Инженерный вестник Дона, №2 (2026)
ivdon.ru/ru/magazine/archive/n1y2026/10748

Сравнительный анализ классических алгоритмов машинного обучения для детекции фишинговых ссылок

Д.С. Джурук

Восточно-Сибирский институт МВД России, Иркутск

Аннотация: В статье представлены результаты сравнительного эксперимента по оценке классических алгоритмов машинного обучения для детекции фишинговых ссылок. На исследуемом датасете алгоритм случайного леса показал максимальную эффективность. Ключевой вывод: в некоторых случаях классические методы могут быть эффективной альтернативой, потенциально предлагая преимущества с точки зрения объяснимости решений и вычислительной эффективности на этапе эксплуатации.
Ключевые слова: фишинг, кибербезопасность, информационная безопасность, машинное обучение, случайный лес, обнаружение фишинговых атак.

Введение

Фишинг остается одной из наиболее распространенных и экономически эффективных киберугроз, наносящей значительный материальный и репутационный ущерб [1,2]. Актуальность проблемы подчеркивается статистикой: вероятность успеха атаки при массовой рассылке достигает 90%, а её рост, по данным отраслевых отчётов, составляет около 18% в год. Эволюция атак, включая распространение модели «фишинг как услуга», требует перехода от традиционных реактивных методов (чёрные списки, статические правила) к системам, способным адаптироваться к новым угрозам [3]. Именно этим обусловлен активный интерес исследователей к методам искусственного интеллекта и машинного обучения [4].

В современных работах наблюдается разнородность подходов. Значительная часть исследований концентрируется на разработке сложных моделей глубокого обучения, таких как многослойные перцептроны или рекуррентные сети. Однако эти модели критикуют за свойство «чёрного

ящика», существенные вычислительные затраты и сильную зависимость от объёма размеченных данных, что ограничивает их практическое внедрение в реальных системах безопасности, где важны как эффективность, так и объяснимость решений. Параллельно существует направление, применяющее классические, интерпретируемые алгоритмы машинного обучения, такие как случайный лес или градиентный бустинг. Несмотря на убедительные результаты, сохраняется дефицит комплексных сравнительных исследований, которые бы системно оценивали модели не только по итоговой метрике точности, но и по критически важным для практики критериям: вычислительной эффективности (время обучения и предсказания) и способности предоставить аналитику практическую интерпретацию через анализ важности признаков. Существующий пробел усугубляется тем, что ряд публикаций не предоставляет полного набора метрик для корректного сравнения [5].

Целью данного исследования является проведение сравнительного анализа эффективности и практической применимости классических интерпретируемых алгоритмов машинного обучения для задачи классификации фишинговых ссылок. Научная новизна работы заключается в сознательном смещении исследовательского фокуса с безоговорочной максимизации точности на анализ практического компромисса между точностью, скоростью работы и объяснимостью модели. Проведённое исследование предоставляет не только количественные результаты сравнения, но и формирует структурированные рекомендации по выбору алгоритма в зависимости от целевого сценария: от быстрого прототипирования и образовательных целей до развертывания в высоконагруженных производственных системах. Данный подход

способствует преодолению разрыва между академическими исследованиями и практическими нуждами индустрии кибербезопасности.

Обзор современных методов детектирования фишинга

Эволюция фишинговых атак закономерно стимулирует развитие разнообразных методов их обнаружения, которые можно систематизировать по объекту анализа и применяемому алгоритмическому аппарату. В рамках настоящего обзора представляется целесообразным классифицировать существующие подходы на три широкие категории: методы, основанные на анализе ссылок и структуры веб-страницы; методы, фокусирующиеся на семантическом анализе текстового контента; а также традиционные методы не основанные на машинном обучении и гибридные решения.

Наиболее распространенным и технически осуществимым направлением является извлечение и анализ признаков из URL-адреса и сопутствующей ему информации. Этот подход базируется на гипотезе о существовании статистически значимых различий в структурных, сетевых и поведенческих характеристиках легитимных и фишинговых ресурсов.

Исследование Лукмановой К.А. и Картака В.М. с использованием многослойного персептрона на 30 признаках показало точность 88% на тестовой выборке. Однако, используемая авторами модель, демонстрируя хорошую эффективность, функционирует как «чёрный ящик», предоставляя минимальные возможности для интерпретации того, какие именно признаки оказались решающими в конкретном случае классификации [6]. Работа Лапиной М.А. и других демонстрирует высокую эффективность ансамблевых методов: метод градиентного бустинга для деревьев решений достиг 97.34% точности, случайный лес — 96.60%. Однако сфокусировавшись на максимизации точности, была оставлена без внимания интерпретируемость и производительность моделей [7]. В исследовании

Беловой Е.И. и Хамитова Р.М. применена простая нейросеть с двумя признаками, но отсутствуют количественные метрики, что затрудняет сравнение и несколько ограничивает возможности для глубокого анализа полученных результатов [5].

Второе крупное направление связано с анализом текстовой составляющей фишинговых атак — содержимого электронных писем или текста на веб-странице. Данные методы особенно актуальны для противодействия целевым фишинговым кампаниям, где социальная инженерия играет ключевую роль. Исследование Болдырихина Н.В. и Ядреца Э.А. посвящено сравнению различных архитектур рекуррентных нейронных сетей для детектирования фишинговых писем на русском языке [8]. Однако авторы справедливо отмечают ключевое ограничение — малый объем обучающей выборки, что ставит под вопрос способность модели к обобщению на широком массиве данных. Кроме того, подобные модели требуют сложной предобработки текста (токенизация, лемматизация, векторное представление) и значительных вычислительных ресурсов, что может быть избыточным для задач, где достаточно анализа лишь URL-адреса, содержащегося в письме. Другие исследования возможностей алгоритмов глубокого обучения для защиты от фишинговых атак также показывают перспективность данного направления, но отмечают проблемы с интерпретируемостью и вычислительной сложностью [9].

Параллельно с развитием методов машинного обучения продолжают существовать и эволюционировать традиционные, детерминированные подходы. Их основу составляют сигнатурные методы (чёрные списки), эвристические правила и статический анализ. Работа Афанасьевой Н.С., Елизарова Д.А. и Мызниковой Т.А. является примером такого подхода [3]. Главное достоинство такого решения — прозрачность и полная

контролируемость процесса принятия решений, а также высокая скорость работы. Тем не менее, принципиальным недостатком является неспособность к обнаружению неизвестных, новых фишинговых ресурсов, отсутствующих в чёрных списках, и уязвимость к обходу заранее заданных эвристик [10].

Таким образом можно сделать вывод, что задача детектирования фишинга успешно решается как с помощью сложных нейросетевых моделей, так и посредством классических алгоритмов машинного обучения. Кроме того, существует выраженный перекос в сторону оптимизации единственной метрики — общей точности модели, в ущерб анализу других важных параметров. Также, прослеживается дефицит исследований, которые бы проводили комплексное сравнение моделей, включающее не только метрики качества, но и метрики производительности, а также давали бы глубокую интерпретацию работы модели через анализ важности признаков.

Методология эксперимента

В качестве основы исследования использовался публичный датасет «Phishing Websites Dataset». Его выбор обусловлен репрезентативным объёмом (88 647 экземпляров), структурированным признаковым пространством (исходно 118 признаков) и открытой доступностью, что соответствует подходам, используемым в других исследованиях [11]. Признаки включали синтаксические характеристики URL (подсчет специфических символов в различных компонентах адреса), сетевые параметры (наличие и количество IP-адресов, на которые разрешается домен, количество серверов системы доменных имен и что особенно важно, возраст домена) и бинарные индикаторы (наличие сертификата SSL (англ. Secure Sockets Layer — уровень защищённых сокетов), индексация в поисковых системах, использование URL-сервисов для сокращения ссылок, а также является ли домен IP-адресом). Распределение классов оказалось

несбалансированным: 65.4% легитимных и 34.6% фишинговых URL, что было учтено при оценке моделей.

Предобработка данных состояла из нескольких шагов. Сначала были удалены 13 постоянных (константных) признаков, не вносивших дискриминативной информации. Пропущенные значения, закодированные как -1, были сохранены как валидный числовой индикатор отсутствия компонента URL (например, отсутствия параметров запроса). Для устранения влияния разного масштаба признаков применялось стандартное масштабирование с расчетом параметров только на обучающей выборке во избежание утечки данных. Итоговый набор данных был разделен в соотношении 70/30 на обучающую (62 052 записи) и тестовую (26 595 записей) выборки с фиксированным начальным значением для обеспечения воспроизводимости.

Для сравнительного анализа были выбраны три классических, но методологически различных алгоритма, представляющих спектр подходов к классификации: логистическая регрессия, решающее дерево, случайный лес. Для всесторонней оценки моделей использовалась система метрик, охватывающая как классические критерии качества бинарной классификации, так и прагматические метрики производительности.

Результаты эксперимента

Первичная оценка эффективности алгоритмов проводилась на основе четырёх ключевых метрик: общей точности, точности предсказания фишинга, полноты и сбалансированной F1-меры. Сводные результаты для моделей логистической регрессии, решающего дерева и случайного леса представлены в Таблице 1.

Таблица № 1

Сводные метрики качества классификации на тестовой выборке

Модель	Общая точность	Точность предсказания	полнота	Сбалансированная F1-мера
Логистическая регрессия	0.9236	0.8496	0.9464	0.8954
Решающее дерево	0.9287	0.8515	0.9616	0.9032
Случайный лес (Random Forest)	0.9692	0.9528	0.9583	0.9556

Анализ данных таблицы позволяет сделать ряд принципиальных выводов. Случайный лес продемонстрировал безусловное лидерство по трём из четырёх метрик. Высочайшая общая точность свидетельствует о минимальной совокупной ошибке модели. Особенно значимо превосходство по показателю точности предсказания, которое на 10 процентных пунктов выше, чем у двух других моделей, это критически важно для минимизации ложных блокировок легитимных ресурсов в реальной системе.

Наиболее интересное наблюдение связано с метрикой полноты, превышающие 94%. Это указывает на то, что все три модели, особенно древовидные, обладают выдающейся способностью выявлять именно фишинговые атаки, минимизируя опасные ложно-отрицательные срабатывания. Логистическая регрессия, показала вполне конкурентоспособный результат. Это подтверждает, что даже простейшая линейная модель способна выявить существенную линейно разделимую составляющую в признаковом пространстве фишинговых URL.

Для более глубокого понимания природы ошибок каждой модели и их потенциального операционного воздействия был проведён анализ матриц ошибок. На рис. 1 приведены данные для тестовой выборки объёмом 26 595 экземпляров (17 401 легитимных и 9 194 фишинговых ссылок).

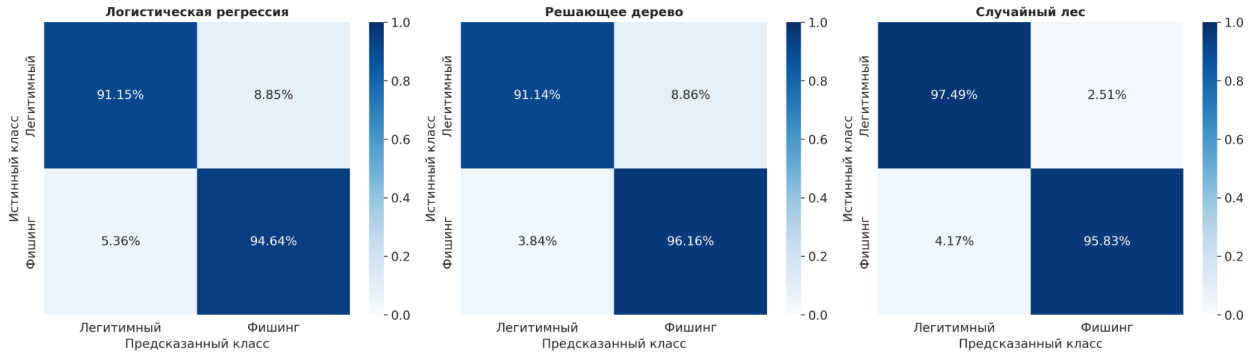

Рис. 1. - Матрицы ошибок для тестируемых моделей

Ложно-отрицательные ошибки наиболее критичны. Решающее дерево показало наименьшее их число (3,84%). Случайный лес был близок (4,17%), тогда как логистическая регрессия пропустила существенно больше атак (5,36%). Ложно-положительные срабатывания создают операционные издержки. Случайный лес здесь лидирует, допустив лишь 2,51% ошибок, что в 3,5 раза меньше, чем у других моделей. Это радикальное сокращение ложных тревог при сохранении чувствительности к угрозам делает его оптимальным для промышленных систем, снижая нагрузку на аналитиков [12].

Анализ показывает, что решающее дерево, стремясь максимизировать обнаружение угроз, «жертвует» точностью, генерируя множество предупреждений. Случайный лес, достигает более надёжного и консервативного решения: он фильтрует большую часть «шумных» срабатываний, характерных для отдельных деревьев, сохраняя при этом высокую чувствительность к реальным угрозам. Это делает его предпочтительным выбором для реальных систем, где важно снижение нагрузки на аналитиков, занимающихся расследованием инцидентов.

Прагматическая ценность модели определяется не только её точностью, но и вычислительной эффективностью. В контексте кибербезопасности, особенно для систем проверки URL в реальном времени, время отклика

является критическим параметром. На рис. 2 представлены результаты замера времени.

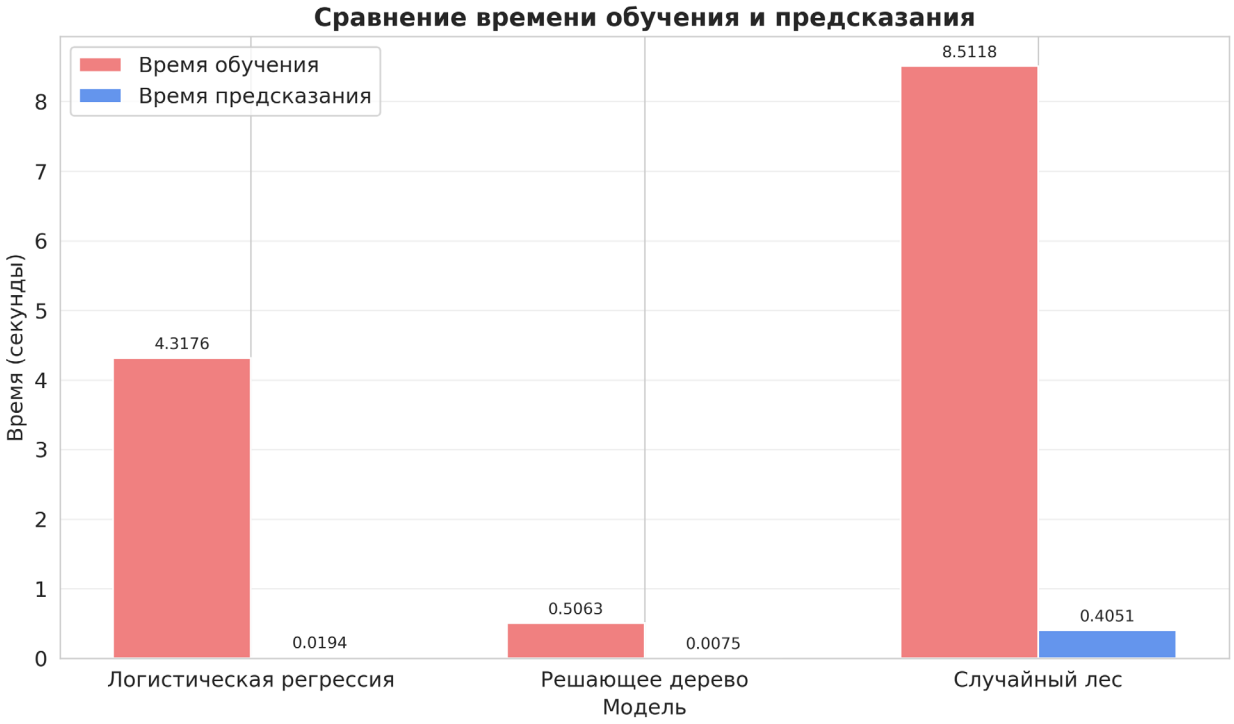

Рис. 2. — Метрики производительности моделей

Анализ производительности выявляет выраженный компромисс между точностью и скоростью. Решающее дерево является абсолютным лидером по обеим временным метрикам. Его обучение заняло менее секунды (0,51 с), а предсказание для тестовой выборки — около 7,5 миллисекунд. Это делает его идеальным кандидатом для сценариев, где требуются мгновенные предсказания на устройствах с ограниченными ресурсами, либо для задач, где важна максимальная прозрачность и скорость обучения.

Логистическая регрессия также показала отличное время предсказания (порядка 20 мс), хотя её обучение заняло существенно больше времени (4,31 с) из-за итеративного процесса оптимизации.

Случайный лес, демонстрирующий наивысшую точность, ожидаемо оказался наиболее требовательным к ресурсам. Его обучение (построение 100

деревьев по умолчанию) заняло более 8,5 секунд, что на порядок больше, чем у решающего дерева. Более важно то, что время предсказания для тестовой выборки составило ~ 400 миллисекунд, что примерно в 50 раз медленнее, чем у решающего дерева.

Одной из центральных целей исследования являлась не только демонстрация точности, но и интерпретация результатов. Механизм оценки важности признаков, встроенный в алгоритм случайного леса, позволил выявить, какие именно характеристики URL несут наибольшую дискриминативную силу для различения фишинговых и легитимных ресурсов. На рис. 3 представлен ранжированный список 40 наиболее важных признаков представлен.

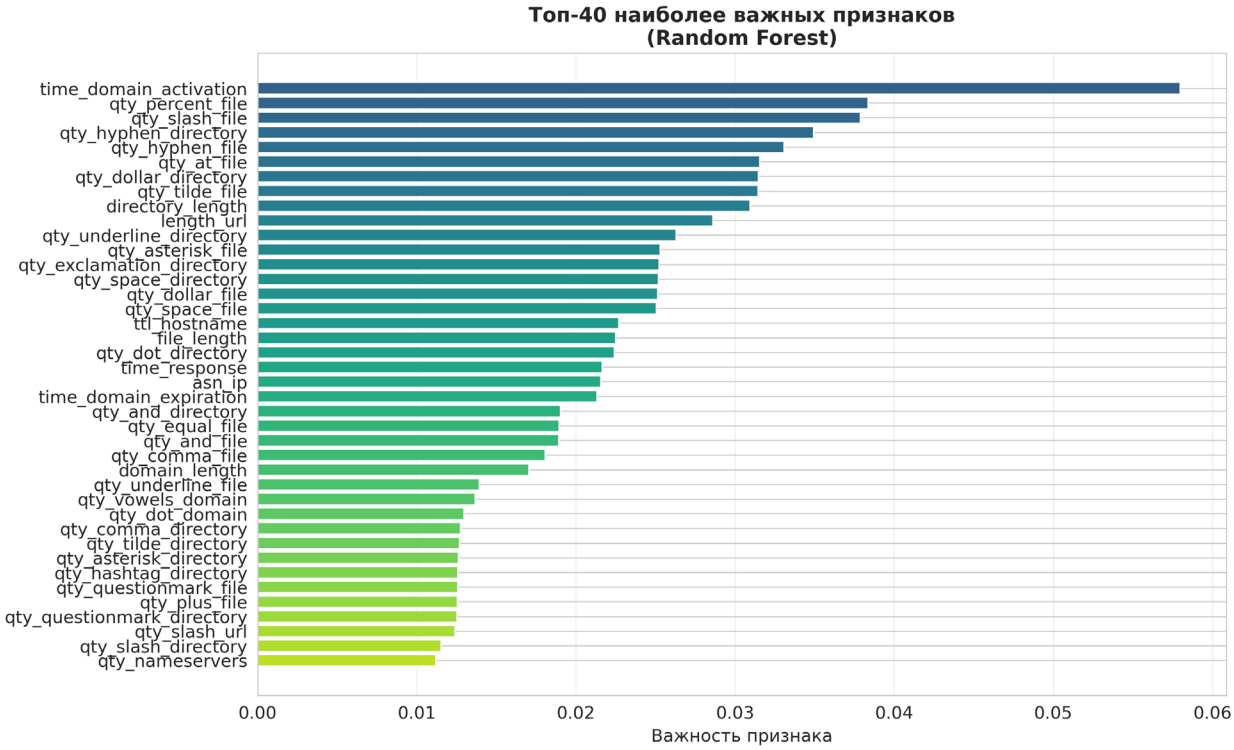

Рис. 3. — Топ-40 наиболее важных признаков по версии модели случайный лес

Анализ важности признаков полностью соответствует выводам предыдущих исследований. Абсолютным лидером, как и ожидалось, стал возраст домена, подтверждая его роль ключевого индикатора краткосрочных

фишинговых кампаний. Высокая значимость синтаксических признаков обфускации (спецсимволы в файлах и директориях) также согласуется с известными тактиками маскировки. При этом снижение относительной важности таких классических маркеров, как SSL-сертификат или IP-адрес в домене, отражает актуальную эволюцию методов фишеров. Модель корректно выделила наиболее релевантные и современные паттерны угроз.

Обсуждение результатов

Наиболее показательным является прямое сравнение эффективности классического алгоритма случайного леса с результатами более сложных нейросетевых моделей. Модель многослойного персептрона, представленная в работе Лукмановой К.А. и Картака В.М., достигла точности 88% на тестовой выборке, используя 30 признаков [6]. Наша модель, обученная на 98 признаках, продемонстрировала точность 96.92%. Это превосходство почти в 9 процентных пунктов нельзя объяснять исключительно разницей в объёме признакового пространства. Оно наглядно демонстрирует, что для задачи классификации структурированных признаков URL ансамблевый метод на основе деревьев может быть существенно эффективнее стандартной полносвязной нейронной сети. Более того, ансамблевый подход не требует трудоёмкого подбора архитектуры, что упрощает разработку и снижает затраты на настройку.

Результаты нашей работы находятся в полном согласии с выводами исследования Лапиной М.А. и соавторов, где ансамблевые методы также показали лидерство [7,11]. В их работе случайный лес достиг точности 96.60%, а метод градиентного бустинга для деревьев решений — 97.34%. Наш результат для случайного леса (96.92%) практически идентичен этим данным, что служит убедительным свидетельством воспроизводимости и надёжности высоких показателей данного алгоритма на крупных датасетах.

Ключевой вывод, подтверждённый обоими исследованиями, заключается в том, что сложные модели глубокого обучения не имеют решающего преимущества перед правильно настроенными ансамблями деревьев для классификации фишинговых URL, что также отмечается в других сравнительных исследованиях [13].

Параллельно наше исследование выявило ограниченную эффективность сверхпростых моделей, что иллюстрирует работа Беловой Е.И. и Хамитова Р.М., [5]. Наш эксперимент с логистической регрессией, которая также является линейной моделью, показал, что даже при доступе ко всем 98 признакам линейный классификатор уступает нелинейным методам. Это подтверждает гипотезу о нелинейной природе взаимосвязей между признаками фишингового URL.

На основе анализа компромисса «точность-производительность-интерпретируемость» сформулированы практические рекомендации. Для образовательных целей и прототипирования оптимально решающее дерево благодаря наглядности и скорости. В реальных системах, где критичны точность и минимум ложных срабатываний, безусловно лидирует случайный лес [12].

Вместе с тем проведенное исследование имеет методологические ограничения. Фокус исключительно на URL-признаках не учитывает текстовый контент и социальную инженерию [8,14]. Модель уязвима к концептуальному дрейфу тактик фишеров и не адаптирована под специфику русскоязычных угроз. Также требуется осторожность при автоматическом развертывании из-за риска ложных срабатываний.

Заключение

Проведённое исследование подтверждает, что классические интерпретируемые алгоритмы машинного обучения демонстрируют

высочайшую эффективность в задаче детекции фишинговых ссылок. Модель случайный лес показала превосходство, достигнув показателя общей точности 96.92% и F1-меры 95.56%, что не только соответствует, но и превосходит результаты более сложных нейросетевых подходов. Эксперимент детально охарактеризовал компромисс между точностью, скоростью и интерпретируемостью: случайный лес оптимален для промышленного развёртывания благодаря высочайшей точности предсказания фишинга (95.28%), решающее дерево — для образования и прототипирования.

Ключевым достижением является обеспечение интерпретируемости: анализ важности признаков выявил ведущую роль короткого возраста домена и признаков обфускации, предоставив практикам понятные индикаторы угроз. Научная новизна работы заключается в системном смещении фокуса с максимизации точности на многокритериальный анализ практической пригодности модели. Ограничением исследования является фокус только на URL-признаках, что определяет перспективу — создание гибридной адаптивной системы, объединяющей анализ URL, текстового контента и контекста для противодействия многофакторным фишинговым атакам.

Литература

1. Усачев, С. И., Усачева Е.А. Информационно-коммуникационные технологии в механизме преступной деятельности // Сибирский юридический вестник. – 2024. – № 3(106). – С. 110-117.

2. Селиверстов В.В., Корчагин С.А. Анализ актуальности и состояния современных фишинг-атак на объекты критической информационной инфраструктуры // Инженерный вестник Дона, 2024, № 6 URL:ivdon.ru/ru/magazine/archive/n6y2024/9277.

3. Афанасьева Н.С., Елизаров Д.А., Мызникова Т.А. Классификация фишинговых атак и меры противодействия им // Инженерный вестник Дона, 2022, № 5 URL:ivdon.ru/ru/magazine/archive/n5y2022/7641

4. Мухамадиева К.Б., Муминов Б.Б. Обзор методов обнаружения фишинговых атак на основе искусственного интеллекта // Вестник Донецкого национального университета. Серия Г: Технические науки. – 2021. – № 4. – С. 37-45.

5. Белова Е.И., Хамитов Р.М. Совершенствование методов обнаружения фишинговых атак на базе алгоритмов машинного обучения // Научно-технический вестник Поволжья. – 2023. – № 12. – С. 646-649.

6. Лукманова К.А., Картак В.М. Разработка системы защиты от фишинговых атак с использованием программно-аппаратной реализации методов машинного обучения // Моделирование, оптимизация и информационные технологии. – 2024. – Т. 12, № 4(47). URL: moitvivt.ru/ru/journal/article?id=1738

7. Лапина М.А., Лукьянов Д.А., Лапин В.Г., Кучеров Н.Н. Исследование методов машинного обучения для обнаружения сайтов-мошенников // Известия ЮФУ. Технические науки. – 2025. – № 4(246). – С. 250-262.

8. Болдырихин Н.В., Ядрец Э.А. Обнаружение фишинговых электронных писем с помощью рекуррентных нейронных сетей // Вопросы кибербезопасности. – 2025. – № 4(68). – С. 134-141.

9. Корнюхина С.П., Лапонина О.Р. Исследование возможностей алгоритмов глубокого обучения для защиты от фишинговых атак // International Journal of Open Information Technologies. – 2023. – Т. 11, № 6. – С. 163-174.

10. Гордиенко В.В., Жданов Д.М. Методы защиты от социальной инженерии и фишинга. Их достоинства и недостатки // Auditorium. – 2024. – № 2(42). – С. 45-49.

11. Adeyemi Onih, V. Phishing Detection Using Machine Learning: A Model Development and Integration // International Journal of Scientific and Management Research. – 2024. – Vol. 07, No. 04. – pp. 27-63. – DOI: 10.37502/ijsmr.2024.7403.

12. Павлычев А.В., Кузьминец К.В. Выявление фишинговых интернет-доменов с помощью алгоритмов машинного обучения в режиме потоковой обработки данных // Вопросы кибербезопасности. – 2025. – №2(66). – С.141-153.

13. Khan, Amir and Ahmed, Dr. Muqeem and Fathima, Afrah, Enhanced Phishing Detection Using Machine Learning Algorithms: A Comparative Study of Random Forest, SVM, and Logistic Regression Models (March 24, 2025). Proceedings of the International Conference on Innovative Computing & Communication (ICICC 2024).

14. Крепак И.П. Разбор проведения интернет-агрессором атаки целевого фишинга для дальнейшей эксплуатации чувствительных данных // Инженерный вестник Дона, 2025, № 3. URL: ivdon.ru/ru/magazine/archive/n3y2025/9946.

References

1. Usachev S.I., Usacheva E.A., Sibirskij juridicheskij vestnik. 2024. № 3. pp. 110-117.

2. Seliverstov V.V., Korchagin S.A., Inzhenernyj vestnik Dona, 2024, № 6. URL: ivdon.ru/ru/magazine/archive/n6y2024/9277.

3. Afanas'eva N.S., Elizarov D.A., Myznikova T.A., Inzhenernyj vestnik Dona, 2022, № 5. URL:ivdon.ru/ru/magazine/archive/n5y2022/7641

4. Muhamadieva K.B., Muminov B.B., Vestnik Doneckogo nacional'nogo universiteta. Serija G: Tehnicheskie nauki. 2021. № 4. pp. 37-45.

5. Belova E.I., Hamitov R.M., Nauchno-tehnicheskij vestnik Povolzh'ja. 2023. № 12. pp. 646-649.

6. Lukmanova K.A., Kartak V.M., Modelirovanie, optimizacija i informacionnye tehnologii. 2024. № 4. URL: moitvivt.ru/ru/journal/article?id=1738

7. Lapina M.A., Luk'janov D.A., Lapin V.G., Kucherov N.N., Izvestija JuFU. Tehnicheskie nauki. 2025. № 4. pp. 250-262.

8. Boldyrihin N.V., Jadrec Je.A., Voprosy kiberbezopasnosti. 2025. № 4. pp. 134-141.

9. Kornjuhina S.P., Laponina O.R., International Journal of Open Information Technologies. 2023. № 6. pp. 163-174.

10. Gordienko V.V., Zhdanov D.M., Auditorium. 2024. № 2. pp. 45-49.

11. Adeyemi Onih, V., International Journal of Scientific and Management Research. 2024. Vol. 07, No. 04. P. 27-63. DOI: 10.37502/ijsmr.2024.7403.

12. Pavlychev A.V., Kuz'minec K.V., Voprosy kiberbezopasnosti. 2025. №2. pp. 141-153.

13. Khan, Amir and Ahmed, Dr. Muqeem and Fathima, Afrah, Enhanced Phishing Detection Using Machine Learning Algorithms: A Comparative Study of Random Forest, SVM, and Logistic Regression Models (March 24, 2025). Proceedings of the International Conference on Innovative Computing & Communication (ICICC 2024).

14. Krepak I.P., Inzhenernyj vestnik Dona, 2025, № 3/ URL: ivdon.ru/ru/magazine/archive/n3y2025/9946.

Авторы согласны на обработку и хранение персональных данных.

Дата поступления: 7.01.2026

Дата публикации: 6.02.2026