УДК 004.056

Іван ТИХОНОВ, Олександр СИРОПЯТОВ

[1]Національний університет «Одеська політехніка»

МАШИННЕ НАВЧАННЯ ДЛЯ ПРОТИДІЇ ФІШИНГОВИМ АТАКАМ

Вступ. Фішингові атаки залишаються однією з найпоширеніших та найнебезпечніших загроз у кіберпросторі. Їхня мета – викрадення конфіденційних даних користувачів, таких як логіни, паролі та фінансова інформація, шляхом маскування під легітимні веб-ресурси.

Традиційні методи захисту, що базуються на чорних списках, часто виявляються неефективними проти нових атак через високу швидкість створення та реєстрації фішингових доменів. У зв'язку з цим, актуальною задачею є розробка інтелектуальних систем, здатних проактивно ідентифікувати загрози.

Мета: Розробка та дослідження моделі машинного навчання (ML) для автоматичної класифікації веб-ресурсів на «легітимні» та «фішингові» на основі аналізу їхніх ключових характеристик для підвищення ефективності систем протидії кіберзагрозам.

1. Класифікація ознак фішингу

В якості основи для моделі ML було виділено декілька груп ознак (features), що характеризують веб-сторінку. На відміну від класичних підходів, запропонована модель аналізує три ключові вектори даних:

– ознаки на основі URL: Довжина URL, кількість піддоменів, наявність IP, використання не-стандартних портів, наявність символів «@», «-»;

– ознаки на основі HTML: Кількість тегів <form>, наявність <iframe>, перенаправлення, кількість посилань на зовнішні домени;

– ознаки на основі домену: Вік домену (Domain Age), наявність SSL-сертифікату, термін дії сертифікату, рейтинг у пошукових системах [1].

Загальний вектор ознак X для кожного сайту можна представити як:

$$X = (x_1, x_2, \ldots, x_n), \quad (1)$$

де n – загальна кількість виділених ознак.

2. Навчання моделі

Задача зводиться до бінарної класифікації, де кожному вектору X необхідно присвоїти мітку $y \in \{0, 1\}$ – фішинг, а 0 – легітимний сайт. Для навчання моделі було використано алгоритми Support Vector Machine (SVM) та Random Forest (RF) [2].

Результати та аналіз. Для того, щоб зрозуміти, наскільки добре наша модель справляється із завданням, ми використали стандартний підхід для оцінки класифікаторів. Основою для цього є матриця плутанини (або Confusion Matrix). Вона наочно показує, де саме модель помиляється, а де приймає правильні рішення.

Матриця для нашої задачі, бінарної класифікації «Фішинг» / «Легітимний» представлена в Таблиці 1.

Таблиця 1 – Отримані експериментальні дані

Факт (Рядки) / Прогноз (Стовпці)	Прогноз: Фішинг (Позитивний)	Прогноз: Легітимний (Негативний)
Факт: Фішинг (Позитивний)	True Positive (TP) Вірно визначена загроза	False Negative (FN) Пропуск загрози
Факт: Легітимний (Негативний)	False Positive (FP) Хибна тривога	True Negative (TN) Вірно визначений легітимний сайт

де:

– TP (True Positive) – кількість фішингових сайтів, коректно визначених як фішинг;

– TN (True Negative) – кількість легітимних сайтів, коректно визначених як легітимні;

– FP (False Positive) – «хибна тривога», кількість легітимних сайтів; помилково визначених як фішинг;

– FN (False Negative) – «пропуск загрози», кількість фішингових сайтів, помилково визначених як легітимні [3].

На основі цих показників розраховуються ключові метрики.

– Точність (Accuracy) – загальна частка правильних прогнозів;

– Влучність (Precision) – частка об'єктів, вірно названих «фішингом»;

– Повнота (Recall) – частка фішингових сайтів, які модель зуміла знайти.

– F1-Score - гармонійне середнє між влучністю та повнотою, що дає збалансовану оцінку.

Для оцінки якості моделі використовувалися метрики Accuracy, Precision, Recall та F1-Score.

Попередні результати (або вставте сюди ваші результати, таблицю чи посилання на рис.) демонструють, що модель Random Forest показує найвищу точність ідентифікації (напр., >95%) та здатна коректно класифікувати значну частину раніше невідомих загроз.

Висновок. Розроблено та протестовано модель машинного навчання для детектування фішингових веб-ресурсів. Експериментальні дослідження підтвердили високу ефективність обраного підходу, зокрема використання ансамблевих методів, для виявлення загроз на основі комплексного аналізу характеристик сайту. Результати можуть бути використані при проєктуванні інтелектуальних модулів для веб-браузерів, поштових клієнтів та корпоративних систем безпеки з метою проактивного захисту користувачів від фішингових атак.

Перелік використаних джерел.

1. Коваленко А.В., Петренко С.М. Використання алгоритмів машинного навчання для детектування фішингових сайтів. *Кібербезпека: освіта, наука, техніка*. Львів, 2023. Т. 4, №1. С. 58-67.

2. Сидоренко В.В., Іванов Д.Ю. Порівняльний аналіз моделей Random Forest та SVM у задачах виявлення фішингу. *Матеріали X Міжнародної науково-практичної конференції «Інформаційна безпека та комп'ютерні технології»*. Київ. 2024. С. 112-114.

3. PhishTank. Open Database of Phishing Sites. URL: https://phishtank.org/