DOI: 10.6060/snt.20258303.0006
УДК 303.732

СИСТЕМА ФИЛЬТРАЦИИ ЛОЖНОПОЛОЖИТЕЛЬНЫХ РЕЗУЛЬТАТОВ OSINT С ИСПОЛЬЗОВАНИЕМ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ОЦЕНКИ ПРИНАДЛЕЖНОСТИ ЦИФРОВЫХ АКТИВОВ КОМПАНИЙ

Астраханцева И.А, Брюшинин А.О., Астраханцев Р.Г., Котенев Т.Е.

Астраханцева Ирина Александровна (ORCID 0000-0003-2841-8639), Астраханцев Роман Геннадьевич (ORCID 0000-0001-9880-2826), Котенев Тимофей Евгеньевич (ORCID 0009-0003-4963-2156)
Ивановский государственный химико-технологический университет,
г. Иваново, Россия. 153000, Ивановская область, г. Иваново, пр. Шереметевский, д. 7.
E-mail: i.astrakhantseva@mail.ru, rgastrakhantsev@gmail.com, kotenev@me.com
Брюшинин Антон Олегович (ORCID 0009-0004-0453-6766)
Российский химико-технологический университет им. Д.И. Менделеева,
г. Москва, Россия. 125047, Московская область, г. Москва, Миусская площадь, 9.
E-mail: anton.bryushinin@yandex.ru

Разведка на основе открытых источников (OSINT) является краеугольным камнем современного аудита информационной безопасности и операций красных команд (red teaming), обеспечивая выявление цифровых активов организации с использованием общедоступных данных. Однако существующие инструменты OSINT часто генерируют избыточное количество ложноположительных результатов при попытке связать IP-адреса и доменные имена с искомыми организациями. В статье предлагается система фильтрации на основе методов машинного обучения, существенно снижающая уровень ложных срабатываний в выборках, полученных с использованием OSINT. Система формирует высокоразмерные векторы признаков на основе данных WHOIS, метаданных DNS, SSL-сертификатов и содержимого веб-страниц. С использованием ансамблевого классификатора на основе градиентного бустинга достигается высокая точность атрибуции при сохранении устойчивости к разнообразным типам инфраструктуры. Эксперимент проведен на более чем 30 000 реальных цифровых объектов, собранных от семи организаций, и показал значительное повышение точности атрибуции, подтверждая практическую применимость предлагаемого подхода для обеспечения информационной безопасности.

Ключевые слова: OSINT, машинное обучение, атрибуция цифровых активов, ложноположительные результаты, классификация IP-адресов, доменная атрибуция, данные WHOIS, DNS-записи, SSL-сертификаты, градиентный бустинг, автоматизация кибербезопасности, извлечение информации

ML-BASED FILTERING SYSTEM FOR FALSE-POSITIVE OSINT RESULTS IN CORPORATE ASSET ATTRIBUTION

Astrakhantseva I.A., BryushininA.O., Astrakhantsev R.G., Kotenev T.E.

Astrakhantseva Irina Aleksandrovna (ORCID 0000-0003-2841-8639), Astrakhantsev Roman Gennadievich (ORCID 0000-0001-9880-2826), Kotenev Timofey Evgenievich (ORCID 0009-0003-4963-2156)
Ivanovo State University of Chemistry and Technology,
Ivanovo, Russia. 153000, Ivanovo region, Ivanovo, Sheremetevsky Ave., 7.
E-mail: i.astrakhantseva@mail.ru, rgastrakhantsev@gmail.com, kotenev@me.com
Bryushinin Anton Olegovich (ORCID 0009-0004-0453-6766)
Russian University of Chemical Technology DI. Mendeleev,
Moscow, Russia. 125047, Moscow region, Moscow, Miusskaya square, 9.
E-mail: anton.bryushinin@yandex.ru

Open-Source Intelligence (OSINT) has become a cornerstone of modern cybersecurity auditing and red teaming operations, enabling the identification of an organization's digital assets using publicly available data. However, existing OSINT tools often produce a high number of false posi-

tives when attempting to associate IP addresses and domain names with target organizations. This paper proposes a machine learning-based filtering system that significantly reduces false positive rates in OSINT-derived datasets. The system generates high-dimensional feature vectors using WHOIS data, DNS metadata, SSL certificates, and website content. By employing an ensemble classifier based on gradient boosting, the approach achieves high attribution accuracy while maintaining robustness across diverse infrastructure types. The system was evaluated on more than 30,000 real-world digital assets collected from seven organizations and demonstrated a substantial improvement in attribution accuracy, confirming the practical applicability of the proposed solution for enhancing cybersecurity operations.

Keywords: OSINT, machine learning, digital asset attribution, false positives, IP address classification, domain attribution, WHOIS data, DNS records, SSL certificates, gradient boosting, cybersecurity automation, information extraction

ВВЕДЕНИЕ

Цифровая трансформация организаций значительно расширила их внешнюю атакуемую поверхность. Это делает проактивную оценку защищённости важнейшим элементом современных практик кибербезопасности. Среди таких подходов разведка на основе открытых источников (OSINT) стала ключевой методологией для сбора разведывательной информации из общедоступных источников, включая реестры доменных имён, DNS-записи, журналы прозрачности SSL-сертификатов и содержимое веб-страниц. Сегодня инструменты OSINT играют важную роль в тестировании на проникновение, моделировании атак и операциях по киберразведке [1]. Согласно последним отчётам IBM X-Force и ENISA более 60% векторов начального доступа в имитационных атаках задействуют публично доступные цифровые ресурсы, что подчеркивает критическую роль OSINT в ранних этапах моделирования атак. Несмотря на стратегическую значимость, процессы обнаружения активов с помощью OSINT сопровождаются серьёзными операционными трудностями, в частности, высоким уровнем ложноположительных результатов при попытке связать IP-адреса и домены с искомыми организациями [2].

Такие ложные срабатывания создают целый ряд рисков. С операционной точки зрения они приводят к избыточной нагрузке на систему реагирования на инциденты, неоправданным усилиям аналитиков и снижению эффективности работы центров мониторинга информационной безопасности (SOC). С точки зрения бизнеса неверно идентифицированные ресурсы могут повлечь за собой нарушения нормативных требований, ошибки в приоритизации уязвимостей и даже нарушения соглашений об уровне сервиса (SLA), особенно в регулируемых отраслях, таких как финансы, здравоохранение и критическая информационная инфраструктура.

Существующие инструменты OSINT, такие как SpiderFoot, Amass и OneForAll, ориентированы в первую очередь на широкий охват обнаружения, но не имеют встроенных механизмов фильтрации или валидации принадлежности ресурсов. В результате ответственность за проверку ложноположительных срабатываний часто ложится на аналитиков. Это снижает масштабируемость процессов и повышает риск субъективных ошибок. Настоящая работа предлагает решение этой проблемы в виде системы постобработки OSINT-результатов, основанной на методах машинного обучения, которая повышает точность атрибуции обнаруженных ресурсов. Предлагаемый подход формирует структурированные векторы признаков для IP-адресов и доменных имён на основе анализа данных WHOIS, записей DNS, SSL-сертификатов и веб-контента. Эти векторы классифицируются с использованием модели обучения с учителем для определения принадлежности ресурса к целевой организации.

Основной научный вклад исследования заключается в следующем:

- разработана специализированная система инженерии признаков для задач валидации атрибуции в OSINT-аналитике;
- проведена сравнительная оценка нескольких алгоритмов классификации, показавшая преимущество градиентного бустинга при несбалансированных выборках;
- выполнена верификация подхода на данных семи реальных организаций, подтверждающая значительный рост точности при практическом применении.

ОБЗОР ЛИТЕРАТУРЫ

Интеграция OSINT-методологий в процессы обеспечения кибербезопасности стимулировала развитие широкого спектра инструментов и фреймворков для внешнего обнаружения цифровых активов. Инструменты, такие как SpiderFoot,

Amass, SubFinder, OneForAll и BBOT (Black Box OSINT Tool), получили популярность благодаря своей модульной архитектуре, возможностям как пассивной, так и активной разведки, а также поддержке множества источников данных, включая реестры WHOIS, DNS-записи, SSL-сертификаты и общедоступную веб-разведку [3, 4].

SpiderFoot предоставляет более 200 модулей для сбора данных, что позволяет интеграцию с разнообразными внешними источниками информации. Он широко применяется в разведывательных цепочках благодаря расширяемости и наличию визуального интерфейса. Amass, поддерживаемый OWASP, специализируется на DNS-энумерации и сопоставлении автономных систем (ASN), обеспечивая масштабное обнаружение доменов и поддоменов. SubFinder и OneForAll - это API-ориентированные инструменты, оптимизированные для высокой скорости и автоматизации. BBOT дополнительно предлагает генерацию поддоменов с применением обработки естественного языка (NLP) и поддержку мультицелевого сканирования [5, 6].

Несмотря на эффективность в расширенном сборе данных, указанные инструменты не оснащены механизмами валидации принадлежности ресурсов. Их выводы, как правило, включают большие массивы доменов и IP-адресов без контекстной проверки принадлежности.

В результате специалисты по кибербезопасности вынуждены вручную фильтровать или игнорировать нерелевантные данные, что делает процесс подверженным ошибкам и субъективным суждениям.

В последние годы предпринимались попытки преодолеть это с помощью методов искусственного интеллекта. Была предложена модель контекстной ситуационной осведомлённости на основе OSINT, проведен систематический обзор и показано недостаточное использование ИИ-технологий для атрибуции цифровых активов в рамках OSINT. Другие исследования были посвящены профилированию посетителей инфраструктуры с применением OSINT, а также обзору классификационных моделей угроз на основе открытых данных [7-12]. Однако подавляющее большинство существующих решений ориентированы на поведенческую аналитику, детектирование угроз или мониторинг ситуации, но не на прямую атрибуцию цифровой инфраструктуры. Пересечение OSINT и машинного обучения по-прежнему недостаточно изучено в контексте валидации принадлежности доменов и IP-адресов конкретным организациям. Для систематизации авторами предлагается сравнительная таблица 1 ключевых инструментов OSINT с точки зрения их возможностей обнаружения, точности принадлежности и поддержки ИИ-усиления.

Большинство платформ ориентированы на максимально широкий охват и сбор разведывательной информации, при этом практически не уделяя внимания точности атрибуции.

Наше решение напрямую устраняет этот недостаток за счёт добавления структурированного слоя классификации на основе признаков, который позволяет существенно повысить точность валидации OSINT-результатов после их получения [13-15].

Таблица 1

Сравнение инструментов OSINT по задачам обнаружения и атрибуции
Table 1. Comparison of OSINT tools for detection and attribution tasks

Инструмент	Источники данных	Преимущества	Ограничения	Поддержка ИИ
SpiderFoot	WHOIS, DNS, журналы CT	Модульная структура, поддержка ~200 источников	Высокий уровень ложных срабатываний, отсутствие фильтрации	Нет
Amass	DNS, ASN, веб-архивы	Эффективная пассивная и активная DNS-разведка	Ограниченный интерфейс, отсутствие ранжирования	Нет
OneForAll	Публичные API, пассивный DNS	Высокая скорость, интеграция с открытыми данными	Нет валидации результатов, ориентирован на китайские ресурсы	Нет
BBOT	NLP, DNS, SSL, скриншоты	Расширяемость, многозадачность, генерация поддоменов	Высокие ресурсоемкость, низкая объяснимость	Частичная
Настоящее исследование	Все вышеперечисленные источники	Фильтрация на основе ML, атрибуционная классификация	Требуется размеченная обучающая выборка	Да

МЕТОДОЛОГИЯ

Предлагаемая система организована как модульный классификационный конвейер, состоящий из трёх основных этапов:

Сбор данных OSINT с использованием внешних инструментов.

Извлечение и преобразование структурированных индикаторов.

Контролируемая классификация кандидатных цифровых активов.

Каждый ресурс, IP-адрес или доменное имя, представляется в виде высокоразмерного вектора признаков, сформированного на основе метаданных WHOIS, записей DNS, информации из SSL-сертификатов и содержимого веб-страниц. Полученные векторы передаются обученной модели классификации для определения принадлежности ресурса целевой организации.

Для моделирования реалистичных сценариев тестирования на проникновение были собраны выборки с использованием SpiderFoot - модульного OSINT-фреймворка, поддерживающего более 200 источников данных. Инструмент был настроен на извлечение записей WHOIS, DNS, данных прозрачности сертификатов и HTML-контента.

В качестве целевых объектов были выбраны семь организаций, представляющих различные сектора (финансы, телекоммуникации, образование). В результате было собрано 17 235 IP-адресов и 12 344 доменных имени, из которых 6 367 были вручную верифицированы как действительно принадлежащие целевым организациям. Эта размеченная выборка использовалась как для обучения, так и для оценки моделей.

Для захвата структурных, контекстуальных и семантических индикаторов, указывающих на принадлежность ресурса организации, была реализована система инженерии признаков, формирующая 62 бинарных и категориальных признака. Эти признаки были сгруппированы в две основные категории: признаки, основанные на IP-адресах, и признаки, основанные на доменных именах. Для каждого обнаруженного IP-адреса система анализирует поля WHOIS, включая inetnum, descr, org-name и netname. На основании наличия ключевых слов, связанных с организацией, и схожести с ранее подтверждёнными записями формируются бинарные индикаторы. Дополнительно извлекаются следующие признаки:

Принадлежность IP к тем же подсетям (/24, /16), что и известные адреса организации;

Наличие и корректность обратной PTR-записи (reverse DNS);

Привязка SSL-сертификатов с известными субъектами или организациями-эмитентами.

Эти индикаторы нормализуются по всей выборке и включаются в объединённое векторное пространство признаков.

Для доменов алгоритм анализирует более широкий набор атрибутов, включая:

- DNS-записи различных типов: A, NS, MX, TXT, CNAME;

- Поля SSL-сертификатов: альтернативные имена (SAN), имя организации-эмитента;

- Индикаторы из веб-контента: хэши favicon, заголовки страниц (title), копирайт в подвале страницы.

В процессе извлечения эти признаки токенизируются и кодируются в виде бинарных флагов. Например, признаки *«совпадение favicon с проверенным доменом»* или *«CNAME-ссылка на домен организации»* считаются высоконадёжными индикаторами принадлежности. Текстовые поля нормализуются через хэширование и фильтрацию ключевых слов (табл. 2).

Таблица 2

Примеры признаков, использованных для атрибуции цифровых активов в OSINT
Table 2. Examples of features used for digital asset attribution in OSINT

Название признака	Тип	Источник	Описание
has_org_keyword_whois	Бинарный	WHOIS	Наличие ключевого слова в полях org-name, descr или netname
shared_netblock	Бинарный	WHOIS / IP	IP-адрес принадлежит той же подсети (/24 или /16), что и проверенные
ptr_record_exists	Бинарный	DNS	Обратная PTR-запись существует и указывает на известный домен
shared_favicon_hash	Бинарный	Веб-контент	Совпадение хэша favicon с известным ресурсом организации
cname_matches_verified	Бинарный	DNS	CNAME-запись совпадает с доменом, подтверждённым как принадлежащий
cert_org_matches	Бинарный	SSL-сертификат	Имя организации в сертификате совпадает с целевой

Эти признаки составляют основу процесса классификации. Дополнительные производные признаки формировались на основе порогов совместного появления и эвристик междоменных связей. Все признаки были закодированы в виде векторов фиксированной длины, включающих бинарные и категориальные элементы.

Например, если в поле WHOIS обнаружено ключевое слово организации, то присваивается значение 1, иначе 0. Если CNAME совпадает с известным доменом, то присваивается значение 1, иначе 0.

Для повышения устойчивости к различиям в форматах и структуре данных:

WHOIS-записи нормализуются, отсутствующие поля заполняются значением 0.

DNS-записи предварительно обрабатываются для устранения шумов и дубликатов.

Итоговые векторы по IP и доменам объединяются в единую матрицу признаков, которая используется как для обучения модели, так и для дальнейшей классификации новых объектов.

Для оценки пригодности различных моделей к задаче бинарной классификации были протестированы четыре алгоритма контролируемого обучения:

Дерево решений (Decision Tree) - легко интерпретируемо, но подвержено переобучению.

Логистическая регрессия (Logistic Regre-ssion) - устойчива и эффективна при линейных границах.

Метод k-ближайших соседей (k-NN) - интуитивно понятен, но чувствителен к размерности пространства.

Градиентный бустинг (Gradient Boosting) - ансамблевая модель, способная улавливать нелинейные зависимости.

Все модели были реализованы с использованием библиотеки scikit-learn и обучены на 70% размеченной выборки. Оставшиеся 30% использовались для тестирования.

Основной метрикой эффективности выступала площадь под кривой Precision-Recall (PR-AUC), наиболее релевантная для задач с дисбалансом классов.

Результаты показали, что модель градиентного бустинга стабильно превосходит альтернативы как по точности (Precision), так и по полноте (Recall), демонстрируя высокую устойчивость к неоднородности данных и шумам.

Система фильтрации была реализована на языке Python 3.11 с использованием модульной архитектуры.

В рамках проекта использовались следующие основные библиотеки и инструменты:

1) pandas, numpy - предобработка данных и операции с матрицами;

2) scikit-learn - построение и обучение моделей машинного обучения;

3) whois, dnspython, socket, ssl, tldextract - извлечение данных OSINT (WHOIS, DNS, SSL);

4) favicon, requests, BeautifulSoup - парсинг веб-контента и хэширование favicon-иконок.

Все зависимости являются кроссплатформенными и распространяются с открытым исходным кодом, что обеспечивает воспроизводимость экспериментов и переносимость проекта в различные вычислительные среды.

Алгоритм построен по модульному принципу и включает пять функциональных компонентов, каждый из которых отвечает за отдельный этап обработки данных:

- source/: исходные данные в формате XLSX, содержащие результаты OSINT-сканирования (например, из SpiderFoot);
- mapping/: извлечённые метаданные из WHOIS, DNS и веб-источников;
- vectors/: числовые векторы признаков, сформированные для IP-адресов и доменов;
- ml/: скрипты обучения моделей и сериализованные (сохранённые) классификаторы;
- main.py: управляющий модуль, координирующий сбор данных, генерацию признаков и классификацию.

Модули слабо связаны между собой и могут быть протестированы независимо, что облегчает расширение системы, внедрение в CI-процессы и интеграцию в другие инструменты. Общая структура компонентов представлена на рис. 1, в котором отображены связи между модулями сбора признаков, их преобразования и финальной классификации. Каждый модуль работает в составе общей конвейерной схемы классификации. Такая архитектура упрощает масштабирование и поддержку, а также позволяет безболезненно добавлять новые источники данных, мультиязычные домены или механизмы объяснимого ИИ (XAI) для повышения интерпретируемости модели. При запуске на стандартной рабочей станции (процессор Intel Core i7, 16 ГБ ОЗУ, SSD-накопитель) полная обработка OSINT-выборки объёмом около 30 000 ресурсов занимает менее 10 минут. Извлечение признаков и построение векторов реализованы с возможностью пакетной (batch) параллелизации, что обеспечивает высокую масштабируемость решения.

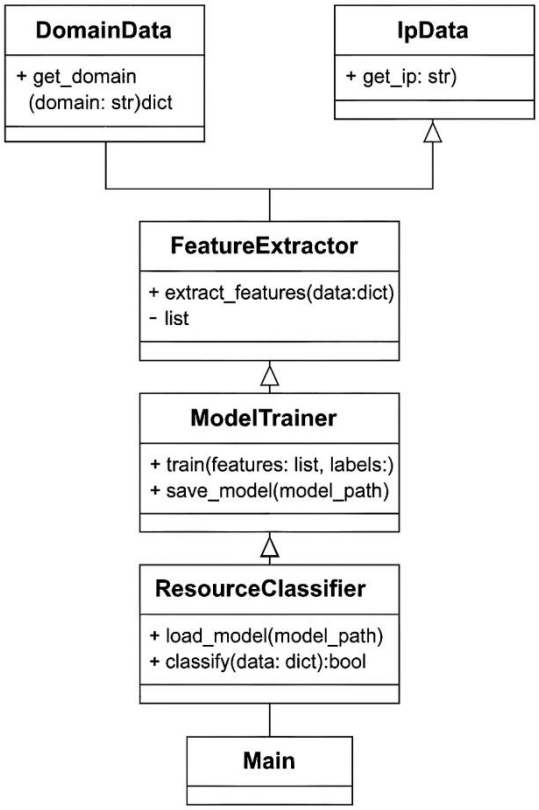

Рис. 1. UML-диаграмма архитектуры системы фильтрации OSINT-атрибуции
Fig. 1. UML diagram of OSINT attribution filtering system architecture

Для обеспечения воспроизводимости исследований весь исходный код, включая конфигурационные файлы, ноутбуки Jupyter и обученные модели, упакован в готовый к публикации архив.

Предлагаемый алгоритм может быть легко интегрирован в существующие инструменты разведки и мониторинга безопасности. Возможные сценарии внедрения включают:

- использование в качестве **предобработчика** для инструментов автоматизированной OSINT-разведки (например, SpiderFoot или Amass);
- встраивание в **платформы центров мониторинга безопасности (SOC)** для предварительной фильтрации подозрительных активов;
- применение в **процессах первичного аудита инфраструктуры** в рамках услуг MSSP (поставщиков управляемой информационной безопасности).

Результаты классификации и векторы признаков могут экспортироваться в табличных форматах (.xlsx, .csv) либо передаваться через REST API (в будущих версиях). Также планируется реализация лёгкого графического интерфейса для поддержки интерактивных расследований, выполняемых специалистами по кибербезопасности.

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Для оценки вкладов отдельных признаков в общее качество классификации была проведена оценка важности отдельных признаков. В ходе эксперимента группы взаимосвязанных признаков поочерёдно исключались из входного вектора, после чего измерялось снижение качества модели по метрикам PR-AUC и F1-оценке. Это позволило количественно определить важность каждой категории признаков - WHOIS, DNS, SSL и веб-контента. Результаты показали, что **атрибуты SSL-сертификатов**, особенно название организации и альтернативные имена (Subject Alternative Names, SANs), являются одними из наиболее значимых признаков. Во многих инфраструктурах компании используют унифицированные схемы именования и сертификации, что даёт устойчивые сигналы принадлежности. Включение этих признаков позволило повысить PR-AUC до 11% по сравнению с базовой моделью. Также высокую предсказательную силу продемонстрировали **поля WHOIS**, в частности descr, org-name и mnt-by, содержащие наименования организаций и операторов сетей.

Эти поля, как правило, более устойчивы и консервативны по сравнению с динамически из-

меняемыми DNS-конфигурациями или временными IP-адресами. Анализ важности признаков с использованием SHAP-показателей подтвердил, что org-name и descr входят в топ-5 по влиянию на результат классификации.

Веб-контентные признаки, такие как хэши favicon и заголовки HTML-страниц, показали высокую дискриминативную способность в тех случаях, когда компания использует стандартизированные шаблоны и визуальную айдентику. Хэши favicon выступили надёжными бинарными идентификаторами, позволяя объединять связанные ресурсы даже без сетевых признаков.

Напротив, **DNS-признаки** (MX, CNAME, NS) продемонстрировали умеренную полезность и склонность к ложноположительным совпадениям. Это объясняется тем, что такие элементы часто являются общими для множества организаций из-за использования внешних сервисов - облачных платформ, почтовых шлюзов (Google Workspace, Microsoft 365), CDN и др. (табл. 3, рис. 2).

Полученные результаты позволяют сделать вывод, что признаки, отражающие **владение ресурсами (WHOIS, SSL)** и **визуальную идентичность (favicon, title)**, обладают большей устойчивостью и надёжностью для атрибуции, чем маршрутизируемые сетевые артефакты, подверженные изменениям.

Для выбора наилучшей модели, обеспечивающей фильтрацию ложноположительных OSINT-результатов, были протестированы четыре алгоритма машинного обучения: градиентный бустинг, логистическая регрессия, метод k-ближайших соседей (k-NN) и дерево решений.

Оценка проводилась по четырём ключевым метрикам: **точность (Precision)**, **полнота (Recall)**, **F1-мера** и **PR-AUC** - последняя особенно важна для несбалансированных выборок (табл. 4).

Таблица 3

Вклад групп признаков в качество модели (ΔPR-AUC)
Table 3. Contribution of feature groups to model quality (ΔPR-AUC)

Группа признаков	Снижение PR-AUC при удалении	Относительная значимость
SSL-сертификаты	−0,11	Очень высокая
WHOIS-метаданные	−0,09	Высокая
Веб-контент (HTML)	−0,07	Средняя
DNS-записи	−0,04	Низкая

Таблица 4

Метрики оценки качества классификаторов
Table 4. Metrics for assessing the quality of classifiers

Модель	Точность (Precision)	Полнота (Recall)	F1-мера	PR-AUC
Градиентный бустинг	0,912	0,996	0,952	0,94
Логистическая регрессия	0,822	0,984	0,896	0,77
Метод k-ближайших соседей (k-NN)	0,722	0,957	0,823	0,71
Дерево решений	0,670	0,920	0,775	0,71

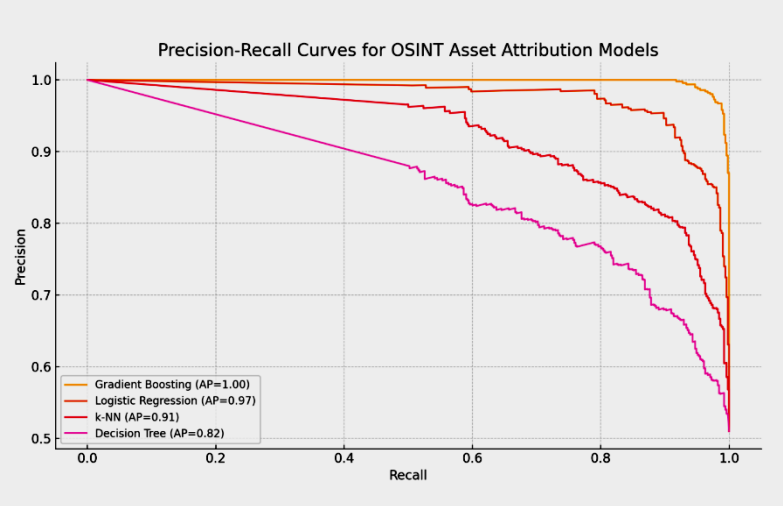

Рис. 2. PR-кривые для всех протестированных классификаторов
Fig. 2. PR curves for all tested classifiers

Модель градиентного бустинга демонстрирует заметное превосходство по всем уровням полноты, особенно в зонах высокой чувствительности (recall), что свидетельствует о её устойчивости к дисбалансу классов.

Графики подтверждают численные результаты. Градиентный бустинг обеспечивает как высокую точность, так и полноту, практически не допуская потерь в чувствительности при росте recall. Логистическая регрессия и k-NN показывают разумные показатели, но теряют точность на сложных объектах. Дерево решений демонстрирует наименьшую устойчивость и склонно к переобучению.

Хотя агрегированные метрики и PR-кривые дают общее представление о качестве модели, **матрицы ошибок (confusion matrix)** позволяют глубже понять, какие типы ошибок совершаются классификаторами (рис. 3).

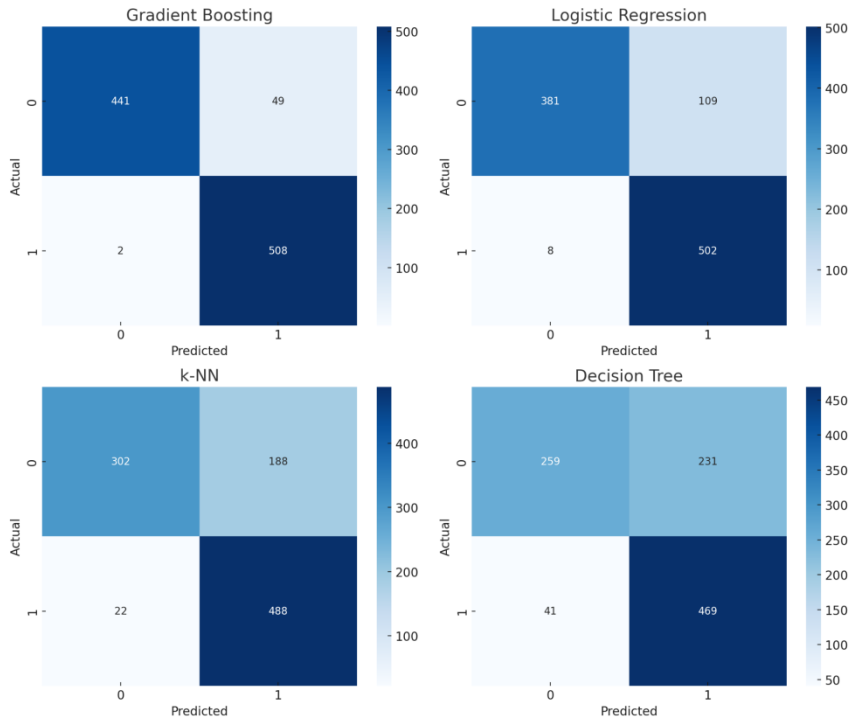

Рис. 3. Матрицы ошибок для каждой из протестированных моделей
Fig. 3. Error matrices for each of the tested models

Модель градиентного бустинга показывает диагональную доминанту, что указывает на высокую точность и сбалансированность предсказаний. Градиентный бустинг обеспечивает наилучший баланс между полнотой и точностью. Модель допускает минимальное количество ложных срабатываний и пропусков. Это делает её особенно надёжной в задачах, где ошибочная атрибуция может привести к операционным или регуляторным рискам. Логистическая регрессия показала схожие характеристики, но с немного более высоким уровнем ложноположительных результатов. Это ожидаемое поведение для моделей, стремящихся максимизировать recall в ущерб точности. Модели k-NN и дерево решений показывают менее устойчивые результаты. Особенно деревья решений склонны к ошибочной классификации негативных примеров как позитивных, что может привести к включению чужих ресурсов в периметр организации - крайне нежелательная ситуация при аудитах безопасности. Ложноположительные ошибки в большинстве случаев связаны с инфраструктурой общего пользования, а именно: облачные IP-диапазоны (например, AWS), общие записи CNAME от CDN и др. Ложноотрицательные ошибки, как правило, обусловлены приватностью WHOIS-записей, регистрацией через прокси или нетипичной структурой DNS. Такие наблюдения подтверждают ценность мультимодальных признаков, например, комбинация SSL-сертификатов, favicon и текстов WHOIS даёт более устойчивую картину принадлежности. Высокая точность модели градиентного бустинга позволяет сократить нагрузку на аналитиков в SOC и повысить производительность в сценариях, где требуется обработка больших объёмов OSINT-данных.

Для оценки способности модели к обобщению и её применимости в реальных условиях мы развернули предложенную систему на крупной выборке примерно из 30 000 цифровых ресурсов. Данные были собраны с использованием инструмента SpiderFoot с расширенными модулями в контексте одного из российских финансовых учреждений.

Результаты классификации следующие:

- **IP-адреса:** 112 ресурсов классифицированы как принадлежащие организации; 99 - подтверждено вручную, 13 - ложноположительные, 0 - ложноотрицательные.

- **Домены:** 365 ресурсов классифицированы как принадлежащие; 357 - подтверждено вручную, 8 - ложноположительные, 0 - ложноотрицательные.

Если рассматривать SpiderFoot, то исходный результат содержал более 85% ложноположительных объектов. Таким образом, предлагаемый алгоритм позволяет снизить уровень ложных срабатываний более чем в 10 раз.

Это значительно сокращает трудозатраты на ручную валидацию. Предлагаемую модель возможно использовать как предфильтр в OSINT-инструментах, уменьшающий шум и избыточные результаты до стадии анализа, модуль постобработки в платформах центров мониторинга (SOC) для уточнения угроз или часть процедур первичного аудита и инвентаризации ИТ-периметра в рамках MSSP-услуг.

ДИСКУССИИ И ОГРАНИЧЕНИЯ

Несмотря на высокую эффективность предлагаемый алгоритм имеет ряд ограничений, которые следует учитывать при её практическом применении и дальнейшем развитии.

WHOIS-записи значительно различаются по полноте и структуре в зависимости от регистратора и юрисдикции. Использование сервисов приватной регистрации и маскировка персональных данных (например, в соответствии с требованиями GDPR) могут затруднить извлечение надёжных атрибуционных признаков.

Широкое применение автоматических удостоверяющих центров (например, Let’s Encrypt) и повторное использование wildcard-сертификатов снижает уникальность информации в SSL и усложняет различение между связанными и нерелевантными ресурсами. Облачные платформы и CDN (сети доставки контента) порождают быстро меняющиеся конфигурации. Автоматическое масштабирование, балансировка нагрузки и временная аренда IP-диапазонов могут приводить к тому, что анализ «снимка состояния» даёт устаревшие или неточные результаты. Для надёжной атрибуции требуется учитывать временной контекст.

Злоумышленники могут использовать техники обфускации, такие как доменное фронтирование, переадресация по цепочкам и аффилированные домены. Такие приёмы затрудняют выявление связей между ресурсами и могут нарушать как ручную, так и автоматическую атрибуцию.

Модель требует наличия обучающей выборки с метками принадлежности, поддержание которой требует усилий. Изменения в инфраструктуре организаций, слияния и ребрендинг могут приводить к устареванию признаков и снижению точности. Это подчеркивает необходимость в разработке полуавтоматических методов обучения, активного обучения или верификации «в человеке». Текущая модель была протестирована в основном на организациях из российской инфраструктуры. Не исключено, что производительность может снижаться при работе с мультиалфавитными доменами, локализованными WHOIS-записями или инфраструктурой из других юрисдикций. Будущие исследования должны охватывать разные языковые и геополитические контексты.

ВЫВОДЫ

В настоящем исследовании решается одна из ключевых проблем OSINT-анализов в области кибербезопасности – это высокая доля ложноположительных результатов при атрибуции IP-адресов и доменных имён. Путём интеграции методов машинного обучения с систематизированным извлечением признаков из открытых источников была разработана модульная система, способная значительно повысить точность автоматизированной идентификации цифровых активов [16, 17]. Предлагаемый подход формирует богатые по структуре векторы признаков на основе метаданных WHOIS, DNS-записей, содержимого SSL-сертификатов и веб-страниц. Экспериментальная проверка на выборке из семи организаций и более 30 000 ресурсов показала, что модель на основе градиентного бустинга достигла PR-AUC 0.94, превзойдя традиционные классификаторы по всем основным метрикам. В условиях реального тестирования система снизила уровень ложноположительных результатов с более чем 85% до менее чем 5%, при этом не допустив ни одного ложного отрицания. Эти результаты подтверждают высокую практическую применимость подхода в задачах тестирования на проникновение, киберразведки и обеспечения соответствия нормативным требованиям.

Несмотря на достигнутые результаты, остаётся ряд перспективных направлений, таких как поддержка международных доменных имён (IDN) и мультиалфавитных структур, интеграция с источниками оперативной угрозной разведки (threat intelligence feeds), включение временных параметров для учёта эволюции цифровой инфраструктуры, разработка модулей объяснимого ИИ (XAI) для интерпретируемости результатов и реализация REST API и визуального интерфейса для работы аналитиков.

Таким образом, предложенный алгоритм закрывает практический разрыв между пассивной разведкой и высокоточными аналитическими инструментами, обеспечивая масштабируемую и воспроизводимую фильтрацию OSINT-данных.

Этот фундамент может быть использован для развития интеллектуальных платформ атрибуции в задачах защиты цифрового периметра, мониторинга соответствия и анализа киберрисков.

Авторы заявляют об отсутствии конфликта интересов, требующего раскрытия в данной статье.

The authors declare the absence a conflict of interest warranting disclosure in this article.

ЛИТЕРАТУРА

1. **Pastor-Galindo J., Nespoli P., Gómez Mármol F., Martínez Pérez G.** The not yet exploited goldmine of OSINT: opportunities, open challenges and future trends. *IEEE Access.* 2022. V. 10. P. 10282–10304. DOI: 10.1109/ACC ESS. 2020.2965257.
2. **Evangelista J., Sassi R., Romero M., Napolitano D.M.** Systematic literature review to investigate the application of open source intelligence (OSINT) with artificial intelligence. *Journal of Applied Security Research.* 2020. V. 15, N 1. P. 1–23. DOI: 10.1080/19361610.2020.1761737.
3. **Sasaki T., Yoshioka K., Matsumoto T.** Who are you? OSINT-based profiling of infrastructure honeypot visitors. Proc. of the 2023 Int. Symposium on Digital Forensics and Security (ISDFS). 2023. P. 1–6. DOI: 10.1109/ISDFS 581 41.2023.10131856.
4. **Hwang Y.-W., Lee I.-Y., Kim H., Lee H., Kim D.** Current status and security trend of OSINT. Wireless Communications and Mobile Computing. 2022. Article ID 1290129. 14 p. DOI: 10.1155/2022/1290129.
5. **Best C.** OSINT, the Internet and Privacy. European Intelligence and Security Informatics Conference. 2012. P. 4. DOI: 10.1109/EISIC.2012.71.
6. **Lee S., Shon T.** Open source intelligence-based cyber threat inspection framework for critical infrastructures. Future Technologies Conference (FTC). 2016. P. 1030–1033. DOI: 10.1109/FTC.2016.7821730.
7. **Jain M., Jain G., Vasavada J., Patel P., Ojha V.** Information security and auditing for distributed network. European Intelligence and Security Informatics Conference. 2012. P. 364–367. DOI: 10.1109/MSNA.2012.6324595.
8. **Bryushinin A.O., Dushkin A.V. and Melshiyan M.A.** Automation of the Information Collection Process by Osint Methods for Penetration Testing During Information Security Audit. 2022 Conference of Russian Young Researchers in Electrical and Electronic Engineering (ElConRus), Saint Petersburg, Russian Federation. 2022. P. 242-246. DOI: 10.1109/ElConRus54750.2022.9755812.
9. **Astrakhantsev R., Chuhno A., Dmukh A. [et al.].** Differences with high probability and impossible differentials for the KB-256 cipher. Journal of Computer Virology and Hacking Techniques. 2024. V. 20, N 3. P. 525-531. DOI 10.1007/s11416-024-00532-2. EDN IQFPUV.
10. **Самсонова А.И., Суворов И.А., Астраханцев Р.Г.** Интеллектуальная система анализа пользовательской активности в защищенных операционных средах. *Известия высших учебных заведений. Серия: Экономика, финансы и управление производством [Ивэкофин].* 2025. № 1(63). С. 79-85. DOI 10.6060/ivecofin.2025631.715. EDN ZCDOVL.

REFERENECES

1. **Pastor-Galindo J., Nespoli P., Gómez Mármol F., Martínez Pérez G.** The not yet exploited goldmine of OSINT: opportunities, open challenges and future trends. *IEEE Access.* 2022. V. 10. P. 10282–10304. DOI: 10.1109/ACCESS.2020.2965257.
2. **Evangelista J., Sassi R., Romero M., Napolitano D.M.** Systematic literature review to investigate the application of open source intelligence (OSINT) with artificial intelligence. *Journal of Applied Security Research.* 2020. V. 15, N 1. P. 1–23. DOI: 10.1080/19361610.2020.1761737.
3. **Sasaki T., Yoshioka K., Matsumoto T.** Who are you? OSINT-based profiling of infrastructure honeypot visitors. Proc. of the 2023 Int. Symposium on Digital Forensics and Security (ISDFS). 2023. P. 1–6. DOI: 10.1109/ISDFS58141. 2023.10131856.
4. **Hwang Y.-W., Lee I.-Y., Kim H., Lee H., Kim D.** Current status and security trend of OSINT. Wireless Communications and Mobile Computing. 2022. Article ID 1290129. 14 p. DOI: 10.1155/2022/1290129.
5. **Best C.** OSINT, the Internet and Privacy. European Intelligence and Security Informatics Conference. 2012. P. 4. DOI: 10.1109/EISIC.2012.71.
6. **Lee S., Shon T.** Open source intelligence-based cyber threat inspection framework for critical infrastructures. Future Technologies Conference (FTC). 2016. P. 1030–1033. DOI: 10.1109/FTC.2016.7821730.
7. **Jain M., Jain G., Vasavada J., Patel P., Ojha V.** Information security and auditing for distributed network. European Intelligence and Security Informatics Conference. 2012. P. 364–367. DOI: 10.1109/MSNA.2012.6324595.
8. 8. **Bryushinin A.O., Dushkin A.V. and Melshiyan M.A.** Automation of the Information Collection Process by Osint Methods for Penetration Testing During Information Security Audit. 2022 Conference of Russian Young Researchers in Electrical and Electronic Engineering (ElConRus), Saint Petersburg, Russian Federation. 2022. P. 242-246. DOI: 10.1109/ElConRus54750.2022.9755812.
9. **Astrakhantsev R., Chuhno A., Dmukh A. [et al.].** Differences with high probability and impossible differentials for the KB-256 cipher. Journal of Computer Virology and Hacking Techniques. 2024. V. 20, N 3. P. 525-531. DOI 10.1007/s11416-024-00532-2. EDN IQFPUV.
10. **Samsonova A.I., Suvorov I.A., Astrakhancev R.G.** Intelligent system for analyzing user activity in secure operating environments. *Ivecofin*. 2025. N 1(63). P. 79-85. DOI 10.6060/ivecofin.2025631.715. EDN ZCDOVL.
11. **Astrakhantseva I.A., Astrakhantsev R.G., Usoltsev S.D. [et al.].** K-means analysis of spectral properties of BODIPY dye during compression on air - water interface: construction of dataset for effective clustering. *Liquid Crystals and their Ap-*

11. **Astrakhantseva I.A., Astrakhantsev R.G., Usoltsev S.D. [et al.].** K-means analysis of spectral properties of BODIPY dye during compression on air - water interface: construction of dataset for effective clustering. *Liquid Crystals and their Application.* 2024. V. 24, N 2. P. 43-53. DOI 10.18083 LCAppl.2024.2.43. EDN CXAIQV.
12. **Астраханцева И.А., Горев С.В., Астраханцев Р.Г.** Фрактальный анализ в оценке эффективности и надежности сложных технических систем. *Современные наукоемкие технологии. Региональное приложение.* 2023. № 4(76). С. 60-68. DOI 10.6060/snt.20237604.0008. EDN NBDYHR.
13. **Зимнуров М.Ф., Астраханцева И.А.** Методология создания многосвязных структур данных с применением LLM в рабочих проектах. *Современные наукоемкие технологии. Региональное приложение.* 2025. № 1(81). С. 76-83. DOI 10.6060/snt.20258101.0009. EDN STBGGB.
14. **Бобков С.П., Астраханцева И.А., Гущин А.А. Бобкова Е.С., Астраханцев Р.Г., Шутов Д.А.** Использование дискретного вероятностного подхода для моделирования проточных трубчатых реакторов. *Изв. вузов. Химия и хим. технология.* 2025. Т. 68. № 2. С. 96-101. DOI 10.6060/ivkkt.20256802.7109.
15. **Бобков С.П., Астраханцев Р.Г., Самарский А.А., Павлова Е.А.** Имитационная модель движения вещества в технологическом аппарате. *Известия высших учебных заведений. Серия: Экономика, финансы и управление производством [Ивэкофин]*. 2024. № 2(60). С. 32-40. DOI 10.6060/ivecofin.2024602.681. EDN QOQPLV.
16. **Telegin F.Y., Karpova V.S., Makshanova A.O. [et al.].** Solvatochromic Sensitivity of BODIPY Probes: A New Tool for Selecting Fluorophores and Polarity Mapping. *International Journal of Molecular Sciences.* 2023. Vol. 24, N 2. P. 1217. DOI 10.3390/ijms24021217. EDN ZDSWJC.
17. **Астраханцева И.А., Котенев Т.Е., Горев С.В.** Интеграция методов линейной регрессии и случайного леса в системы интеллектуальной поддержки управления криогенными транспортными системами. *Известия высших учебных заведений. Серия: Экономика, финансы и управление производством [Ивэкофин]*. 2024. № 3(61). С. 81-90. DOI 10.6060/ivecofin.2024613.692.

plication. 2024. V. 24, N 2. P. 43-53. DOI 10.18083/ LCAppl .2024. 2.43. EDN CXAIQV.
12. **Astrakhantseva I.A., Gorev S.V., Astrakhantsev R.G.** Fractal analysis in assessing the efficiency and reliability of complex technical systems. *Modern high technology. Regional application.* 2023. N 4(76). P. 60-68. DOI 10.6060/snt. 2023 7604. 0008. EDN NBDYHR.
13. **Zimnurov M.F., Astrakhantseva I.A.** Methodology for creating multi-linked data structures using LLM in working projects. *Modern high technology. Regional application.* 2025. N 1(81). P. 76-83. DOI 10.6060/snt.20258101.0009. EDN STBGGB.
14. **Bobkov S.P., Astrakhantseva I.A., Gushchin A.A. Bobkova E.S., Astrakhantsev R.G., Shutov D.A.** Using a discrete probabilistic approach for modeling flow tubular reactors. *ChemChemTech [Izv. Vyssh. Uchebn. Zaved. Khim. Khim. Tekhnol.]*. 2025. V. 68. N 2. P. 96-101. DOI 10.6060/ivkkt.20256802.7109. EDN OWCXJD.
15. **Bobkov S.P., Astrakhancev R.G., Samarskiy A.A., Pavlova E.A.** Simulation model of substance movement in a technological apparatus. *Ivecofin.* 2024. N 2(60). P. 32-40. DOI 10.6060/ivecofin.2024602.681. EDN QOQPLV.
16. **Telegin F.Y., Karpova V.S., Makshanova A.O. [et al.].** Solvatochromic Sensitivity of BODIPY Probes: A New Tool for Selecting Fluorophores and Polarity Mapping. *International Journal of Molecular Sciences.* 2023. V. 24, N 2. P. 1217. DOI 10.3390/ijms24021217. EDN ZDSWJC.
17. **Astrakhantseva I.A., Kotenev T.E., Gorev S.V.** Integration of linear regression and random forest methods into intelligent control support systems for cryogenic transport systems. *Ivecofin.* 2024. N 3(61). P. 81-90. DOI 10.6060/ivecofin.2024613.692.

Поступила в редакцию(Received) 10.05.2025
Принята к опубликованию (Accepted) 05.07.2025