Электронный научный журнал "Математическое моделирование, компьютерный и натурный эксперимент в естественных науках" http://mathmod.esrae.ru/
URL статьи: **mathmod.esrae.ru/50-206**
Ссылка для цитирования этой статьи:
Назаров А.А., Губенков А.А. Использование машинного обучения для защиты от фишинговых атак: сравнение с традиционными механизмами // Математическое моделирование, компьютерный и натурный эксперимент в естественных науках. 2025. №2

УДК 004.891.3 DOI:10.24412/2541-9269-2025-2-3-7

ИСПОЛЬЗОВАНИЕ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ЗАЩИТЫ ОТ ФИШИНГОВЫХ АТАК: СРАВНЕНИЕ С ТРАДИЦИОННЫМИ МЕХАНИЗМАМИ

Назаров А.А.[1], Губенков А.А.[2]
[1] Саратовский государственный технический университет имени Гагарина Ю.А., Россия, Саратов, nazandre1233@gmail.com
[2]Саратовский государственный технический университет имени Гагарина Ю.А., Россия, Саратов, agubenkov@mail.ru

USING MACHINE LEARNING FOR PROTECTION AGAINST PHISHING ATTACKS: A COMPARISON WITH TRADITIONAL MECHANISMS

Nazarov A.A.[1], Gubenkov A.A.[2]
[1]Yuri Gagarin State Technical University of Saratov, Russia, Saratov, azandre1233@gmail.com
[2]Yuri Gagarin State Technical University of Saratov, Russia, Saratov, agubenkov@mail.ru

Аннотация. Рассмотрена проблема фишинговых атак и недостатки традиционных методов их обнаружения, основанных на статических списках. Для повышения эффективности защиты предложено использовать методы машинного обучения (ML), которые автоматически извлекают признаки из URL-адресов. Проведен сравнительный анализ ML-модели (точность 0.97) и 145 сервисов проверки фишинговых ссылок, из которых только 18 показали точность выше 0.55, а лучший – 0.83. Описаны используемые признаки, методика оценки и экспериментальные результаты, подтверждающие преимущество ML-подхода в обнаружении фишинга.

Ключевые слова: фишинг, машинное обучение, информационная безопасность, обнаружение фишинга, анализ URL-адресов.

Abstract. The problem of phishing attacks and the limitations of traditional detection methods based on static lists are examined. To improve protection efficiency, a machine learning (ML)-based approach is proposed, which automatically extracts features from URLs. A comparative analysis was conducted between the ML model (accuracy: 0.97) and 145 phishing detection services, of which only 18 achieved an accuracy above 0.55, with the best reaching 0.83. The study describes the extracted features, evaluation methodology, and experimental results, confirming the

advantage of the ML approach in phishing detection.

Keywords: phishing, machine learning, information security, phishing detection, URL analysis.

Фишинговые атаки стали одной из самых серьёзных угроз для современных информационных систем [1]. С каждым годом злоумышленники совершенствуют свои техники, совершенствуют способы обхода традиционных систем защиты. Фишинг распространяется через электронную почту, SMS-сообщения и социальные сети, что приводит к значительным финансовым потерям и утечкам персональных данных [2]. Необходимость оперативно реагировать на изменяющиеся тактики атакующих обусловливает необходимость применения методов, способных к самообучению и адаптации.

Методы машинного обучения привлекают внимание исследователей в области информационной безопасности [3-10] благодаря своей способности автоматически извлекать релевантные признаки из сырых данных (например, URL-адресов, HTML-кода) и быстро адаптироваться к новым шаблонам атак. Учитывая динамическую природу фишинговых сайтов, ML-подходы способны обеспечить более высокую точность обнаружения атак по сравнению с традиционными методами. В данной статье проводится сравнительный анализ качества работы модели машинного обучения и ряда сервисов, основанных на общедоступных механизмах обнаружения фишинга.

Для обучения и тестирования модели использовался готовый набор данных [11], содержащий большое количество URL-адресов с извлечёнными признаками.

Основой классификатора является XGBClassifier, который учитывает как нелинейные, так и линейные зависимости между признаками, что особенно актуально для задач обнаружения фишинга. В качестве метрики оптимизации используется logloss, а управляющий параметр random_state фиксирован для обеспечения воспроизводимости результатов.

Для повышения качества модели применяется RandomizedSearchCV. Были настроены следующие гиперпараметры:

- n_estimators (число деревьев): [50, 100, 200]
- max_depth (максимальная глубина деревьев): [3, 5, 7]
- learning_rate (скорость обучения): [0.01, 0.1, 0.2]
- subsample (доля объектов для каждого дерева): [0.8, 0.9, 1.0]
- colsample_bytree (доля признаков для каждого дерева): [0.8, 0.9, 1.0]
- gamma (минимальное снижение потерь для разделения узла): [0, 0.1, 0.2]
- min_child_weight (минимальное количество объектов в листе): [1, 3, 5]

Подбор оптимальных значений ведется по схеме 5-кратной стратифицированной кросс-валидации, что позволяет учитывать дисбаланс классов и исключить переобучение.

Поскольку датасет характеризуется дисбалансом между легитимными и фишинговыми URL, применяется метод SMOTE (Synthetic Minority Over-sampling Technique) для искусственного увеличения выборки для класса меньшинства. SMOTE генерирует синтетические образцы, что позволяет обеспечить равномерное распределение классов в обучающей выборке и повысить устойчивость модели к ошибкам при предсказании.

В рамках тестирования были проведены следующие измерения:

- Модель машинного обучения, обученная на указанном наборе данных с использованием широкого спектра признаков, показала точность 0.97. Высокий показатель обусловлен возможностью автоматического извлечения релевантных признаков и оперативной адаптацией модели к изменениям в характере атак. Матрица ошибок представлена на рисунке 1.

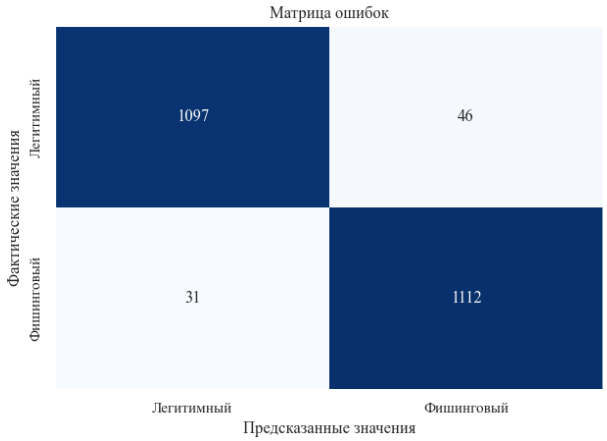

Рисунок 1. Матрица ошибок рассматриваемой модели машинного обучения

- Проведён сравнительный анализ 145 сервисов проверки фишинговых ссылок. Из них только 18 продемонстрировали точность выше 0.55, а наилучший достиг точности 0.83. Это свидетельствует о том, что традиционные методы и эвристические алгоритмы имеют существенные ограничения при работе в условиях быстрого изменения тактик злоумышленников. Результат измерений представлен на рисунке 2.

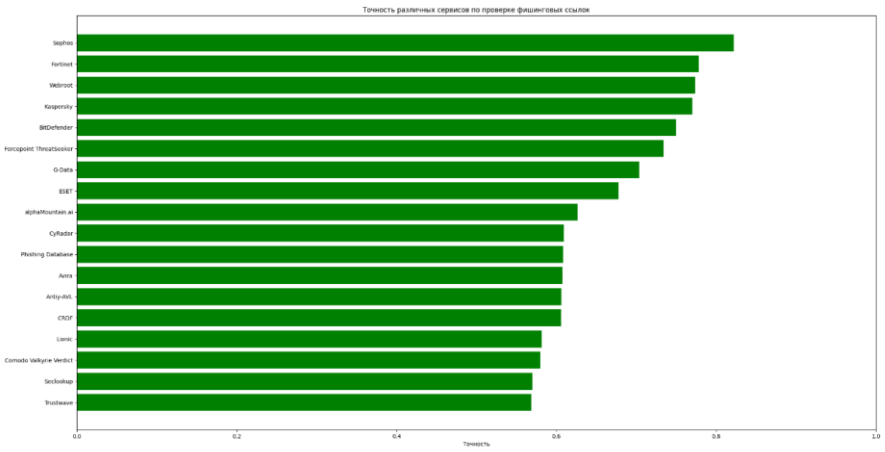

Рисунок 2. Диаграмма точности рассматриваемых традиционных механизмов обнаружения фишинговых атак

Полученные результаты подтверждают, что ML-подход позволяет значительно повысить эффективность обнаружения фишинговых атак благодаря динамической адаптации и многоаспектному анализу параметров URL.

Литература

1. Кобец П.Н. Фишинговые атаки как один из самых распространенных видов киберпреступности и меры по противодействию // Научный портал МВД России. 2023. №1 (61). URL: https://cyberleninka.ru/article/n/fishingovye-ataki-kak-odin-iz-samyh-rasprostranennyh-vidov-kiberprestupnosti-i-mery-po-protivodeystviyu.
2. Muminov K., Ziyodjon O. Social Engineering, Human Factor in Cybersecurity // Al-Farg'oniy avlodlari. 2024. №3. URL: https://cyberleninka.ru/article/n/social-engineering-human-factor-in-cybersecurity.
3. Mutaz Abdel Wahed, Enhanced machine learning algorithm for detection and classification of phishing attacks // International Journal of Open Information Technologies. 2025. №1. URL: https://cyberleninka.ru/article/n/enhanced-machine-learning-algorithm-for-detection-and-classification-of-phishing-attacks.
4. Pratikkumar A. Barot, Sunil A. Patel, Jethva H.B. Evaluation of performance measures for reliable and secure phishing detection system // RT&A. 2023. №4 (76). URL: https://cyberleninka.ru/article/n/evaluation-of-performance-measures-for-reliable-and-secure-phishing-detection-system.
5. Корнюхина С. П., Лапонина О. Р. Исследование возможностей алгоритмов глубокого обучения для защиты от фишинговых атак // International Journal of Open Information Technologies. 2023. №6. URL: https://cyberleninka.ru/article/n/issledovanie-vozmozhnostey-algoritmov-glubokogo-obucheniya-dlya-zaschity-ot-fishingovyh-atak.

6. Gylychdurdyyev D., Otuzova B. Using ai to detect fake emails and phishing attacks // Символ науки. 2024. №12-2-1. URL: https://cyberleninka.ru/article/n/using-ai-to-detect-fake-emails-and-phishing-attacks.
7. Saleem Raja A.S., Pradeepa G., Mahalakshmi S., Jayakumar M.S. Natural language based malicious domain detection using machine learning and deep learning // Научно-технический вестник информационных технологий, механики и оптики. 2023. №2. URL: https://cyberleninka.ru/article/n/natural-language-based-malicious-domain-detection-using-machine-learning-and-deep-learning.
8. А А. Отахонов Обнаружение и оценка фишинговых url-адресов с использованием алгоритмов машинного обучения // Al-Farg'oniy avlodlari. 2024. №4. URL: https://cyberleninka.ru/article/n/obnaruzhenie-i-otsenka-fishingovyh-url-adresov-s-ispolzovaniem-algoritmov-mashinnogo-obucheniya.
9. Brij Bhooshan Gupta, Akshat Gaurav, Razaz Waheeb Attar, Varsha Arya, Ahmed Alhomoud, Kwok Tai Chui «Optimized Phishing Detection with Recurrent Neural Network and Whale Optimizer Algorithm», ScienceDirect, 2024. doi: 10.32604/cmc.2024.050815.
10. Abdulla Al-Subaiey, Mohammed Al-Thani, Naser Abdullah Alam, Kaniz Fatema Antora, Amith Khandakar, SM Ashfaq Uz Zaman, Novel interpretable and robust web-based AI platform for phishing email detection, ScienceDirect, 2024. doi: 10.1016/j.compeleceng.2024.109625.
11. Hannousse, Abdelhakim; Yahiouche, Salima «Web page phishing detection», Mendeley Data, V3, 2021. doi: 10.17632/c2gw7fy2j4.3