МОДИФИКАЦИЯ МЕТОДА АНАЛИЗА ПОЛЬЗОВАТЕЛЕЙ ПЛАТФОРМ ЭЛЕКТРОННОЙ КОММЕРЦИИ

Н.О. Ставцев, *магистрант*
Национальный исследовательский университет «МЭИ», филиал в г. Смоленск (Россия, г. Смоленск)

DOI: 10.24411/2411-0450-2020-10509

Аннотация. *В статье описаны современные условия деятельности платформ электронной коммерции. Выявлены проблемы используемых методов для анализа пользователей платформ электронной коммерции с целью выявления мошеннических транзакций. Разработано два метода анализа пользователей с целью выявления мошеннических платежей в зависимости от доступного набора данных: построение системы на основе метода случайного леса и автоэнкодера.*

Ключевые слова: *платформы электронной коммерции, мошеннические транзакции, машинное обучение, случайный лес, автоэнкодер.*

Платформы электронной коммерции выступают инструментом для большинства организаций с позиции ведения и переноса своего бизнеса в сеть интернет. При этом данные организации выступают клиентами ИТ-организации предоставляющей услуги платформы электронной коммерции – торговой площадки. Подобного рода ИТ-организации накапливают на своих информационных ресурсах персональные данные как юридических, так и физических лиц, а кроме того организуют процесс электронных взаиморасчетов. В современных условиях кибертерроризма и электронного мошенничества [1], организации предоставляющие услуги платформ электронной коммерции должны тщательно следить за рисками реализации кибератак и уровнем информационной безопасности сервиса. Данные действия необходимы как с позиции поддержания высокого имиджа организации и доверия клиентов (как реализующих товар на данной платформе, так и выступающих покупателем) за счет безопасности проведения платежей и сохранности персональных данных, так и с экономической позиции – предотвращение возможных убытков от проведенной мошеннической операции. Следовательно, для выявления мошеннических действий на платформах электронной коммерции, организации нуждаются в оптимальной инновационной технологии.

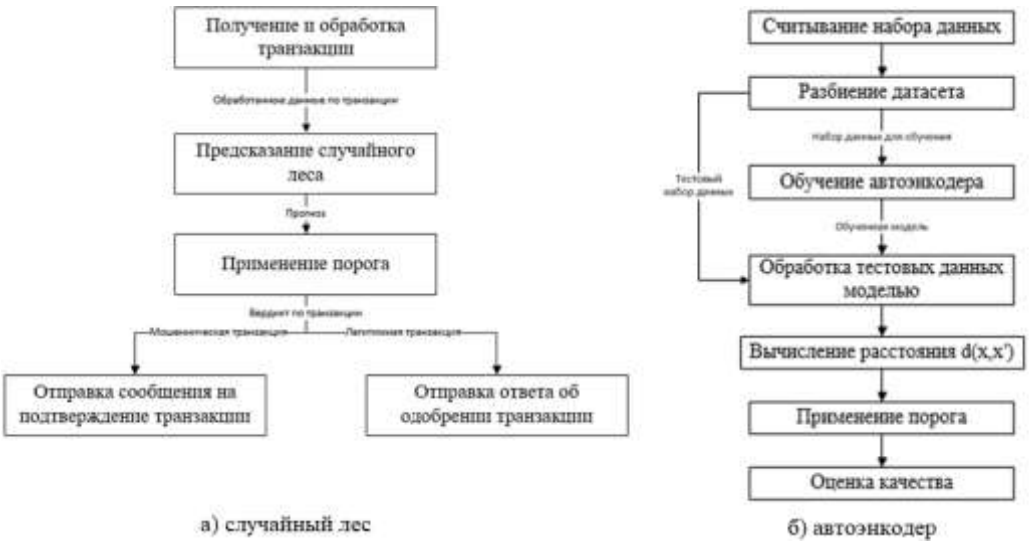
Для анализа пользователей с целью обнаружения мошеннических действий традиционно используют экспертные системы, содержащие множество логических выражений и статистических правил, направленных на выявление подозрительных действий [2]. Однако, данный подход обладает высоким риском ложных срабатываний, так как аналитики сервиса электронной коммерции способны проверить вручную лишь малую часть подозрительных ситуаций, а их расследование занимает много времени и отнимает значительные ресурсы, блокировка легитимных операций, ошибочно принятых за мошеннические, создает неудобства для клиентов и снижает доверие к сервису. Кроме того, при использовании экспертных систем для анализа пользователей невозможно вручную обнаружить новые схемы мошенничества и выявить все закономерности. Особенности профилей злоумышленников и схемы атак для различных каналов мошенничества зачастую кардинально отличаются.

С целью снижения риска ложных срабатываний авторами [3-6] отмечается возможность использования методов машинного обучения совместно со статистическими правилами. Данный подход позволит сократить количество случаев ошибочного определения легитимных операций как мошеннические, и увеличить число успешно выявленных действительно

мошеннических операций. Алгоритмы машинного обучения позволяют обнаружить неочевидные для человека зависимости, быстро анализируя огромные объемы данных, а также, обладая способностью обучаться, смогут распознавать даже новые схемы мошенничества. Наиболее эффективными для выявления мошеннических действий являются ансамблевые методы машинного обучения, состоящие из ансамбля различных классификаторов, что значительно повышает качество данных моделей.

На основе данного подхода к выявлению мошеннических транзакций на платформах электронной коммерции разработано множество алгоритмов. В зависимости от доступного набора данных для обучения можно выделить два пути развития предлагаемого метода: в наборе данных содержится достаточное количество мошеннических транзакций или количество транзакций слишком мало. При достаточном количестве мошеннических операций в основу разрабатываемого метода закладывается один из методов обучения с учителем. С помощью которого можно произвести обучение модели и вычислить вероятность мошеннических и легитимных транзакции, применяя модель к новым транзакциям для определения их легитимности. Для данного случая предлагается использовать случайный лес, так как он является ансамблевым методом машинного обучения, снижает риски несбалансированности классов, увеличивают точность самого анализа. При недостаточном количестве мошеннических операций, мошенничество рассматривается как отклонение или как аномальное значение, так как есть только образцы легитимных транзакций. Поэтому для выявления отклонения предлагается использовать метод изолированного леса, а для выявления аномального значения метод автоэнкодера.

На рисунке 1 представлены методы работы случайного леса и автоэнкодера. Модели случайного леса, с реальными данными о транзакциях с такими процессами, как обработка поступившей транзакции, прогнозирование моделью, применение выбранного порога, для получения конечного прогноза. Если данная транзакция моделью посчиталась мошеннической, то техническому специалисту и пользователю отправляется письмо для подтверждения транзакции владельцу банковской карты отправляется письмо, чтобы он подтвердил легитимность операции. Автоэнкодер представляет собой сеть, состоящую из входного и выходного слоя, содержащих n узлов. Между ними от одного до нескольких скрытых слоёв и в середине находится самый маленький слой, который состоит из h узлов ($h < n$). Таким образом, нейросеть обучается воспроизводить входной вектор x в виде выходного вектора x'.



а) случайный лес б) автоэнкодер

Рис. 1. Модели случайного леса и автоэкодера для анализа пользователей электронной коммерции

В задаче выявления мошенничества автоэнкодер будет обучаться только на образцах класса легитимных транзакций. При выявлении аномальных транзакций автоэнкодер воспроизводит входной x в выходного x', причём для аномалий мы получим не оптимальный результат [7]. Разницу между x и x' можно оценить за счет измерения расстояния по формуле (1).

$$d(x, x') = \frac{1}{n}\sum_{i=1}^{n}(x_i - x_i')^2 \quad (1)$$

Решение, к какому классу относится транзакция происходит за счет использования к расстоянию $d(x, x')$ порогового значения δ. Транзакция (x) является мошеннической в соответствии с правилом определения аномалий. Можно выбрать такое пороговое значение, чтобы выявлять мошенничество только в очевидных случаях или подобрать в зависимости от ситуации (формула (2), формула (3)).

$$x \rightarrow \text{"legitimate" IF} d(x, x') \leq \delta \quad (2)$$
$$x \rightarrow \text{"fraudulent" IF} d(x, x') > \delta \quad (3)$$

Таким образом, алгоритм базирующийся на автоэнкодере будет иметь следующие процессы: деление датасета на обучающий и тестовый набор данных, обучение нейросети, обученная сеть обрабатывает тестовые данные, вычисление расстояния $d(x, x')$, применение выбранного порога δ, для получения конечного прогноза и оценка результата.

Следовательно, было разработано два метода анализа пользователей с целью выявления мошеннических платежей в зависимости от доступного набора данных, а именно построение системы на основе метода случайный лес, когда есть достаточное количество образцов обоих видов транзакций или автоэнкодера, когда количество образцов мошеннических транзакций слишком мало. Данные методы позволят снизить риск несбалансированности классов и увеличить точность самого анализа пользователей платформ электронной коммерции.

Библиографический список

1. Analysis of cyber attack and incident data from IBM's worldwide security services operations // IBM 2015 Cyber Security Intelligence Index. 2015. – [Электронный ресурс]. – Режим доступа: https://essextec.com/wp-content/uploads/2015/09/IBM-2015-Cyber-Security-Intelligence-Index_FULL-REPORT.pdf (дата обращения: 20.05.2020).

2. Ускенбаева Р.К., Бектемысова Г.У. Применение больших данных в электронной коммерции: Перспективы и проблемы // Colloquium-journal. – 2019. – №2-1 (26). – С. 7-11.

3. Гультяева Т.А. Методы статистического обучения в задачах регрессии и классификации: монография / Т.А. Гультяева, А.А. Попов, А.С. Саутин; Новосибирский государственный технический университет. – Новосибирск: Новосибирский государственный технический университет, 2015. – 323 с.

4. Машинное обучение против кредитных рисков // Хабр. – [Электронный ресурс]. – Режим доступа: https://habr.com/ru/company/vtb/blog/417739/ (дата обращения: 20.05.2020).

5. Кузьмина С.В., Ефимов А.И. Актуальные методы машинного обучения в области классификации // Актуальные проблемы современной науки и производства. – 2018. – С. 34-38.

6. Жуков Д.А., Клячкин В.Н., Кувайскова Ю.Е. Сравнительный анализ методов машинного обучения при прогнозировании состояния технического объекта // Радиоэлектронная техника. – 2017. – № 1. – С. 189.

7. Нечахин В.А., Пищик Б.Н. Применение методов глубинного обучения для обнаружения вторжений // Вестник Новосибирского государственного университета. Серия: Информационные технологии. – 2019. – №17 (2). – С. 114-121.

MODIFICATION OF THE METHOD FOR ANALYZING USERS OF E-COMMERCE PLATFORMS

N.O. Stavtsev, *Graduate Student*
National Research University «Moscow Power Engineering Institute» (MPEI), Smolensk Branch
(Russia, Smolensk)

Abstract. *The article describes the current conditions of e-Commerce platforms. Problems with the methods used for analyzing users of e-Commerce platforms in order to detect fraudulent transactions were identified. We have developed two methods for analyzing users in order to detect fraudulent payments depending on the available data set: building a system based on the random forest method and an autoencoder.*

Keywords: *e-Commerce platforms, fraudulent transactions, machine learning, random forest, autoencoder.*