УДК: 004.056 DOI 10.33111/mise.103.18

Шкоденко Т.В.,
аспірант кафедри системного аналізу та кібербезпеки,
КНЕУ ім.Вадима Гетьмана

Shkodenko T.V.,
graduate student of the department system analysis and cyber security, KNEU named after Vadim Hetman

ПРО ЕКОНОМЕТРИЧНУ МОДЕЛЬ ОЦІНЮВАННЯ СИСТЕМИ ЗАХИСТУ ЕЛЕКТРОННОГО БІЗНЕСУ

ABOUT ECONOMETRIC MODEL FOR THE EVALUATION OF AN ELECTRONIC BUSINESS PROTECTION SYSTEM

Анотація. *Розвиток цифрових технологій стає невід'ємною частиною електронного бізнесу, а безпека інформації перетворюється на ключовий елемент, який забезпечує довіру клієнтів та стабільність компанії. Тому захист конфіденційних даних у електронному бізнесі являється квінтесенцією, яка визначає здатність підприємства протидіяти загрозам кібербезпеки та зберегти свою репутацію. В статті представлені матеріали дослідження і формування моделі машинного навчання ідентифікації фішингових вебсайтів для оцінювання систем захисту інформації в сегменті електронного бізнесу. Оскільки фішинг, як одна із основних небезпечних кіберзагроз, може призвести до значних фінансових втрат користувачів і організацій, то застосування моделі із заздалегідь оптимально підібраними параметрами, дає можливість оцінити стан і дії легітимних сайтів від фішингових з високим ступенем точності. Така точність критично важлива у випадку застосування моделі в реальних умовах, оскільки помилкова ідентифікація може призвести до невиправданих фінансових і репутаційних витрат. Шляхом аналізу результатів моделювання, своєчасного виявлення та блокування фішингових вебсайтів, компанія забезпечує захист особистих даних користувачів і корпоративної інформації, тим самим знижуючи потенційні втрати від шахрайства.*
Ключові слова: *моделювання, вебсайт, фішинговий ресурс, електронний бізнес, модель, машинна модель, регресія, алгоритм.*

Abstract. *The development of digital technologies is becoming an integral part of e-business, and information security is turning into a key element that ensures customer trust and company stability. Therefore, the protection of confidential data in electronic business is the quintessence that determines the ability of the enterprise to counter cyber security threats and preserve its reputation. The article presents research materials and the formation of a machine learning model for identifying phishing websites for evaluating information protection systems in the e-business segment. Since phishing, as one of the main and dangerous cyber threats, can lead to significant financial losses for users and organizations, the use of a model with previously optimally selected parameters makes it possible to assess the state and actions of legitimate sites from phishing ones with a high degree of accuracy. Such*

accuracy is critically important in the case of applying the model in real conditions, since false identification can lead to unjustified financial and reputational costs. By analyzing simulation results, timely detection and blocking of phishing websites, the company ensures protection of users' personal data and corporate information, thereby reducing potential losses from fraud.
Keywords: *modeling, website, phishing resource, e-business, model, machine model, regression, algorithm.*

Вступ. Сучасний рівень розвитку технологій таких як RPA, AI, ML та BigData дозволяє створювати комплексні IT-рішення для оптимізації бізнес-процесів в різноманітних рівнях не тільки великого, але і в сегментах середнього і малого електронного бізнесу. Режим реального часу створює умови для отримання доступу до власних операцій, які відкривають нові можливості для аналітики, зростання та побудови більш потужних зв'язків з клієнтами. Так, за експертними оцінками до 2024 року прогнозується за рахунок перебудови операційних процесів в електронному бізнесі в різних галузях економіки знизити видатки до 30%. Цей результат може бути досягнутий за рахунок економії трудозатрат і часу зменшення кількості помилок, прозорості бізнес-процесів, залучення штучного інтелекту, розвинутої системи захисту інформації.

Постановка проблеми. Забезпечення відповідного рівня захисту інформаційних ресурсів у сфері електронного бізнесу та їхньої інформаційної безпеки потребує точної оцінки параметрів інформаційних систем, які визначають їх функціонування при впливі різних інформаційних загроз. З одного боку, великий клас таких параметрів являються характеристиками, які описуються виключно якісно (без чітких числових значень) і найскладніша серед них — оцінка уразливостей інформаційної системи. З іншої сторони, в умовах протидії інформаційної системи групам зловмисників потрібне застосування великої кількості механізмів захисту, які мають різну природу, умови функціонування та орієнтовані на захист інформаційних ресурсів різноманітних видів. Все зазначене ставить завдання комплексної оцінки як самої інформаційної системи, так і системи захисту інформації. Внаслідок складності формалізації деяких процесів, які протікають у інформаційному просторі та залежать від людського фактора (наприклад, потенційні можливості й дії зловмисника, фішингові сайти), успішне вирішення такого завдання можливе тільки із застосуванням моделювання. Розвиток цифрових технологій стає невід'ємною частиною електронного бізнесу, а безпека інформації перетворюється на ключовий елемент, який забезпечує довіру клієнтів та стабільність компанії. Захист конфіденційних даних у

електронному бізнесі являється квінтесенцією, що визначає здатність підприємства встояти перед загрозами кібербезпеки та зберегти свою репутацію. В сучасних умовах, коли інформація може представлятися новою валютою, ефективний захист даних стає вирішальним фактором для успіху в електронному бізнесі, гарантуючи цілісність та доступність життєво важливих ресурсів. Відповідальне ставлення до інформаційної безпеки в електронному бізнесі не тільки захищає компанію від кіберзагроз, але й служить основою для побудови довгострокових відносин з клієнтами на принципах довіри та надійності. Забезпечення безпеки інформації у електронному бізнесі це не лише технічний аспект, але й стратегічне рішення, яке сприяє зростанню, інноваціям та конкурентоспроможності на ринку. Останні публікації з цього напряму свідчить про постійно зростаючу кількість кібератак та загроз, які пов'язані з витоками даних (Таблиця 1).

Аналіз останніх досліджень і публікацій. Теоретичні аспекти і практичні заходи застосування моделювання були запропоновані в декількох наукових працях. Так, роботи [16,17] наводять приклад Байєсівської мережі на основі поточної моделі графа безпеки. Марківська модель змінної довжини, яка фіксує особливості траєкторій атаки, що дозволяє передбачити ймовірні наступні дії при поточних атаках, проаналізовано в [18]. Слід зазначити, що дані методи враховують лише вразливості в мережі, але не виявляють реальних відмінностей між типами зловмисників. В інших роботах це питання розглядалося шляхом моделювання можливостей опонентів [19] або застосування методології теорії ігор [20] для моделювання нападника і захисника. Жоден із цих методів не моделює зловмисника на основі інформації, яку зловмисник отримує під час атаки, хоча вона відіграє важливу роль у прийнятті рішень щодо самої атаки. Визначення фішингових сайтів виступає важливою сферою кібербезпеки, яка зосереджена на виявленні та зменшенні впливу шахрайських сайтів, призначених для обману користувачів з метою надання конфіденційної інформації, наприклад, паролів або номерів кредитних карток. В цій галузі на сьогоднішній день найчастіше розглядають різноманітні методи машинного навчання та штучного інтелекту для автоматичного виявлення подібних шкідливих сайтів[10-15]. Основні підходи до визначення фішінгових сайтів включають такі кроки: аналіз вмісту веб-сайтів, структури URL-адрес та інших параметрів (наявність субдомену) та метаданих на наявність ознак фішингу. Методи ж машинного моделювання включають алгоритми глибокого вивчення, які здатні вчитися та гнучко адаптуватися до еволюції фішингових стратегій.

Таблиця 1

СТАТИСТИЧНІ ДАНІ ВПЛИВУ КІБЕРЗАГРОЗ НА ДІЯЛЬНІСТЬ ОРГАНІЗАЦІЙ

Дата	Організація, на яку це вплинуло	Деталі порушення	Кількість постраждалих користувачів	Джерело
Листопад 2023 року	Infosys	Подія безпеки, що впливає на американський підрозділ, програми недоступні	Все ще розслідується	[4]
Жовтень 2023 року	Indian Council of Medical Research	Витік медичних даних зміг призвести до витоку тестів на Covid та інших даних про здоров'я 815 мільйонів громадян	Близько 815 мільйонів	[4]
Жовтень 2023 року	Okta	Доступ зловмисника з використанням викрадених облікових даних	Не визначено	[4]
Жовтень 2023 року	23andMe	Атака з підміною облікових даних, що компрометує генетичні дані клієнтів	Не визначено	[4]
Серпень 2023 року	Forever 21	Порушення даних з доступом до імен, дат народження, банківської інформації та номерів соціального страхування	500,000	[4]
Липень 2023 року	Maximus	Витік даних, що зачіпає медичні дані громадян США	8-11 мільйонів	[4]
Червень 2023 року	Bryan Cave/Mondelez	Скомпрометована особиста інформація співробітників	51,110	[4]

Дата	Організація, на яку це вплинуло	Деталі порушення	Кількість постраждалих користувачів	Джерело
Червень 2023 року	Reddit	Загроза від банди вимагачів BlackCat з викраденими конфіденційними даними	Не визначено	[4]
Квітень 2023 року	T-Mobile	Витік даних, що впливає на контактну інформацію клієнтів, ідентифікаційні картки та SSN	800	[4]
Квітень 2023 року	Pizza Hut/KFC (Yum! Brands)	Атака вірусу-вимагача з викриттям персональних даних	Не визначено	[4]
Травень 2023 року	MOVEit	Використана вразливість призвела до значного витоку даних	17,5 мільйонів	[5]
Квітень 2023 року	Enzo Biochem	Атака вірусу-здирника, що впливає на особисту інформацію	2,5 мільйона	[5]
Березень 2023 року	PharMerica	Витік даних, що компрометує особисту інформацію, на яку претендує група вимагачів «Money Message»	6 мільйонів	[5]
2023 (загалом)	Multiple Sectors	Загальна кількість витоків даних та порушених записів	694 порушення, 612,4 млн записів	[5]

Виклад основного матеріалу. Виходячи з аналізу останніх публікацій *[13, 14]* застосування методів економетричного моделювання допомагає у спрощенні та пришвидшенні процесів вимірювання й оцінювання придатності сайту до *використання*, а також підвищує ефективність систем та заходів щодо захисту інформації. Виконати аналіз даних набору пропонується за допомогою [2]. На основі моделі логістичної регресії та її параметрів (Таблиця 2) існує можливість отримати дані з високою точністю результатів: чи являється сайт фішинговим за набором наведених змінних або навпаки визначити, що сайт не є фішинговим.

Таблиця 2

ОПИС ЗМІННИХ ТА ДІЙ ЗА НАБОРОМ ДАНИХ

Змінна	Опис дій
Index	Порядковий номер рядка в наборі даних.
having_IPhaving_IP_Address	Якщо IP-адреса використовується як альтернатива доменному імені в URL-адресі, наприклад, «http://125.98.3.123/fake.html», то користувачі можуть бути впевнені, що хтось намагається викрасти їхню особисту інформацію. Іноді IP-адреси навіть перетворюється на шістнадцятковий код наприклад: «http://0x58.0xCC.0xCA.0x62/2/paypal.ca/index.html».
URLURL_Length	Зловмисники (фішери) можуть використовувати довгі URL-адреси, щоб приховати сумнівну частину в адресному рядку. Наприклад: http://federmacedoadv.com.br/3f/aze/ab51e2e319e51502f416dbe46b773a5e/?cmd=home&dispatch=11004d58f5b74f8dc1e7c2e8dd4105e811004d58f5b74f8dc1e7c2e8dd4105e8@phishing.website.html Для забезпечення точності дослідження, було визначено довжину URL-адреси у наборі даних і виведено середню довжину URL-адреси. Результати показали, що якщо довжина URL більше або дорівнює 54 символам, то URL класифікується як фішинговий. Проаналізувавши набір даних, було визначено 1220 URL-адрес, довжина яких дорівнює 54 або більше символам, що становить 48,8% від загального обсягу цього набору даних. Це правило було оновлене за допомогою методу, заснованого на частоті статистичної вибірки і таким чином, було підвищено його точність.
Shortining_Service	Скорочення URL-адреси — метод у «всесвітній павутині», за допомогою якого URL-адреса може бути значно меншою за довжиною, але переадресовує на потрібну веб-сторінку. Цей процес досягається за допомогою «HTTP-перенаправлення» на коротке доменне ім'я, яке посилається на веб-сторінку з довгою URL-адресою. Наприклад, URL-адресу «http://portal.hud.ac.uk/» можна скоротити до «bit.ly/19DXSk4».

Продовження табл. 2

Змінна	Опис дій
having_At_Symbol	Використання символу «@» в URL-адресі призводить до того, що браузер ігнорує все, що передує символу «@», і реальна адреса часто слідує за символом «@».
double_slash_redirecting	Наявність послідовність символів «//» в URL-адресі означає, що користувач буде перенаправлений на інший веб-сайт. Прикладом таких URL-адрес є: «http://www.legitimate.com//http://www.phishing.com». Було досліджено місце, де з'являється «//». Якщо URL-адреса починається з «HTTP», це означає, що «//» має з'явитися в шостій позиції. Однак, якщо URL-адреса використовує «HTTPS», то «//» має з'явитися в сьомій позиції.
Prefix_Suffix	Символ — рідко використовується в легальних URL-адресах. Фішери, як правило, додають до доменного імені префікси або суфікси, розділені знаком (-), щоб користувачі відчували, що вони мають справу з легальною веб-сторінкою. Наприклад, http://www.Confirme-paypal.com/.
having_Sub_Domain	Припустимо, існує таке посилання: http://www.hud.ac.uk/students/. Доменне ім'я може включати домен верхнього рівня з кодом країни (ccTLD), у даному прикладі це «uk». Частина «ac» — це скорочення від «academic», комбінація «ac.uk» являється домену другого рівня (ДВУ), а «hud» — власна назва домену. Щоб створити правило для вилучення цієї ознаки, по-перше треба прибрати (www.) з URL-адреси, яка фактично є піддоменом сама по собі, по-друге необхідно вилучити (ccTLD).
SSLfinal_State	Наявність HTTPS дуже важлива для створення враження про легітимність веб-сайту, але цього явно недостатньо. Автори [12-15] пропонують перевіряти сертифікат, призначений для HTTPS, включаючи ступінь довіри до емітенту сертифіката та вік сертифіката. Центри сертифікації, які постійно входять до списку найбільш надійних, включають «GeoTrust, GoDaddy, Network Solutions, Thawte, Comodo, Doster і VeriSign». Крім того, тестування набору даних, виявило, що мінімальний вік надійного сертифіката становить від двох і більше років.
Domain_registeration_length	Довіру визивають домени, що сплачені більше ніж на один рік.
Favicon	Наявність зображення favicon на сайті.
port	Номер порту в адресі сайту.

Закінчення табл. 2

Змінна	Опис дій
HTTPS_token	Наявність параметру token.
Request_URL	Адреса посилання URL.
URL_of_Anchor	Адреса якоря в URL (якщо є).
Links_in_tags	Наявність посилань в тегах.
SFH	Наявність на сайті даних щодо розмітки параметру SFH.
Submitting_to_email	Чи відправляє сайт дані користувачам за email?
Abnormal_URL	Чи є адреса URL підозрілою?
Redirect on_mouseover	Чи виконується переадресація при наведенні миші користувачем?
RightClick	Чи заблокована функція натискання на праву кнопку миші на сайті?
popUpWidnow	Чи відкривається на сайті pop up вікна?
Iframe	Чи є на сайті вбудовані теги iframe?
age_of_domainь DNSRecord	Вік DNS запису доменної зони.
web_traffic	Кількість переходів з Інтернет на сайт.
Page_Rank	SEO параметр Page Rank.
Google_Index	SEO параметр Google Index.
Links_pointing_to_page	SEO параметр кількість зовнішніх посилань на сайт.
Statistical Report	Наявність статистичного звіту.
Result	Результат перевірки сайту (значення 1 означає — такий ресурс є фішинговим, а 0 — напроти не є фішинговим).

Процес моделювання відбувався за допомогою алгоритмів бібліотеки **sklearn** [1] та мови програмування **Python** [9], в рамках яких побудовано **логістичну регресію**, також відому як **logit** та **MaxEnt** класифікатор.

В умовах багатокласовості алгоритм навчання використовує схему «один проти решти» (OvR), в якому опція ‘multi_class’ має значення ‘ovr’, і використовує перехресні ентропійні втрати, якщо опція ‘multi_class’ має значення ‘multinomial’. (В поточній версії бібліотеки опцію ‘multinomial’ підтримують лише

розв'язувачі 'lbfgs', 'sag', 'saga' та 'newton-cg'). Цей клас *sklearn.linear_model.LogisticRegression* реалізує регуляризовану логістичну регресію, за допомогою бібліотеки 'liblinear', алгоритмів-розв'язувачів та параметру 'newton-cg', 'sag', 'saga' та 'lbfgs'. Процес регуляризації застосовується за замовчуванням і тому він може обробляти як щільні, так і розріджені вхідні дані. Для оптимальної продуктивності застосовуються C-впорядковані масиви або CSR-матриці, які містять 64-розрядні числа з плаваючою комою; будь-який інший формат вхідних даних буде перетворено і скопійовано. Алгоритми-розв'язувачі та параметр 'newton-cg', 'sag' та 'lbfgs' підтримують лише L2 регуляризацію з первинним формулюванням або без регуляризації. В той же час розв'язувач 'liblinear' підтримує як L1, так і L2 регуляризації з подвійним формулюванням лише для штрафу L2. Регуляризація методу Elastic-Net підтримується лише розв'язувачем 'saga'.

Метод лінійної регресії було обрано для моделювання, оскільки він оптимізує процес вибору визначення сайта до фішингових.

Здатність класифікаційної моделі розрізняти сайти між класами можливо за допомогою *ROC* кривої — графіка (Рис.1).

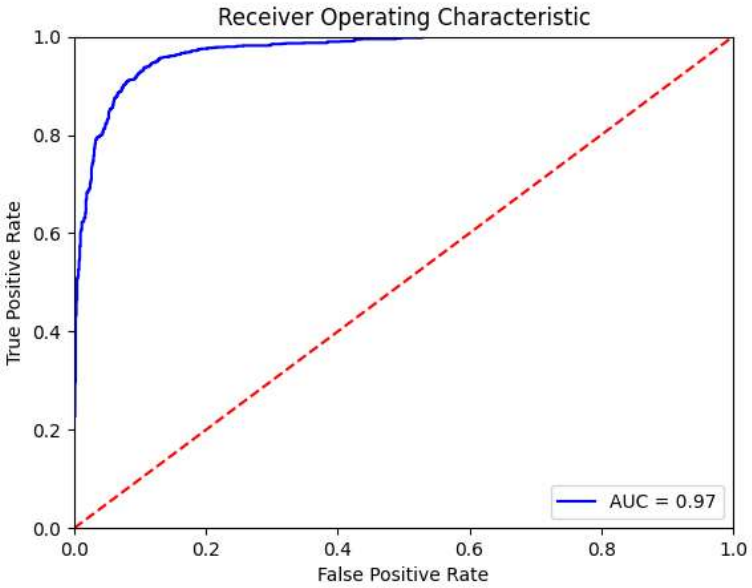

Рис. 1. ROC-крива основної моделі класифікації

В цьому графіку вісь Y (вертикальна) *ROC* кривої показує *Чутливість* (Sensitivity, True Positive Rate), а вісь X (горизонтальна) показує *специфічність* (Specificity, False Positive Rate). *Чутливість* — відсоток справжніх позитивних результатів (правильно ідентифікованих позитивних випадків), а с*пецифічність* відсоток хибних позитивних результатів (неправильно ідентифікованих негативних випадків).

AUC (Area Under the ROC Curve): AUC — площина під ROC кривою, яка вимірюється від 0 до 1, де 1 означає ідеальний класифікатор, а 0.5 — не кращий за випадковий вибір. Вищий показник AUC свідчить про кращу здатність моделі розрізняти між позитивними та негативними класами. Наприклад, AUC, рівний 0.8, означає, що 80% ймовірності того, що модель вибере випадковий позитивний випадок вище за випадковий негативний випадок.

За результатами програмної реалізації моделі (табл. 3) статистично значимими виявилися такі змінні: **5** (double_slash_redirecting) — 0.32, **7** (having_Sub_Domain) — 0.77 та **13** (Request_URL) — 0.71. За параметром Accuracy (точність), найкращим виявився алгоритм **saga** зі значенням **AUC**: 0.97.

Для кожного алгоритму оптимізації у масиві [‘lbfgs’, ‘liblinear’, ‘newton-cg’, ‘newton-cholesky’, ‘sag’, ‘saga’] код виводить AUC для відповідної моделі. Різні алгоритми оптимізації можуть впливати на швидкість навчання та конвергенцію моделі. За результатами моделювання було підібрано наступні оптимальні параметри для моделі: Best parameters: {‘C’: 1, ‘solver’: ‘saga’}

Висновки. Розвиток цифрових технології стає невід’ємною частиною електронного бізнесу, а безпека інформації перетворюється на ключовий елемент, який забезпечує довіру клієнтів та стабільність роботи компанії. Захист конфіденційних даних у електронному бізнесі є квінтесенцією, що визначає здатність підприємства встояти перед загрозами кібербезпеки та зберегти свою репутацію. В сучасних умовах, коли інформацію можна сприймати ніби вона є новою валютою, ефективний захист даних стає вирішальним фактором для успіху в електронному бізнесі, гарантуючи цілісність та доступність життєво важливих ресурсів. Відповідальне ставлення до інформаційної безпеки в електронному бізнесі не тільки захищає компанію від кіберзагроз, але й служить основою для побудови довгострокових відносин з клієнтами на принципах довіри та надійності. Забезпечення безпеки інформації у електронному бізнесі це не лише технічний аспект, але й стратегічне рішення, яке сприяє зростанню, інноваціям та конкурентоспроможності бізнесу на ринку.

Таблиця 3

ФРАГМЕНТ ПРОГРАМНОЇ РЕАЛІЗАЦІЇ МОДЕЛІ

Рядок	Параметр	Значення	Опис
32	confusion_matrix(test_y, y_pred)	[[850 106] [77 1178]]	
33	classification_report(test_y, y_pred)	precision recall f1-score support -1 0.92 0.89 0.90 956 1 0.92 0.94 0.93 1255 accuracy 0.92 2211 macro avg 0.92 0.91 0.92 2211 weighted avg 0.92 0.92 0.92 2211	
34	Accuracy accuracy_score(test_y, y_pred)	0.9172320217096337	
35	F1 score f1_score(test_y, y_pred)	0.9279243796770383	
37	AUC roc_auc_score(y_score=y_pred_p[:,1], y_true=test_y-1)	0.9706129457067129	
48	Significant variable index	5	Індекс статистично значимої змінної набору даних
49	Significant variable value	0.3219159924659166	Значення статистично значимої змінної набору даних
48	Significant variable index	7	Індекс статистично значимої змінної набору даних

Закінчення табл. 3

Рядок	Параметр	Значення	Опис
49	Significant variable value	0.771302758626852	Значення статистично значимої змінної набору даних
48	Significant variable index	13	Індекс статистично значимої змінної набору даних
49	Significant variable value	0.7138059931019559	Значення статистично значимої змінної набору даних
75	roc_auc_score(y_score=y_pred_p[:,1], y_true=test_y)		
	1 — AUC	0.9771724816216305	
	0.5 — AUC	0.9771374752037875	
	0.1 — AUC	0.9770099518245011	
	0.01 — AUC	0.975706379502909	
	0.001 — AUC	0.9706129457067129	
	0.0001 — AUC	0.9619388554568339	
	1e-07 — AUC	0.9564828551901181	
83	roc_auc_score(y_score=y_pred_p[:,1], y_true=test_y)		
	lbfgs — AUC	0.9771724816216305	
	liblinear — AUC	0.9770841320908833	
	newton-cg — AUC	0.977168314190935	
	newton-cholesky — AUC	0.977168314190935	
	sag — AUC	0.9771699811632132	
	saga — AUC	0.9771691476770741	

В статті розроблена модель машинного навчання для ідентифікації фішингових вебсайтів. Оскільки фішинг є одним із основних варіантів сучасних кіберзагроз інформаційним активам, то застосування моделі, заснованої на оптимально підібраним параметрам (в нашому випадку ‘C’: 1, ‘solver’: ‘saga’), демонструє високий рівень точності (AUC: 0.97) завдяки якій можна суттєво знизити рівень небезпеки інформаційних активів. Така точність критично важлива у випадку застосування моделі в реальних умовах, оскільки помилкова ідентифікація може призвести до невиправданих втручань у роботу легітимних вебсайтів або, навпаки, пропуску шкідливих сайтів.

Бібліографічні посилання

1. Опис sklearn.linear_model.LogisticRegression — документація бібліотеки scikit-learn 1.3.2 URL: https://scikit-learn.org/stable/modules/generated/ sklearn.linear_model.LogisticRegression.html
2. Phishing website dataset. [Електронний ресурс]: https://www.kaggle.com/datasets/akashkr/phishing-website-dataset
3. Код на мові Python для статті розміщений на Github. [Електронний ресурс]: https://github.com/podlom/kneu-article-regression
4. [Електронний ресурс]: https://tech.co/
5. [Електронний ресурс]: https://www.udel.edu/
6. Грищук Р.В. Теоретичні основи моделювання процесів нападу на інформацію методами теорій диференціальних ігор та диференціальних перетворень: Монографія / Р. В. Грищук. — Житомир : Рута, 2010. — 280 с.: іл. ISBN 978-617-581-033-0
7. Kirlappos, I. and Sasse, M.A., 2012. Security education against phishing: A modest proposal for a major rethink. IEEE Security and Privacy Magazine, 10(2), pp.24-32.
8. Aggarwaly, A., Rajadesingan, A. and Kumaraguru, P., 2012. PhishAri:Automatic Realtime Phishing Detection on Twitter, Seventh IEEE APWG eCrime Researchers Summit (eCRS), Las Croabas, Puerto Rico, pp. 22-25, Available at: (Accessed 25 November 2016)
9. Завантажити інтерпретатор мови Python 3.* [Електронний ресурс]: https://www.python.org/downloads/
10. Asadullah Safi, Satwinder Singh, A systematic literature review on phishing website detection techniques, Journal of King Saud University — Computer and Information Sciences, Volume 35, Issue 2, 2023, Pages 590-611, ISSN 1319-1578, https://doi.org/10.1016/j.jksuci.2023.01.004 . (https://www.sciencedirect.com/science/article/pii/S1319157823000034)
11. Mahdi Bahaghighat, Majid Ghasemi, Figen Ozen, A high-accuracy phishing website detection method based on machine learning, Journal of Information Security and Applications, Volume 77, 2023, 103553, ISSN

2214-2126, https://doi.org/10.1016/j.jisa.2023.103553 . (https://www.sciencedirect.com/science/article/pii/S2214212623001370)

12. L. Jovanovic et al., «Improving Phishing Website Detection using a Hybrid Two-level Framework for Feature Selection and XGBoost Tuning,» in Journal of Web Engineering, vol. 22, no. 3, pp. 543-574, May 2023, doi: 10.13052/jwe1540-9589.2237.

13. R. Zieni, L. Massari and M. C. Calzarossa, «Phishing or Not Phishing? A Survey on the Detection of Phishing Websites,» in IEEE Access, vol. 11, pp. 18499-18519, 2023, doi: 10.1109/ACCESS.2023.3247135.

14. Alazaidah Raed, Al-Shaikh Ala'a, Almousa Mohammad, Khafajeh Hayel, Samara Ghassan, Alzyoud Mazen, Al-shanableh Najah, Almatarneh Sattam, Publication date: 2024/01/01, Page 119, Website Phishing Detection Using Machine Learning Techniques, Volume: 13, DOI: 10.18576/jsap/130108, Journal: Journal of Statistics Applications & Probability

15. Adebowale Moruf Akin, Lwin Khin T., Hossain M. A., Intelligent phishing detection scheme using deep learning algorithms, Publication date: 2024/01/14, DOI: 10.1108/JEIM-01-2020-0036, URL: https://doi.org/10.1108/JEIM-01-2020-0036

16. Qin, X., Lee, W. (2004) Attack plan recognition and prediction using casual networks. Proceeding of 20th Annual Computer Security Applications Conference. Tucson, 370-379. Doi: http://doi.org/10.1109/ scac.2004.7

17. Xie, P., Li, J. H., Ou, X., Liu, P., Levy, R. (2010). Using bayesian networks for cyber security analysis. Proceedings of 2010 IEEE/IFIP International Conference on Dependable Systems and Networks (DSN). Chicago, 211-220. Doi: http://doi.org/10.1109/dsn.2010.5544924

18. Fava, D.S., Byers, S.R., Yang, S.J. (2008). Projecting Cyberattacks Through Variable-Length Markov Models. IEEE Transactions on Information Forensics and Security, 3 (3), 359-369. Doi: http://doi.org/10.1109/tifs.2008.924605

19. Stotz, A., Sudit, M. (2007). Information fusion engine for real-time decisionmaking: A perceptual system for cyber attack tracking. Proceedings of 2007 10th International Conference on Information Fusion. Quebec, 1-8. Doi: http://doi.org/10.1109/icif.2007.4408113

20. Wang, B., Cai, J., Zhang, S., Li, J. (2010). A network security assessment model based on attack-defence game theory. Proceedings of 2010 International Conference on Computer Application and System Modeling (ICCASM). Taiyuan, 3, V3-639. Doi: http://doi.org/10.1109/ iccasm.2010.5620536

Статтю подано до редакції: 22.11.2023