация**	**n (test)**	**Признаков**	**Доля аномалий, %**	**Источник**
Credit Card Fraud	Транзакции платежной КИС	15 000	30	0,17	ULB MLG [10]
Payment IS sessions	Сессия (окна по 30 операций)	1 424	91	4,32	Агрегация журнала [10]

Код на Python 3.14 (scikit-learn, XGBoost 2.x). Разбиение 70 % / 15 % / 15 % (обучение / валидация / тест), random_state = 42. Для транзакций – стратификация по классу, для сессий – хронологический порядок: модель учится на «прошлом» и проверяется на «будущем». SMOTE в основных прогонах не использовали; отдельно прогнали Isolation Forest с SMOTE, чтобы зафиксировать эффект балансировки.

Формирование сессионного набора Payment IS sessions. Транзакционный журнал ULB [10] отсортирован по полю Time. Непрерывные блоки по W = 30 операций образуют сессию i. Для вектора признаков одной транзакции x (30 числовых полей: V1–V28, Time, Amount) вычисляются агрегаты по окну: среднее $\bar{x_i}$, стандартное отклонение s_i и максимум m_i по каждому признаку. Итоговый вектор сессии $z_i = concat(\bar{x_i}, s_i, m_i, W) \in R^{91}$. Метка сессии $y_i = max_t Class_{i,t}$ (аномальная, если хотя бы одна операция в окне мошенническая). Получено 9 493 сессии (4,32 % аномалий); при хронологическом разбиении 70/15/15 на тесте – 1 424 объекта. Это упрощённая модель сессии платёжной КИС, а не выгрузка промышленной ERP.

Описание алгоритмов

Модифицированный Z-критерий (MAD). Для числового признака считаем отклонение от медианы:

$$M_i = \frac{0{,}6745 \cdot |x_i - \tilde{x}|}{\text{MAD}}, \quad \text{MAD} = \text{median}_j \left| x_j - \tilde{x} \right| \quad (2)$$

где $\tilde{x}$ – медиана признака по обучающей выборке. Коэффициент 0,6745 приводит MAD к масштабу стандартного отклонения σ при нормальном распределении $\left(\frac{1}{\Phi^{-1}(0{,}75)}\right)$; порог $|M_i| > 3{,}5$ соответствует классическому модифицированному Z-критерию [1, 2]. Для распределения с тяжёлыми хвостами финансовых данных MAD остаётся робастной оценкой масштаба; при сильной асимметрии порог уточняют на валидации (как и для остальных методов). Аномалия фиксируется, если $|M_i| > 3{,}5$ хотя бы по одному признаку (берём максимум по признакам). Считается быстро и годится как первый грубый фильтр.

Isolation Forest. Оценка аномальности через глубину изоляции в случайном лесу:

$$s(x,n) = 2^{-\frac{E[h(x)]}{c(n)}}, \quad c(n) = 2H_{n-1} - \frac{2(n-1)}{n} (3)$$

Использовали 100 деревьев и подвыборку 256 объектов. Разметка не нужна.

XGBoost. Градиентный бустинг с логистической потерей; вес редкого класса задаётся через scale_pos_weight. Параметры: 200 деревьев, скорость обучения 0,05, глубина 6. Метод обучения с учителем – без истории подтверждённых мошенничеств не обучить. Для разбора вклада признаков в решение модели в службе контроля применяют методы объяснимого машинного обучения, в частности SHAP [8].

Автокодировщик последовательностей (SeqAE). На окнах длины T = 10 обучали MLP-автокодировщик минимизировать ошибку восстановления:

$$\mathcal{L}(\theta) = \frac{1}{T}\sum_{t=1}^{T} |\widehat{x^{(t)}} - x^{(t)}|_2^2 + \lambda|\theta|_2^2 (4)$$

Метрики. Кроме Cost (1) считали F1, точность, полноту и PR-AUC:

$$F_1 = \frac{2PR}{P+R}, \quad P = \frac{TP}{TP+FP}, \quad R = \frac{TP}{TP+FN} (5)$$

Порог для каждого алгоритма подбирали на валидации по максимуму F1.

Гибридная каскадная архитектура

На рис. 1 показана целевая трёхступенчатая схема MAD → IF → SeqAE. В эксперименте реализованы и оценены отдельно: MAD only – статистический фильтр; IF only – Isolation Forest по всему потоку; MAD+IF – объединение кандидатов: запись попадает в подмножество, если её MAD-

score ≥ q_{MAD} или нормированный IF-score ≥ q_{IF}; MAD+IF+SeqAE – к MAD+IF-кандидатам дополнительно применяется SeqAE; итоговый score – maximum(MAD, IF, SeqAE) внутри маски кандидатов.

Доли отсечения q_{MAD} и q_{IF} (квантили 0–1) подбирались на валидационной выборке перебором с шагом 0,01 по сетке $q_{MAD} \in [0,85; 0,96]$, $q_{IF} \in [0,80; 0,96]$ с выбором пары, максимизирующей F1; порог бинаризации score – также по валидации. На тесте получено $q_{MAD} = 0,85$, $q_{IF} = 0,80$ (верхние 15 % по MAD и 20 % по IF). SeqAE в полном каскаде не улучшил F1 на обоих наборах (см. табл. 2–3): на транзакциях окно T = 10 не несёт содержательной динамики; на сессиях агрегаты уже «сжимают» последовательность.

Рис. 1. Гибридная каскадная архитектура MAD → IF → SeqAE

Результаты

На тестовой выборке Credit Card Fraud (0,17 % мошенничества) лидирует XGBoost: F1 = 0,808, PR-AUC = 0,775, Cost@100 = 505 (табл. 2). При накопленной истории подтверждённых инцидентов модель с обучением на размеченных данных даёт наименьший ожидаемый ущерб. Среди методов без учителя на этом наборе лучше всего Isolation Forest без SMOTE – F1 = 0,174, Cost@100 = 1679.

MAD почти не пропускает аномалии (полнота близка к 1), но точность крайне низкая: F1 = 0,009, Cost@100 = 5767. Для автоматической блокировки

операций такой режим неприемлем – поток ложных тревог перегрузит контролёров. SeqAE на отдельных транзакциях не сработал (F1 = 0,000): окно из 10 проводок не отражает содержательную динамику.

Добавление SMOTE к Isolation Forest ухудшило F1 с 0,174 до 0,125 и подняло Cost@100 с 1679 до 2149 – на наших данных «парадокс изоляции» проявился явно.

На Payment IS sessions (4,32 % аномалий) XGBoost снова впереди: F1 = 0,767, PR-AUC = 0,735. Isolation Forest при подобранном пороге не вышел на ненулевой F1; MAD даёт F1 = 0,091 за счёт высокой полноты. Каскад MAD+IF уменьшает число кандидатов на 58,4 % для транзакций и на 32,3 % для сессий.

Таблица 2 – Сравнение методов (тестовая выборка, собственные эксперименты)

Метод	**Набор**	**F1**	**PR-AUC**	**Cost@100**	**Комментарий**
MAD	Credit Card	0,009	0,004	5767	Ловит почти всё, но сыплет ложными тревогами
Isolation Forest	Credit Card	0,174	0,230	1679	Лучший вариант без разметки
XGBoost	Credit Card	0,808	0,775	505	Минимальный ущерб при наличии меток
SeqAE	Credit Card	0,000	0,007	2600	На точечных транзакциях бесполезен
MAD+IF	Credit Card	0,009	0,004	5767	F1 как у MAD; −58,4 % объёма (q_MAD=0,85; q_IF=0,80)
MAD+IF+SeqAE	Credit Card	0,009	0,004	5767	SeqAE не повышает F1; время ×130 vs MAD+IF
IF + SMOTE	Credit Card	0,125	0,086	2149	Балансировка вредит Isolation Forest
MAD	Sessions	0,091	0,047	1213	Умеренное качество на агрегатах
Isolation Forest	Sessions	0,000	0,067	4916	Порог не дал приемлемой полноты
XGBoost	Sessions	0,767	0,735	1604	Наилучший результат на сессиях
MAD+IF	Sessions	0,091	0,048	1213	−32,3 % объёма кандидатов
MAD+IF+SeqAE	Sessions	0,091	0,048	1213	SeqAE не повышает F1

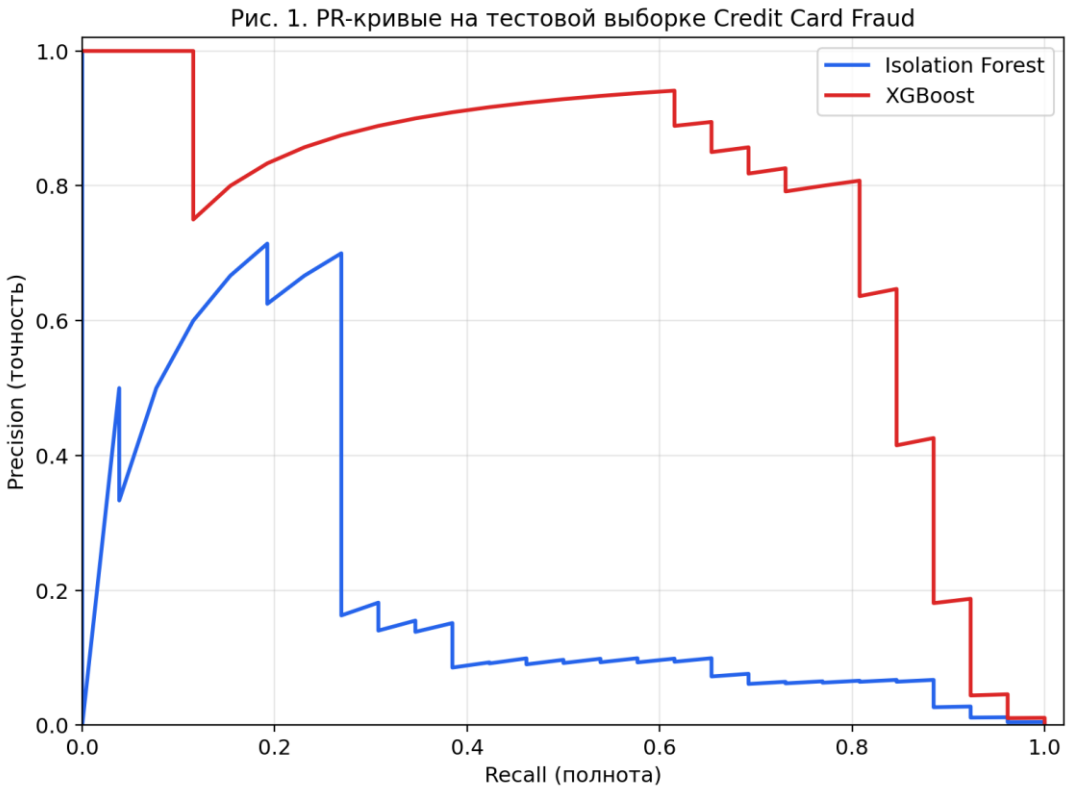

Рис. 2. PR-кривые методов Isolation Forest и XGBoost на тестовой выборке Credit Card Fraud (собственные эксперименты)

На рис. 2 XGBoost держит более высокую точность при той же полноте. Isolation Forest без меток сохраняет PR-AUC = 0,230 – разумный компромисс, пока накопленной разметки мало.

Поэтапный анализ каскада (табл. 3) показывает: на транзакциях максимальный F1 у IF only (0,174). Каскад MAD+IF F1 не улучшает, зато отсеивает 58,4 % потока до тяжёлого анализа; MAD+IF+SeqAE F1 также не растёт, время – 2077 мс/1000 записей. Выигрыш каскада – не в метрике качества, а в сокращении объёма для контролёра и вычислительных ресурсов.

Таблица 3 – Поэтапный анализ каскада (F1, время на 1000 записей, мс)

Набор	**Вариант**	**F1**	**мс/1000**
Credit Card	MAD only	0,009	0,7
Credit Card	IF only	0,174	12,1
Credit Card	MAD+IF	0,009	15,9
Credit Card	MAD+IF+SeqAE	0,009	2077
Sessions	MAD only	0,091	4,1
Sessions	IF only	0,000	77,1
Sessions	MAD+IF	0,091	89,0
Sessions	MAD+IF+SeqAE	0,091	3478

Пояснение к табл. 3. Измерено календарное время оценивания 1000 записей на тесте (обучение IF – на обучающей выборке). Каскад MAD+IF медленнее IF only, потому что последовательно выполняет MAD и IF по всему потоку; выигрыш каскада – в сокращении числа записей, передаваемых аналитику или тяжёлой модели (SeqAE, ручная проверка), а не в ускорении IF. Накладные расходы передачи данных между этапами в замер не входили.

Обсуждение

Цифры согласуются с [4, 10]: там, где есть разметка, модель с обучением на размеченных данных выигрывает и по F1, и по показателю ущерба. Isolation Forest имеет смысл как первичный фильтр или на старте проекта, когда подтверждённых кейсов мало. При $C_{\mathrm{FN}}/C_{\mathrm{FP}} = 100$ порядок методов по F1 совпадает с порядком по функции (1), поэтому F1 можно использовать как рабочую прокси при первичном выборе алгоритма.

Согласованность названий: в табл. 2–3 MAD+IF – реализованный двухступенчатый каскад; MAD+IF+SeqAE – добавление автокодировщика к кандидатам. Целевая схема на рис. 1 сохранена как архитектурная модель

службы контроля; SeqAE включён в эксперимент отдельной строкой, чтобы показать отсутствие прироста F1 при текущей постановке.

Практически это выглядит так. Если организация ведёт реестр подтверждённых мошеннических операций – разумно обучать XGBoost и пересматривать модель по мере накопления данных. Если разметки нет – Isolation Forest без SMOTE; синтетическое уравнивание классов в нашем эксперименте только ухудшило результат. Чтобы не перегружать аналитиков, перед детальной проверкой имеет смысл поставить лёгкий каскад MAD + IF: он пропускает дальше около 40 % записей вместо всего потока.

Ограничения. Оба набора открытые; сессии мы собрали сами из ULB-транзакций – это не журнал 1С или отраслевой ERP. Пороги каскада q_{MAD} и q_{IF} подобраны на валидации; перенос на другие КИС требует повторного подбора. SeqAE в виде MLP на коротких окнах не заменяет LSTM-AE для многошаговых сценариев [6]. Перенос на предприятия АПК (зерновая логистика, элеваторы, внутренний учёт) нужно проверять на их собственных данных.

Заключение

Собран единый протокол сравнения статистических методов и методов машинного обучения для журналов КИС с метриками, учитывающими стоимость ошибок. На тесте XGBoost даёт F1 = 0,81 (транзакции) и 0,77 (сессии) при наименьшем Cost@100. Isolation Forest без SMOTE – лучший метод без учителя на транзакциях (F1 = 0,17); с SMOTE падает до 0,12. Каскад MAD+IF сокращает поток на углублённый анализ на 58 % (транзакции) и 32 % (сессии); MAD+IF+SeqAE качество не повышает. Сопоставление с подходами [13] задаёт направление следующего этапа –

проверка ансамблевых схем на тех же журналах КИС при имитации дрейфа во времени.

Работа относится к п. 9, 10 и 13 паспорта 5.2.2. Дальше имеет смысл прогнать тот же протокол на реальных журналах ERP предприятий АПК и заменить MLP-AE на LSTM-автокодировщик для многошаговых сценариев злоупотреблений.

Список использованных источников

1. Montgomery R. M. Techniques for Outlier Detection: A Comprehensive View // Preprints.org. – 2024. – DOI: 10.20944/preprints202410.1603.v1. – URL: https://doi.org/10.20944/preprints202410.1603.v1

2. Yaro A. S., Maly F., Prazak P., Maly K. Outlier Detection Performance of a Modified Z-Score Method in Time-Series RSS Observation With Hybrid Scale Estimators // IEEE Access. – 2024. – Vol. 12. – P. 12785–12796. – DOI: 10.1109/ACCESS.2024.3356731. – URL: https://doi.org/10.1109/ACCESS.2024.3356731

3. Prajesha T. M., Veni S. An Efficient Outlier Detection Using Isolation Forest Based on Robust Scaling and PCA // Indian Journal of Science and Technology. – 2023. – Vol. 16, No. 29. – P. 2244–2251. – DOI: 10.17485/IJST/v16i29.638. – URL: https://doi.org/10.17485/IJST/v16i29.638

4. Al-Shehari T. et al. Insider Threat Detection Model Using Anomaly-Based Isolation Forest Algorithm // IEEE Access. – 2023. – Vol. 11. – P. 118170–118185. – DOI: 10.1109/ACCESS.2023.3326750. – URL: https://doi.org/10.1109/ACCESS.2023.3326750

5. Ness S. et al. Anomaly Detection in Network Traffic Using Advanced Machine Learning Techniques // IEEE Access. – 2025. – Vol. 13. – P. 16133–16149. – DOI: 10.1109/ACCESS.2025.3526988. – URL: https://doi.org/10.1109/ACCESS.2025.3526988

6. Wei Y. et al. LSTM-Autoencoder based Anomaly Detection for Indoor Air Quality Time Series Data // arXiv:2204.06701. – 2022. – URL: https://arxiv.org/abs/2204.06701

7. Smiałkowski T., Czyzewski A. Autoencoder Application for Anomaly Detection in Power Consumption // IEEE Access. – 2023. – Vol. 11. – P. 124150–124162. – DOI: 10.1109/ACCESS.2023.3330135. – URL: https://doi.org/10.1109/ACCESS.2023.3330135

8. Lundberg S. M., Lee S.-I. A Unified Approach to Interpreting Model Predictions // Advances in Neural Information Processing Systems. – 2017. – Vol. 30. – P. 4765–4774. – URL: https://proceedings.neurips.cc/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html

9. Phua C., Lee V., Smith-Miles K., Gayler R. A Comprehensive Survey of Data Mining-based Fraud Detection Research // Clayton School of Information Technology, Monash University. – 2010. – URL: https://arxiv.org/abs/1009.6119

10. Bhattacharyya S. et al. Data mining for credit card fraud: A comparative study// Decision Support Systems. – 2011. – Vol. 50, No. 3. – P. 602–613. – DOI: 10.1016/j.dss.2010.08.006. – URL: https://doi.org/10.1016/j.dss.2010.08.006

11. Федеральный закон от 07.08.2001 № 115-ФЗ «О противодействии легализации (отмыванию) доходов, полученных преступным путём, и финансированию терроризма» (ред. от 2024 г.) // Собрание законодательства РФ. – URL: https://www.consultant.ru/document/cons_doc_LAW_32834/ Transl.: Federal'nyj zakon ot 07.08.2001 № 115-FZ «O protivodejstvii legalizatsii (otmyvaniyu) dokhodov, poluchennykh prestupnym putem, i finansirovaniyu terrorizma».

12. Elkan C. The Foundations of Cost-Sensitive Learning // Proceedings of the 17th International Joint Conference on Artificial Intelligence (IJCAI). – 2001. – P. 973–978. – URL: https://cseweb.ucsd.edu/~elkan/rescale.pdf

13. Paldino G. M., Lebichot B., Le Borgne Y.-A. et al. The role of diversity and ensemble learning in credit card fraud detection // Advances in Data Analysis and Classification. – 2024. – Vol. 18, No. 1. – P. 193–217. – URL: https://doi.org/10.1007/s11634-022-00515-5