УДК 004.8:004.056

ОБНАРУЖЕНИЕ АНОМАЛИЙ НА ОСНОВЕ BIG DATA ДЛЯ ВЫЯВЛЕНИЯ МОШЕННИЧЕСТВА В ЦИФРОВЫХ СИСТЕМАХ

М.М. Ходжамаммедов
преподаватель кафедры информационных технологий
Государственного энергетического института Туркменистана
mekanhoja2021@gmail.com

М.М. Ходжамаммедов
Окончил Государственного энергетического института Туркменистана. Область научных интересов связана с разработкой методов и алгоритмов построения информационно-компьютерных систем, организацией учебного и научно-исследовательского процессов в техническом университете.

Аннотация. В данной работе рассматриваются теоретические и прикладные аспекты обнаружения аномалий в условиях обработки больших данных с целью выявления мошеннической активности в цифровых системах. Актуальность темы обусловлена экспоненциальным ростом объемов данных и усложнением методов совершения мошенничества. В статье анализируются современные подходы, основанные на статистических методах, алгоритмах машинного обучения и глубоких нейронных сетях. Особое внимание уделяется роли распределенных вычислительных систем и потоковой обработки данных. Представлен сравнительный анализ методов и их эффективности в задачах обнаружения мошенничества.

Ключевые слова: Big Data, обнаружение аномалий, мошенничество, машинное обучение, нейронные сети, анализ данных, кибербезопасность.

Введение. Цифровая трансформация современного общества сопровождается стремительным ростом объемов генерируемых данных, что приводит к формированию принципиально новых подходов к их обработке и анализу. В финансовых, телекоммуникационных и информационных системах ежедневно обрабатываются миллионы транзакций, каждая из которых может потенциально содержать признаки мошенничества. В таких условиях традиционные методы анализа данных оказываются недостаточно эффективными из-за ограниченной масштабируемости и неспособности выявлять сложные зависимости. Мошенничество в цифровой среде принимает все более сложные и скрытые формы, включая использование автоматизированных скриптов, социальной инженерии и распределенных атак. Это требует применения интеллектуальных методов анализа, способных адаптироваться к изменяющимся условиям и выявлять аномалии в режиме реального времени. Обнаружение аномалий является ключевым направлением в области анализа данных, направленным на выявление отклонений от нормального поведения системы.

Теоретические основы обнаружения аномалий. Понятие аномалии в анализе данных определяется как наблюдение, существенно отличающееся от большинства других наблюдений. Такие отклонения могут свидетельствовать о наличии ошибок, редких событий или целенаправленных мошеннических действий. Формализация задачи обнаружения аномалий связана с построением модели нормального поведения и последующим выявлением отклонений от этой модели. Существует несколько подходов к формализации данной задачи. Вероятностный подход основывается на предположении о распределении данных и позволяет вычислять вероятность принадлежности наблюдения к нормальному классу. Геометрический подход рассматривает данные в пространстве признаков и определяет аномалии как точки, находящиеся на значительном расстоянии от плотных областей. Информационный подход использует меры энтропии и сложности для выявления нестандартных паттернов. Особенностью задач обнаружения мошенничества является высокая степень несбалансированности данных, при которой доля мошеннических операций составляет крайне малую часть общего объема. Это усложняет процесс обучения моделей и требует применения специальных методов обработки данных.

Методы машинного обучения для обнаружения аномалий. Современные методы машинного обучения играют ключевую роль в решении задач обнаружения аномалий. Их преимущество заключается в способности автоматически выявлять скрытые закономерности и адаптироваться к изменяющимся данным. Методы обучения с учителем предполагают наличие размеченных данных и позволяют строить классификационные модели. Однако в задачах обнаружения мошенничества такие данные часто ограничены или неполны. В связи с этим широкое распространение получили методы обучения без учителя, которые не требуют предварительной разметки. Алгоритмы кластеризации позволяют выявлять структуру данных и определять аномалии как объекты, не принадлежащие ни одному из кластеров. При этом эффективность данных методов зависит от выбора метрики расстояния и параметров модели. Особое внимание уделяется алгоритму Isolation Forest, который основан на принципе изоляции редких наблюдений. Данный метод демонстрирует высокую эффективность при работе с большими объемами данных и не требует значительных вычислительных ресурсов. Нейросетевые подходы, в частности автоэнкодеры, позволяют моделировать сложные нелинейные зависимости. Принцип их работы основан на реконструкции входных данных и анализе ошибки восстановления. Аномалии характеризуются высокой ошибкой реконструкции, что позволяет эффективно их выявлять.

Глубокое обучение и анализ временных рядов. С развитием технологий глубокого обучения появились новые возможности для анализа данных, представленных в виде временных рядов. В задачах обнаружения мошенничества важную роль играет учет временной динамики, поскольку поведение пользователей изменяется во времени. Рекуррентные нейронные сети и архитектуры типа LSTM позволяют моделировать последовательности и выявлять отклонения в поведении. Такие модели способны учитывать контекст и выявлять

сложные паттерны, недоступные для традиционных методов. Применение сверточных нейронных сетей также находит применение в анализе структурированных данных, где они используются для выявления локальных закономерностей.

Инфраструктура Big Data. Эффективная реализация методов обнаружения аномалий невозможна без использования современных технологий Big Data. Распределенные вычислительные системы обеспечивают обработку данных в условиях высокой нагрузки и позволяют масштабировать алгоритмы. Платформы Hadoop и Apache Spark являются основой для построения систем анализа данных. Они обеспечивают распределенное хранение и обработку информации, что позволяет эффективно работать с большими объемами данных. Особое значение имеет потоковая обработка данных, реализуемая с использованием технологий Kafka и Spark Streaming. Это позволяет анализировать данные в режиме реального времени и оперативно реагировать на выявленные аномалии.

Практическое применение в финансовых системах. В финансовых организациях системы обнаружения мошенничества интегрируются в процесс обработки транзакций и функционируют в режиме реального времени. Они анализируют множество параметров, включая географическое положение пользователя, частоту операций, тип устройства и поведенческие характеристики. Современные системы используют гибридные подходы, объединяющие несколько методов анализа данных. Это позволяет повысить точность и снизить количество ложных срабатываний. Важным аспектом является баланс между безопасностью и удобством пользователей.

Проблемы и ограничения. Несмотря на значительные достижения, системы обнаружения аномалий сталкиваются с рядом проблем. К ним относятся высокая сложность моделей, необходимость больших вычислительных ресурсов и ограниченность размеченных данных. Кроме того, мошенники постоянно адаптируют свои методы, что требует регулярного обновления моделей.

Проблема интерпретируемости моделей также является актуальной, особенно в финансовой сфере, где требуется объяснение принятых решений.

Перспективы развития. Дальнейшее развитие технологий обнаружения аномалий связано с интеграцией методов искусственного интеллекта, развитием объяснимых моделей и использованием гибридных архитектур. Особое внимание уделяется автоматизации процессов обучения и адаптации моделей. Перспективным направлением является использование графовых моделей для анализа связей между объектами, а также внедрение федеративного обучения для обеспечения конфиденциальности данных.

Заключение. В результате проведенного анализа можно сделать вывод о том, что технологии Big Data в сочетании с методами машинного и глубокого обучения являются эффективным инструментом для выявления мошенничества.

Их применение позволяет существенно повысить уровень безопасности цифровых систем и снизить риски финансовых потерь.

Дальнейшие исследования должны быть направлены на повышение точности моделей, снижение вычислительных затрат и улучшение интерпретируемости результатов.

Список литературы

[1] Banerjee A., Kumar V. Anomaly Detection: A Survey // ACM Computing Surveys. 2009. Aggarwal C.C. Outlier Analysis. Springer, 2017.

[2] Liu F.T., Ting K.M., Zhou Z.-H. Isolation Forest // ICDM. 2008.

[3] Breunig M.M. et al. LOF: Identifying Density-Based Local Outliers // SIGMOD. 2000.Goodfellow I., Bengio Y., Courville A. Deep Learning. MIT Press, 2016.

Авторский вклад

Мекан Ходжамамедов Мерданович – постановка задач научных исследований, разработка методологии оценки эффективности генеративной политики при оптимизации проектных решений, а также руководство проведением этих исследований.

ANOMALY DETECTION BASED ON BIG DATA FOR IDENTIFYING FRAUD IN DIGITAL SYSTEMS

M.M. Hojamammedov
Lecturer, Department of Information Technologies, Turkmen State Energy Institute
mekanhoja2021@gmail.com

Abstract. This work examines the theoretical and applied aspects of anomaly detection in the context of big data processing for identifying fraudulent activity in digital systems. The relevance of the topic is driven by the exponential growth of data volumes and the increasing complexity of fraud techniques. The article analyzes modern approaches based on statistical methods, machine learning algorithms, and deep neural networks. Special attention is paid to the role of distributed computing systems and stream data processing. A comparative analysis of methods and their effectiveness in fraud detection tasks is presented.

Keywords: Big Data, anomaly detection, fraud, machine learning, neural networks, data analysis, cybersecurity.