УДК 004.056
О. Борух; М. Карпінський, Ph.D
(Тернопільський національний технічний університет імені Івана Пулюя, Україна)

РЕАЛІЗАЦІЯ АВТОМАТИЗОВАНОГО ІНСТРУМЕНТУ ДЛЯ ВИЯВЛЕННЯ ФІШИНГОВИХ ВЕБ-САЙТІВ

UDC 004.056
O. Borukh, student; M. Karpinskyi, Ph.D

IMPLEMENTATION OF AN AUTOMATED TOOL FOR DETECTING PHISHING WEBSITES

Фішингові веб-сайти становлять значну загрозу сучасному інформаційному суспільству, обманюючи користувачів і викрадаючи їхні конфіденційні дані. За даними Anti-Phishing Working Group, кількість фішингових атак щорічно зростає, охоплюючи фінансові установи, технологічні компанії та звичайних користувачів. У 2023 році частка фішингових інцидентів перевищила 36% усіх кіберзагроз, демонструючи потребу у впровадженні сучасних технологій для протидії таким атакам.

Сучасні методи боротьби з фішингом, такі як чорні списки, мають обмеження, пов’язані з їхньою реактивністю та неефективністю проти нових загроз. Це обумовлює потребу у впровадженні автоматизованих інструментів, які здатні виявляти нові патерни та мінімізувати затримки у реагуванні.

Машинне навчання відкриває нові можливості для активного та автоматизованого виявлення фішингових веб-сайтів. Алгоритми, такі як Random Forest, XGBoost і Support Vector Machines, дозволяють аналізувати структуру URL-адрес, вміст сторінок і поведінкові характеристики для ідентифікації потенційно небезпечних сайтів. У цій роботі розроблено інструмент на основі Python, який інтегрує методи машинного навчання для автоматизованого аналізу веб-ресурсів.

Для навчання моделі використано відкриті бази даних, зокрема PhishTank та OpenPhish, які містять актуальні фішингові сайти, а також легітимні ресурси з Alexa Top Sites. Попередня обробка даних включала виділення ключових ознак: довжина URL, наявність IP-адреси в домені, кількість спеціальних символів тощо. Модель була оцінена за допомогою метрик точності, повноти та F1-score.

Тестування продемонструвало високу точність інструменту, що перевершує ефективність традиційних методів, таких як чорні списки. Алгоритм Random Forest показав найкращі результати для задач класифікації URL-адрес. Додатково було протестовано XGBoost, який забезпечив найкращу швидкодію при роботі з великими обсягами даних.

Запропонований інструмент може бути використаний як окремий захисний механізм або інтегрований у загальні системи кібербезпеки, забезпечуючи своєчасне виявлення та блокування загроз. Для досягнення найвищої ефективності рекомендується регулярне оновлення навчальних даних та впровадження додаткових заходів, таких як багатофакторна автентифікація. Подальші дослідження можуть включати використання нейронних мереж для покращення точності, а також створення інтерактивного інтерфейсу для аналітиків із кібербезпеки.

Література

1. URL: https://phishtank.org/
2. URL: https://openphish.com/
3. Breiman L. Random Forests. Machine Learning, 2001.
4. Jason Brownlee. Machine Learning Mastery with Python, 2022.
5. Anti-Phishing Working Group (APWG) Report, 2023.