МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ В ЗАДАЧАХ КОНТРОЛЯ КРИПТОВАЛЮТНЫХ ТРАНЗАКЦИЙ

Феклин В.Г.[1], Соловьев В.И.[2], Корчагин С.А.[3], Царегородцев А.В.[4]

Цель работы: *разработка методики контроля за оборотом цифровых финансовых активов, иных цифровых прав и цифровой валюты для противодействия коррупции на основе анализа криптовалютных транзакций.*

Методы исследования: *методы анализа, сравнения, обобщения, структурной декомпозиции из теории системного анализа, методы машинного обучения.*

Полученный результат: *проведен анализ технологических возможностей контроля за оборотом цифровых финансовых активов, иных цифровых прав, цифровой валюты. Предложена новая методика контроля за оборотом цифровых финансовых активов, иных цифровых прав и цифровой валюты для противодействия коррупции на основе анализа криптовалютных транзакций с использованием методов машинного обучения. Проведено сравнение и оценка точности различных методов машинного обучения: логистическая регрессия, случайный лес, ансамблевые методы. Разработан программный прототип, позволяющий проводить интеллектуальный анализ и контроль криптовалютных транзакций.*

Научная новизна: *предложена новая методика анализа контроля за оборотом цифровых финансовых активов, иных цифровых прав и цифровой валюты для противодействия коррупции на основе анализа криптовалютных транзакций, основанная на технологиях обработки больших данных и методах машинного обучения.*

Вклад соавторов: *Феклин В.Г. — анализ технологических возможностей контроля за оборотом цифровых финансовых активов, иных цифровых прав, цифровой валюты, разработка алгоритмов; Соловьев В.Г. — разработка методики контроля за оборотом цифровых финансовых активов, общее руководство проектом; Корчагин С.А. — подготовка и анализ данных, программная реализация методов интеллектуального анализа и контроля криптовалютных транзакций; Царегородцев А.В. — разработка методологии анализа криптовалютных транзакций.*

Ключевые слова: *цифровые активы, криптовалютные транзакции, интеллектуальная система, методы машинного обучения.*

DOI:10.21681/2311-3456-2023-4-2-11

Введение

Контроль за оборотом цифровых финансовых активов и цифровых валют, в настоящее время, является важной и актуальной темой, о чем свидетельствует ряд исследований, например, [1-3]. Поскольку транзакции с криптовалютами могут быть осуществлены без участия банков, это делает их привлекательными для тех, кто пытается скрыть свои финансовые операции от контролирующих органов или совершить другие незаконные действия. Одной из главных проблем, связанных с цифровыми финансовыми активами, является их анонимность и невозможность контроля за нецелевым использованием. Это может стать причиной незаконных операций, мошенничества и финансирования терроризма. Мошеннические схемы и нарушения закона, проводимые с использованием цифровых финансовых активов рассмотрены в работах [4-6]. Контроль за оборотом цифровых финансовых активов становится необходимым для обеспечения

1 Феклин Вадим Геннадьевич, кандидат физико-математических наук, доцент, декан Факультета информационных технологий и анализа больших данных Финансового университета при Правительстве Российской Федерации, Москва, Россия. E-mail: vfeklin@fa.ru

2 Соловьев Владимир Игоревич, доктор экономических наук, доцент, профессор Департамента анализа данных и машинного обучения Финансового университета при Правительстве Российской Федерации, Москва, Россия. E-mail: vsoloviev@fa.ru

3 Корчагин Сергей Алексеевич, кандидат физико-математических наук, доцент, ведущий научный сотрудник Института цифровых технологий, доцент Департамента анализа данных и машинного обучения, Финансового университета при Правительстве Российской Федерации, Москва, Россия. E-mail: sakorchagin@fa.ru

4 Царегородцев Анатолий Валерьевич, доктор технических наук, профессор, руководитель Департамента информационной безопасности Финансового университета при Правительстве Российской Федерации, Москва, Россия. E-mail: anvtsaregorodtsev@fa.ru

безопасности граждан и борьбы с преступностью [7]. В последние годы, количество незаконных действий с использованием криптовалют только увеличивается [8-9]. С развитием технологий и ростом популярности криптовалюты и других цифровых активов, возникает необходимость в их регулировании со стороны государства и других организаций. Анализ криптовалютных транзакций может способствовать получению сведений, позволяющих выявлять транзакции, которые нарушают законы или противоречат официальной политике государства в области отношения к криптовалютам [10]. Такой анализ может помочь правительствам и уполномоченным государственным ведомствам выявлять связи между подозрительными транзакциями и определенными лицами, а также раскрывать незаконные схемы обращения и использования цифровых финансовых активов и цифровой валюты. Таким образом, контроль за оборотом цифровых финансовых активов, иных цифровых прав и цифровой валюты является важным инструментом противодействия коррупции и другим незаконным финансовым операциям, а регулярный анализ криптовалютных транзакций может помочь идентифицировать законные и незаконные операции с цифровыми активами.

2. Методы и алгоритмы

Общая схема выявления мошеннических операций представлена на рис. 1.

Для поиска подозреваемых лиц, совершающих незаконные криптовалютные транзакции, применяются различные методы и технологии, включая мониторинг криптовалютных бирж [11-12], сбор информации из открытых источников [12-13], анализ блокчейна [14-15], проведения расследований [16-17] и пр. Данные действия находятся в компетенции правоохранительных органов и не являются предметом исследования настоящей работы. Мы остановимся на 3-м (анализ транзакций) и 4-м (выявление мошенничества) блоках приведенной на рис. 1 схемы. После того, как компетентные органы выявили подозреваемое лицо и установили перечень криптовалютных кошельков, принадлежащих подозреваемому лицу, встает задача анализа транзакций, которые проходили через эти счета. Зачастую набор транзакций представляет собой большой объем данных. Для проведения их анализа необходимы автоматизированные средства обработки таких данных, включая технологии искусственного интеллекта. В работе предлагается методология анализа криптовалютных транзакций, включающая в себя:

- корреляционный анализ – для определения связей между различными криптовалютными транзакциями и выявления потенциальных злоумышленников;
- кластерный анализ – для группировки криптовалютных транзакций по различным параметрам, например, объем транзакции, время, местоположение, это позволяет выявить подозрительные группы транзакций;
- факторный анализ – для анализа взаимосвязи между признаками и выделение факторов, влияющих на криптовалютные транзакции;
- графовый анализ – для анализа связей между узлами (адресами кошельков) в сети блокчейн и выявление подозрительных транзакций, осуществляемых между ними;
- методы классификации машинного обучения – для анализа больших объемов данных, обучения на основе примеров и автоматического распознавания мошеннических транзакций.

В текущем исследовании реализован один из подходов, предлагаемых в методологии, который основан на методах машинного обучения для классификации мошеннических операций.

Алгоритм работы автоматизированной системы для выявления подозрительных транзакций представлен на рис. 2.

Исходными данными для построения моделей машинного обучения был набор данных для обнаружения мошенничества по транзакциям в сети Ethereum [18]. Криптовалюта Ethereum – одна из самых популярных криптовалют, наряду с Биткоином и несколькими другими альткойнами [19]. Etherium обладает следующими свойствами: широкий спектр применения, высокая скорость обработки транзакций, поддержка

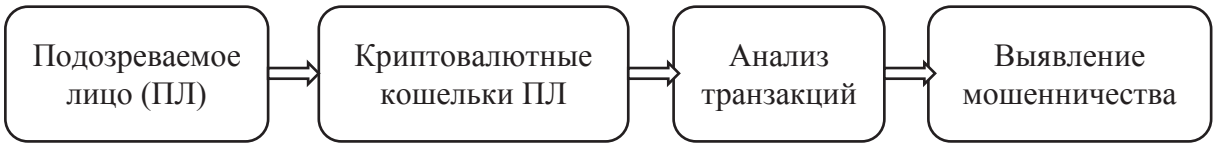

Рис. 1. Общая схема выявления мошеннических операций

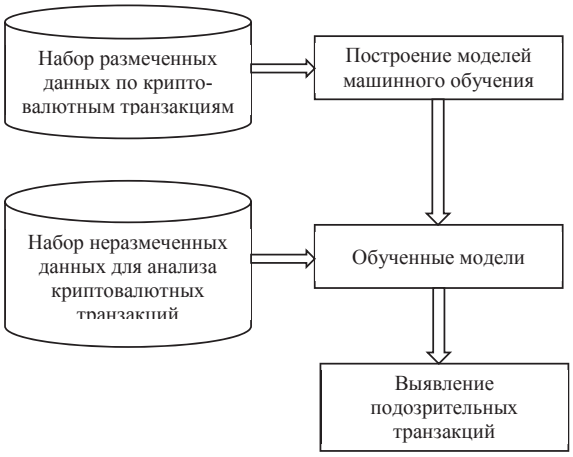

Рис. 2. Алгоритм работы автоматизированной системы для выявления подозрительных транзакций

большим количеством бирж, высокий уровень защиты конфиденциальности. Исследования [20-21] показали, что данная сеть довольно часто используется мошенниками и правонарушителями, поэтому в качестве примера для анализа транзакций мы взяли эту криптовалюту. Набор данных для обучения моделей содержит следующие признаки:

- Index – порядковый номер строки;
- Address – адрес учетной записи Ethereum;
- FLAG – флаг: является ли транзакция мошеннической или нет;
- Avg min between sent tnx – среднее время между отправленными транзакциями для аккаунта в минутах;
- Avg min between received tnx – среднее время между полученными транзакциями для аккаунта в минутах;
- Time Diff between first and_last (Mins) – разница во времени между первой и последней транзакцией;
- Sent_tnx – общее количество отправленных обычных транзакций;
- Received_tnx – общее количество полученных обычных транзакций;
- NumberofCreated_Contracts – общее количество созданных контрактных транзакций;
- UniqueReceivedFrom_Addresses – всего уникальных адресов, с которых были получены транзакции;
- UniqueSentTo_Addresses – всего уникальных адресов, с которых были отправлены транзакции;
- MinValueReceived – минимальное значение в эфире, когда-либо полученное;
- MaxValueReceived – максимальное значение в эфире, когда-либо полученное;
- AvgValueReceived – среднее значение в эфире, когда-либо полученное;
- MinValSent – минимальное значение эфира, когда-либо отправленное;
- MaxValSent – максимальное количество эфира, когда-либо отправленное;
- AvgValSent – среднее значение эфира, когда-либо отправленное;
- MinValueSentToContract – минимальная стоимость эфира, отправленная на контракт;
- MaxValueSentToContract – максимальное количество эфира, отправленного на контракт;
- AvgValueSentToContract – средняя стоимость эфира, отправленного на контракты;
- TotalTransactions(IncludingTnxtoCreate_Contract) – общее количество транзакций;
- TotalEtherSent – общее количество эфира, отправленного на адрес аккаунта;
- TotalEtherReceived – общее количество эфира, полученное для адреса аккаунта;
- TotalEtherSent_Contracts – общее количество эфира, отправленного на адреса контрактов;
- TotalEtherBalance – общий баланс эфира после проведенных транзакций;
- TotalERC20Tnxs – общее количество транзакций по передаче токенов ERC20;
- ERC20TotalEther_Received – общее количество транзакций, полученных токеном ERC20 в эфире;
- ERC20TotalEther_Sent – общее количество транзакций, отправленных токеном ERC20 в эфире;
- ERC20TotalEtherSentContract – общий перевод токена ERC20 на другие контракты в эфире;
- ERC20UniqSent_Addr – количество транзакций с токенами ERC20, отправленных на уникальные адреса учетных записей;
- ERC20UniqRec_Addr – количество транзакций с токенами ERC20, полученных с уникальных адресов;
- ERC20UniqRecContractAddr – количество транзакций с токенами ERC20, полученных с уникальных адресов контрактов;
- ERC20AvgTimeBetweenSent_Tnx – среднее время между отправленными транзакциями с токеном ERC20 в минутах;
- ERC20AvgTimeBetweenRec_Tnx – среднее время между транзакциями, полученными токеном ERC20, в минутах;
- ERC20AvgTimeBetweenContract_Tnx – среднее время токена ERC20 между отправленными транзакциями токена;

— ERC20MinVal_Rec – минимальное значение в эфире, полученное от транзакций с токенами ERC20 для учетной записи;
— ERC20MaxVal_Rec – максимальное значение в эфире, полученное от транзакций с токенами ERC20 для учетной записи;
— ERC20AvgVal_Rec – среднее значение в эфире, полученное от транзакций с токенами ERC20 для учетной записи;
— ERC20MinVal_Sent – минимальное значение в эфире, отправленное из транзакций с токенами ERC20 для учетной записи;
— ERC20MaxVal_Sent – максимальное значение в эфире, отправленное из транзакций с токенами ERC20 для учетной записи;
— ERC20AvgVal_Sent – среднее значение в эфире, отправленное из транзакций с токенами ERC20 для учетной записи;
— ERC20UniqSentTokenName – количество переданных уникальных токенов ERC20;
— RC20UniqRecTokenName – количество полученных уникальных токенов ERC20;
— ERC20MostSentTokenType – наиболее часто отправляемый токен для учетной записи через транзакцию ERC20;
— ERC20MostRecTokenType – наиболее популярный токен для учетной записи через транзакции ERC20.

Построение моделей машинного обучения проходило на языке Python с использованием библиотек pandas, numpy, matplotlib, seaborn, scikit-learn. В качестве моделей машинного обучения использовались: логистическая регрессия, случайный лес, XGBoost.

Данные методы используются для классификации и прогнозирования, каждый метод имеет свои преимущества и недостатки. Логистическая регрессия – метод, используемый для вероятностной классификации, когда относительный риск зависит от набора независимых переменных [22]. Данный метод может помочь в прогнозировании, когда нужно предсказать наличие или отсутствие определенного события, на основе данных о предыдущих событиях. Основным преимуществом этого метода является простота его реализации и легкость интерпретации результатов. Однако, логистическая регрессия обычно применяется только для данных с линейным порядком и малым количеством признаков. Случайный лес – метод, который использует множество деревьев решений также для задач классификации и прогнозирования [23]. Он позволяет улучшить точность прогнозирования, а также уменьшить переобучение модели. Его преимущества включают возможность работы с большим количеством признаков и возможность обработки нелинейных данных. Однако, случайный лес может быть медленным в расчетах и требуются дополнительные усилия для подбора корректной модели. XGBoost – это метод градиентного бустинга, использующий деревья решений для классификации и прогнозирования [24]. Он может обрабатывать любые типы данных, включая категориальные признаки, что является большим преимуществом. Также XGBoost может работать с большим числом признаков, что может быть полезно при работе с большими наборами данных. Однако, XGBoost может быть сложным в использовании. Анализ моделей машинного обучения и подбор наиболее оптимального метода применительно к задаче выявления мошеннических транзакций является одной из ключевых задач настоящего исследования. Достоверность правильного обнаружения подозрительной транзакции определялась точностью (precision), характеризующей способность алгоритма обнаруживать именно нужные объекты и полнотой (recall) обнаружения [25], характеризующей способность алгоритма находить нужный объект в полном объеме:

$$precision = \frac{TP}{TP + FP},$$

$$recall = \frac{TP}{TP + FN},$$

где TP – число правильных решений о наличии объекта;
FP – число ошибок второго рода;
FN – число ошибок первого рода.

Для интегральной характеристики точности и полноты использовалась F-мера [26]. Данная метрика рассчитывается в соответствии с выражением:

$$F = \left(1 + \beta^2\right) \frac{precision \times recall}{\beta^2 \times precision + recall}.$$

3. Результаты вычислительных экспериментов

3.1. Первичный анализ данных и их предобработка

На первом этапе исследований проводился первичный анализ данных и их предобработка. Набор размеченных данных по криптовалютным транзакциям, который будет использоваться для обучения, содержит 9841 записей и 48 признаков. Типы данных, используемых в признаках, распределены следующим образом: float64 – 39 признаков, int64 – 7 признаков, object – 2 признака. На следующем этапе

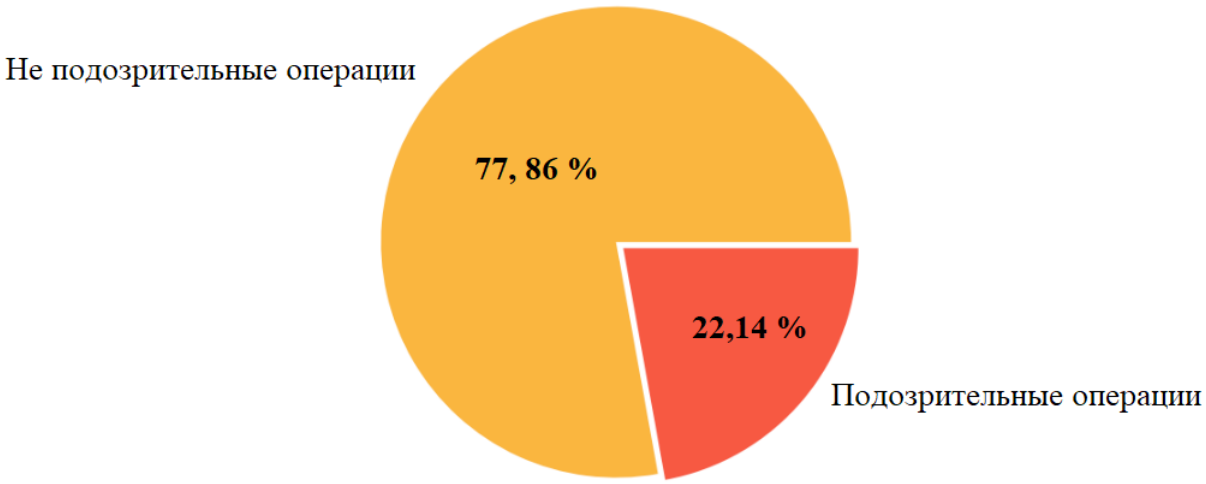

Рис. 3. Распределение категорий подозрительных и не подозрительных операций в обучающем датасете

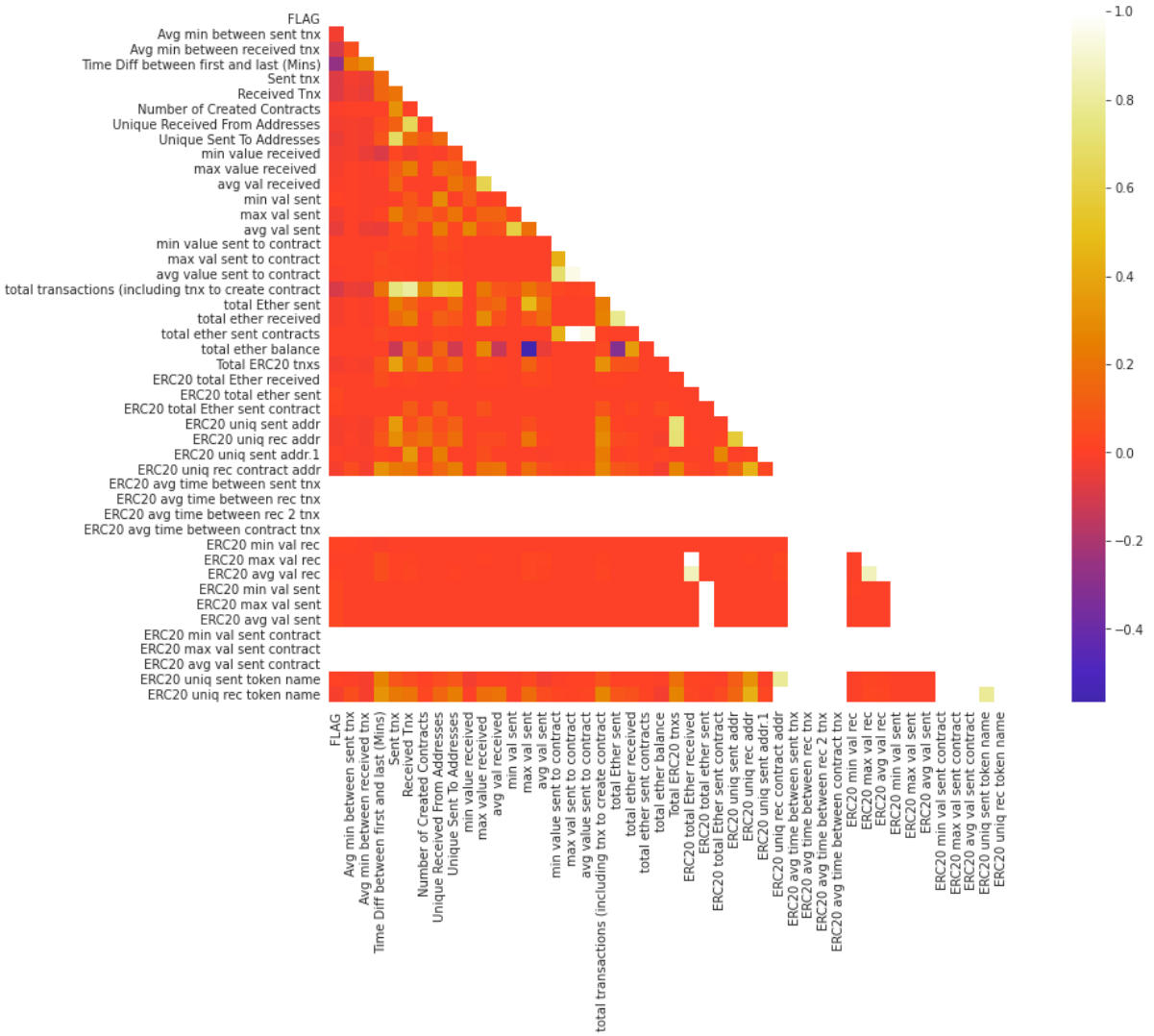

Рис. 4. Общий вид корреляционной карты

для большей эффективности проведения вычислений объектные переменные переводим в категорию объектов типа данных – dtype. После проведения описательных статистик проводим анализ распределения в обучающем датасете по категориям: подозрительные и не подозрительные операции. Как видим из рис. 3 – 77,6 % транзакций являются не подозрительными, а 22, 14 – подозрительными.

Важным аспектом при проведении исследования является наличие зависимостей между различными признаками. Для установления таких зависимостей была построена корреляционная карта, общий вид которой представлен на рис. 4. Как видим из рисунка, в целом, большинство признаков слабо коррелируют между собой, однако есть признаки, которые заслуживают внимание – это в первую очередь пары параметров с высоким уровнем корреляции, включая отрицательную корреляцию, а также те параметры, которые наибольшим образом коррелируют с параметром – флаг, который указывает на то, является ли транзакция мошеннической или нет. Анализ данных параметров позволяет установить следующие закономерности:

- для мошеннических операций характерна малая разница во времени между первой и последний транзакцией;
- средний объем отправленной криптовалюты в мошеннических операциях превышает средний объем отправленной криптовалюты в операциях, проведенных без нарушения закона;
- общее количество транзакций в мошеннических операциях больше, чем в не мошеннических;
- в мошеннических операциях используется большее количество уникальных адресов, с которых отправляются транзакции.

После первичного анализа данных была проведена их предобработка для последующего машинного обучения. Данные были разделены на тренировочный (80%) и тестовый набор (20%), нормализованы функции обучения, проведено устранение дисбаланса классов.

3.2. Построение моделей машинного обучения

Для построения прогнозных моделей были использованы следующие методы машинного обучения: логистическая регрессия, случайный лес, XGBoost.

Модель, основанная на логистической регрессии, показала следующие результаты:

Таблица 1

Метрики для модели, основанной на логистической регрессии

	Точность	Полнота	F-мера
0	0,96	0,89	0,92
1	0,68	0,88	0,77

На рис. 5 изображена матрица неточности для модели логистической регрессии.

Рис. 5. Матрица неточности для модели логистической регрессии

Исходя из рис. 5 видим – модель логистической регрессии правильно идентифицировала 373 (TP) случая мошенничества из 422 (P). Также указанная модель установила метку «мошенничество» в 171 (FP) из 1547, когда эти случаи на самом деле не были мошенничеством. Имея дело со сценарием обнаружения мошенничества, наибольшую значимость имеют транзакции, которые на самом деле были мошенническими, но которые рассматривались нашей моделью как законные операции (FN - 49), т.е. мы столкнулись с ошибкой второго типа (в [27, 28] также называют «неверными отрицательными утверждениями»).

Далее был построен классификатор на основе модели случайного леса.

Результаты работы модели случайного леса для исследуемых данных показали следующие результаты (таблица 2).

Таблица 2

Метрики для модели, основанной на случайном лесе

	Точность	Полнота	F-мера
0	0,99	0,98	0,98
1	0,93	0,95	0,94

Матрица неточности для модели случайного леса выглядит следующим образом – рис. 6.

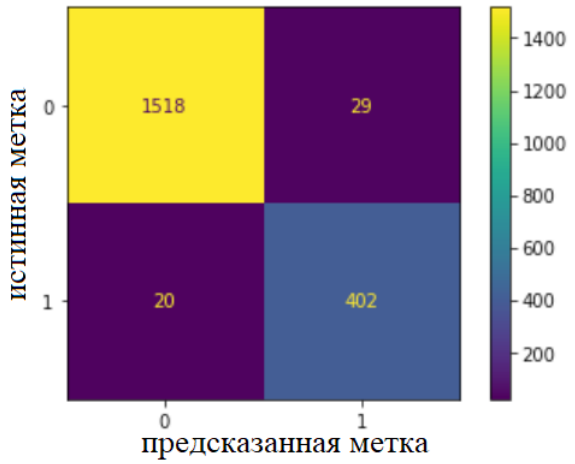

Рис. 6. Матрица неточности для модели случайного леса

Классификатор на основе модели случайного леса дает более точные результаты. Число ошибок первого и второго рода значительно снижены, что увеличивает точность и полноту. Используя модель случайного леса для исследуемых данных, не удалось выявить только 20 случаев мошеннических транзакций из 9841.

Далее была построена ансамблевая модель на основе XGBoost. В табл. 3 представлены результаты для данной модели.

Таблица 3

Метрики для модели, основанной на XGBoost

	Точность	Полнота	F-мера
0	0,99	0,98	0,98
1	0,95	0,96	0,96

На рис. 7 представлена матрица неточности для модели XGBoost.

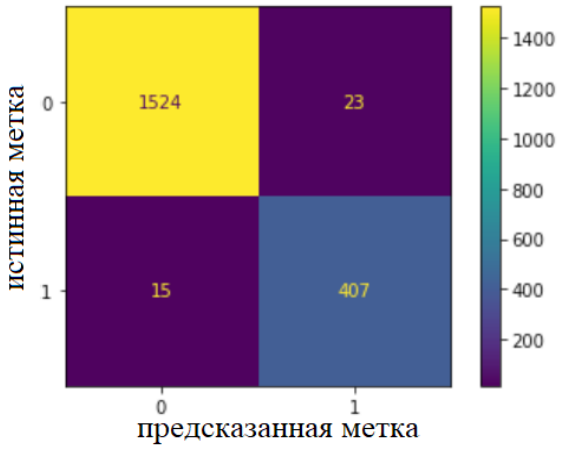

Рис.7. Матрица неточности для модели XGBoost

Результаты ансамблевой модели XGBoost показывают, что она работает лучше, чем случайный лес.

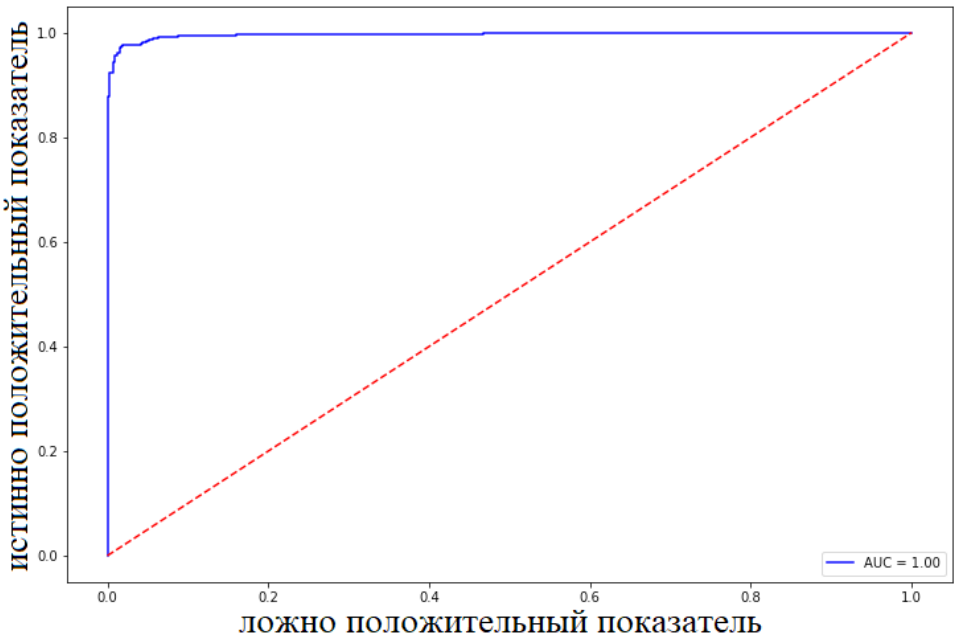

Рис. 8. ROC-кривая для модели XGBoost

Модель ошиблась в 22 случаях, определив транзакции как мошеннические, хотя на самом деле они таковыми не были. Что касается выявления мошеннических транзакций, модель XGBoost пропустила 16 транзакций из 9841, что также является наилучшим результатом. В ходе исследования была проведена попытка подбора гиперпараметров для улучшения данной модели, однако, подбор гиперпараметров не привел к улучшению её работы. Для более наглядной оценки качества классификации построена ROC-кривая, которая показана на рис. 8.

Поскольку модель XGBoost показала наилучшие результаты именно она была использована в серверной части прототипа автоматизированной системы по контролю криптовалютных транзакций.

Заключение

В работе рассмотрены технологические возможности контроля за оборотом цифровых финансовых активов, иных цифровых прав, цифровой валюты на примере сети Ethereum. Предложен алгоритм автоматизированной системы по выявлению мошеннических транзакций, методология их анализа, а также реализован ряд моделей машинного обучения, позволяющих такие операции идентифицировать. Наилучшие результаты по обнаружению мошеннических транзакций криптовалюты Ethereum показала ансамблевая модель XGBoost, которая и легла в основу прототипа автоматизированной системы. В исследовании было показано, что методы машинного обучения являются эффективным инструментом для контроля криптовалютных транзакций. Они позволяют автоматически обнаруживать аномальные операции, которые могут быть связаны с различными незаконными действиями. Методы машинного обучения также могут быть использованы для анализа моделей поведения участников рынка криптовалют и прогнозирования ценовой динамики, что не входило в задачу настоящего исследования, однако, представляет интерес и может послужить объектом дальнейших исследований. Тем не менее, рассмотренные в работе методы не являются универсальными и требуют тщательной настройки и обучения для каждого конкретного случая.

Статья подготовлена по результатам исследований, выполненных за счет бюджетных средств по государственному заданию Финуниверситета.

Литература

1. Татоян А. А. Экономико-правовая природа цифровых финансовых активов //Образование и право. – 2022. – №. 1. – С. 107-111.
2. Сорока Э. Ю. Правовая природа цифровых финансовых активов в законодательстве Российской Федерации //Вопросы российского и международного права. – 2021. – Т. 11. – №. 9-1. – С. 84.
3. Shestak V., Kiseleva A., Kolesnikov Y. Taxation Issues for Digital Financial Assets //Social Science Computer Review. – 2021. – С. 08944393211003919.
4. Соловьев В.И., Конторович В.К., Феклин В.Г. О возможности осуществления контроля за оборотом цифровых финансовых активов //Проблемы экономики и юридической практики Учредители: ООО" Издательский дом" Юр-ВАК". – 2022. – Т. 18. – №. 5. – С. 242-247.
5. Симаков А. А., Неелов В. В. Схемы преступлений с использованием криптовалюты //Закон и право. – 2020. – №. 5. – С. 106-109.
6. Bartoletti M. et al. Cryptocurrency scams: analysis and perspectives //Ieee Access. – 2021. – Т. 9. – С. 148353-148373.
7. Царегородцев А.В., Романовский С.В., Волков С.Д., Самойлов В.Е. Управление рисками информационной безопасности цифровых продуктов финансовой экосистемы организации //Моделирование, оптимизация и информационные технологии. – 2020. – Т. 8. – №. 4(31). – Доступно по: https://moitvivt.ru/ru/journal/pdf?id=888 DOI:10.26102/2310- 6018/2020.31.4.038.
8. Соловьев В. И., Конторович В. К., Феклин В. Г., Лавров Д. А. Контроль за совершением правонарушений в сфере криптовалют / // РИСК: Ресурсы, Информация, Снабжение, Конкуренция. – 2022. – № 4. – С. 156-160.
9. Mackenzie S. Criminology towards the metaverse: Cryptocurrency scams, grey economy and the technosocial //The British Journal of Criminology. – 2022. – Т. 62. – №. 6. – С. 1537-1552.
10. Пелисова И. П. Использование криптовалюты при совершении преступлений, предусмотренных статьями 174-175 УК РФ //Современные закономерности и тенденции развития наук криминального цикла. – 2020. – С. 161-163.
11. Kim D., Bilgin M. H., Ryu D. Are suspicious activity reporting requirements for cryptocurrency exchanges effective? //Financial Innovation. – 2021. – Т. 7. – №. 1. – С. 1-17.
12. Roberts H. et al. Media cloud: Massive open source collection of global news on the open web //Proceedings of the International AAAI Conference on Web and Social Media. – 2021. – Т. 15. – С. 1034-1045.
13. Gasser R. et al. Cottontail DB: an open-source database system for multimedia retrieval and analysis //Proceedings of the 28th ACM International Conference on Multimedia. – 2020. – С. 4465-4468.
14. Соловьев В.И., Федоткина О.П., Феклин В.Г., Коровин Д.И. Технологические возможности контроля за оборотом цифровых финансовых активов / // Современная наука: актуальные проблемы теории и практики. Серия: Экономика и право. – 2022. – № 11. – С. 87-93.

15. Гарипов Р. И., Максимова Н. Н. Анализ методических подходов к оценке эффективности блокчейна //Управление в современных системах. – 2020. – №. 1 (25). – С. 13-17.
16. Никитин П. В. и др. Распознавание эмоций по аудио сигналам как один из способов борьбы с телефонным мошенничеством // Программные системы и вычислительные методы. – 2022. – №. 3. – С. 1-13.
17. 17. Velasco C. Cybercrime and Artificial Intelligence. An overview of the work of international organizations on criminal justice and the international applicable instruments //ERA Forum. – Berlin/Heidelberg: Springer Berlin Heidelberg, 2022. – Т. 23. – №. 1. – С. 109-126.
18. Jung E. et al. Data mining-based ethereum fraud detection //2019 IEEE International Conference on Blockchain (Blockchain). – IEEE, 2019. – С. 266-273.
19. Wu J. et al. Who are the phishers? phishing scam detection on ethereum via network embedding //IEEE Transactions on Systems, Man, and Cybernetics: Systems. – 2020. – Т. 52. – №. 2. – С. 1156-1166.
20. Chen L. et al. Phishing scams detection in ethereum transaction network //ACM Transactions on Internet Technology (TOIT). – 2020. – Т. 21. – №. 1. – С. 1-16.
21. Senaviratna N., A Cooray T. M. J. Diagnosing multicollinearity of logistic regression model //Asian Journal of Probability and Statistics. – 2019. – Т. 5. – №. 2. – С. 1-9.
22. Kamps J., Trozze A., Kleinberg B. Cryptocurrencies: Boons and curses for fraud prevention //A Fresh Look at Fraud. – Routledge, 2022. – С. 192-219.
23. Андриянов Н. А., Дементьев В. Е., Ташлинский А. Г. Обнаружение объектов на изображении: от критериев Байеса и Неймана–Пирсона к детекторам на базе нейронных сетей EfficientDet //Компьютерная оптика. – 2022. – Т. 46. – №. 1. – С. 139-159.
24. Chen T. et al. Xgboost: extreme gradient boosting //R package version 0.4-2. – 2015. – Т. 1. – №. 4. – С. 1-4.
25. Carvalho D. V., Pereira E. M., Cardoso J. S. Machine learning interpretability: A survey on methods and metrics //Electronics. – 2019. – Т. 8. – №. 8. – С. 832.
26. Soleymani R., Granger E., Fumera G. F-measure curves: A tool to visualize classifier performance under imbalance //Pattern Recognition. – 2020. – Т. 100. – С. 107146.
27. Petritoli E., Leccese F., Spagnolo G. S. Inertial Navigation Systems (INS) for Drones: Position Errors Model //2020 IEEE 7th International Workshop on Metrology for AeroSpace (MetroAeroSpace). – IEEE, 2020. – С. 500-504.
28. Childs A. M. et al. Theory of trotter error with commutator scaling //Physical Review X. – 2021. – Т. 11. – №. 1. – С. 011020.

MACHINE LEARNING METHODS FOR CONTROL OF CRYPTOCURRENCY TRANSACTIONS

Feklin V.G.[5], Soloviev V.I.[6], Korchagin S.A.[7], Tsaregorodtsev A.V.[8]

The purpose of the work: *to develop a methodology for controlling the circulation of digital financial assets, other digital rights and digital currency to combat corruption based on the analysis of cryptocurrency transactions.*

Research methods: *methods of analysis, comparison, generalization, structural decomposition from the theory of system analysis, machine learning methods.*

Result obtained: *an analysis of the technological possibilities of controlling the circulation of digital financial assets, other digital rights, digital currency was carried out. A new method for controlling the circulation of digital financial assets, other digital rights and digital currency to combat corruption is proposed based on the analysis of cryptocurrency transactions using machine learning methods. A comparison and evaluation of the accuracy of various machine learning methods was carried out: logistic regression, random forest, ensemble methods. A software prototype has been developed that allows for intellectual analysis and control of cryptocurrency transactions.*

Scientific novelty: *a new method for analysing control over the circulation of digital financial assets, other digital rights and digital currency to combat corruption based on the analysis of cryptocurrency transactions, based on big data processing technologies and machine learning methods, is proposed.*

Contribution of co-authors: *Feklin V.G. — analysis of technological capabilities to control the circulation of digital financial assets, other digital rights, digital currency, development of algorithms; Solovyov V.G. — development of*

5 Vadim G. Feklin, Ph.D., Dean of the Faculty of Information Technology and Big Data Analysis of the Financial University under the Government of the Russian Federation, Moscow, Russia. E-mail: vfeklin@fa.ru

6 Vladimir I. Soloviev, Dr.Sc., Professor, Department of Data Analysis and Machine Learning, Financial University under the Government of the Russian Federation, Moscow, Russia. E-mail: vsoloviev@fa.ru

7 Sergeiy A. Korchagin, Ph.D., Senior Researcher, Institute of Digital Technologies, Associate Professor, Department of Data Analysis and Machine Learning, Financial University under the Government of the Russian Federation, Moscow, Russia. E-mail: sakorchagin@fa.ru

8 Anatoly V. Tsaregorodtsev, Dr.Sc. (Eng), Professor, Head of Information Security Department, Financial University under the Government of the Russian Federation, Moscow, Russia. E-mail: anvtsaregorodtsev@fa.ru

a method for controlling the circulation of digital financial assets, general project management; Korchagin S.A. — preparation and analysis of data, software implementation of methods for intellectual analysis and control of cryptocurrency transactions; Tsaregorodtsev A.V. — development of a methodology for analysis of cryptocurrency transactions.

Keywords: *digital assets, cryptocurrency transactions, intelligent system, machine learning methods.*

References

1. Tatoyan A. A. Ekonomiko-pravovaya priroda cifrovyh finansovyh aktivov //Obrazovanie i pravo. – 2022. – №. 1. – S. 107-111.
2. Soroka E. YU. Pravovaya priroda cifrovyh finansovyh aktivov v zakonodatel'stve Rossijskoj Federacii //Voprosy rossijskogo i mezhdunarodnogo prava. – 2021. – T. 11. – №. 9-1. – S. 84.
3. Shestak V., Kiseleva A., Kolesnikov Y. Taxation Issues for Digital Financial Assets //Social Science Computer Review. – 2021. – C. 08944393211003919.
4. Soloviev V.I., Kontorovich V.K., Feklin V. G. O vozmozhnosti osushchestvleniya kontrolya za oborotom cifrovyh finansovyh aktivov // problemy ekonomiki i yuridicheskoj praktiki Uchrediteli: OOO "Izdatel'skij dom" YUr-VAK". – 2022. – T. 18. – №. 5. – S. 242-247.
5. Simakov A. A., Neelov V. V. Skhemy prestuplenij s ispol'zovaniem kriptovalyuty //Zakon i pravo. – 2020. – №. 5. – S. 106-109.
6. Bartoletti M. et al. Cryptocurrency scams: analysis and perspectives //Ieee Access. – 2021. – T. 9. – C. 148353-148373.
7. Tsaregorodtsev A.V., Romanovsky S.V., Volkov S.D., Samoilov V.E. Upravlenie riskami informatcionnoi bezopasnosti tsifrovih produktov finansovoi ekosistemy organizatsii //Modelirovanie, optimizatsiya i informatsionnie tekhnologii. – 2020. – T. 8. – №. 4(31). – Dostupno po: https://moitvivt.ru/ru/journal/pdf?id=888 DOI:10.26102/2310- 6018/2020.31.4.038.
8. Soloviev V. I., Kontorovich V. K., Feklin V. G., Lavrov D. A. Kontrol' za sovershenie pravonarushenij v sfere kriptovalyut / // RISK: Resursy, Informaciya, Snabzhenie, Konkurenciya. – 2022. – № 4. – S. 156-160.
9. Mackenzie S. Criminology towards the metaverse: Cryptocurrency scams, grey economy and the technosocial //The British Journal of Criminology. – 2022. – T. 62. – №. 6. – C. 1537-1552.
10. Pelisova I. P. Ispol'zovanie kriptovalyuty pri sovershenii prestuplenij, predusmotrennyh stat'yami 174-175 UK RF //Sovremennye zakonomernosti i tendencii razvitiya nauk kriminal'nogo cikla. – 2020. – S. 161-163.
11. Kim D., Bilgin M. H., Ryu D. Are suspicious activity reporting requirements for cryptocurrency exchanges effective? //Financial Innovation. – 2021. – T. 7. – №. 1. – C. 1-17.
12. Roberts H. et al. Media cloud: Massive open source collection of global news on the open web //Proceedings of the International AAAI Conference on Web and Social Media. – 2021. – T. 15. – C. 1034-1045.
13. Gasser R. et al. Cottontail DB: an open source database system for multimedia retrieval and analysis //Proceedings of the 28th ACM International Conference on Multimedia. – 2020. – C. 4465-4468.
14. Soloviev V.I., Fedotkina O.P., Feklin V.G., Korovin D.I. Tekhnologicheskie vozmozhnosti kontrolya za oborotom cifrovyh finansovyh aktivov / // Sovremennaya nauka: aktual'nye problemy teorii i praktiki. Seriya: Ekonomika i pravo. – 2022. – № 11. – S. 87-93.
15. Garipov R. I., Maksimova N. N. Analiz metodicheskih podhodov k ocenke effektivnosti blokchejna //Upravlenie v sovremennyh sistemah. – 2020. – №. 1 (25). – S. 13-17.
16. Nikitin P. V. i dr. Raspoznavanie emocij po audio signalam kak odin iz sposobov bor'by s telefonnym moshennichestvom //Programmnye sistemy i vychislitel'nye metody. – 2022. – №. 3. – S. 1-13.
17. Velasco C. Cybercrime and Artificial Intelligence. An overview of the work of international organizations on criminal justice and the international applicable instruments //ERA Forum. – Berlin/Heidelberg: Springer Berlin Heidelberg, 2022. – T. 23. – №. 1. – C. 109-126.
18. Jung E. et al. Data mining-based ethereum fraud detection //2019 IEEE International Conference on Blockchain (Blockchain). – IEEE, 2019. – C. 266-273.
19. Wu J. et al. Who are the phishers? phishing scam detection on ethereum via network embedding //IEEE Transactions on Systems, Man, and Cybernetics: Systems. – 2020. – T. 52. – №. 2. – C. 1156-1166.
20. Chen L. et al. Phishing scams detection in ethereum transaction network //ACM Transactions on Internet Technology (TOIT). – 2020. – T. 21. – №. 1. – C. 1-16.
21. Senaviratna N., A Cooray T. M. J. Diagnosing multicollinearity of logistic regression model //Asian Journal of Probability and Statistics. – 2019. – T. 5. – №. 2. – C. 1-9.
22. Kamps J., Trozze A., Kleinberg B. Cryptocurrencies: Boons and curses for fraud prevention //A Fresh Look at Fraud. – Routledge, 2022. – C. 192-219.
23. Andriyanov N. A., Dement'ev V. E., Tashlinskij A. G. Obnaruzhenie ob"ektov na izobrazhenii: ot kriteriev Bajesa i Nejmana–Pirsona k detektoram na baze nejronnyh setej EfficientDet //Komp'yuternaya optika. – 2022. – T. 46. – №. 1. – S. 139-159.23. Chen T. et al. Xgboost: extreme gradient boosting //R package version 0.4-2. – 2015. – T. 1. – №. 4. – C. 1-4.
24. Chen T. et al. Xgboost: extreme gradient boosting //R package version 0.4-2. – 2015. – T. 1. – №. 4. – C. 1-4.
25. Carvalho D. V., Pereira E. M., Cardoso J. S. Machine learning interpretability: A survey on methods and metrics //Electronics. – 2019. – T. 8. – №. 8. – C. 832.
26. Soleymani R., Granger E., Fumera G. F-measure curves: A tool to visualize classifier performance under imbalance //Pattern Recognition. – 2020. – T. 100. – C. 107146.
27. Petritoli E., Leccese F., Spagnolo G. S. Inertial Navigation Systems (INS) for Drones: Position Errors Model //2020 IEEE 7th International Workshop on Metrology for AeroSpace (MetroAeroSpace). – IEEE, 2020. – C. 500-504.
28. Childs A. M. et al. Theory of trotter error with commutator scaling //Physical Review X. – 2021. – T. 11. – №. 1. – C. 011020.