УДК 004.056:004.8

ВИЯВЛЕННЯ ФІШИНГОВИХ САЙТІВ ЗА ДОПОМОГОЮ МАШИННОГО НАВЧАННЯ

Трохименко Ю.О.
e-mail:yuliia.trokhymenko@nure.ua
Науковий керівник – доц., к.т.н., Філіппенко І.В.
Харківський національний університет радіоелектроніки, каф. ЕОМ
м. Харків, Україна

Phishing attacks are constantly increasing on the internet today. Fraudsters create fake websites imitating banking and payment services in order to obtain users' confidential data. But machine learning can help detect phishing websites faster and more effectively. This paper examines classifiers used to identify new, unlisted threats: Random Forest, XGBoost, and Support Vector Machines (SVM).

Фішинг є однією з найбільш тривожних проблем у світі, що постійно змінюється. Зростаюче використання Інтернету призвело до нового способу крадіжки даних, відомого як кіберзлочинність. Кіберзлочинність означає крадіжку приватної інформації та порушення конфіденційності через комп’ютери. Основною технікою є фішинг. Фішинг через URL-адреси (уніфіковані покажчики ресурсів) є одним із найпоширеніших типів, і його основна мета – викрасти дані у користувача, коли користувач отримує доступ до шкідливого веб-сайту. Виявлення шкідливої URL-адреси є серйозною проблемою.

Традиційні методи виявлення фішингових сайтів ґрунтуються на використанні чорних списків (blacklists) та сигнатурного аналізу. Однак такі підходи мають ряд недоліків, оскільки нові шахрайські сайти можуть з'являтися швидше, ніж оновлюються бази даних. У зв'язку з цим актуальним напрямом досліджень використання методів машинного навчання для автоматичного виявлення фішингових ресурсів.

Методи машинного навчання дозволяють аналізувати різні характеристики веб-сайтів та виявляти ознаки, характерні для сторінок фішингу. Серед ознак, що найчастіше використовуються, можна виділити URL-адрес, структуру HTML-коду сторінки, наявність підозрілих скриптів, а також параметри домену. На основі цих характеристик формується набір ознак, що використовується для навчання моделей класифікації.

Ключові ознаки фішингових посилань:

- незначні відмінності від оригіналу (наприклад, privat24-ua.com замість privat24.ua);
- використання схожих символів, наприклад, латинська «I» замість «l» або цифра «0» замість літери «o»;
- використання імені бренду на початку довгої адреси (наприклад, google.com.security-update.net);
- незвичайна доменна зона, замість .com або .ua використовуються дивні комбінації, особливо у безкоштовних зонах;
- довгі та зашифровані посилання.

У задачах виявлення фішингу широко застосовують такі алгоритми машинного навчання, як Random Forest, Support Vector Machine, XGBoost і та інші. Ці алгоритми здатні ефективно аналізувати великі обсяги даних та виявляти приховані закономірності. Результати сучасних досліджень показують, що використання методів машинного навчання дозволяє досягати точності виявлення фішингових сайтів на рівні 95-99%.

Таблиця 1 - Порівняльний аналіз продуктивності моделей для виявлення фішингу

Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC
Random Forest	97%	0.96%	0.97	0.96	0.98
SVM	96%	0.95%	0.96	0.95	0.97
XGBoost	95%	0.94%	0.95	0.94	0.96

Тому модель машинного навчання може бути створена за допомогою всіх алгоритмів, розглянутих вище, а для тестування та навчання модель машинного навчання та 80% набору даних були використані для навчання та 20% для тестування. Random Forest, SVM та інших – це методи машинного навчання, які використовуються для аналізу, щоб визначити, чи є така URL-адреса шахрайською чи ні.

На основі проведеного аналізу можна сказати, що машинне навчання є ефективним інструментом в пошуку та аналізі фішингових сайтів. Розробка інтелектуальних систем виявлення фішингових сайтів дозволяє підвищити рівень захисту користувачів мережі Інтернет та знизити ризики витоку конфіденційної інформації, адже добре навчена модель швидко знаходить шкідливі сайти серед мільйона інших сайтів по всьому Інтернету.

Список використаних джерел:

1. Prashantha G. R., Meghana G. R., Bhoomika P. M., Prajwal C. T., Shivasali Rajendra Babu «Detection Of Phishing Using Machine Learning» URL:https://www.ijert.org/detection-of-phishing-using-machine-learning?utm_ source=chatgpt.com (дата звернення 06.03.2026).

2. SK Hasane Ahammad, Sunil D. Kale, Gopal D. Upadhye, Sandeep Dwarkanath Pande, E Venkatesh Babu, Amol V. Dhumane, Mr. Dilip Kumar Jang Bahadur «Phishing URL detection using machine learning methods» URL: https://www.sciencedirect.com/science/article/abs/pii/S0965997822001892? utm_source=chatgpt.com (дата звернення 07.03.2026).

3. Ozgur Koray Sahingoz, Ebubekir Buber, Onder Demir, Banu Diri «Machine learning based phishing detection from URLs» URL: https://www.sciencedirect.com/science/article/abs/pii/S0957417418306067?utm_source= chatgpt.com (дата звернення 07.03.2026).

Riduana Adneen Adrita Adneen Adrita, Md Mahenur Islam and Md Sabiul Islam «Comparative Analysis of Machine Learning Models for Phishing Website Detection Using URL and Web Features» URL: https://easychair.org/publications/preprint/GC3s/open (дата звернення 08.03.2026).