БЕЗОПАСНОСТЬ ЦИФРОВЫХ ТЕХНОЛОГИЙ. – 2025. – № 1 (116). – 70–85

МЕТОДЫ И СИСТЕМЫ ЗАЩИТЫ ИНФОРМАЦИИ, ИНФОРМАЦИОННАЯ БЕЗОПАСНОСТЬ

УДК 004.056.5 DOI: 10.17212/2782-2230-2025-1-70-85

ПРИМЕНЕНИЕ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ОБНАРУЖЕНИЯ ФИШИНГОВЫХ ПИСЕМ: АЛГОРИТМЫ И ПОДХОДЫ*

А.Б. АРХИПОВА[1], В.А. ИЖИК[2]

[1] *630073, г. Новосибирск, пр. Карла Маркса, 20, Новосибирский государственный технический университет, доцент кафедры защиты информации. E-mail: arhipova@corp.nstu.ru*

[2] *630073, г. Новосибирск, пр. Карла Маркса, 20, Новосибирский государственный технический университет, лаборант кафедры защиты информации. E-mail: avtf@corp.nstu.ru*

В статье рассматривается подход к обнаружению фишинговых писем с использованием методов машинного обучения. Основное внимание уделяется построению и сравнению моделей классификации, таких как логистическая регрессия, случайный лес и градиентный бустинг. Основное внимание уделяется тому, каким образом модели машинного обучения могут быть встроены на различные уровни системы безопасности – от сетевых протоколов до приложений и систем хранения данных. Предложен метод предобработки текстовых данных, включающий извлечение ключевых признаков. Проанализированы основные меры защиты от почтового фишинга и методы отбора признаков для повышения точности моделей на основе методов фильтрации, методов обертки и методов вложения. В статье сформулированы требования к разработке антифишинговой системы, включая учет языковых особенностей текста Разработан прототип программного обеспечения, предполагающего внедрение комплексного гибридного подхода к анализу содержимого на основе алгоритмов машинного обучения. Программное обеспечение позволяет гибко формировать стратегию поведения пользователя в случае преднамеренной фишинговой атаки.

Ключевые слова: защита информации, архитектура безопасности, модели угроз, фишинг, машинное обучение, случайный лес, градиентный бустинг

* Статья получена 31 января 2025 г.

ВВЕДЕНИЕ

Современные информационные системы сталкиваются с постоянно возрастающей угрозой кибератак, что требует разработки более продвинутых и гибких мер безопасности. Одной из наиболее распространенных и опасных атак является фишинг, который представляет собой попытку злоумышленников получить конфиденциальную информацию (пароли, банковские данные и др.) путем обмана пользователей. Ущерб от таких атак ежегодно исчисляется миллиардами долларов, что делает фишинг одной из ключевых угроз информационной безопасности.

Особую актуальность проблема фишинга приобретает в условиях роста объемов электронной переписки и увеличения сложности атак. Злоумышленники активно используют современные технологии, такие как создание дубликатов реальных сайтов и автоматическая генерация писем, что делает их атаки сложными для обнаружения традиционными средствами защиты.

В этих условиях разработка интеллектуальных систем анализа и обнаружения фишинговых писем с использованием методов машинного обучения является не просто актуальной задачей, а насущной необходимостью для обеспечения защиты данных и предотвращения финансовых потерь [1, с. 2].

Целью исследования в настоящей работе является разработка системы, способной эффективно анализировать и обнаруживать фишинговые письма с применением методов машинного обучения. Основное внимание уделяется тому, каким образом модели машинного обучения могут быть встроены на различные уровни системы безопасности – от сетевых протоколов до приложений и систем хранения данных.

1. МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ И ИХ ПРИМЕНЕНИЕ В КИБЕРБЕЗОПАСНОСТИ

Фишинговые письма представляют собой одну из сложных задач классификации, требующую учета разнообразных текстовых, структурных и технических характеристик [2, с. 3]. Современные методы машинного обучения включают широкий спектр алгоритмов, которые эффективно решают эту задачу. Ниже приведены наиболее часто используемые из них.

1.1. ГРАДИЕНТНЫЙ БУСТИНГ (GRADIENT BOOSTING)

Градиентный бустинг, представленный такими популярными библиотеками, как XGBoost, LightGBM и CatBoost, является одним из самых мощных методов для задач классификации фишинга [3, с. 11].

Преимущества градиентного бустинга:

- высокая точность благодаря возможности учитывать сложные нелинейные зависимости между признаками;
- хорошая работа с разнородными данными, включая текстовые и числовые характеристики.

Градиентный бустинг используют для анализа текстовых и HTML-характеристик писем, таких как частота ключевых слов, длина URL, наличие подозрительных доменов и метатегов.

1.2. СЛУЧАЙНЫЙ ЛЕС (RANDOM FOREST)

Алгоритм случайного леса представляет собой ансамблевый метод, объединяющий множество деревьев решений. Метод является устойчивым к переобучению за счет использования множества случайных подвыборок данных. Преимуществом метода является также возможность интерпретируемости результатов по важности признаков.

Случайный лес активно используется для ранней фильтрации подозрительных писем на основе базовых характеристик, таких как домен отправителя, длина заголовка и наличие вложений [4, с. 5].

1.3. НЕЙРОННЫЕ СЕТИ (NEURAL NETWORKS)

Современные нейронные сети, включая сверточные сети (CNN), рекуррентные сети (RNN) и трансформеры (например, BERT), обеспечивают новые уровни точности в задачах анализа фишинга.

Преимущества метода:

- высокая гибкость и способность обрабатывать неструктурированные данные;
- возможность извлечения скрытых паттернов в текстовых данных.

Нейронные сети особенно эффективны для анализа текста писем, включая их тональность, стиль написания и синтаксические особенности [5, с. 18].

2. РОЛЬ МЕТОДОВ ОТБОРА ПРИЗНАКОВ ДЛЯ ПОВЫШЕНИЯ ТОЧНОСТИ МОДЕЛЕЙ

Ключевым аспектом повышения эффективности машинного обучения является грамотный отбор признаков, так как избыточные или нерелевантные данные могут снижать точность модели и увеличивать ее сложность [6, с. 5].

Роль методов отбора признаков заключается в следующем:

- повышение точности за счет исключения нерелевантных признаков снижения шума в данных и улучшения предсказательной способности модели;
- сокращение времени обучения при работе с меньшим числом признаков, что уменьшает вычислительные затраты;
- интерпретируемость результатов за счет облегчения понимания, какие именно характеристики играют ключевую роль в классификации.

Методы отбора признаков для повышения точности моделей:

1) методы фильтрации (Filter Methods), которые оценивают признаки независимо от алгоритма классификации, используя такие метрики, как корреляция, информация взаимности или статистические тесты;

2) методы обертки (Wrapper Methods), которые подбирают оптимальный набор признаков на основе оценки модели, например, методом рекурсивного исключения признаков (RFE);

3) методы вложения (Embedded Methods), игнорирующие отбор признаков непосредственно в процесс обучения модели, как это реализовано в градиентном бустинге или линейных моделях с регуляризацией (LASSO).

3. ПРЕДОБРАБОТКА ДАННЫХ ДЛЯ АНАЛИЗА ФИШИНГОВЫХ ПИСЕМ

Предобработка данных является важнейшим этапом при создании систем классификации фишинговых писем. Качество входных данных напрямую влияет на производительность моделей машинного обучения. В задачах анализа фишинга особое внимание уделяется обработке текстовой информации, выделению ключевых признаков и учету многоязычности.

Ключевые признаки из текстов писем являются основой для построения эффективных моделей классификации. Современные подходы включают как классические методы обработки текста, так и методы, основанные на глубоких нейронных сетях (рис. 1).

1. Токенизация текста. Метод подразумевает деление текста на отдельные слова, фразы или символы для дальнейшего анализа.

Пример: разделение строки "Click here for urgent action" на токены ["Click", "here", "for", "urgent", "action"], точная токенизация повышает эффективность моделей на этапе анализа синтаксических и лексических характеристик.

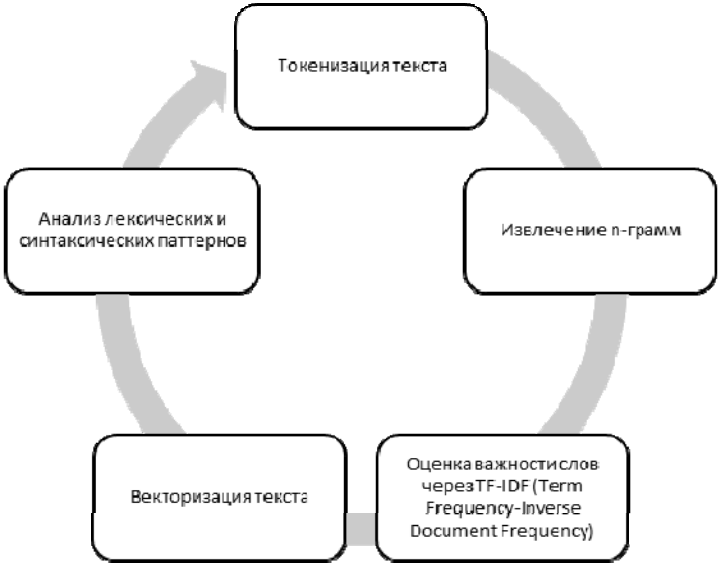

Рис. 1. Методы предобработки данных для анализа фишинговых писем

2. Извлечение *n*-грамм. Моделирование последовательностей из *n* слов (*n*-грамм) направлено на выявление частотных паттернов, характерных для фишинговых писем.

Пример: для фразы "Click here for urgent action" биграммы будут ["Click here", "here for", "for urgent", "urgent action"], использование биграмм и триграмм позволяет обнаруживать манипулятивные фразы, характерные для фишинга.

3. Оценка важности слов TF-IDF (Term Frequency-Inverse Document Frequency). TF-IDF вычисляет значимость слова в контексте всего корпуса текстов.

Пример: слова, такие как "urgent" или "free", могут получать высокий вес в фишинговых письмах по сравнению с нейтральными словами.

4. Векторизация текста. Преобразование текста в числовой формат для подачи в модель машинного обучения.

Примеры методов векторизации текста:

- Bag-of-Words (BoW): отражает частоту слов в документе;
- Word Embeddings, Word2Vec и GloVe: создают плотные векторы слов, учитывающие контекст;

- Transformer-based embeddings: BERT и его аналоги формируют контекстуализированные векторы, позволяющие учитывать скрытые зависимости между словами.

5. Анализ лексических и синтаксических паттернов. Выявление подозрительных фраз ("click here", "urgent attention required") и грамматических ошибок, часто встречающихся в фишинговых письмах. Анализ тональности и структуры писем может выявить, что фишинговые сообщения часто содержат сочетания высокой срочности и формального стиля [7, с. 5].

4. ОБРАБОТКА МНОГОЯЗЫЧНОГО ТЕКСТА ДЛЯ ВЫЯВЛЕНИЯ ФИШИНГОВЫХ ПИСЕМ

В условиях глобализации фишинговые атаки распространяются на разных языках, что требует учета языковых особенностей в процессе анализа.

4.1. ИДЕНТИФИКАЦИЯ ЯЗЫКА ТЕКСТА

- Использование библиотек, таких как LangDetect или FastText, для автоматического определения языка письма.
- Пример: фишинговое письмо, написанное на испанском языке, может быть автоматически классифицировано как "es" для дальнейшей обработки.

4.2. УНИФИКАЦИЯ ТЕКСТА

- Перевод текста на единый язык (например, английский) с помощью инструментов автоматического перевода (Google Translate API, DeepL).

4.3. ОБОГАЩЕНИЕ ДАННЫХ МНОГОЯЗЫЧНЫМИ EMBEDDING-МОДЕЛЯМИ

Использование мультилингвистических моделей, таких как mBERT или XLM-R, которые обучены на текстах разных языков.

Преимущества:

– учитывают сложные языковые зависимости;
– обеспечивают переносимость моделей на новые языки.

4.4. УЧЕТ СПЕЦИФИЧЕСКИХ ЯЗЫКОВЫХ ПАТТЕРНОВ

- Для разных языков фишинговые письма могут содержать уникальные лексические и синтаксические особенности.

Пример: в русском языке популярными являются фразы «Ваш счет заблокирован» или «Срочно подтвердите данные». Выявление таких шаблонов помогает повысить эффективность анализа.

5. ИССЛЕДОВАНИЕ АЛГОРИТМОВ И МОДЕЛЕЙ МАШИННОГО ОБУЧЕНИЯ ДЛЯ КЛАССИФИКАЦИИ ФИШИНГОВЫХ ПИСЕМ

Эффективная классификация фишинговых писем требует применения и анализа различных моделей машинного обучения, каждая из которых обладает своими преимуществами и ограничениями. В данной области популярны как простые линейные модели, так и более сложные ансамблевые подходы. Кроме того, всё чаще применяются гибридные модели, которые позволяют сочетать сильные стороны разных алгоритмов для повышения точности классификации.

5.1. ЛОГИСТИЧЕСКАЯ РЕГРЕССИЯ, ГРАДИЕНТНЫЙ БУСТИНГ, СЛУЧАЙНЫЙ ЛЕС И ИХ СРАВНИТЕЛЬНЫЙ АНАЛИЗ

Модели машинного обучения различаются по сложности, требованиям к данным и эффективности. Для классификации фишинговых писем наиболее часто применяются следующие алгоритмы.

5.1.1. Логистическая регрессия представляет собой простой и интерпретируемый алгоритм, хорошо подходящий для базовых задач классификации.

Преимущества:

– низкая вычислительная сложность;

– возможность интерпретации весов признаков, что полезно для анализа значимости слов или фраз в письмах.

Ограничение: плохо справляется с нелинейными зависимостями между признаками.

5.1.2. Случайный лес (Random Forest) является ансамблевым алгоритмом, состоящим из множества решающих деревьев, результат которого получается путем усреднения предсказаний.

Преимущества:

– высокая точность для небольших и средних наборов данных;

– интегрированный механизм оценки важности признаков.

Ограничение: более высокая вычислительная сложность по сравнению с логистической регрессией.

5.1.3. Градиентный бустинг (Gradient Boosting Machines) представляет собой алгоритм, который последовательно улучшает ошибки предыдущих моделей (базовые деревья решений).

Преимущества:

– высокая точность для сложных задач;

– эффективность в обработке несбалансированных данных, часто встречающихся в задачах классификации фишинга.

Ограничения:

– длительное время обучения;

– чувствительность к выбору гиперпараметров.

5.2. ИСПОЛЬЗОВАНИЕ ГИБРИДНЫХ МОДЕЛЕЙ ДЛЯ ПОВЫШЕНИЯ ТОЧНОСТИ

Гибридные модели сочетают в себе сильные стороны нескольких подходов для улучшения производительности классификации фишинговых писем [8, с. 9], их построение основано на выполнении последовательности этапов.

На первом этапе логистическая регрессия используется для выделения первичных признаков. На втором этапе градиентный бустинг анализирует эти признаки, выявляя сложные зависимости. Итоговая точность гибридной модели может стать выше, чем у чистого градиентного бустинга.

В литературе встречается вариант сочетания элементов, например, за счет объединения случайного леса для обработки текстовых признаков и нейронных сетей типа LSTM для анализа структуры текста.

Гибридные модели отличает ряд преимуществ:

– улучшение точности за счет комбинирования линейных и нелинейных подходов;

– устойчивость к недостаткам отдельных сегментов;

– возможность обработки разных типов данных (текст, метаданные, вложения).

6. ПРИМЕНЕНИЕ НЕЙРОННЫХ СЕТЕЙ ДЛЯ АНАЛИЗА ТЕКСТА

Современные нейронные сети, особенно архитектуры с комплексным подходом, играют ключевую роль в задачах обработки текстов, включая анализ фишинговых писем. Они обеспечивают высокую точность и гибкость при работе с многоязычными и сложными данными.

Одним из подходов к обнаружению фишинга является анализ тональности текста письма. Этот метод основывается на выявлении эмоционального окраса, уровня доверительности или намерений, что может быть индикатором мошеннических действий [9, с. 5].

Фишинговые письма часто содержат агрессивные призывы к действию (например, «Срочно выполните оплату») или, наоборот, излишне обнадеживающие фразы (например, «Вы выиграли приз»). Модели нейронных сетей помогают выделять сложные зависимости между словами, что недоступно для традиционных методов анализа. Например, они могут определить, что фраза «Ваш аккаунт временно заблокирован» указывает на фишинг, даже если она формально корректна.

Модели анализа тональности выделяют такие паттерны и классифицируют их как потенциально опасные.

Для обработки фишинговых писем на русском языке применяются специализированные модели, такие как RuBERT (Bidirectional Encoder Representations from Transformers) и его производные, которые адаптированы к особенностям русского языка [10, с. 4]. BERT позволяет обрабатывать текст в двух направлениях, что критически важно для понимания контекста тональности.

Особенности RuBERT:

– разработан на базе оригинальной модели BERT, но обучен на русскоязычных текстах;

– учитывает грамматические и синтаксические особенности русского языка, такие как сложная морфология и изменение формы слов;

– модель выделяет ключевые паттерны, такие как прямые призывы к действию («Немедленно выполните перевод»), и фразы, вызывающие эмоцию срочности («Только сегодня вы можете восстановить доступ»).

В литературе встречаются также другие русскоязычные модели:

– DeepPavlov BERT: модель на основе BERT, обученная на больших корпусах текстов на русском языке;

– RuGPT: используется для генерации и анализа текста, в том числе для выявления аномалий в формулировках, характерных для фишинговых писем.

В задачах анализа фишинга алгоритмы семейства BERT содержат последовательные этапы от преобразования исходного текста письма в векторный формат с помощью эмбеддингов, выделения ключевых слов и фраз, характеризующие мошеннические письма до итоговой классификации по принципу безопасности.

Нейронные сети, такие как RuBERT и другие модели на основе комплексного подхода, являются эффективным инструментом для анализа текста фишинговых писем. Они обеспечивают высокую точность за счет глубокого

понимания контекста и тональности. Применение таких технологий, особенно в многоязычной среде, способствует значительному снижению рисков информационной безопасности.

7. ПРОГРАММНАЯ АРХИТЕКТУРА ПРИЛОЖЕНИЯ ПО ОБНАРУЖЕНИЮ ФИШИНГОВЫХ ПИСЕМ

Обеспечение безопасности пользователя посредством анализа входящих сообщений на предмет присутствия вредоносной активности в различных каналах связи возможно в среде, в которой используются информационные технологии и в которой происходит обмен информацией между пользователями сети Интернет и телекоммуникационными линиями связи. Возможно предложить решение, обеспечивающее детектирование входящих фишинговых сообщений в различных адаптациях: как расширение для браузера; защищенный браузер, включающий в себя механизмы защиты; почтовый агент, реализующий проверку писем, поступающих на почтовый сервер; отдельное прикладное программное обеспечение.

В рамках исследования было реализовано монолитное приложение. Монолитная архитектура программного обеспечения имеет ряд преимуществ. Во-первых, его легче разрабатывать и тестировать, поскольку все компоненты находятся в одной кодовой базе. Во-вторых, его проще развертывать, поскольку всё приложение может быть развернуто как единое целое. В-третьих, его легче масштабировать, поскольку всё приложение может быть реплицировано для обработки увеличившегося трафика. В-четвертых, такое приложение обладает более высокой производительностью, поскольку отсутствуют накладные расходы, связанные с межкомпонентной связью. Наконец, его легче обезопасить, поскольку между компонентами нет внешних интерфейсов.

Всю архитектуру можно разделить на две основные компоненты: пользовательский интерфейс (Frontend) и серверная часть (Backend). Ключевыми элементами каждого компонента приложения является модуль отрисовки сообщений и кластер модулей анализа сообщений соответственно. Важно отметить, что всё взаимодействие двух главных модулей между собой осуществляется с помощью программного интерфейса приложения (API), обеспечиваемого модулем маршрутизации. Также стоит упомянуть один из основных по функционалу модулей, а именно компонент взаимодействия с серверами IMAP для скачивания писем на сервер пользователя. Что касается взаимодействия с пользователем, в программе содержится модуль обработки пользовательского ввода, а также модуль сохранения и шифрования учетных данных пользователя.

На рис. 2 изображена объединяющая все модули архитектура приложения.

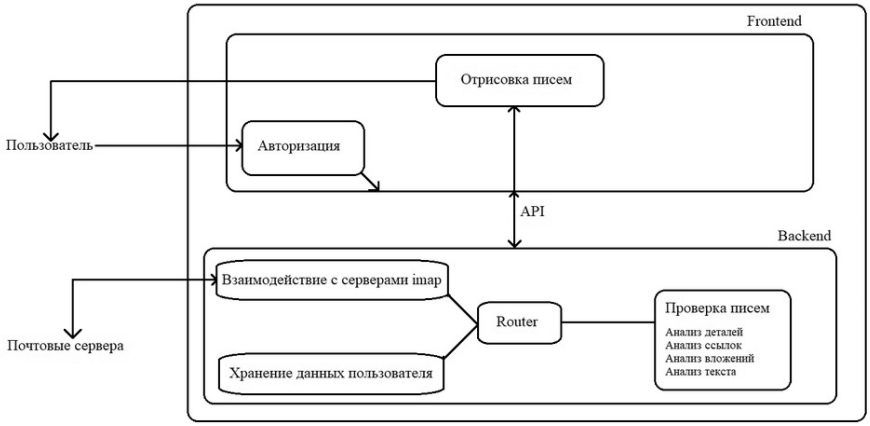

Рис. 2. Архитектура приложения

Если рассматривать процесс определения фишинговой активности в письме с точки зрения подхода с машинным обучением гибридных моделей, то проверка в прототипе программного обеспечения осуществлялась на основе анализа по трем компонентам: 1) анализ заголовков письма, 2) анализ вложений, 3) анализ ссылок. Проверка показала высокую точность работы приложения (85 %) и высокую скорость обработки результатов.

ЗАКЛЮЧЕНИЕ

Использование машинного обучения для борьбы с фишинговыми угрозами значительно повышает уровень безопасности в интернет-среде. Нейронные сети и гибридные модели являются эффективным решением в обработке текстов и обнаружении фишинговых атак, особенно в сочетании с многоязычными подходами. Перспективы развития технологий машинного обучения в этой области обещают значительные улучшения, обеспечивая более точную и быструю защиту данных, особенно в глобальном масштабе. Важно отметить, что постоянное совершенствование методов анализа и их адаптация к новым угрозам остается ключевым аспектом для борьбы с фишингом в будущем.

СПИСОК ЛИТЕРАТУРЫ

1. *Гущина А.А., Пчелинцева Н.В., Шацкий В.А.* Применение искусственного интеллекта в обеспечении безопасности данных // Инженерное обеспечение инновационных технологий в АПК: материалы Международной научно-практической конференции, 26–28 октября 2021 года. – Мичуринск, 2021. – С. 79–81. – EDN NJVTOM.

2. *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. – EDN JHFVLY.

3. *Шульгин С.Г.* Отбор переменных для анализа и прогнозирования нестабильности с помощью моделей градиентного бустинга // Системный мониторинг глобальных и региональных рисков: ежегодник. Т. 9. Социально-политическая и экономическая дестабилизация: анализ страновых и региональных ситуаций в мир-системном аспекте. – Волгоград, 2018. – С. 115–153. – EDN JFUHQP.

4. *Konstantinov A.V.* Deep gradient boosting for regression // Computing, Telecommunications and Control. – 2021. – Vol. 14 (3). – P. 7–19. – DOI: 10.18721/JCSTCS.14301. – EDN CZBFQN.

5. *Астраханцева И.А., Герасимов А.С.* Прогнозирование региональной инфляции на основе гибридной модели машинного обучения: градиентный бустинг и случайный лес // Научные труды Вольного экономического общества России. – 2023. – Т. 243, № 5. – С. 200–226. – DOI: 10.38197/2072-2060-2023-243-5-200-226. – EDN JATZXG.

6. *Козлова Н.Ш., Довгаль В.А.* Анализ применения искусственного интеллекта и машинного обучения в кибербезопасности // Вестник Адыгейского государственного университета. Серия: Естественно-математические и технические науки. – 2023. – № 3 (326). – С. 65–72. – DOI: 10.53598/2410-3225-2023-3-326-65-72. – EDN CYUKLH.

7. *Фадеева А.А., Сиякина В.В., Салахутдинов Э.Р.* Применение искусственных нейронных сетей для анализа тональности текста // Социально-экономические и технические системы: исследование, проектирование, оптимизация. – 2020. – № 1 (84). – С. 164–171. – EDN RLMKMG.

8. *Торкунова Ю.В., Милованов Д.В.* Оптимизация нейронных сетей: методы и их сравнение на примере интеллектуального анализа // International Journal of Advanced Studies. – 2023. – Т. 13, № 4. – С. 142–158. – DOI: 10.12731/2227-930X-2023-13-4-142-158. – EDN SFIPKW.

9. *Яцино С.Л.* Сравнение типов искусственных нейронных сетей для семантического анализа текста // Научное сообщество студентов XXI столетия. Технические науки: электронный сборник статей по материалам LXV студенческой Международной научно-практической конференции. Т. 5 (64). – Новосибирск, 2018. – С. 239–243. – EDN XOGGJF.

10. *Голубев А.А., Лукашевич Н.В.* Исследование моделей нейронных сетей типа BERT для анализа тональности текстов на русском языке // Научно-техническая информация. Серия 2, Информационные процессы и системы. – 2021. – № 1. – С. 32–41. – DOI: 10.36535/0548-0027-2021-01-3. – EDN DQHLBO.

11. *Архипова А.Б., Нечаев Д.А.* Технология формирования интегрированной антифишинговой системы в цифровом обществе // Вестник СибГУТИ. – 2023. – Т. 17, № 2. – С. 93–103. – DOI: 10.55648/1998-6920-2023-17-2-93-103.

12. *Архипова А.Б., Листаров Р.Е.* Формирование методики безопасного программирования для разработки веб-приложений // Безопасность цифровых технологий. – 2023. – № 4 (111). – С. 64–81. – DOI: 10.17212/2782-2230-2023-4-64-81.

13. *Архипова А.Б., Нечаев Д.А.* К вопросу построения модели фишинговой атаки на базе теории некооперативных игр // Первспектива-2021: материалы IX Всероссийской молодежной школы-семинара по проблемам информационной безопасности, Красноярск, 30 сентября – 3 октября 2021 г. – М., 2022. – С. 6–12.

Архипова Анастасия Борисовна, доцент кафедры защиты информации Новосибирского государственного технического университета. Основное направление научных исследований – математическое моделирование в информационной безопасности, оценка качества социально значимой деятельности. E-mail: arhipova@corp.nstu.ru

Ижик Владислав Александрович, лаборант кафедры защиты информации Новосибирского государственного технического университета. Область научных интересов – разработка систем обнаружения утечки информации. E-mail: avtf@corp.nstu.ru

DOI: 10.17212/2782-2230-2025-1-70-85

Using machine learning methods to detect phishing emails: algorithms and approaches*

A.B. Arkhipova[1], V.A. Izhik[2]

[1] *Novosibirsk State Technical University, 20 Karl Marx Avenue, Novosibirsk, 630073, Russian Federation, Associate Professor of the Department of Information Security. E-mail: arhipova@corp.nstu.ru*

[2] *Novosibirsk State Technical University, 20 Karl Marx Avenue, Novosibirsk, 630073, Russian Federation, laboratory assistant of Information Security Department. E-mail: avtf@corp.nstu.ru*

The article discusses the approach to detecting phishing emails using machine learning methods. The focus is on building and comparing classification models such as logistic regression, case forest, and gradient boosting. The main focus is on how machine learning models can be built into various levels of security systems, from network protocols to applications and storage systems. Disclosed is a method of preprocessing text data, involving extraction of key features. Analyzed the main measures of protection against mail phishing. Methods for selecting features to increase the accuracy of models based on filtering methods, wrapper methods and embedding methods are analyzed. The article formulates the requirements for the development of an anti-phishing system, including taking into account the linguistic features of the text. A pro-totype of software has been developed, which involves the introduction of a comprehensive hybrid approach to content analysis based on machine learning algorithms. The software allows you to flexibly shape the user's behavior strategy in the event of a deliberate phishing attack.

Keywords: information protection, security architecture, threat models, phishing, machine learning, random forest, gradient boosting

REFERENCES

1. Gushchina A.A., Pchelintseva N.V., Shatsky V.A. [Application of artificial intelligence in ensuring data security]. *Inzhenernoe obespechenie innovatsionnykh tekhnologii v APK* [Engineering support of innovative technologies in the agro-industrial complex]. Materials of the International Scientific and Practical Conference, October 26–28, Michurinsk, 2021, pp. 79–81. (In Russian).

2. Korniukhina S.P., Laponina O.R. Issledovanie vozmozhnostei algoritmov glubokogo obucheniya dlya zashchity ot fishingovykh atak [Research of the capabilities of deep learning algorithms to protection against phishing attacks]. *Interna-*

* *Received 31 January 2025.*

tional Journal of Open Information Technologies, 2023, vol. 11 (6), pp. 163–174. (In Russian).

3. Shul'gin S.G. Otbor peremennykh dlya analiza i prognozirovaniya nestabil'nosti s pomoshch'yu modelei gradientnogo bustinga [Selection of variables for analysis and prediction of non-stability using gradient boosting models]. *Sistemnyi monitoring global'nykh i regional'nykh riskov*. T. 9. *Sotsial'no-politicheskaya i ekonomicheskaya destabilizatsiya: analiz stranovykh i regional'nykh situatsii v mir-sistemnom aspekte* [System monitoring of global and regional risks. Vol. 9. Socio-political and economic destabilization: analysis of country and regional situations in the world-systemic aspect]. Volgograd, 2018, pp. 115–153.

4. Konstantinov A.V. Deep gradient boosting for regression problems. *Computing, Telecommunications and Control*, 2021, vol. 14 (3), pp. 7–19. DOI: 10.18721/JCSTCS.14301.

5. Astrakhantseva I.A., Gerasimov A.S. Prognozirovanie regional'noi inflyatsii na osnove gibridnoi modeli mashinnogo obucheniya: gradientnyi busting i sluchainyi les [Regional inflation prediction based on hybrid machine learning model: gradient boosting and random forest]. *Nauchnye trudy Vol'nogo ekonomicheskogo obshchestva Rossii = Scientific works of the Free economic society of Russia*, 2023, vol. 243, no. 5, pp. 200–226. DOI: 10.38197/2072-2060-2023-243-5-200-226.

6. Kozlova N.Sh., Dovgal V.A. Analiz primeneniya iskusstvennogo intellekta i mashinnogo obucheniya v kiberbezopasnosti [Analysis of the use of artificial intelligence and machine learning in cybersecurity]. *Vestnik Adygeiskogo gosudarstvennogo universiteta. Seriya: Estestvenno-matematicheskie i tekhnicheskie nauki = Bulletin of Adyghe State University. Series: Natural-Mathematical and Technical Sciences*, 2023, no. 3 (326), pp. 65–72. DOI: 10.53598/2410-3225-2023-3-326-65-72.

7. Fadeeva A.A., Siyakina V.V., Salakhutdinov E.R. Primenenie iskusstvennykh neironnykh setei dlya analiza tonal'nosti teksta [Application of artificial neural networks for sentiment analysis for text]. *Sotsial'no-ekonomicheskie i tekhnicheskie sistemy: issledovanie, proektirovanie, optimizatsiya = Social-economic and technical systems: research, design and optimization*, 2020, no. 1 (84), pp. 164–171.

8. Torkunova Yu.V., Milovanov D.V. Optimizatsiya neironnykh setei: metody i ikh sravnenie na primere intellektual'nogo analiza [Optimization of neural networks: methods and their comparison using the example of text mining]. *International Journal of Advanced Studies*, 2023, vol. 13 (4), pp. 142–158. DOI 10.12731/2227-930X-2023-13-4-142-158.

9. Yatsino S.L. [Comparison of types of artificial neural networks for se-mantic text analysis]. *Nauchnoe soobshchestvo studentov XXI stoletiya. Tekhnicheskie nauki* [Scientific community of students of the XXI century. Technical sciences]. Electronic collection of articles on materials of the LXV student international scien-

tific and practical conference. Vol. 5(64). Novosibirsk, 2018, pp. 239–243. (In Russian).

10. Golubev A.A., Loukashevitch N.V. Issledovanie modelei neironnykh setei tipa BERT dlya analiza tonal'nosti tekstov na russkom yazyke [Investigation of BERT-type neural network models for sentiment analysis of Russian texts]. *Nauchno-tekhnicheskaya informatsiya. Seriya 2, Informatsionnye protsessy i sistemy*, 2021, no. 1, pp. 32–41. DOI: 10.36535/0548-0027-2021-01-3.

11. Arkhipova A.B., Nechaev D.A. Tekhnologiya formirovaniya integrirovannoi anti-fishingovoi sistemy v tsifrovom obshchestve [Technology for the formation of an integrated anti-phishing system in a digital society]. *Vestnik SibGUTI = The Herald of the Siberian State University of Telecommunications and Information Science*, 2023, vol. 17no. 2, pp. 93–103. DOI: 10.55648/1998-6920-2023-17-2-93-103.

12. Arkhipova A.B., Listarov R.E. Formirovanie metodiki bezopasnogo programmirovaniya dlya razrabotki veb-prilozhenii [Formation of secure programming methods for development web applications]. *Bezopasnost' tsifrovykh tekhnologii = Digital Technology Security*, 2023, no. 4 (111), pp. 64–81. DOI: 10.17212/2782-2230-2023-4-64-81.

13. Arkhipova A.B., Nechaev D.A. [On the issue of building a model of a phishing attack based on the theory of non-cooperative games]. *Pervspektiva-2021* [First perspective – 2021]. Proceedings of the IX All-Russian Youth School-Seminar on Information Security Issues, Krasnoyarsk, September 30 – October 3, 2021. Moscow, 2022, pp. 6–12. (In Russian).

Для цитирования:

Архипова А.Б., Ижик В.А. Применение методов машинного обучения для обнаружения фишинговых писем: алгоритмы и подходы // Безопасность цифровых технологий. – 2025. – № 1 (116). – С. 70–85. – DOI: 10.17212/2782-2230-2025-1-70-85.

For citation:

Arkhipova A.B., Izhik V.A. Primenenie metodov mashinnogo obucheniya dlya obnaruzheniya fishingovykh pisem: algoritmy i podkhody [Using machine learning methods to detect phishing emails: algorithms and approaches]. *Bezopasnost' tsifrovykh tekhnologii = Digital Technology Security*, 2025, no. 1 (116), pp. 70–85. DOI: 10.17212/2782-2230-2025-1-70-85.