УДК 004.056

ОБНАРУЖЕНИЕ МОШЕННИЧЕСТВА С КРЕДИТНОЙ КАРТОЙ С ИСПОЛЬЗОВАНИЕМ МЕТОДОВ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ДАННЫХ И ОБЗОР ИХ ХАРАКТЕРИСТИК

Каримова Диана Дулатовна
karimovad@yahoo.com
Магистрант Администратор по управлению и защите компьютерных систем и сетей на предприятиях ЕНУ им. Л.Н.Гумилева, Нур-Султан, Казахстан
Научный руководитель – Б.Ергеш

В нынешнюю эпоху кредитные карты играют очень важную роль в экономике. Они стали неотъемлемой частью как в рабочей среде, так и в бытовой жизни благодаря развитию Интернета. Хотя использование кредитных карт дает огромные преимущества при осторожном и ответственном использовании, мошеннические действия все же могут привести к значительному кредитному и финансовому ущербу. Было предложено множество методов противодействия растущему мошенничеству с кредитными картами. И хотя все эти методы преследуют одну и ту же цель - избежать мошенничества с кредитными картами; у каждого есть свои недостатки, достоинства и характеристики, которые стоило бы изучить для понимания какой же метод является наиболее эффективным.

1. Мошенничество с кредитной картой

Для начала стоит определить что есть «мошенничество с кредитной картой». Незаконное использование кредитной карты или информации тем или иным образом связанной с нею без ведома владельца называется мошенничеством с кредитными картами. Различные приемы мошенничества с кредитными картами относятся в основном можно подразделить на две группы: прикладное и поведенческое мошенничество [1]. Главным представителем первой группы является мошенничество с заявками – такой тип определяется, когда мошенники применяют новые карты от банка или компаний-эмитентов, используя ложную или иную информацию. Несколько заявок могут быть поданы одним пользователем с одним набором пользовательских данных, также называемое мошенничество с дублированием, или другим пользователем с идентичными данными, также с использование личных данных.

Поведенческое мошенничество, с другой стороны, имеет четыре основных типа: украденная/потерянная карта, кража почты, поддельная карта и мошенничество «владелец карты отсутствует». Мошенничество с украденной/утерянной картой происходит, когда мошенники крадут кредитную карту или получают доступ к утерянной карте. Мошенничество с кражей почты происходит, когда мошенник получает кредитную карту по почте или личную информацию из банка до того, как та дойдет до фактического владельца карты [2]. Как при подделке, так и при мошенничестве «владелец карты отсутствует» данные кредитной карты получаются без ведома владельцев карт. В первом случае удаленные транзакции могут осуществляться с использованием реквизитов карты по почте, телефону или Интернету. В последнем случае поддельные карты изготавливаются на основе карточной информации.

2. Трудности обнаружения мошенничества с кредитными картами

Не секрет, что уже существуют системы обнаружения мошенничества, но стоит взять во внимание то, что даже самые лучшие такие системы все же натыкаются на ряд трудностей и проблем, которые должны быть преодолены для сохранения безопасности ценных данных. К таким трудностям можно отнести:

- Несбалансированные данные: данные об обнаружении мошенничества с кредитными картами имеют несбалансированный характер. Это означает, что очень небольшой процент всех транзакций по кредитным картам является мошенническим. Из-за этого обнаружение мошеннических операций очень затруднено и неточно.
- Различная важность неправильной классификации: в задаче обнаружения мошенничества разные ошибки неправильной классификации имеют разное значение. Ошибочная классификация обычной транзакции как мошенничества не так опасна, как обнаружение мошеннической транзакции в обычном режиме. Потому что в первом случае ошибка классификации будет выявлена при дальнейших исследованиях.
- Перекрывающиеся данные: многие транзакции могут считаться мошенническими, в то время как на самом деле они являются нормальными (ложноположительными), и наоборот, мошеннические транзакции также могут казаться законными, хоть такими и не являются (ложноотрицательными). Следовательно, получение низкого уровня

ложноположительных и ложноотрицательных результатов является ключевой задачей систем обнаружения мошенничества [3, 4].

- Отсутствие адаптируемости: алгоритмы классификации обычно сталкиваются с проблемой обнаружения новых типов нормальных или мошеннических шаблонов. Контролируемые и неконтролируемые системы обнаружения мошенничества неэффективны при обнаружении новых моделей нормального и мошеннического поведения соответственно.
- Стоимость обнаружения мошенничества: система должна учитывать как стоимость обнаруженного мошенничества, так и стоимость его предотвращения. Например, прекращение мошеннической транзакции на несколько долларов не приносит дохода [5, 6].
- Отсутствие стандартных метрик: нет стандартного критерия оценки для определения и сравнения результатов систем обнаружения мошенничества.

3. Существующие системы обнаружения мошенничества, использующие методы интеллектуального анализа данных

Система нечеткой логики задействовала действительную политику оценки мошенничества с использованием оптимальных пороговых значений. Результатом работы данной системы являются вероятность мошенничества и причины, по которым то или иное событие является мошенническим. Эта система предсказывала несколько лучшие результаты, чем аудиторы. Еще одна логическая система использовала два подхода для имитации рассуждений экспертов по мошенничеству:

1) модель обнаружения, использующая неконтролируемую нейронную сеть для поиска взаимосвязей в данных и для поиска кластеров, затем выявляются шаблоны в кластерах;

2) модель обнаружения нечетких аномалий, в которой использовался алгоритм Ванга-Менделя, чтобы выяснить, как поставщики медицинских услуг совершали мошенничество против страховых компаний [7].

Методология горячих точек выполняла трехэтапный процесс:

1) алгоритм кластеризации k-средних используется для обнаружения кластеров, потому что другие алгоритмы кластеризации имеют тенденцию быть дорогостоящими с вычислительной точки зрения там, где наборы данных очень большие;

2) алгоритм C4.5, результирующее дерево решений может быть преобразовано в набор правил и сокращено;

3) инструменты визуализации для оценки правил, построения статистических сводок сущностей, связанных с каждым правилом [8].

Грэм Уильямс в своей работе «Интеллектуальный анализ данных с эволюционными горячими точками. Архитектура для изучения интересных открытий» расширил методологию горячих точек, с целью использования енетических алгоритмов для генерации и исследования правил [9].

Модель кредитного мошенничества Гротта предлагала метод классификации с атрибутом «мошенничество/законно» и кластеризацию, за которой следует метод классификации без атрибута «мошенничество/законно». Самоорганизующаяся карта характеристик Кохонена использовалась для категоризации исков о травмах, связанных с автомобилем, в зависимости от размера подозрений в мошенничестве. Затем достоверность карты характеристик оценивалась с использованием алгоритма обратного распространения и нейронных сетей с прямой связью. Результат показал, что метод был более надежным и последовательным по сравнению с оценкой мошенничества [10].

Методы классификации оказались очень эффективными при обнаружении мошенничества и, следовательно, могут применяться для категоризации данных о преступности. Модель распределенного интеллектуального анализа данных использует реалистичную модель затрат для оценки C4.5, CART и наивных байесовских моделей классификации. Метод был применен к операциям с кредитными картами. Подход нейронного анализа данных использует правила ассоциации на основе правил для анализа символьных данных и нейронную сеть с радиальной базисной функцией для анализа

аналоговых данных. В рамках подхода обсуждается важность использования нечисловых данных при обнаружении мошенничества. Было обнаружено, что результаты ассоциативных правил повышают точность прогнозов [11].

Программное обеспечение SAS Enterprise Miner зависит от правил ассоциации, обнаружения кластеров и методов классификации для обнаружения мошеннических заявлений. Исследование байесовской сети убеждений (BBN) и искусственной нейронной сети (ANN) использовали алгоритм STAGE для BBN при обнаружении мошенничества и алгоритм обратного распространения ANN. Результаты исследования показывают, что BBN обучались намного быстрее, но были медленнее при применении к новым экземплярам [12].

4. Методы интеллектуального анализа данных предназначенные для обнаружения мошенничества

Интеллектуальный анализ данных относится к извлечению или «добыче» знаний из большого количества данных. Существует ряд методов интеллектуального анализа данных, таких как кластеризация, нейронные сети, регрессия, а также большое количество моделей прогнозирования. Однако, анализируя описанные ранее системы обнаружения мошенничества можно выделить методы интеллектуального анализа данных, которые могут считаться наиболее важными и подходящими для обнаружения мошенничества. Наилучшими моделями считаются байесовская сеть, дерево решений и искусственные нейронные сети.

Байесовские сети убеждений предоставляют графическую модель причинно-следственных связей, на основе которых прогнозируются вероятности относительности к классу, которая представляет является ли данный случай законным или мошенническим. Наивная байесовская классификация предполагает, что атрибуты экземпляра независимы от целевого атрибута. Цель состоит в том, чтобы назначить новый экземпляр классу, который имеет наивысшую апостериорную вероятность. Алгоритм очень эффективен и может обеспечить лучшую точность прогнозирования по сравнению с деревьями решений C4.5 и обратным распространением. Однако, когда атрибуты избыточны, точность прогнозирования снижается [13].

Деревья решений – это методы машинного обучения, которые выражают независимые атрибуты и зависимый атрибут в древовидной структуре, которая далее представляет собой набор решений. Правила классификации, извлеченные из деревьев решений, представляют собой выражения IF-THEN, в которых предварительные условия объединены логическим AND, где все тесты должны принести положительный результат для генерации всех правил до единого. Алгоритм C4.5 используется для разделения данных на сегменты на основе и для создания описательных правил классификации, которые можно использовать для классификации нового экземпляра. C4.5 хорошо справляется с прогнозами и выявлениями модели преступности. Он генерирует правила из деревьев и обрабатывает числовые атрибуты, пропущенные значения, сокращение и оценку частоты ошибок. C4.5 работает немного лучше, чем CART и ID3 с точки зрения точности прогнозов. Этапы обучения и классификации в данной модели раскрываются весьма быстро. Однако снижение производительности может произойти, когда C4.5 применяется к большим наборам данных. C5.0 показывает незначительные улучшения в индукции дерева решений [10, 13].

Искусственные нейронные сети представляют собой сложные математические уравнения с суммированием, экспонентами и параметрами для копирования нейронов (Берри и др., 2000). Нейронные сети обратного распространения могут обрабатывать большое количество экземпляров с устойчивостью к зашумленным данным и могут классифицировать шаблоны, на которых они не были обучены. Данный вариант является лучшим выбором для тех, кому результаты намного более важны других критериев. В остальном же обратное распространение выглядит весьма проигрышно так как требует долгих часов обучения, обширного тестирования, сохранения таких параметров, как количество скрытых нейронов, скорость обучения и т.д. [10]

5. Метод Random Forest

Не так давно в 2019 году было предложено использование алгоритма «случайного леса» для решения проблемы обнаружения мошенничества с кредитными картами. Лоуренс Б, Селина Б. и Пракаш Б. рассмотрели новую методологию и создали программу, которая может рассматриваться как часть платежного шлюза, которая проверяет, является ли транзакция мошеннической.

Модуль идентификации мошенничества по их методологии работает следующим образом:

- Платежный шлюз должен предоставить информацию о кредитной карте, такую как номер карты, дату истечения срока действия и т.д.
- Продавец же будет предоставлять такую информацию, как почтовый адрес, номер продажи, дата и время доставки, и т.д.
- Платежный шлюз должен направить соответствующие данные в Программу обнаружения мошенничества.
- Программа по обнаружению мошенничества будет использовать эффективные методы интеллектуального анализа данных для обучения и получения результатов с использованием лучшего алгоритма классификации.
- Окончательный результат транзакции, мошеннический он или законный, будет доставлен на платежный шлюз.

Результат, включающий решение и другие связанные детали, будет предоставлен администратором платежного шлюза в окончательном отчете пользовательского интерфейса продавцу. По итогу лучшим алгоритмом классификации был назван алгоритм «случайного леса», он же и использовался для оценки данных держателя карты наряду с информацией о кредитной карте и местонахождении транзакция для моделирования реальных транзакций. Производительность «случайного леса» была намного выше остальных, что позволило с максимальной точностью определить степень мошенничества.

По результатам нескольких экспериментов (Таблица 1) было доказано, что алгоритм «случайного леса» хоть и не на много, но все же был точнее тех методов интеллектуального анализа данных, которые закрепились в истории обнаружения мошенничества, как «самые лучшие и точные».

	Меры оценки			
Алгоритмы	Точность	Прецизионность	Полнота	F-мера
Байесовские сети	80%	85%	62%	75%
Деревья решений	83%	83%	80%	81%
Опорные вектора	86%	85%	85%	85%
Случайный лес	88%	87%	80%	88%

Таблица 1. Сравнительный анализ алгоритмов

Результаты, занесенные в Таблицу 1, были получены с помощью тестирования всех 4 алгоритмов на наборе данных полученного с сайта «kaggle.com». Выбранный набор данных содержал 3075 строк и 11 ключевых функций. Мерами, используемыми для оценки производительности, являются точность (как в смысле близости к истинному значению, так и в смысле близости измерения друг к другу), полнота и F-мера. Результаты экспериментов доказывают, что производительность алгоритма «случайного леса» с большей вероятностью указывает на точную степень мошенничества.

Однако, данный метод, а точнее программа, использующая этот метод, ориентирована в первую очередь на отраслевую розничную торговлю и будет выгодна розничному торговцу за счет минимизации затрат, которые торговец должен нести, если транзакция станет мошеннической. Система полагается на данные о продажах и продажах продавца и данные, имеющиеся в платежном шлюзе, для обработки транзакции, которая снижает возможность выявления мошенничества. Данная схема может стать более эффективной в предотвращении мошеннических транзакций в будущем, если розничные торговцы будут делиться

тенденциями использования кредитной карты владельца, а банк-эмитент поделится историей использования держателя карты, не затрагивая конфиденциальность владельца [14].

Заключение

На сегодняшний день финансовые организации расширяют доступность финансовых средств за счет использования инновационных услуг, таких как кредитные карты, банкоматы, услуги интернет-банкинга и мобильного банкинга. Кроме того, наряду с быстрым развитием электронной коммерции, использование кредитной карты стало удобством и необходимой частью финансовой жизни. Потому ни в коем случае нельзя игнорировать такие проблемы как мошенничество и кража личных данных через кредитные карты. Были рассмотрены как сами трудности опасности обнаружения мошенничества с кредитными картами, так и уже существующие системы обнаружения мошенничества. Были изучены различные методы интеллектуального анализа данных, которые закрепились в сфере безопасности как самые лучшие и точные: байесовская сеть, дерево решений и искусственные нейронные сети. Также были найдены исследования по использованию метода «случайного леса», который неожиданно оказался точнее, хоть и в весьма специфических условиях, таких как со стороны розничного торговца отрасли, отдавая предпочтение именно розничному торговцу, минимизируя расходы, которые торговец должен нести, если транзакция становится мошеннической.

Список использованных источников

1. Linda Delamaire, Hussein Abdou, John Pointon, «Credit card fraud and detection techniques: a review», Banks and Bank Systems, Volume 4, Issue 2, 2009
2. SamanehSorournejad, Zahra Zojaji, Reza Ebrahimi Atani, Amir Hassan Monadjemi, « A Survey of Credit Card Fraud Detection Techniques: Data and Technique Oriented Perspective», c.3
3. Salvatore J. Stolfo, David W. Fan, Wenke Lee and Andreas L. Prodromidis; "Credit Card Fraud Detection Using Meta-Learning: Issues and Initial Results"; Department of Computer ScienceColumbia University; 1997.
4. Andreas L. Prodromidis and Salvatore J. Stolfo; "Agent-Based Distributed Learning Applied to Fraud Detection"; Department of Computer Science- Columbia University; 2000
5. Maes S. Tuyls K. Vanschoenwinkel B. and Manderick B.; "Credit Card Fraud Detection Using Bayesian and Neural Networks"; Vrije University Brussel – Belgium; 2002.
6. Salvatore J. Stolfo, Wei Fan, Wenke Lee and Andreas L. Prodromidis; "Cost-based Modeling for
7. Cox, E. (1995). A Fuzzy System for Detecting Anomalous Behaviors in Healthcare Provider Claims. In Goonatilake, S. & Treleaven, P. (eds.) Intelligent Systems for Finance and Business, 111-134. John Wiley
8. Williams, G. J. and Huang, Z.(1997). ‘Mining the Knowledge: Mine the Hot Spots Methodology for Mining Large Real World Databases’, 10th Australian Joint Conference on Artificial Intelligence, Published in Lecture Notes in Artificial Intelligence, Springer-Verlag, December, Perth, Western Australia.
9. Williams, G.(1999). ‘Evolutionary Hot Spots Data Mining: An Architecture for Exploring for Interesting Discoveries’, Proceedings of the 3rd Pacific-Asia Conference in Knowledge Discovery and Data Mining, Beijing, China
10. Bhowik R. (2008). “Data Mining Techniques in Fraud Detection”, The Journal of Digital Forensics, Security and Law, Volume 3, Number 2, p.36
11. Brause, R., Langsdorf, T. & Hepp, M. (1999). Neural Data Mining for Credit Card Fraud Detection. Proc. of 11th IEEE International Conference on Tools with Artificial Intelligence
12. SAS e-Intelligence (2000). Data Mining in the Insurance industry: Solving Business problems using SAS Enterprise Miner Software, White Paper
13. Chan, P., Fan, W., Prodromidis, A. & Stolfo, S. (1999). Distributed Data Mining in Credit Card Fraud Detection. IEEE Intelligent Systems 14: 67-74.

14. Lawrence B., Saleena B., Prakash B. (2020) Credit Card Fraud Detection Using Data Mining Techniques. Seybold Report 15(9):2431 - 2436