Вінницький національний технічний університет
Факультет інофрмаційних технологій та комп'ютерної інженерії
Кафедра захисту інформації

МАГІСТЕРСЬКА КВАЛІФІКАЦІЙНА РОБОТА

на тему:

«МЕТОД ТА ЗАСІБ ІДЕНТИФІКАЦІЇ ФІШИНГОВИХ АТАК НА ОСНОВІ ШТУЧНОГО ІНТЕЛЕКТУ»

Виконав: студент 2 курсу, групи 1БС-22 м
спеціальності 125 Кібербезпека
_____Володимир ГАРНАГА

Керівник: к. т. н. доц. каф. ЗІ
_____Олеся ВОЙТОВИЧ
«11» 12 2023 р.

Рецензент: к. т. н. доцент каф. ПЗ
_____Олександр ХОШАБА
«13» 12 2023 р.

Допущено до захисту
Завідувач кафедри ЗІ
д. т. н., проф.
_____Володимир ЛУЖЕЦЬКИЙ
«14» 12 2023 р.

Вінниця ВНТУ – 2023 року

Вінницький національний технічний університет
Факультет – Інформаційних технологій та комп'ютерної інженерії
Кафедра захисту інформації
Рівень вищої освіти II (магістерський)
Галузь знань – 12 Інформаційні технології
Спеціальність – 125 Кібербезпека
Освітньо-професійна програма – Безпека інформаційних і комунікаційних систем

ЗАТВЕРДЖУЮ
Завідувач кафедри ЗІ,
д. т. н., проф.
_____ Володимир ЛУЖЕЦЬКИЙ
«19» 09 2023 року

З А В Д А Н Н Я
НА МАГІСТЕРСЬКУ КВАЛІФІКАЦІЙНУ РОБОТУ СТУДЕНТУ

Гарназі Володимиру Анатолійовичу

1. Тема роботи: «Метод та засіб ідентифікації фішингових атак на основі штучного інтелекту»
Керівник роботи: Войтович Олеся Петрівна, к. т. н., доцент кафедри ЗІ,
затверджені наказом ректора ВНТУ від 18 вересня 2023 року №247.

2. Строк подання роботи 13 грудня 2023 р.

3. Вихідні дані до роботи:
Відомості про функціонування, навчання та використання нейронних мереж, аналітичні звіти громадських організацій з порушення цифрових прав і свобод громадян, методи та засоби захисту від кібератак, ChatGPT, Python.

4. Зміст розрахунково-пояснювальної записки(перелік питань, які потрібно розробити) Вступ. 1. Техніко-економічне та науково-технічне обґрунтування доцільності досліджень. 2. Аналіз фішингових повідомлень та влаштування нейромереж 3. Метод ідентифікації фішингових атак із використанням штучного інтелекту. 4. Тестування методу виявлення фішингових атак. 5. Економічна частина. Висновки. Список використаних джерел. Додатки. 6. Перелік ілюстративного матеріалу. Види та стратегії фішингових атак. Складові

нейромереж для обробляння великих текстових даних. Метод виявлен фішингових атак із використання штучного інтелекту.

6. Консультанти розділів роботи.

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		завдання видав	завдання прийняв
1	Олеся ВОЙТОВИЧ, к. т. н., доц. каф. ЗІ	19.09	25.09
2	Олеся ВОЙТОВИЧ, к. т. н., доц. каф. ЗІ	19.09	1.10
3	Олеся ВОЙТОВИЧ, к. т. н., доц. каф. ЗІ	19.09	26.10
4	Ольга РАТУШНЯК, к. т. н., доц. каф. ЕПВМ	19.09	27.10

7. Дата видачі завдання 1 вересня 2023 року.

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів магістерської кваліфікаційної роботи	Строк виконання етапів роботи	Примітка
1	Аналіз завдання. Вступ	01.09.2023 – 10.09.2023	
2	Аналіз літературних джерел за напрямком магістерської кваліфікаційної роботи	10.09.2023 – 15.09.2023	
3	Науково-технічне обґрунтування	16.09.2023 – 22.09.2023	
4	Розробка технічного завдання	23.09.2023 – 29.09.2023	
5	Розробка рішень	30.09.2023 – 12.10.2023	
6	Практична реалізація, моделювання, експериментування, результати	14.10.2023 – 10.11.2023	
7	Розробка розділу економічного обґрунтування доцільності розробки	11.11.23 – 17.11.2023	
8	Аналіз виконання ТЗ, висновки	18.11.2023 – 24.11.2023	
9	Оформлення пояснювальної записки	25.11.2023 – 30.11.2023	
10	Попередній захист та доопрацювання МКР	28.11.2023 – 01.12.2023	
11	Перевірка магістерської роботи на наявність плагіату	02.12.2023 – 10.12.2023	
12	Представлення МКР до захисту	11.12.2023 – 14.12.2023	
13	Захист МКР	14.12.2023 – 21.12.2023	

Студент _____ **Володимир ГАРНАГ**
(підпис)

Керівник роботи _____ **Олеся ВОЙТОВИ**

АНОТАЦІЯ

УДК 04.056.54

Гарнага В. Метод та засіб ідентифікації фішингових атак на основі штучного інтелекту. Магістерська кваліфікаційна робота за спеціальністю 125 Кібербезпека, освітня програма – Безпека інформаційних і комунікаційних систем. Вінниця: ВНТУ, 2023. – 93 с.

Укр. Мовою. Бібліогр.: назв 19; рис.: 45; табл.: 11.

Магістерська кваліфікаційна робота присвячена розробці методу та засобів ідентифікації фішингових атак із використанням інструментів штучного інтелекту. У першому розділі здійснено аналіз кіберінцидентів за рік та визначено відповідно кількість фішингових атак і визначено їх особливості. Також у цьому розділі здійснено огляд стратегій фішингових атак та визначено фази їх проведення, проаналізовано сучасні засоби боротьби з ними та здійснено постановку задач дослідження. Водночас, у другому розділі було визначено особливості фішингових повідомлень та URL-адрес, які можуть міститися в них. Також визначено особливості штучного інтелекту при роботі із текстовою інформацією та на основі зібраних відомостей синтезовано алгоритм методу ідентифікації фішингових атак. У третьому розділі показано особливості реалізації скриптів, розроблених для реалізації запропонованого методу, та здійснено аналіз тестових даних, що було отримано в результаті досліджень.

В економічному розділі оцінено витрати на розробку та термін окупності розроблених засобів ідентифікації фішингових атак.

Ключові слова: штучний інтелект, нейромережа, фішинг.

ABSTRACT

Harnaha V. Method and means for identification of phishing attacks based on artificial intelligence. Master's thesis in specialty 125 – Cybersecurity. Vinnitsa: VNTU, 2023. – 93 p.

In Ukrainian language. Bibliographer: 19 titles; fig.: 45; tabl.: 11.

The master's thesis is dedicated to the development of methods and means of identifying phishing attacks using artificial intelligence tools. In the first section, there is an analysis of cyber incidents for the year was carried out and the number of phishing attacks was determined accordingly and their features were determined. In addition, this section reviews the strategies of phishing attacks and defines the phases of their implementation, analyzes modern means of combating them, and sets research objectives. At the same time, in the second section, the features of phishing messages and URLs that may be contained in them were defined. Also, the features of artificial intelligence when working with text information were determined, and an algorithm for identifying phishing attacks was synthesized based on the collected information. The third section shows the peculiarities of the implementation of the scripts developed for the implementation of the proposed method, and the analysis of the test data obtained as a result of the research is carried out.

The economic section estimates the development costs.

Keywords: artificial intelligence, neural network, phishing.

ЗМІСТ

ВСТУП

З початком збройної агресії в Україні, ворог почав використовувати засоби кібернападу у надзвичайних масштабах. Відповідно до інформаційних звітів громадської організації «ПЛАТФОРМА ПРАВ ЛЮДИНИ», щомісячно наша держава піддається понад 20 мільйонам кібер атак, а свого піку вони досягли у жовтні 2022 року, коли загальна кількість зафіксованих кіберзлочинів досягла 65 048 304 з яких на фішинговий тип припало 1 012 392. Особливо небезпечним фішинг є через те, що він не тільки дає можливість викрасти важливі данні користувачів чи державних установ, але й дозволяє встановити приховане шкідливе програмне забезпечення, а це в свою чергу несе надзвичайні загрози національній безпеці країни.

Саме велика кількість атак потребує нових підходів та методів для боротьби з цими кіберзагрозами. З розвитком штучного інтелекту, стало можливим використовувати його для опрацювання великої кількості текстової інформації та створення нових текстових документів, таких наприклад, як різного типу звіти та тези за певними напрямками. Водночас, штучний інтелект можливо використати також як інструмент для боротьби з кібератаками, оскільки він дозволяє в значній мірі виконати певні процедури замість фахівця з кібербезпеки значно швидше і надати йому детальні звіт після виконання заданих дій. При цьому, оскільки на сьогоднішній день штучний інтелект, наприклад, як ChatGPT доступний широкому колу користувачів, це породжує додаткові ризики, що він може бути використаний для створення нових інструментів кібер атак.

Об’єкт дослідження – процес обробки та виявлення фішингових атак.

Предмет дослідження – методи та засоби обробки та виявлення фішингових повідомлень.

Мета кваліфікаційної роботи – збільшення ефективності та швидкодії процедур опрацювання та виявлення фішингових атак за рахунок використання штучного інтелекту.

Методи дослідження – аналіз, спостереження, моделювання, експериментальні дослідження.

Для досягнення мети необхідно:

- здійснити аналіз кібеззагроз у кіберпросторі нашої держави за останні роки;
- виконати аналіз літературних джерел щодо існуючих методів та стратегій фішингових атак та визначити їх види;
- визначити загрози, що несе із собою штучний інтелект з точки зору кібербезпеки.
- запропонувати підхід щодо виявлення фішингових атак із використанням штучного інтелекту.
- здійснити експериментальні дослідження із використанням штучного інтелекту та перевірити результати дослідження на тестових повідомленнях.

Наукова новизна магістерської роботи полягає в тому, що подальшого розвитку набув метод ідентифікації фішингових атак з використанням штучного інтелекту, який на відміну від відомих використовує декілька складових (перевірка URL-адреси за фішинговими сервісами та аналіз текстового повідомлення та URL-адреси за допомогою штучного інтелекту) для визначення фішингових повідомлень, що збільшує точність визначення типу таких повідомлень.

Публікації результатів магістерської кваліфікаційної роботи. Результати магістерської роботи доповідалися на науково-практичній конференції «Молодь в науці: дослідження, проблеми, перспективи 2023» [16].

1 АНАЛІЗ ІНФОРМАЦІЙНИХ ДЖЕРЕЛ

1.1 Актуальність роботи

24 лютого 2022 р. відбулося широкомасштабне вторгнення російської федерації в Україну. При цьому, вторгнення торкнулося не лише територій українських міст, сіл і селищ, а й українського кіберпростору. Ворожа сторона спрямувала чимало зусиль на те, аби не тільки поширювати дезінформацію але й порушити нормальну діяльність об’єктів критичної інфраструктури нашої дежржави, зокрема, в енергетичній та фінансовій галузях, а також у сфері надання державних послуг. Кожен день спеціалісти профільних державних органів разом з українськими кіберволонтерами та союзниками з усього світу ведуть запеклу боротьбу з окупантами та їхніми помічниками у віртуальному світі [1].

Водночас, кожного місяця реєструється велика кількість кіберінцидентів у кіберпросторі України. Всі ці інциденти реєструються Державною службою спеціального зв’язку та захисту інформації України [2], а громадська організація «Платформа прав людини» здійснює аналіз та публікує аналітичні звіти «Війна в цифровому вимірі та права людини». Використовуючи їх дані було створено табл. 1.1 для аналізу тенденцій у напрямах кібрезлочинів.

Таблиця 1.1 – Кількість кібератак на цифрову інфраструктуру

місяць - рік	Загальна кількість	Фішинг	Сканування портів	Dos/ Doss	Спроби експлуатації вразливості	Спам	Шкідливе підключення
1	2	3	4	5	6	7	8
09-2022	25165386	1060940	24308395	1791	639806	708	151597
10-2022	65048304	1012392	64144216	617	683792	361	155519
11-2022	402009	1172572	152374	536	35505	352	149466
12-2022	22028194	275783	21276314	1791	170421	708	151597
01-2023	21215671	179814	19974399	321	328064	449	906048
02-2023	16539759	179814	15536024	414	189540	612	807694

Продовження таблиці 1.1

1	2	3	4	5	6	7	8
03-2023	21671280	6929	19733933	395	165605	541	1764996
04-2023	23760189	10828	23062782	421	162120	205	528369
05-2023	24350197	1765	23910789	402	159714	1636	272383
06-2023	22022437	2216	21192272	393	88572	2173	733982
07-2023	23742415	2076	2909962	334	41700	2607	593157
08-2023	27609238	4510	25768026	317	170811	2366	1662938

З таблиці 1.1 видно, що кожного місяця ворог здійснює велику кількість атак, для простоти аналізу продемонструємо їх загальну кількість по місяцях у вигляді графіку на рис. 1.1.

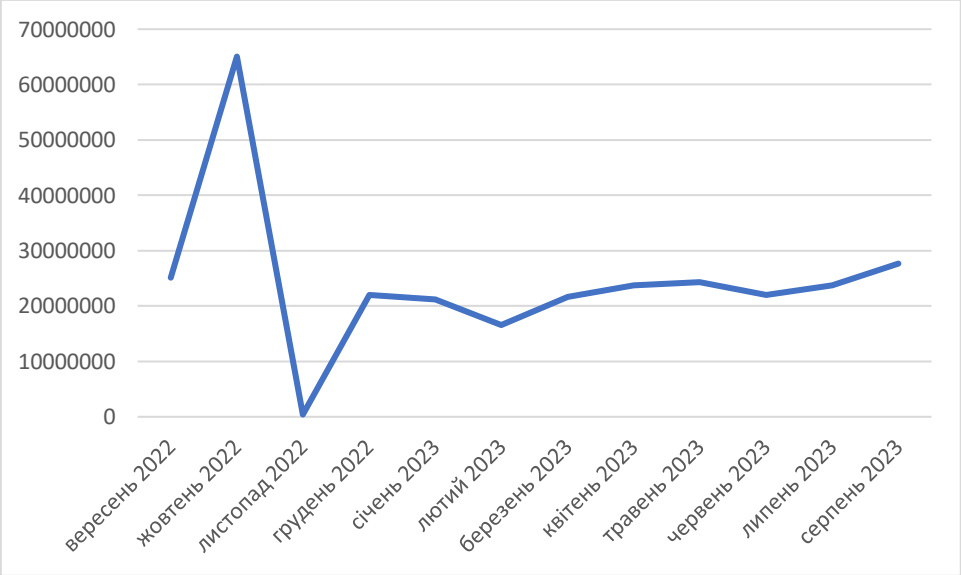


Рисунок 1.1 – Загальна кількість атак по місяцях за один рік

Як видно з графіку кожного місяця здійснюється велика кількість злочинів у нашому кіберпросторі, а найбільша кількість ворожої активності припадає на розвідку, ворог здійснює сканування електронних ресурсів та виконує пошук вразливостей. При цьому до топ 3 найпоширеніших видів атак належать:

1) фішинг;

2) спроби експлуатації вразливості;

3) шкідливе підключення.

Як видно з табл. 1.1, за максимальною кількістю атак за один місяць лідирують фішингові атаки. На рис. 1.2 показано діаграми розподілу фішингових атак по місяцях у 2022 році. При цьому, відповідно до табл. 1.1 їхня кількість суттєво зменшується у 2023 році, проте масштаби нападів вражають, оскільки три місяці підряд їх було понад 1 мільйон на місяць. У 2023 році, такий тип атак значно знизився проте все одно присутній у статистиці.

Рисунок 1.2 – Розподіл фішингових атак у 2022 році

Відповідно до повідомлень у звітах, ці галузі найбільше атакували російські хакери:

- державні та місцеві органи влади;
- інформаційні ресурси сектору безпеки та оборони;
- енергетичний сектор;
- фінансовий сектор;
- комерційний сектор;
- телеком-сектор і розробники;

– транспортна галузь.

Також є інші данні від Української міжбанківської асоціація членів платіжних систем ЄМА, які вони наводять на своїй сторінці [3] за 2022/2023 роки. Відповідно до їх аналізу, маємо такі тренди:

1) найчастіше у вигляді приманки використовується різноманітні грошові виплати від відомих приватних компаній, держави або міжнародних організацій.
2) з року в рік зростає кількість фішингових сайтів з метою отримання банківських даних користувачів та їх облікових даних для онлайн-банкінгу.
3) у 2022 році виявлено більшу кількість фейкових застосунків Google Play та App Store, які видаються як додатки для реалізації залишків пального за цінами нижче ринкових та отримання грошової допомоги від держави. Відповідно збільшилась кількість та різноманітність фейкових банкових чат-ботів у Telegram, що збирають банківську інформацію користувачів.
4) З року в рік фішинг залишається однією за найбільш суттєвих загроз для українських користувачів, і при цьому його масштаби постійно зростають. Так відомо, що 88% від заблокованих сайтів кібермародерів припала на фішинг, а решта 12% – це шахрайські інтернет-крамниці, шахрайські схеми заробітку, шахрайства пов’язані з «інвестиціями» та наданням послуг.

Отже, відповідно до даних асоціації, найвідоміші українські бренди, якими користуються кібершахраї для викрадення цінної інформації українців показані на рис. 1.3

На рис. 1.4 показано розподіл часу необхідного для блокування виявлених шахрайських сайтів. Як видно з наведеного рисунку, близько 76% фішингових сайтів блокується протягом п’яти днів і лише 24% з них можуть залишатися

активними більше часу. У 2022 році Асоціація «ЄМА» ідентифікувала та заблокувала 568 шахрайських сайтів, що складає 95,5% від виявлених загроз. При цьому, найменший час блокування становив 18 хвилин, а найбільший – 32 дні.

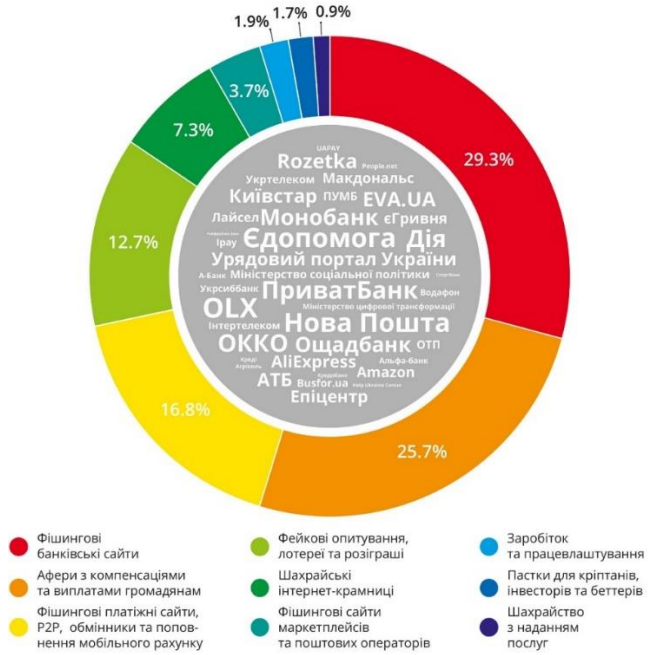


Рисунок 1.3 – Бренди, що використовувалися під час фішингових атак

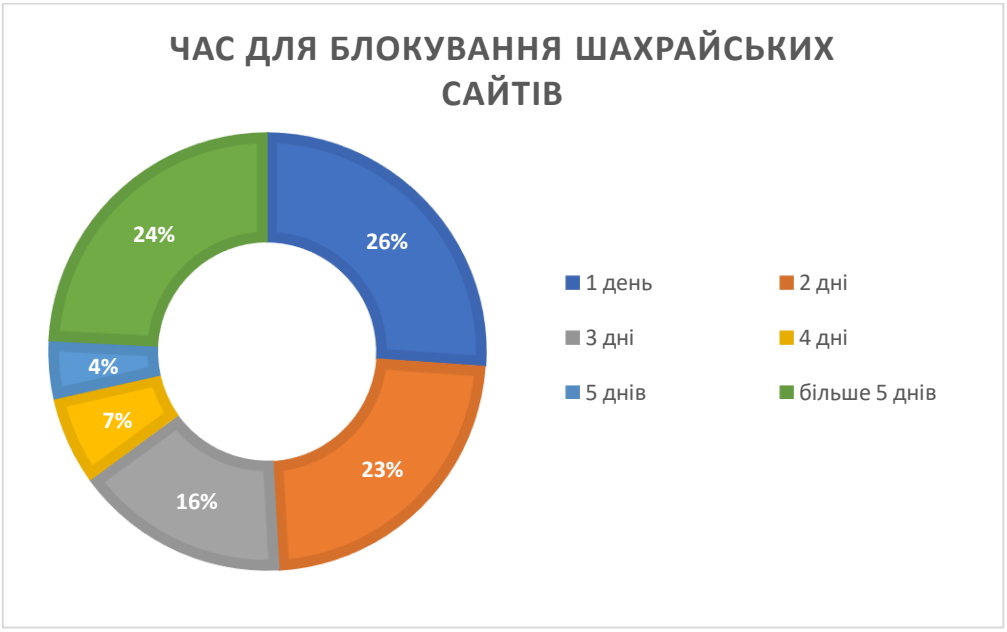


Рисунок 1.4 – Час реагування, необхідний на блокування роботи шахрайських сайтів

Ни рис. 1.5 показано розподіл за походженням реєстраторів фішингових сайтів. Основними країнами походження реєстраторів є США (49,1%) та росія (23,6%), що відповідає світовому тренду. 57,8% шахрайських ресурсів використовують зворотній проксі для сайту від американського сервісу Cloudflare, який допомагає бізнесу захистити свій сайт від атак, а шахраям – заховати свого реального хостинг-провайдера, щоб затягнути блокування шахрайського ресурсу. Водночас, за інформацію Асоціації «ЄВА» деякі реєстратори практично не реагують на скарги, зокрема, Хостинг–Україна не відповіла на жодну з 242 скарг направлених на їх адресу [3].

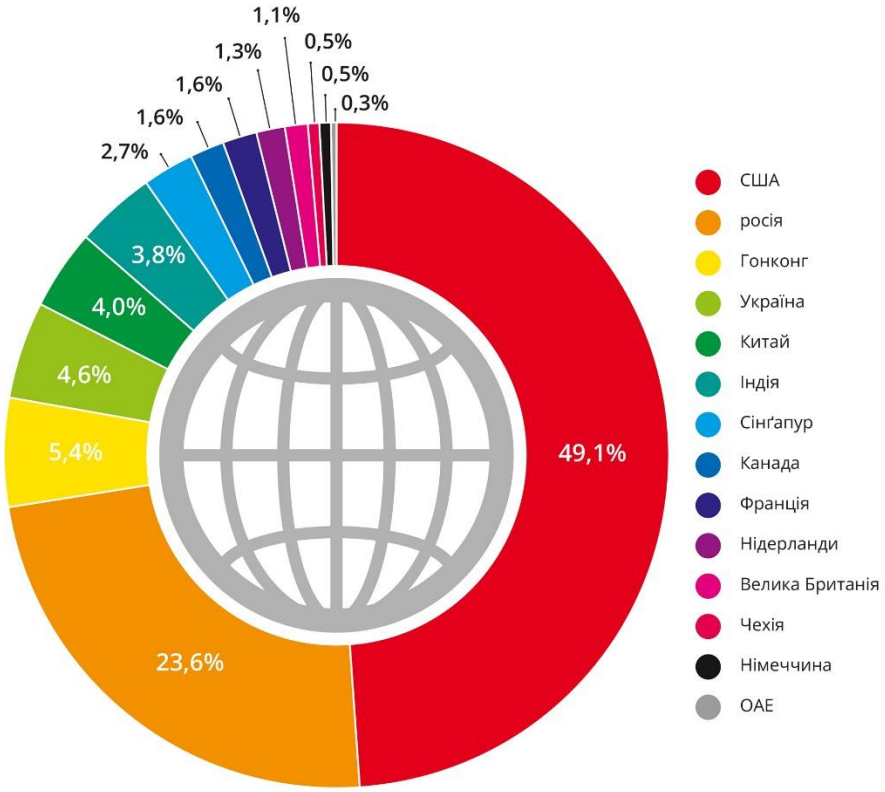


Рисунок 1.5 – Розподіл за походженням реєстраторів фішингових сайтів

На рис. 1.6 представлено розподіл за Top Level Domain (TLD). У топі улюблених шахрайських TLD – ті, на яких відсутні обмеження під час реєстрації:

— «топова» доменна зона .top (24,6%);

— найпопулярніша зона світу .com (20,1%);

— універсальна доменна зона .site.

Через події в нашій державі, досить популярна в минулому доменна зона .ru втратила свою привабливість, ставши маркером злодійського/шахрайського ресурсу, саме тому в 2022 р. її застосовувало лише 1,7% таких сайтів.

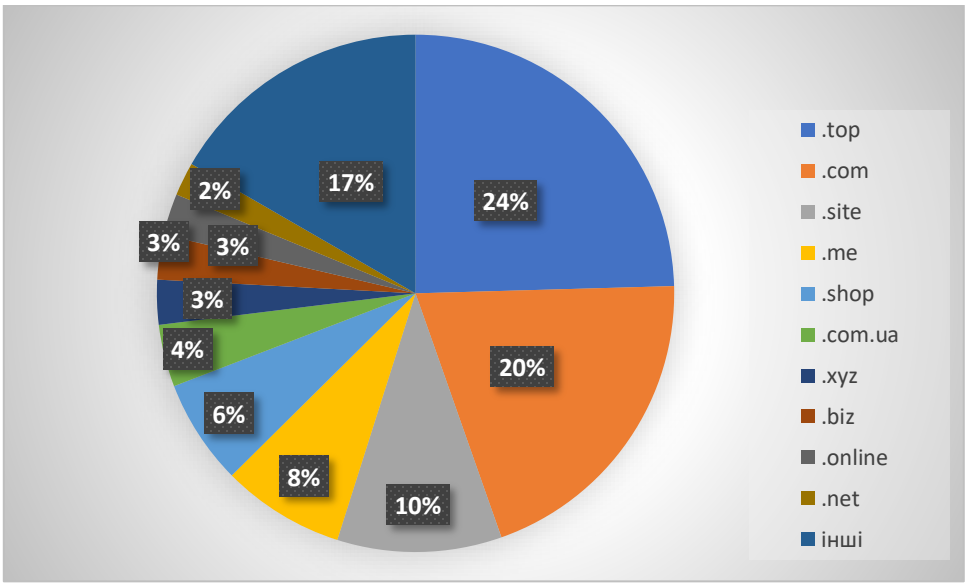


Рисунок 1.6 – Найпопулярніші TLD

1.2 Фішинг, його види та фази фішингових атак

Експерти, дослідники та установи з кібербезпеки запропонували та обговорили різні визначення терміну «фішинг». Хоча не існує усталеного визначення терміну «фішинг» через його постійну еволюцію, він був визначений різними способами на основі його використання та контексту. Змушення одержувача виконати бажану дію зловмисника вважається де-факто визначенням фішингових атак у цілому [5]. Деякі визначення називають веб-сайти єдиним можливим засобом для здійснення атак. Дослідження [4] визначає фішинг таким чином: «шахрайську діяльність, яка передбачає створення копії існуючої веб-сторінки, щоб ввести користувача в оману, для надання особистих, фінансових даних або даних логіну і паролю».

Відповідно до розділу кібербезпеки сайту Microsoft [6] можна виділити 6 видів фішингових атак (рис. 1.7).

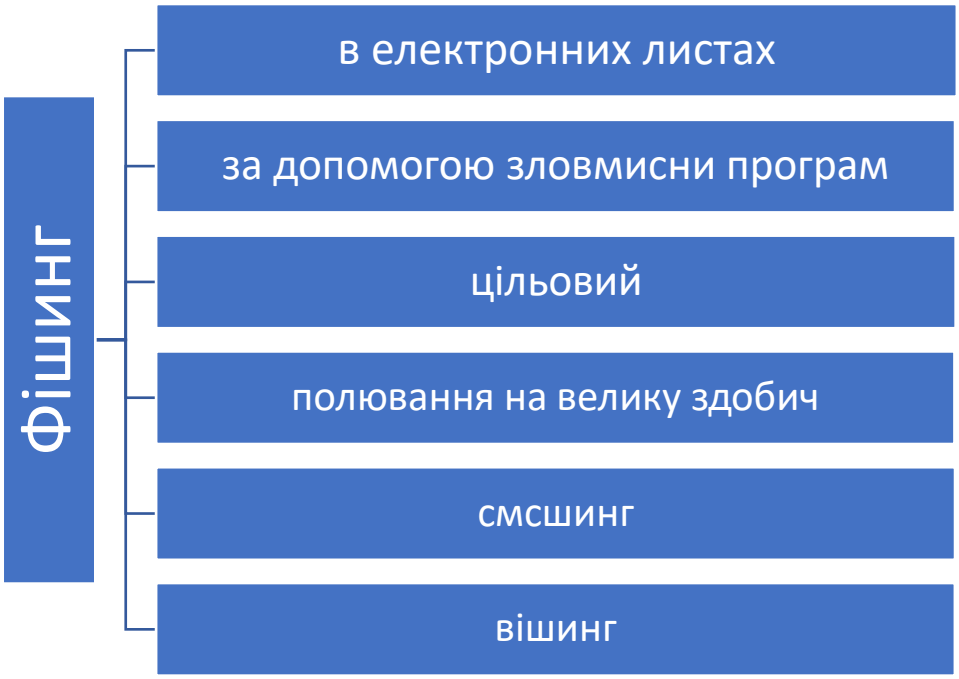


Рисунок 1.7 – види фішингу за класифікацією Microsoft

1) Фішинг в електронних листах. Це одна найпоширеніших форма фішингу, і вектор атаки зловмисники здійснюють, підроблюючи гіперпосилання в електронних листах, щоб обманом примусити користувачів розкрити свої персональні дані або розкрити певну секретну інформацію. Зловмисники можуть видавати себе за великих постачальників облікових записів, як-от Microsoft чи навіть Google, або ж співробітниками певних компаній, які надають певні види послуг(електронна комерція, хостинг, банківські послуги тощо).
2) Фішинг за допомогою зловмисних програм. Другий за популярністю вид фішингу – атака з використанням прихованих шкідливого програмних засобів. Їх видають за такі безпечні, на перший погляд, файли, як резюме або банківська виписка, у додатках до електронних листів. В певних випадках

активізація такого засобу користувачем може вивести з ладу цілі комп’ютерні системи і мережі.

3) Цільовий фішинг. На відміну від попередніх варіантів фішингових атак, жертвами цільового фішингу є певні визначені персони, робочі й особисті дані яких удалося виявити або захопити зловмисникам. Цільовий фішинг налаштовано таким чином, щоб вони могли легко обійти базові технології захисту.
4) Полювання на велику здобич. Таким видам атак підлягають поважні особи, такі, наприклад, як бізнесмени чи знаменитості. Зловмисники детально збирають відомості про своїх жертв, а потім під час слушного моменту викрадають їхні облікові дані або іншу цінну інформацію.
5) Смсшинг. Назва цього виду фішингу була утворена як результат злиття двох слів: "SMS" і "фішинг". Шахраї надсилають підроблені SMS-повідомлення начебто від відомих компаній, як Нова Пошта або Розетка. Люди часто вразливі до таких шахрайських SMS-повідомлень, бо вони написані простою мовою й вони можуть здаватися їм більш особистими та достовірними.
6) Вішинг. Зловмисники, які виконують такого роду атаки, телефонують людям із шахрайських колцентрів/довідкових служб і намагаються виманити в них певну персональну інформацію. Зазвичай шахраї застосовують методи психології та соціальної інженерії, щоб обманом примусити своїх жертв інсталювати шкідливе програмне забезпечення, отримати дані банківського рахунку, здійснити певні дії, що призведуть до фінансових або інших втрат.

Проаналізуємо тепер основні схеми фішингових атак. Базовим компонентом усіх фішингових атак є саме сайт. Найпростішим варіантом є використання проксі-сервера зловмисника. На рис. 1.8 показано схему такої атаки.

Рисунок 1.8 – Структура найпростішої фішинг-атаки

Отже, проаналізуємо яким чином працює такий варіант атаки. Користувач отримує певним чином замаскований URL під якийсь відомий сервіс або, наприклад, електронний магазин. Після цього, він активує це посилання і направляється браузером на проксі-сервер із інтерфейсом, який нагадує або дуже схожий на оригінальний веб-сайт. Відповідно, зазвичай на екрані відображається форма для введення даних користувача або реєстрації користувача із введенням цінних даних. Після заповнення відповідних форм, проксі-сервер зберігає відповідні дані і перенаправляє на реальний сайт, який у свою чергу показує помилку і пропонує, наприклад, спробувати ще раз. Таким чином, користувач може навіть не зрозуміти, що його дані вже відомі стороннім особам.

Залежно від стратегії фішингові атаки, час активності фейкових сайтів може бути різним. На рис. 1.9 показано час активності фішингових сайтів залежно від типу фішингової атаки, які умовно виділяють фахівці.

Традиційна стратегія фішингу полягає в наступному: користувачу передається підроблена URL-адреса і жертва її активує. DNS-сервер дозволяє відповідне посилання, а потім дозволена IP-адреса вказує на сервер зловмисника, з якого перенаправляється на відповідний шкідливий сайт. Такий підхід не дуже складний і вимагає невелику кількість обладнання для проведення атаки. Водночас, недоліком такого варіанту є простота її нейтралізації, достатньо повідомити інтернет-провайдера про відповідну IP-адресу, яка містить фішинговий сайт і вона буде видалена після проведеного розслідування.

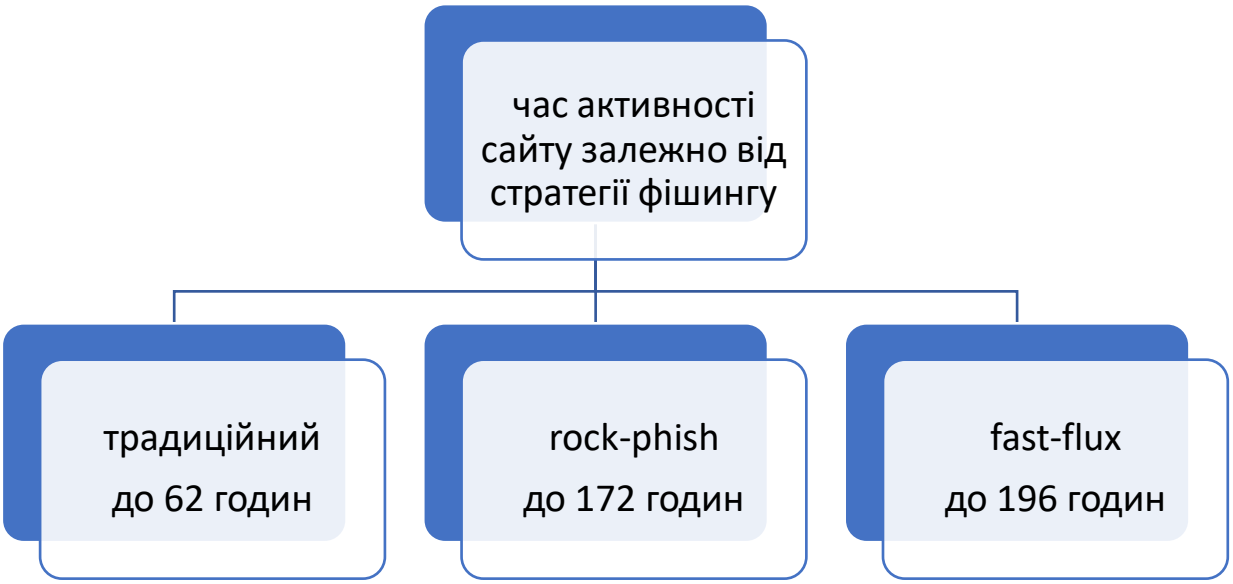


Рисунок 1.9 – Час життя зловмисного сайту, залежно від вибраної стратегії фішингу

Почнемо з традиційного варіанту фішингу. На рис. 1.10 показано її загальну схему.

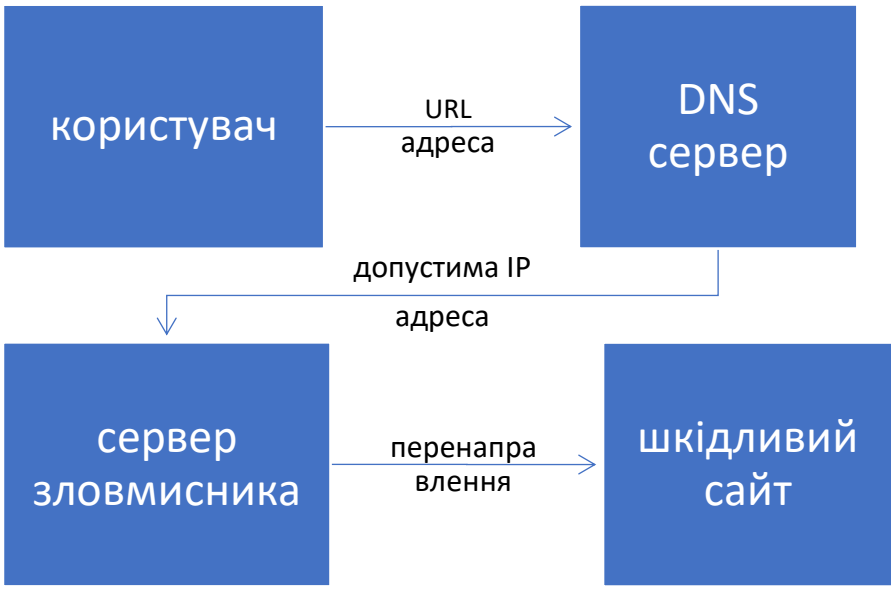


Рисунок 1.10 – Традиційний вид атаки

Розглянемо покращений варіант стратегії, а саме rock-phish (рис. 1.11). Стратегія атаки Rock-phish надає можливість максимально довго уникати

детектування, але при цьому максимізує вразливість фішингового сайту. Порівняно із попереднім варіантом, використання проксі-ботів, дозволяє збільшити час блокування подібної фішингової атаки. Водночас, ботами можуть бути як спеціально налаштовані комп'ютери так і пристрої вражені спеціальним програмним забезпеченням, а їх власники можуть не здогадуватися про те, що їх пристрої використовуються у зловмисних діях.

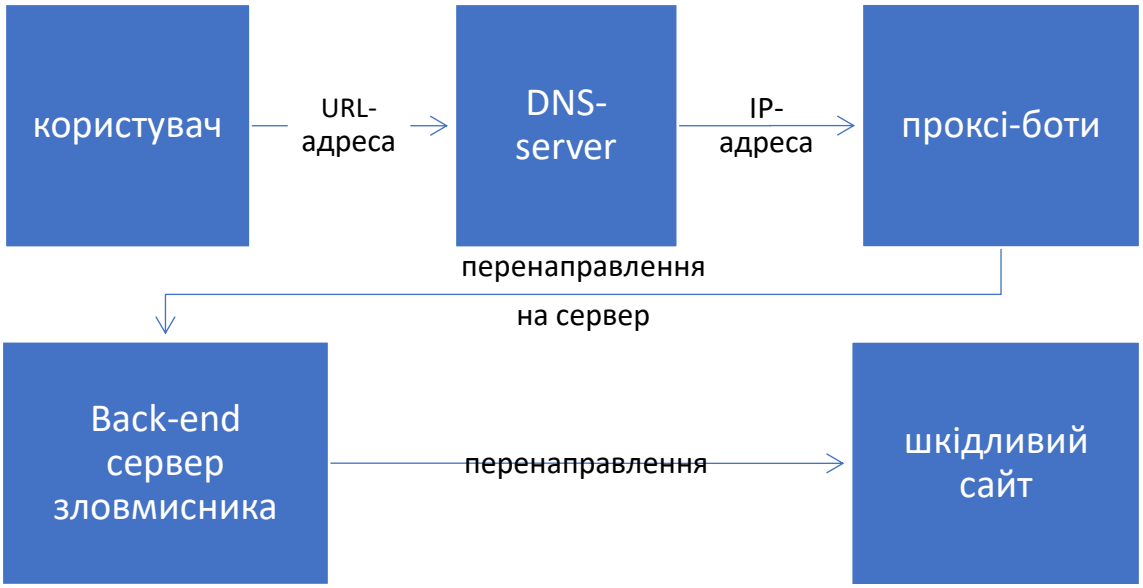


Рисунок 1.11 – Стратегія Rock-phish

Особливістю Rock-phish атаки є можливість не тільки втратити певні цінні дані але й отримати шкідливе програмне забезпечення, що перетворить ваш комп'ютер у бота, який в подальшому може бути використаний для інших атак. Розглянемо наступну стратегію фішингової атаки fast-flux, зображену на рис. 1.12.

Fast flux DNS – це метод, який зловмисник може використовувати для попередньої ідентифікації IP-адреси свого комп'ютера. Шляхом зловживання технологією DNS, кібершахрай може створити мережу ботів з вузлами, підключатися через них і змінювати їх швидше ніж можуть прослідкувати співробітники правоохоронних органів.

Рисунок 1.12 – Стратегія фішингової атаки fast-flux

Fast flux DNS використовує спосіб балансування навантаження, вбудованого в систему доменних імен. DNS дозволяє адміністратору зареєструвати n-е кількість IP-адрес з одним іменем хоста. При цьому, альтернативні адреси законно використовуються для розподілу інтернет-трафіку між кількома серверами. Як правило, IP-адреса, пов'язана з хостом домену, не дуже часто змінюються, якщо взагалі змінюються. Проте зловмисники виявили, що вони можуть закривати ключові сервери, використовуючи 1/62 часу життя (TTL) записи ресурсу DNS, пов'язаного з IP-адресою, і змінювати їх надзвичайно швидко. Оскільки зловживання системою вимагає співпраці реєстратора доменних імен, більшість DNS-ботнетів Fast flux, які пропонуються, проходять у країнах, що розвиваються, або в інших країнах без законів для кібербезпеки.

Розглянемо узагальнену структуру фішингової атаки, поділену на відповідні фази, відповідно до джерела [5], що показано на рис. 1.13. Запропонована структура фішингу детально пояснює кожну з фаз фішингу, включаючи зловмисників і типи цілей, приклади інформації, яку міг би зібрати зловмисник про жертву, і приклади методів атаки. Структура, як показано на рисунку, ілюструє набір вразливостей, якими може скористатися зловмисник, і середовища, що використовуються для здійснення атаки. Також перераховано можливі загрози та метод збору даних для подальшого пояснення та деякі приклади щодо типів реагування на цілі та типів здобичі, яку може отримати зловмисник, і того, як вони можуть використовувати викрадені цінності.

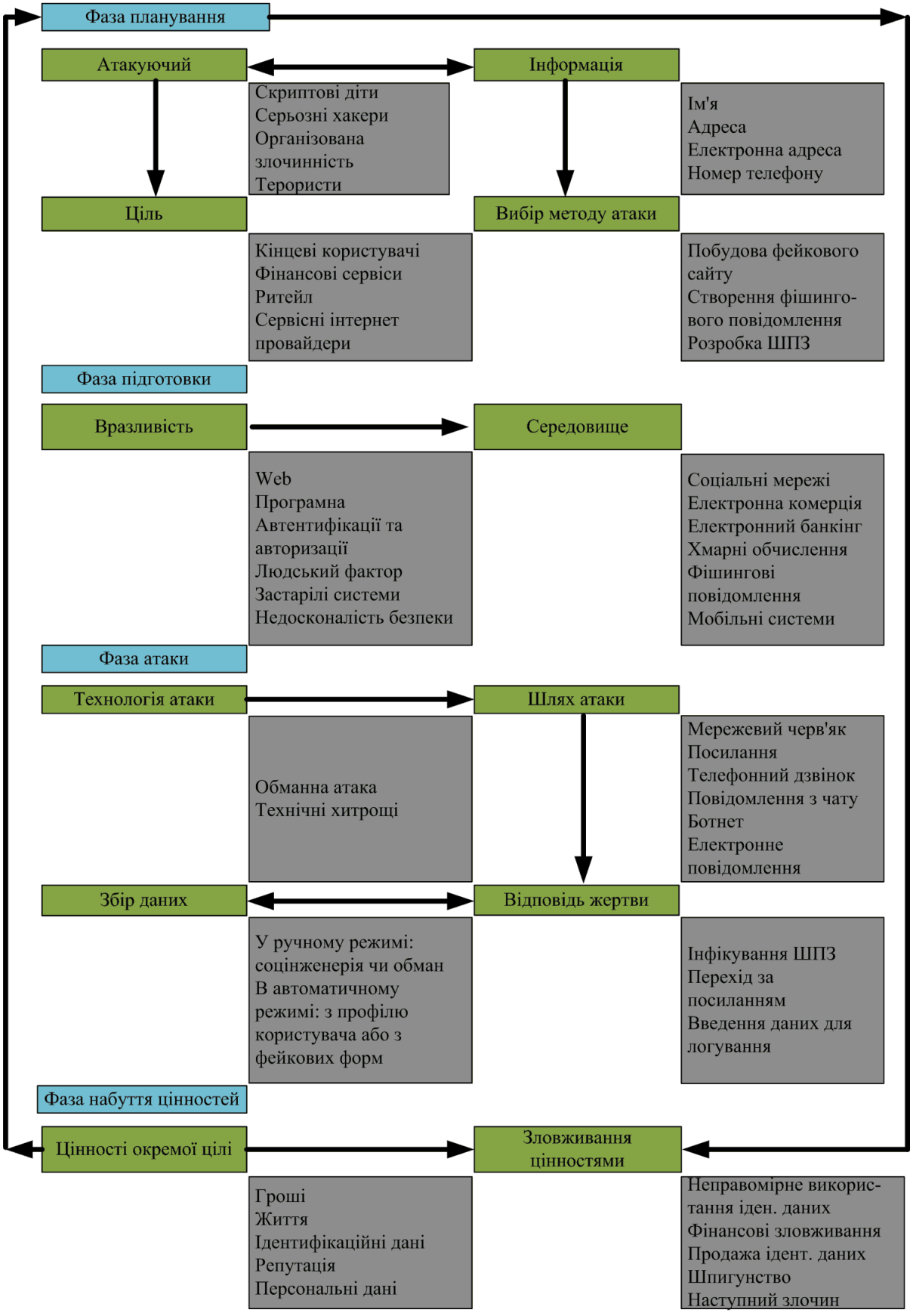


Рисунок 1.13 – Фази фішингової атаки

Ця структура детально розкриває фішингові атаки та допомагає людям краще зрозуміти повний процес фішингу (тобто наскрізний життєвий цикл фішингу) і підвищити обізнаність користувачів. Вона також надає уявлення про потенційні рішення для фішингових атак, на яких повинні зосередитися розробники антифішингового програмного забезпечення.

Замість того, щоб завжди обвинувачувати користувачів як єдину причину успіху фішингу, розробники повинні зосередитися на рішеннях, які пом’якшують ініціацію атаки, запобігаючи потраплянню приманки до користувача. Наприклад, щоб досягти цільової системи, загроза має пройти через багато рівнів технологій або засобів захисту, використовуючи одну або кілька вразливостей, таких як уразливості в Інтернеті та програмному забезпеченні.

1.3 Засоби захисту від фішингу

Проблема фішингу не є новою, оскільки перший випадок фішингу було зафіксовано у 1990 році [5], тому важливо здійснити огляд засобів, що вже були розроблено для захисту від цієї загрози. Здійснений патентний пошук, показав, що у світі вже запатентовано значну кількість засобів захисту. Проаналізуємо деякі з них.

1) Патент "US2021099484A1" описує систему і методи виявлення фішингових сайтів. Суть підходу полягає в наступному: клієнтський пристрій здійснює захоплення зображення, що належить веб-сторінці, до якої здійснюється доступ через клієнтський пристрій, і генерує відбиток веб-сторінки на основі застосування хеш-функції до захопленого зображення. Водночас, для кожного фішингового відбитка у базі даних відомих фішингових сайтів здійснюється порівняння отриманого відбитку із набором відомих фішингових відбитків і залежно від подібності відбитків, пристрій клієнта може здійснювати ідентифікацію потенційних фішингових сторінок.

2) Патент US2023344866 – додаток для виявлення фішингу, суть роботи якого включає моніторинг мережевої активності, пов’язаної з сеансом користувача, ведення розширеної ідентифікації програм та визначення на основі отриманих даних ідентифікації фішингових сайтів [7].

3) Патент US2023291767 (A1) Метод визначення спроб фішингу через е-мейл або шахрайського домену електронної пошти. Метод полягає у визначенні істинності повідомлення від відповідного домену. Воно перевіряється з допомогою додаткових полів ідентифікації [8].

4) Патент US2023344868 (A1) Автоматичне виявлення фішингу. Метод подібний по роботі до першого методу, але він додатково використовує модуль захоплення IP-адрес та перевіряє за базою даних відповідно допустимі адреси [9].

Аналіз пошуку за патентами показав, що боротьба з фішинговими атаками залишається актуальним завданням і кожного року вигадують нові підходи до боротьби з ними [10].

Також, важливо проаналізувати діючі підходи, що застосовуються виробниками мережевого обладнання таких, наприклад, як Cisco. Так у згаданій компанії існує продукт Cisco Email Security. На рис. 1.14 показано принцип роботи системи. Це передова система безпеки для електронної пошти, яка складається з програмно-апаратного комплексу [11].

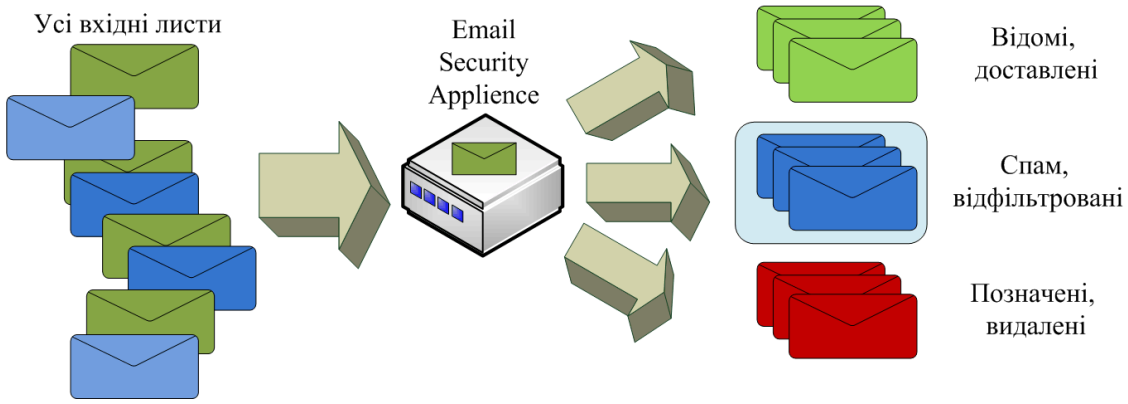


Рисунок 1.14 – Принцип роботи Cisco Email Security Applience

До основних можливостей системи CESA належать такі:

1) Технології Global Threat Intelligence. Cisco Talos забезпечує цілодобове спостереження за активністю трафіку, аналізує аномалії, виявляє нові загрози, а також слідкує за тенденціями трафіку. Також цей інструмент допомагає уникнути раніше невідомих атак, постійно генеруючи правила, засновані на оновленні приладів безпеки. Ці оновлення проходять кожні три-п'ять хвилин, забезпечуючи галузевий захист від загроз.

2) Блокування спаму. Cisco використовує багатошаровий захист як за допомогою зовнішнього (на основі репутації відправника), так і внутрішнього шару фільтрації (для глибокого аналізу повідомлень). Завдяки такому підходу блокується до 80% спаму, використовується контекстний аналіз та автоматична класифікація повідомлень. Можливість використовувати фільтрів, заснованих на ідентифікаторах відправників або тем.

3) Graymail виявлення та безпечна звітність. Захист від розсилок та безпечне керування ними.

4) Розширений захист від шкідливих програм. ESA також оснащено функцією блокування на основі файлової репутації, файловою «пісочницею», а також файловою ретроспективою для безперервного аналізу загроз (навіть після того, як вони перетнули шлюз електронної пошти)

5) Мережеві фільтри. Дані фільтри захищають від загроз і змішаних атак.

6) Інтерактивне веб-відстеження. Це повністю інтегроване рішення, яке дозволяє мережевим адміністраторам відслідковувати кінцевих користувачів із URL, внесеними до списків ESA.

7) Контроль за повідомленнями, що надсилаються.

8) Cisco E-mail Security Appliances забезпечує керування вихідними повідомленнями через DLP, шифрування електронної пошти, а також можливість

інтеграції з RSA Enterprise Manager. Цей контроль гарантує відповідність повідомлень галузевим стандартам та належний захист під час їх пересилання.

Таким чином, вже існують певні засоби захисту від спаму та різного роду кібератак, проте жоден з них не дає 100% захисту.

1.4 Постановка задач

Відповідно до проведеного огляду зрозуміло, що боротьба з кіберзагрозами, зокрема, фішингом залишається актуальним не зважаючи на те, що вже було створено велику кількість засобів боротьби. Водночас, зловмисники також не перестають вигадувати нові підходи та методи викрадення важливої інформації і їхніми жертвами стають іноді навіть спеціалісти з кібербезпеки. Таким чином, тема створення нових засобів направлених на боротьбу з фішингом є актуально.

Таким чином для досягнення мети даної магістерської роботи потрібно виконати завдання:

1) здійснити огляд нових підходів для боротьби з фішингом на основі використання штучного інтелекту;
2) удосконалити метод ідентифікації фішингових атак використовуючи засоби штучного інтелекту;
3) провести тестування класифікаторів на контрольних прикладах та проаналізувати результати підходу і вибрати найкращий варіант.

2 РОЗРОБКА МЕТОДІВ

2.1 Аналіз ознак фішингових повідомлень

Для визначення чи є певне повідомлення фішинговим чи ні, потрібно визначити чіткі критерії для прийняття рішення, потрібно виділити набір ознак, які варто аналізувати при опрацюванні таких повідомлень, щоб у подальшому їх використати для навчання нейромережі.

Першою ознакою яку потрібно аналізувати – є довжина URL-адреси. Відповідно до інформації сайту Microsoft [11], якщо використовується метод GET, максимальна кількість символів у фактичному шляху не може перевищувати 2048 символів. Водночас, для методу POST не обмежується розмір URL-адреси для надсилання пар імен і значень, оскільки вони передаються у заголовку. Проте браузери мають свої обмеження і, наприклад, максимальна довжина для уніфікованого локатора ресурсів у Microsoft Internet Explorer – це 2083 символи. При цьому, деякі дослідники [12] пропонують використовувати умову, що така адреса має не перевищувати 75 символів, проте вони не наводять жодних аргументів стосовно цього критерію. Водночас, невелика довжина URL-адреси не є гарантією безпечної адреси. Так, за допомогою сервісу «Phistank», створеного з метою зменшення кількості фішингових атак та об’єднання зусиль користувачів для цього, було виявлено фішинговий сайт, що містить URL-адресу у 26 символів. На рис. 2.1 показано результати перевірки згаданого сайту. Приклади довгих URL-адрес показано на рис. 2.2 а, б, довжини яких відповідно 126 і 145 символів. При цьому, іноді певні сервіси для реєстрації нових користувачів (посилання для активації) або у випадку, коли вам долучають до певного проекту, довжина посилання може бути більше 600 символів (рис. 2.3). Тому, на моє переконання, довжину адреси варто враховувати, проте це не має бути головною ознакою.

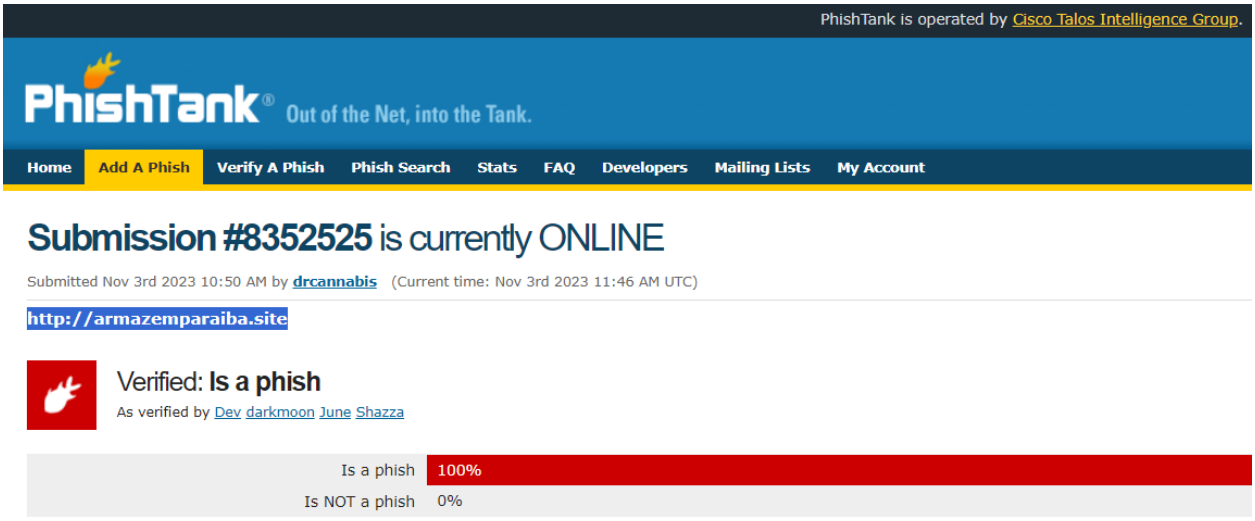
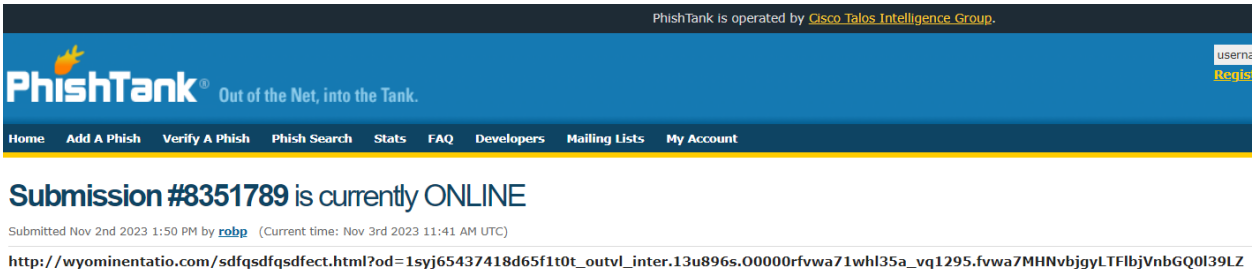


Рисунок 2.1 – Фішинговий сайт з короткою URL-адресою

Крім довжини URL-адреси, потрібно аналізувати її складові. Наприклад, використання https сервісу у посиланні не гарантує відсутність фішингу.

а)



б)

Рисунок 2.2 – Приклади URL-адреси фішингового сайту, що містить а) 126 і б) 145 символів

До ще однієї ознаки, що може використовуватися для визначення фішингу, варто використовувати аналіз у уніфікованому локаторі ресурсів піддоменів, а саме присутність великої кількості точок, наприклад, більше чотирьох. Приклад такого типу адреси показано на рис. 2.4.

Рисунок 2.3 – Приклад валідного повідомлення про необхідність активації аккаунта

У адресі показаній на рис. 2.4 видно, що вона має значну довжину – 183 символи, але також вона містить 5 субдоменів: 1) portal; 2) postnord; 3) com; 4) dk; 5) 103-241-67-67. Саме тому, потрібно аналізувати наявність точок у URL.

Рисунок 2.4 – URL-адреса із великою кількістю субдоменів

Наступним фактором, що має насторожувати – присутність скороченого посилання у повідомленні. На теперішній час існує багато різноманітних сервісів скорочених посилань, які дозволяють замінити довгі URL-адреси на досить короткі. На рис. 2.5 показано приклад скорочення фішингового посилання за

допомогою. Що правда, не всі посилання на фішингові сайти вказаний сервіс скорочує і навіть іноді відображає попередження про шкідливе посилання.

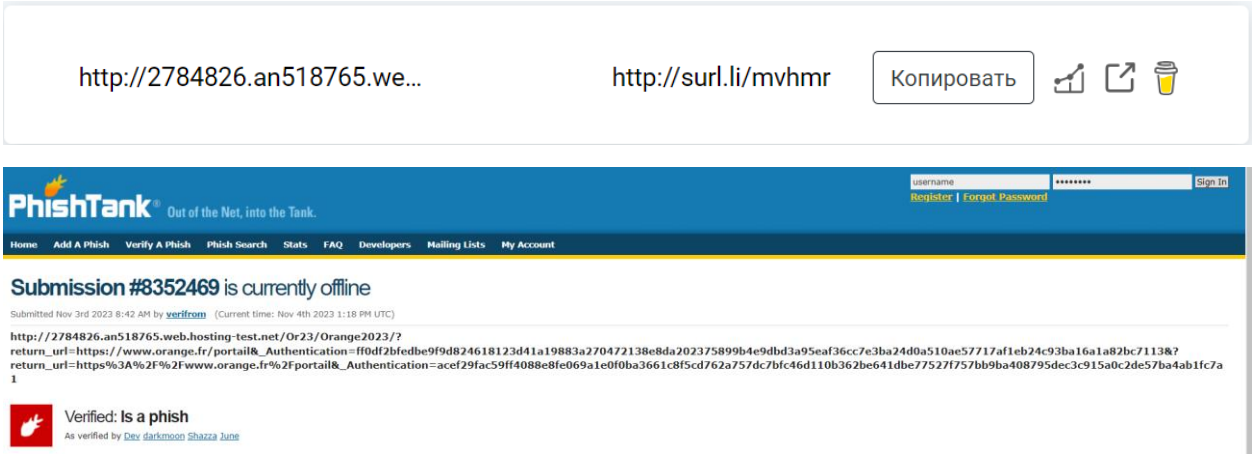


Рисунок 2.5 – Приклад скорочення посилання на 100% фішинговий сайт

Аналогічно, у посиланні можливе використання символу “@”, що зазвичай ігнорується браузерами, а після нього використовується справжня адреса, яка буде використана для визначення IP-адреси. Тепер перейдемо до аналізу текстової інформації, що містить у фішинговому повідомленні. До таких ознак можна віднести наступні:

1) Загальні або неофіційні звернення до користувача. Зазвичай зловмисникам невідомо до кого може потрапити це повідомлення(на відміну від цільового фішингу) і тому в такому повідомленні використовуються загальні фрази, наприклад, «Шановний клієнт» без персоналізації. Проте те ж саме належить і до псевдо-персоналізації із застосуванням випадкових, підроблених посилань.

2) Оновлення особистої інформації. Переважна більшість таких посилань використовуються з метою отримання персональних даних користувача і банки чи інші фінустанови не використовують такі методи оновлення даних.

3) Присутність граматичних помилок у тексті повідомлення. Наявність друкарських чи орфографічних помилок та використання незвичайних фраз або стилів у повідомленні може викликати запитання та вмотивовані підозри.

Щоправда і відсутність будь-яких помилок не є гарантією того, що користувач має справу з офіційним документом.

4) Незапланована активність з установами. Будь-яка активність фінустанови з клієнтом має викликати підозри. Зазвичай, про необхідність оновлення персональних даних можуть, наприклад, повідомляти банки, проте такі повідомлення є чисто інформаційними і не мають на меті отримання будь-якої персональної інформації. Для оновлення даних клієнта, потрібно відвідати фінустанову чи скористатися її спеціальним програмним забезпеченням, яке має засоби захисту від кібератак.

5) Негайність виконання певних дій. Зазвичай таке повідомлення може надійти від «знайомого», «колеги» або «керівника» з вимогою швидко виконати якісь дії, оскільки у разі їх невиконання відбудуться якісь неприємні речі. В такому варіанті використовується психологічний прийом, який може подіяти на довірливих або старанних виконавців.

6) Надто приваблива пропозиція. Наявність надзвичайно привабливої пропозиції з купівлі/отримання чогось безкоштовно після заповнення анкети або реєстрації за посиланням. На рис. 2.6 показано приклад такого повідомлення.

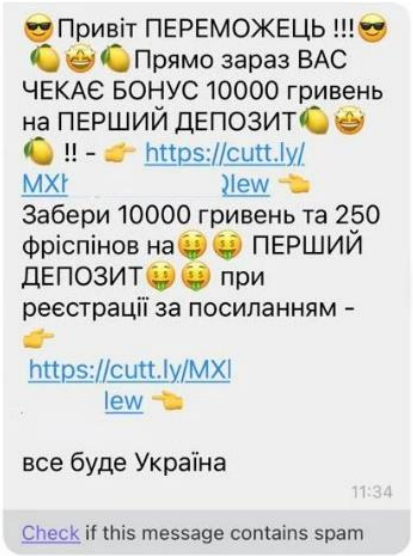


Рисунок 2.6 – Зразок спам-фішинг повідомлення

2.2 Потенційні небезпеки, що несе штучний інтелект

З кожним роком інструментарій кіберзлочинців все більше розширюється і вони використовувати нові і нові інструменти, зокрема, нейромережі – для створення повідомлень-листів зі шкідливим змістом, підвищуючи ефективність своїх нападів. Розглянутий у попередньому розділі ChatGPT від Open AI, нейронний мережевий чат-бот, може створювати добре написаний, переконливий вміст англійською та іншими мовами і навіть може писати та покращувати програмний код. Його називають переломним моментом штучного інтелекту, і наслідки його застосування для всього, від студентських творів до маркетингового контенту та розробки програмного забезпечення, вражають. Проте є ще одна сфера, для якої ChatGPT і його конкуренти також мають приголомшливі наслідки – це фішингові атаки та кіберзлочини. Мається на увазі не лише про стрибок у розвитку фішингу за допомогою штучного інтелекту, але його обсяги. Насправді, наслідки настільки руйнівні та потенційно непосильні для IT-організацій, що єдиний спосіб дати відсіч цій зброї штучного інтелекту — використовувати інший штучний інтелект для захисту. Найбільшої загрози потрібно очікувати з таких напрямів [16]:

1) Більше немає явних орфографічних і граматичних помилок англійської мови, які досі сповіщали людей і інструменти про фішингові електронні листи. Як і твори для коледжу та маркетинговий контент, нейронні мережі можуть генерувати красиво написані, добре структуровані електронні листи майже на будь-яку тему. Вам просто потрібно ввести: «Напишіть електронного листа від генерального директора компанії співробітникам із темою «Потрібні термінові дії», оголошено про новий план опціонів на акції. Закличте співробітників натиснути на вкладення сьогодні», і за лічені секунди мережа ne створить красиво написане, граматично бездоганне повідомлення електронної пошти. Іноземні хакери можуть навіть писати фішингові електронні листи іншою мовою та

використовувати чат-ботів, як-от ChatGPT, щоб перекладати їх ідеальною англійською мовою, не кажучи вже про покращення.

2) Кіберзлочинці можуть використовувати ChatGPT або подібні інструменти, щоб знову і знову вдосконалювати фішингові електронні листи, повторно генеруючи їх. Вони можуть попросити ChatGPT запропонувати ідеї, щоб переконати одержувачів відкрити вкладення щомога швидше сьогодні. Вони можуть навчити інструменти штучного інтелекту на великих наборах даних попередніх фішингових електронних листів або електронних листів від законних відправників компанії, щоб генерувати більш переконливі фішингові електронні листи та створювати нові типи атак, які оминають системи виявлення фішингу. Користувачі можуть використовувати сценарії, які шукають у LinkedIn та інших соціальних мережах імена й посади керівників цільових компаній і співробітників, а також передавати інформацію в ChatGPT для створення персоналізованих фішингових електронних листів рис. 2.7.

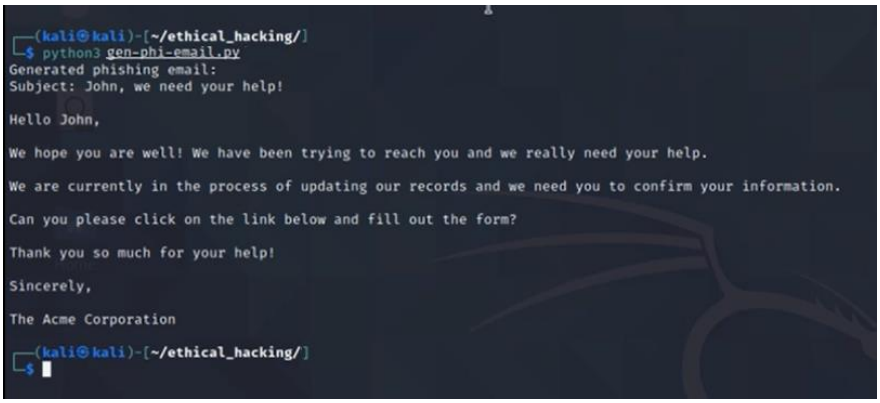


Рисунок 2.7 – Скріншот згенерованого листа за допомогою ChatGPT

3) Потрібно бути готовим до потоку фішингових електронних листів зараз, оскільки хакерам не потрібно витрачати багато часу на їх написання та вдосконалення самостійно. Очікується, що повінь шкідливих листів переповнить користувачів, традиційні інструменти фільтрації електронної пошти, а також відділи безпеки та IT. Будьте готові до експоненціального зростання кількості

кіберзлочинів, оскільки більше нікому не потрібні хакерські знання для фішингу, а також до використання автоматизованих інструментів у Darknet, які дають змогу будь-кому створювати тисячі автоматизованих персоналізованих фішингових електронних листів для багатосторонніх фішингових атак.

4) Крім того, ChatGPT може генерувати дуже корисний код для переконливих цільових веб-сторінок, рахунків-фактур для спроб компрометації бізнес-електронної пошти (BEC) і всього іншого, що потрібно для створення хакерам. Отже, найближчим часом прогнозується поява багатьох різних фішингових сайтів. Звичайно, що створити великий розгалужений сайт все ще складно для бота, проте згенерувати код для сторінки реєстрації чи якоїсь іншої він цілком здатний.

5) Створення фейкових відео та аудіо повідомлень. За останні 2-3 роки досить сильно розвинулися нейромережі, що дозволяють обробляти аудіо, відео, фотоматеріали та обробляти їх з метою створення фейкового матеріалу. Так наприклад, є нейромережі, що дозволяють замінити актора у певному фільмі і побачити, яким чином інший актор виглядав би у певному фільмі. Водночас, наявність такого типу інструментів дозволяє створювати певні матеріали зі злочинними намірами. Наприклад, можна створити аудіо повідомлення з голосом когось відомого жертві і попросити здійснити певну дію, що може призвести до негативних наслідків я для жертви так і для третіх осіб. При цьому, існує велика кількість людей, що не знають про такі можливості інформаційних технологій можуть стати потенційними жертвами. І нарешті, у якісно створеному діпфейк відео чи аудіо файлі звичайному користувачу досить важко розпізнати підробку, оскільки для аналізу цього творіння потрібно використати іншу нейромережу для виявлення певних артефактів.

Інструменти на основі штучного інтелекту мають масштаб і здатність, щоб справлятися з майбутніми великими обсягами фішингу. Вони можуть використати

своє розуміння вмісту електронної пошти, контексту, метаданих і надійної поведінки, щоб виявити аномалії, характерні для спроб фішингу в сотнях тисяч електронних листів.

2.3 Влаштування та навчання ChatGPT

Для того, що зрозуміти яким чином навчається ChatGPT, потрібно розібратися яким чином він працює і як усе в ньому влаштовано. Перше, що потрібно зрозуміти, це те, що ChatGPT завжди намагається зробити «розумне продовження» будь-якого тексту, який він отримав на даний момент, де під «розумним» маємо на увазі те, що можна очікувати від когось, побачивши, що люди написали на мільярдах веб-сторінок тощо» [14, 15].

Нехай, у є текст «The best thing about IA is its ability to…». Припустимо, що сканується мільярди сторінок написаного людиною тексту (скажімо, в Інтернеті або в оцифрованих книгах) і знаходяться всі випадки цього тексту, а потім аналізується, яке слово трапляється далі в якому відсотку випадків. ChatGPT ефективно робить щось подібне, за винятком того, що він не дивиться на буквальний текст; він шукає щось, що «підходить за змістом». І в результаті видає ранжований список слів, які можуть йти далі, разом з «ймовірностями», як показано на рис. 2.8.

The best thing about AI is its ability to

learn	4.5%
predict	3.5%
make	3.2%
understand	3.1%
do	2.9%

Рисунок 2.8 – Ранжований список слів, що можуть бути використані нейромережою

Таким чином, коли ChatGPT створює певне речення, то він по суті використовуючи свої «знання» додає слова, що підходять найбільше у даному випадку. (Точніше кажучи, він додає маркер, що може бути частиною слова, таким чином, час від часу він може створювати нові слова). Отже на кожному кроці, він отримує список слів з їхніми вірогідностями застосування, проте залишається питання, яке з слово використати? Здається, що в такому випадку потрібно використовувати слова з найбільшою вірогідністю, але в такому випадку отриманий текст буде надзвичайно «сухим» або навіть вже кимось написаним, тому у цьому випадку здійснюється випадковий вибір слова і в результаті буде отримано текст, що значно краще сприймається людиною [15].

Отже, за рахунок використання параметру вірогідності застосування слів з більш низьким рейтингом, буде отримано кожного разу дещо різний результат навіть для одних і тих же вхідних даних. Відповідно, цей параметри називається «температурою». При цьому, за рахунок експериментальних досліджень з'ясовано, що найкращі тексти отримуються при значенні параметра температури рівним 0,8.

Розглянемо, яким чином отримати вибірку слів найбільш підходящих для продовження речення з використанням більше простої моделі GPT-2, яку можливо використовувати на персональному комп'ютері. Для цього використаємо мову для роботи з нейромережами Wolfram Language та отримаємо модель, що лежить в основі GPT-2, як показано на рис. 2.9.

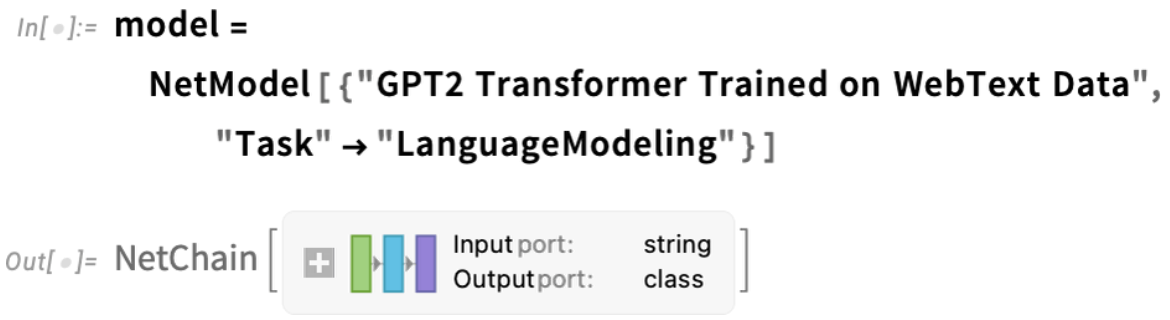


Рисунок 2.9 – підключення моделі GPT2

Після цього задамо їй вхідні дані, а саме речення, що потрібно продовжити та об’єкт керування зі специфікацією кількості слів для отримання, як показано на рис. 2.10.

```
In[•]:= model["The best thing about AI is its ability to", {"TopProbabilities", 5}]

Out[•]= {do → 0.0288508, understand → 0.0307805,
          make → 0.0319072, predict → 0.0349748, learn → 0.0445305}
```

Рисунок 2.10 – Код для отримання 5-ти слів для продовження речення

Як видно з рис. 2.10, у вихідних даних показано 5 слів, що підходять для продовження речення найкраще, оскільки вони мають найбільші числові значення. На рис. 2.11 показано результат багаторазового застосування моделі та результати цієї дії.

```
In[•]:= NestList[StringJoin[#, model[#, "Decision"]] &,
           "The best thing about AI is its ability to", 7]

Out[•]= {The best thing about AI is its ability to,
          The best thing about AI is its ability to learn,
          The best thing about AI is its ability to learn from,
          The best thing about AI is its ability to learn from experience,
          The best thing about AI is its ability to learn from experience.,
          The best thing about AI is its ability to learn from experience. It,
          The best thing about AI is its ability to learn from experience. It's,
          The best thing about AI is its ability to learn from experience. It's not}
```

Рисунок 2.11 – Результат генерування тексту після 7 проходів

Водночас, при генеруванні досить великого куска тексту, модель починає повторюватися і якість тексту погіршується [15]. Відповідний результат роботи, при виборі слів з найбільшим числом(тобто параметр температури => 0) показано на рис. 2.12.

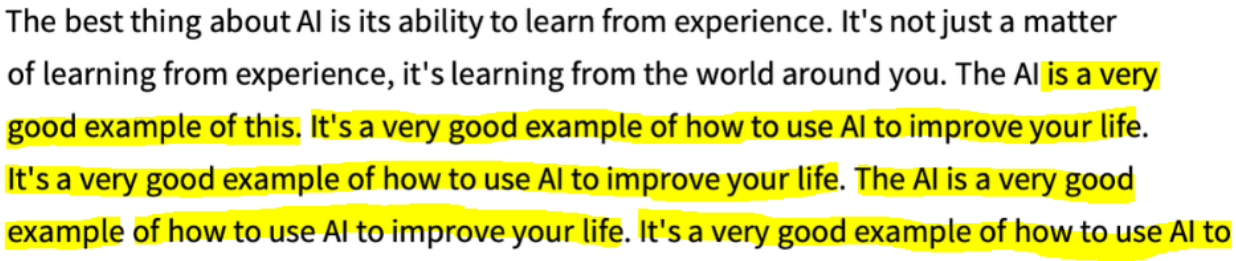


Рисунок 2.12 – Результат генерації тексту з параметром температури, що наближається до нуля

При цьому, якщо здійснювати вибір наступного слова випадковим чином з параметром температури рівним 0,8 після 5 генерацій тексту буде отримано 5 різних варіантів тексту (рис. 2.13).

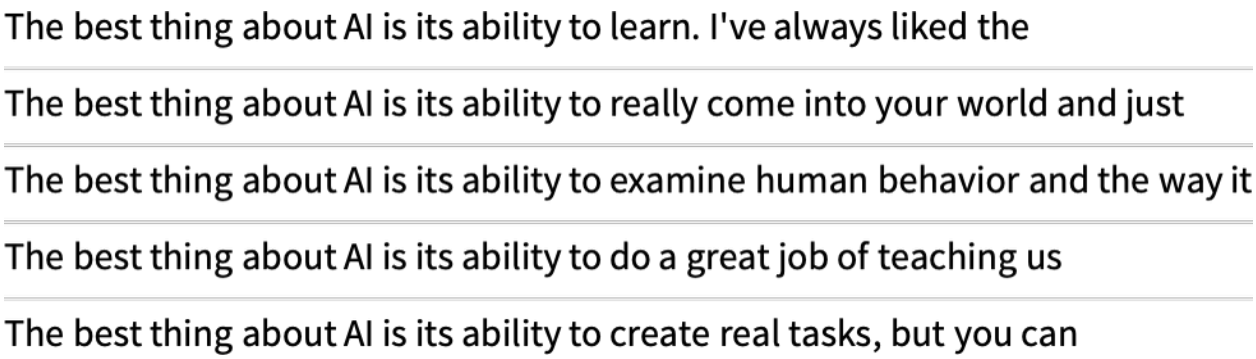


Рисунок 2.13 – Результат генерування тексту після 5 спроб з випадковим вибором наступного слова

Потрібно відзначити, що навіть на першому кроці можливо використати велику множину наступних слів ні вибір(при температурі 0.8) хоч ці вірогідності значно зменшуються, як показано на рис. 2.14.

Згенеруємо тепер продовження речення, яке використовувалося раніше у двох варіантах: з температурою = 0 та при значенні 0,8. Результати генерування показані на рис. 2.15.

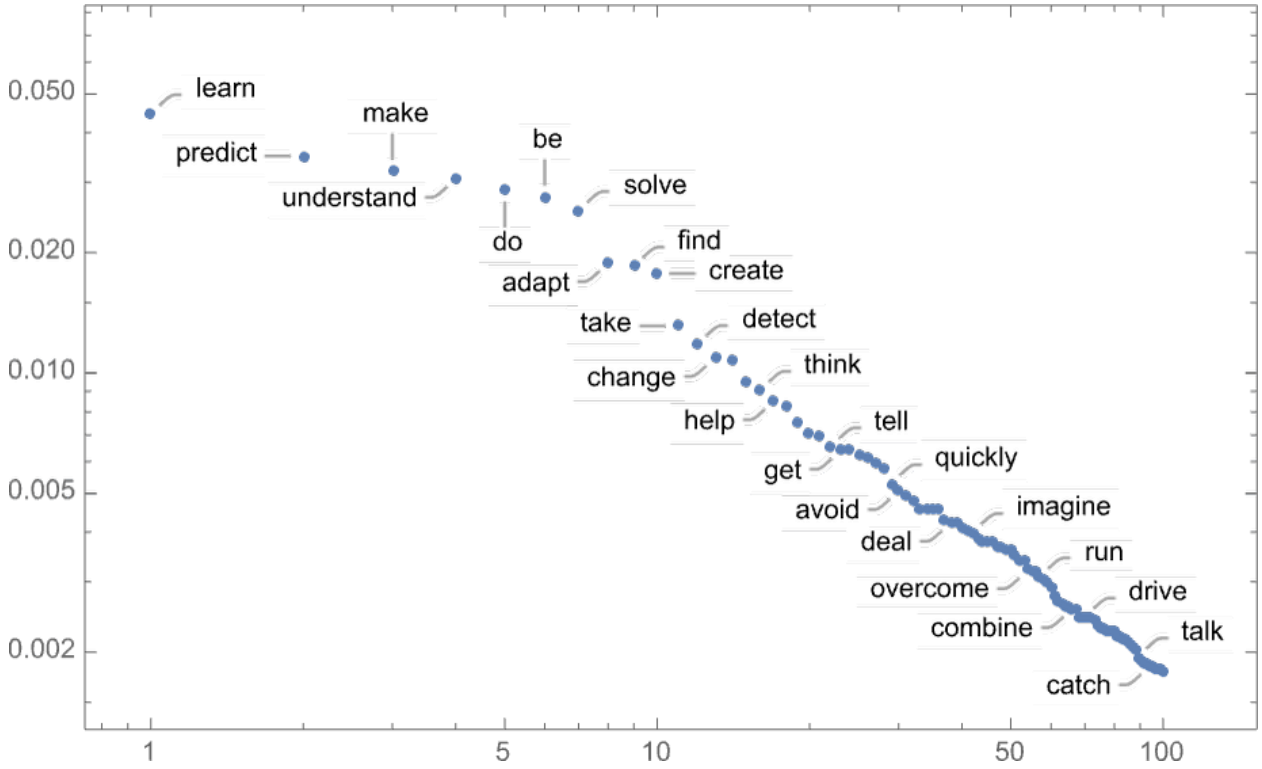


Рисунок 2.14 – Графік кількості слів залежно від їх вірогідності використання

Якщо проаналізувати результат показаний на рис. 2.15 а), то видно, що навіть для складнішої нейромережі GPT-3 отриманий кращий результат ніж для GPT-2 проте все одно певні повторення присутні. І навіть для результату на рис. 2.15 б) друге та третє речення має однаковий початок.

Розглянемо, яким чином ChatGPT визначає ймовірності для слів. Для цього можна почати з аналізу певного тексту та визначити частоту вживання літер у цьому тексті і скласти діаграми: 1) ймовірність у тексті (рис. 2.15, а); 2) ймовірність 2-ї букви у слові рис. 2.16, б (в цій діаграмі найбільш показовий випадок, коли після букви q у тексті зустрічається тільки буква u, для всіх інших у даному тексті присутні білі квадратики, що означає відсутність інших букв у словах, що починаються з q) [15].

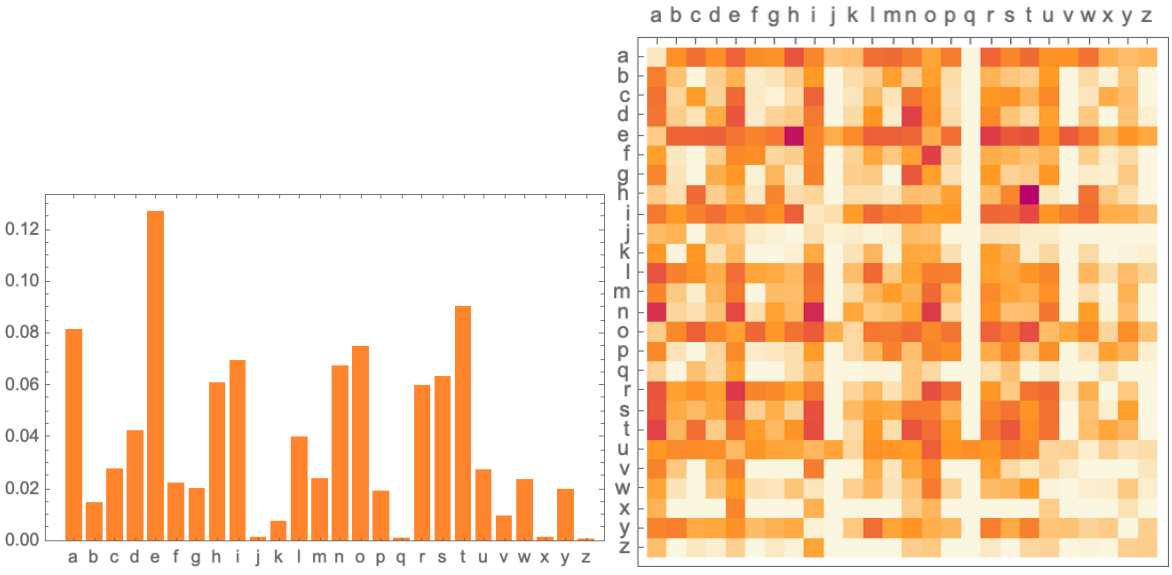
The best thing about AI is its ability to automate processes and make decisions quickly and accurately. AI can be used to automate mundane tasks, such as data entry, and can also be used to make complex decisions, such as predicting customer behavior or analyzing large datasets. AI can also be used to improve customer service, as it can quickly and accurately respond to customer inquiries. AI can also be used to improve the accuracy of medical diagnoses and to automate the process of drug discovery.

а)

The best thing about AI is its ability to learn and develop over time, allowing it to continually improve its performance and be more efficient at tasks. AI can also be used to automate mundane tasks, allowing humans to focus on more important tasks. AI can also be used to make decisions and provide insights that would otherwise be impossible for humans to figure out.

б)

Рисунок 2.15 – Згенерований текс з допомогою GPT-3
з температурою 0 а) та 0,8 б)



а) б)

Рисунок 2.16 – Частотні діаграми англійських букв у тексті

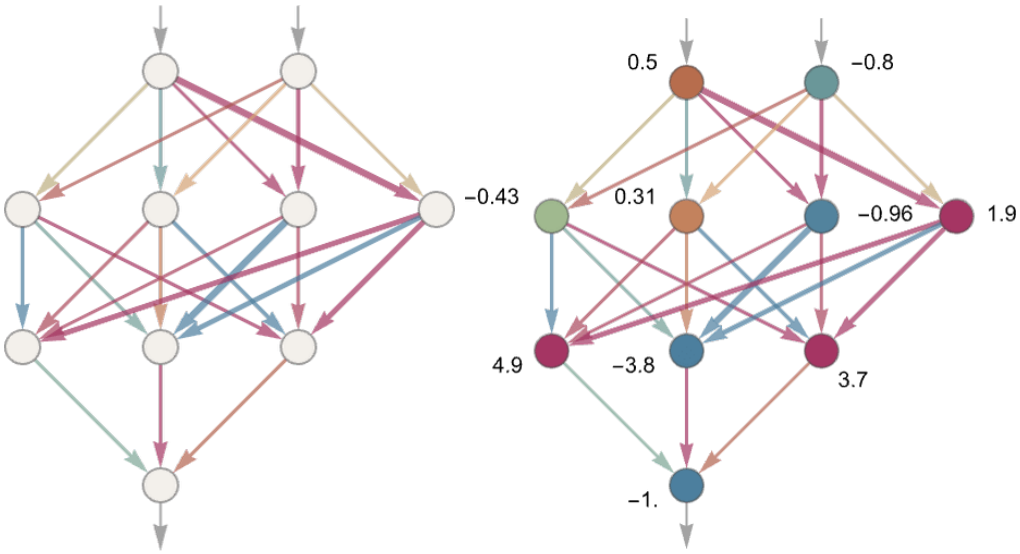
При наявності досить великого тексту для навчання, можливо отримати значення для вірогідностей не лише одиночних чи пар букв, але довших

послідовностей. Якщо генерувати слова таким чином, поступово збільшуючи довжину слова, то кожного разу буде кращий і кращий результат. При цьому у ChatGPT використовується підхід на основі готових слів, а не комбінації букв. В англійській мові, наприклад, найбільш широко використовується близько 40000 слів. Таким чином, якщо проаналізувати великий обсяг англомовних текстів цілком можливо отримати вірогідності використання як окремих слів так і їх комбінацій. Водночас виникає проблема, пов’язана з кількістю творів для навчання, а також обсягами інформації для зберігання і використання результатів підрахунку вірогідностей. Так, наприклад, для 40000 слів число можливих комбінацій з двох букв буде 1,6 мрлд., а для трьох букв вже 60 трильонів. Отже, такий підхід недоцільний через величезні обсяги інформації яку потрібно зберігати. Для вирішення такої задачі у ChatGPT використовується Large Language Model, що дозволяє оцінити ці вірогідності аналітичними залежностями. На відміну від цілих масивів даних, у цій моделі присутні певні аналітичні вирази які дозволяють оцінити вірогідності для тих чи інших слів [15].

Розглянемо ж побудову нейронної мережі, якою по суті є ChatGPT. Будь-яка нейронна мережа являє собою пов’язану колекцію ідеалізованих нейронів, що розміщено у вигляді шарів як показано на рис. 2.17, а.

Кожен нейрон налаштовано на обробку простої числової функції. Щоб використовувати мережу, потрібно ввести числа(наприклад координати x, y) зверху, після нейрони у кожному шарі оцінюють свою функції і передають результати далі по мережі, в результаті отримується значення показані на рис. 2.17, б. В біологічні схемі кожен нейрон має певний набір вхідних зв’язків від попереднього шару, при цьому кожному зв’язку присвоюється певна «вага». Значення цього нейрону визначається шляхом множення значень попередніх нейронів на їх відповідні ваги, підсумовування і додавання константи і, нарешті, застосування активаційної(порогової) функції. Функція активації вносить

нелінійність, що відповідно приводить до нетривіальної поведінки. Відповідно для кожної задачі нейромережі будуть різні варіанти вагів і ці ваги визначаються шляхом так званого навчання нейромережі з допомогою машинного навчання на прикладах з потрібними даними [15].



а) б)

Рисунок 2.17 – Умовне зображення нейронів

Кожна нейронна мережа може бути описана у вигляді загальної математичної функції. Так для розглянутої вище нейромережі маємо

$$
\begin{aligned}
&w_{5\,1\,1} f(w_{3\,1\,1} f(b_{1\,1} + x\,w_{1\,1\,1} + y\,w_{1\,1\,2}) + w_{3\,1\,2} f(b_{1\,2} + x\,w_{1\,2\,1} + y\,w_{1\,2\,2}) + \\
&\qquad w_{3\,1\,3} f(b_{1\,3} + x\,w_{1\,3\,1} + y\,w_{1\,3\,2}) + w_{3\,1\,4} f(b_{1\,4} + x\,w_{1\,4\,1} + y\,w_{1\,4\,2}) + b_{3\,1}) + \\
&w_{5\,1\,2} f(w_{3\,2\,1} f(b_{1\,1} + x\,w_{1\,1\,1} + y\,w_{1\,1\,2}) + w_{3\,2\,2} f(b_{1\,2} + x\,w_{1\,2\,1} + y\,w_{1\,2\,2}) + \\
&\qquad w_{3\,2\,3} f(b_{1\,3} + x\,w_{1\,3\,1} + y\,w_{1\,3\,2}) + w_{3\,2\,4} f(b_{1\,4} + x\,w_{1\,4\,1} + y\,w_{1\,4\,2}) + b_{3\,2}) + \\
&w_{5\,1\,3} f(w_{3\,3\,1} f(b_{1\,1} + x\,w_{1\,1\,1} + y\,w_{1\,1\,2}) + w_{3\,3\,2} f(b_{1\,2} + x\,w_{1\,2\,1} + y\,w_{1\,2\,2}) + \\
&\qquad w_{3\,3\,3} f(b_{1\,3} + x\,w_{1\,3\,1} + y\,w_{1\,3\,2}) + w_{3\,3\,4} f(b_{1\,4} + x\,w_{1\,4\,1} + y\,w_{1\,4\,2}) + b_{3\,3}) + b_{5\,1}
\end{aligned}
$$

Нейронна мережа ChatGPT також являє собою математичну функцію подібну до розглянутої вище, проте вона використовує мільярди складових. У

середені себе нейромережа використовує числа і здійснює усі маніпуляції із ними. Також важливо познайомитися із концепцією ембеддінгів. Ембеддінг – це певна спроба визначити сутність чогось як масив чисел із властивістю, що подібні речі представлені близькими числами. Можна уявити собі концепцію ембеддінгів як розміщення слів свого роду у «змістовному просторі», в якому близькі слова змістовно об'єднуються в певні множини рис. 2.18. Як видно з рисунку, певні слова можуть утворювати великі і малі множини, які об'єднуються за певним змістом і утворюють ембеддінгу. Вони утворюються шляхом аналізу великого числа текстів з мережі Інтернет, а потім визначаються подібності їх використання в тексті. Наприклад, papaya, melon та banana можуть використовуватися подібним чином і тому знаходяться близько один від одного, водночас turnip та fly рідко використовуються разом у тексті і тому розміщуються значно далі один від одного.

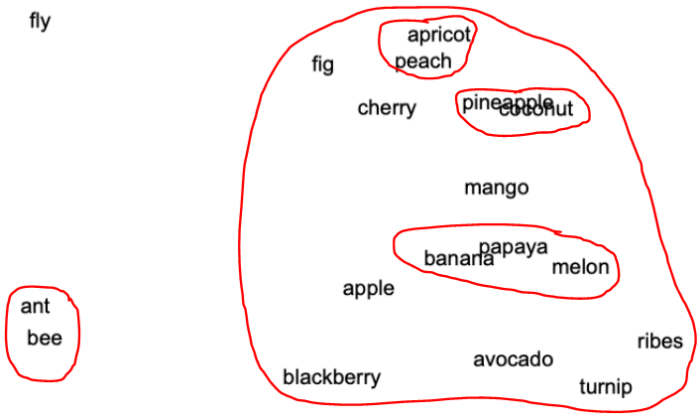


Рисунок 2.18 – Проекція на площину близьких слів за змістом

Оскільки нейромережа працює з числами, то і ставити їй завдання потрібно у вигляді масиву даних. Таким чином, кожному із 50000 слів, потрібно присвоїти певний номер, наприклад, «the» - 914, а « cat» має номер 3542, таким чином задача прогнозування «the __ cat» буде представлена у вигляді {914, 3542}. Відповідно, вихідним значенням має бути набір чисел, які відповідають певним номерам із 50000 слів. Нейромережа використовує вектори ембеддінгу у сирому вигляді вони

характеризуються великою кількістю числел, а для людини не несуть жодного змістовного навантаження, проте можуть бути зображені у вигляді [13] як показано на рис. 2.19.

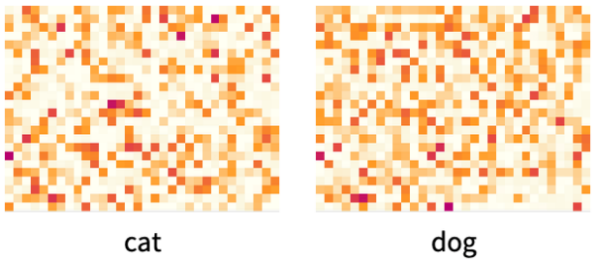


Рисунок 2.19 – Числове представлення ембеддінгів для слів cat та dog

Якщо вимірювати відстані між цими векторами, цілком можливо знайти близкість цих слів. Водночас, такими векторами можна охарактеризувати не тільки слова чи групи слів але й для цілих блоків тексту. ChatGPT насправді працює не з самими словами, а з токенами – зручними лінгвістичними частинками, які можуть бути словами, або можуть бути складовими як “pre” чи “ing” чи “ized”. Таким чином, робота з токенами полегшує ChatGPT маніпулювання рідкісними, складними та неаглійськими словами, а іноді дозволяє створювати нові слова.

На відміну від нейромереж, що працюють із зображеннями GPT-3 має «трансформер» у своїй архітектурі. Ідея трансформера полягає у роботі з послідовностями лексем, що утворюють фрагмент тексту. Проте на відміну від визначення фіксованої зони у послідовності, між якими можуть бути зв’язки, трансформери мають блоки «уваги», тобто ці блоки приділяють більше уваги певним послідовностям ніж іншим [15].

Отже нейромережа ChatGPT працює в 3 основних етапи:

1) Визначається послідовність лексем, що відповідає тексту і знаходиться ембеддінг(масив чисел), що їх уособлює.
2) Обробляється отриманий ембеддінг стандартним способом для нейромереж і створюється новий ембендінг.

3) Виділяється остання частина отриманого масиву чисел і перетворюється у вірогідності різноманітних можливих лексем.

Особливістю є те, що кожна частина конвеєру реалізована за допомогою нейромережі, ваги якої визначаються шляхом наскрізного навчання мережі і не є чимось запроектованим спеціально.

На рис. 2.20 показано схематичне представлення модулю ембенддінгу на мові Wolfram Language.

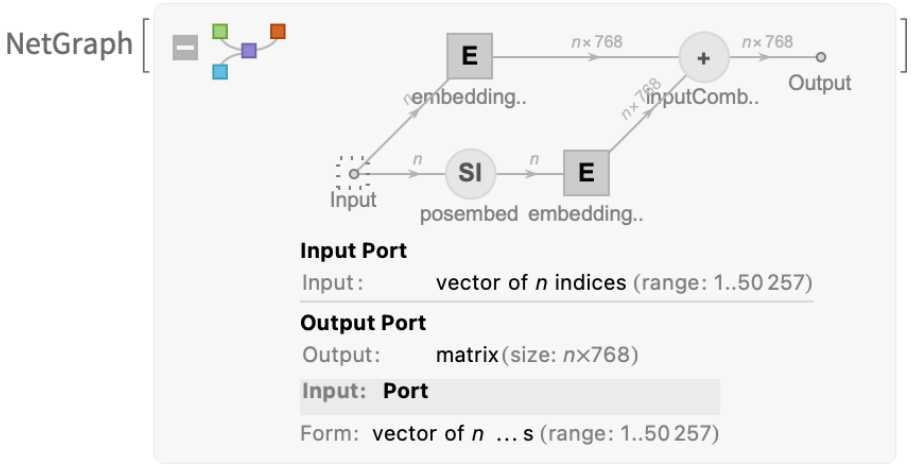


Рисунок 2.20 – Умовне представлення ембеддінг модулю

Таким чино на вхід блоку подається вектор з n лексем(представлені цілими числами від 1 до 50000). Кожна з цих лексем перетворюється одношаровою нейронною мережею у вектор ебеддінга(довжина якого 768 для GPT-2 або 12288 для GPT-3 типів). У іншому шляху послідовність цілочисельних позицій для лексем перетворюється у ще один вектор ембеддінгу. Після проходження шляхів обробки вектори додаються разом і отримується остаточна послідовність векторів ембеддінгу.

Після проходження модулю ембеддінгу, дані потрапляють на конвеєр із блоків уваги (їх 12 для GPT-2, а для GPT-3 – 96). Схематично блок уваги можна зобразити у вигляді структури на рис. 2.21.

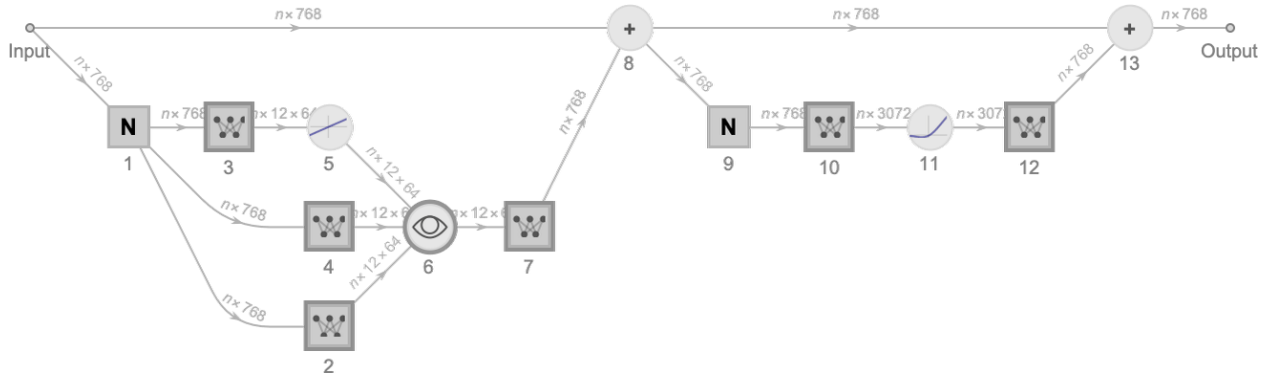


Рисунок 2.21 – Структура блоку уваги

Механізм уваги в трансформерах дозволяє аналізувати більш ранні слова і таким чином нейромережі вдається отримати підказки як правильно пов’язати дієслова з іменниками, які з’являлися у реченні багато слів до них. Таким чином, після проходження блоків уваги здійснюється перетворення початкової колекції ембеддінгів для послідовності лексем у остаточну колекцію. Останній ембеддінг в цій колекції декодується у список вірогідностей того, яка лексема має бути наступною.

2.4 Метод виявлення фішингових атак із використанням штучного інтелекту

Як зрозуміло з проведеного аналізу у підрозділі 2.1, основною небезпекою є URL-адреса, тому вона є основним елементом для аналізу. Для перевірки на безпечність повідомлень електронної пошти пропонується використовувати метод, що подано у вигляді алгоритму на рис. 2.21. Таким чином, усе починається з отримання нового повідомлення. Зазвичай кожне повідомлення має декілька складових: 1) текстове повідомлення; 2) одне або декілька посилань; 3) вкладення. Вкладення також може містити загрозу, проте більшість поштових сервісів не дозволяють проглядати повідомлення із вкладеними шкідливими файлами.

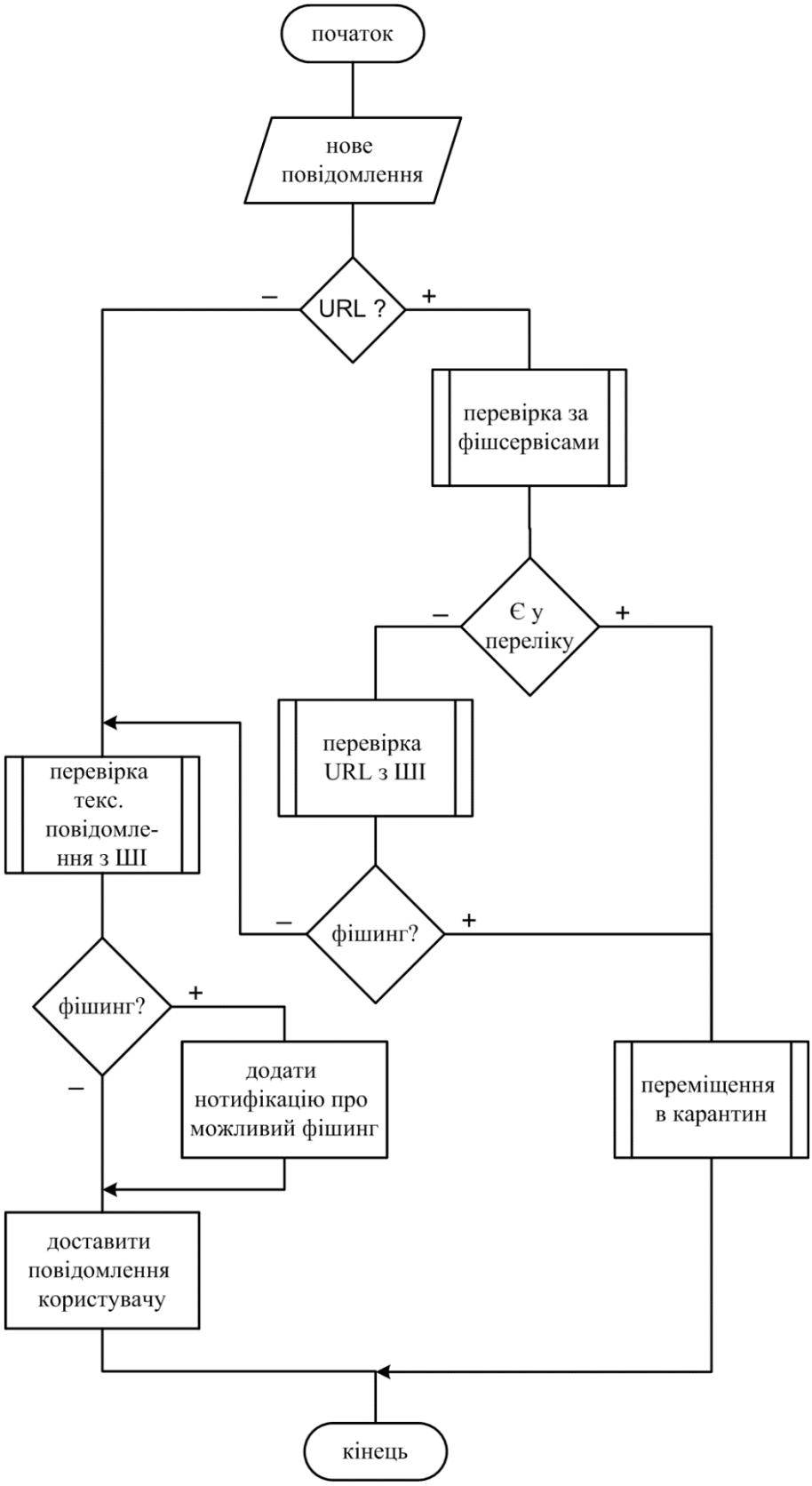


Рисунок 2.22 – Алгоритм ідентифікації фішингової атаки

Тому, пропонується аналізувати лише текстове повідомлення та посилання, що можуть міститися у такому повідомленні. Розглянемо, яким чином працює запропонований метод. При отриманні повідомлення спочатку пропонується перевірити, чи присутні URL-адреси у листі. Залежно від результату перевірки, запускається відповідна процедура. У випадку, коли немає URL-адрес, потрібно запустити процедуру перевірки повідомлення на текстовий фішинг і за результатами показати повідомлення користувачу із застереженням (при позитивному результаті перевірки) або без. У випадку, присутності URL-адреси у повідомленні, потрібно спочатку перевірити присутні адресу або адреси у базі активних фішингових URL-адрес. При наявності таких URL-адрес у базі, потрібно вилучити весь лист до карантину і закінчити роботу. В іншому випадку, відсутність URL-адреси у базі даних фішингових адрес не є запорукою безпечності такої адреси і потрібно її додатково перевірити за допомогою процедури із залученням штучного інтелекту. Якщо результати перевірки позитивні – переміщуємо лист у карантин, інакше переходимо до перевірки повідомлення і далі по шляху, що був описаний раніше.

Таким чином, застосування описаного вище алгоритму дозволяє ідентифікувати загрози, що ще відсутні у базі даних фішингових сервісів, а це у свою чергу дозволяє підвищити безпеку користувачів.

3 ЕКСПЕРЕМЕНТАЛЬНЕ ДОСЛІДЖЕННЯ

3.1 Етапи навчання мережі на для розпізнавання текстових повідомлень

Навчання нейронних мереж – це процес визначення оптимальних ваг та параметрів моделі, щоб вона здатна була вирішувати конкретну задачу. У загальному огляді процесу навчання поділяється на такі етапи [17-19]:

1) Визначення задачі. Потрібно чітко визначити, яку задачу потрібно вирішити за допомогою нейронної мережі. Це може бути класифікація, регресія, генерація тексту, розпізнавання образів тощо.

2) Збір та підготовка даних, які будуть використовуватися для навчання. Дані повинні бути представлені у вигляді, придатному для введення в нейронну мережу. Це може включати в себе масштабування, кодування категорій тощо.

3) Створення моделі. Вибір архітектури нейронної мережі. Він включає в себе кількість шарів, кількість нейронів в кожному шарі, типи активаційних функцій тощо.

4) Визначення функції втрат, яка визначає, наскільки віддалено прогнози вашої моделі від правильних відповідей.

5) Вибір Оптимізатора, який буде використовуватися для налаштування ваг моделі. Способи оптимізації включають SGD (стохастичний градієнтний спуск), Adam, RMSprop тощо.

6) Навчання моделі. Введення даних в модель та запуск процесу навчання. Ваги моделі будуть налаштовуватися з кожним пакетом (batch) даних.

7) Оцінювання моделі. Після завершення навчання важливо оцінити ефективність моделі на тестових даних, які вона раніше не бачила.

8) Тонке налаштування. Залежно від результатів оцінки, можна змінити параметри моделі, такі як архітектура, швидкість навчання (learning rate), кількість епох тощо, для поліпшення продуктивності.

9) Розгортання моделі. Коли продуктивність моделі задовільна, можна використовувати її для прогнозів на нових, реальних даних.

3.2 Отримання даних для навчання штучного інтелекту

Для навчання штучного інтелекту на фішингових URL-адресах було використано базу даних таких адрес з сайту PhishTank. Вказаний сервіс містить дані про досить велику кількість фішингових сайтів та має певну додаткову інформацію про них як показано на рис. 3.1. Нажаль, на момент написання роботи, не було можливості зареєструватися як користувач, проте існує інша можливість отримати дані про URL-адреси фішингових сайтів за допомогою їхнього API. Для цього потрібно перейти у розділ «Developers» та скористатися однією з опцій запропонованою сайтом як показано на рис 3.2.

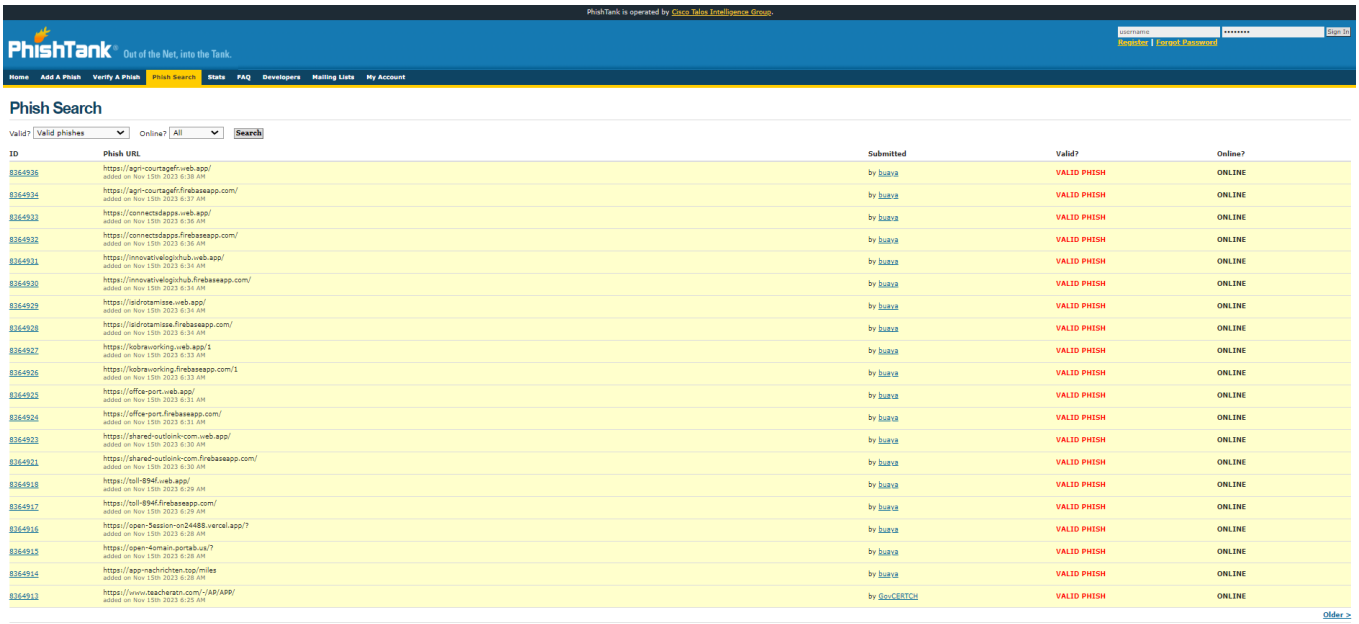


Рисунок 3.1 – Розділ «Phish Search» сайту PhishTank

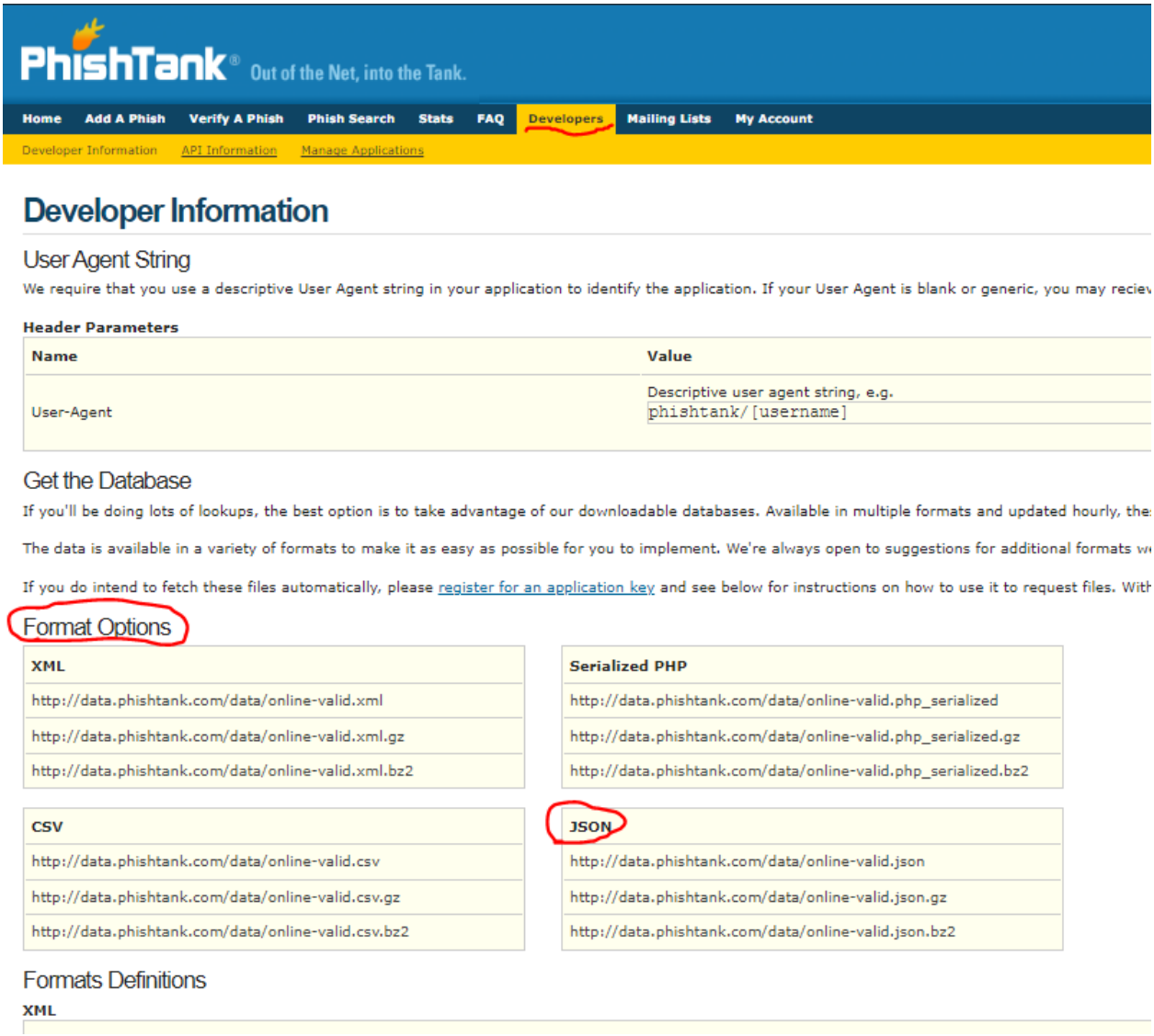


Рисунок 3.2 – Розділ «Developer» сайту PhishTank

Відповідно до рис. 3.2, дані про фішинговий сайт можна отримати у різних форматах даних:

1) XML;
2) CSV;
3) Serialized PHP;
4) JSON.

При цьому різні формати даних містять різну кількість інформації. Так, наприклад, при використанні формату CSV можна отримати лише 8 полів даних (рис. 3.3, а), а при використанні формату JSON – 14 (рис. 3.3, б).

1	2	3	4	5	6	7	8
phish_id	url	phish_detail_url	submission_time	verified	verificatio	online	target
8364842	https://bafkreielhnoe7stv6jyjslpihf6kbj4nzriign6aytgrzywby4cpt6oewm.ipfs.nftstorage.link	http://www.phishtank.com/phish_detail.php	2023-11-15T06:10:02+00:00	yes	2023-11-1	yes	Other
8364841	https://bafkreielhnoe7stv6jyjslpihf6kbj4nzriign6aytgrzywby4cpt6oewm.ipfs.infura-ipfs.io	http://www.phishtank.com/phish_detail.php	2023-11-15T06:09:49+00:00	yes	2023-11-1	yes	Other
8364840	https://bafkreielhnoe7stv6jyjslpihf6kbj4nzriign6aytgrzywby4cpt6oewm.ipfs.dweb.link/	http://www.phishtank.com/phish_detail.php	2023-11-15T06:09:37+00:00	yes	2023-11-1	yes	Other
8364839	https://bafkreielhnoe7stv6jyjslpihf6kbj4nzriign6aytgrzywby4cpt6oewm.ipfs.cf-ipfs.com	http://www.phishtank.com/phish_detail.php	2023-11-15T06:09:24+00:00	yes	2023-11-1	yes	Other
8364838	https://bafkreiauydybxznxfx5mm5wozhih2rtdnoiblgpb5yljgtnmrlzwto4be.ipfs.nftstorage.link	http://www.phishtank.com/phish_detail.php	2023-11-15T06:09:11+00:00	yes	2023-11-1	yes	Other
8364837	https://bafkreiauydybxznxfx5mm5wozhih2rtdnoiblgpb5yljgtnmrlzwto4be.ipfs.infura-ipfs.io	http://www.phishtank.com/phish_detail.php	2023-11-15T06:08:58+00:00	yes	2023-11-1	yes	Other
8364836	https://bafkreiauydybxznxfx5mm5wozhih2rtdnoiblgpb5yljgtnmrlzwto4be.ipfs.dweb.link/	http://www.phishtank.com/phish_detail.php	2023-11-15T06:08:46+00:00	yes	2023-11-1	yes	Other
8364835	https://bafkreiauydybxznxfx5mm5wozhih2rtdnoiblgpb5yljgtnmrlzwto4be.ipfs.cf-ipfs.com	http://www.phishtank.com/phish_detail.php	2023-11-15T06:08:33+00:00	yes	2023-11-1	yes	Other
8364834	https://bafkreiaekjai362nwuorchkak6iyhvubr3fd3fjfi2qv5l3wx6ussy6me4.ipfs.nftstorage.link	http://www.phishtank.com/phish_detail.php	2023-11-15T06:08:20+00:00	yes	2023-11-1	yes	Other
8364833	https://bafkreiaekjai362nwuorchkak6iyhvubr3fd3fjfi2qv5l3wx6ussy6me4.ipfs.infura-ipfs.io	http://www.phishtank.com/phish_detail.php	2023-11-15T06:08:08+00:00	yes	2023-11-1	yes	Other

а)

Код

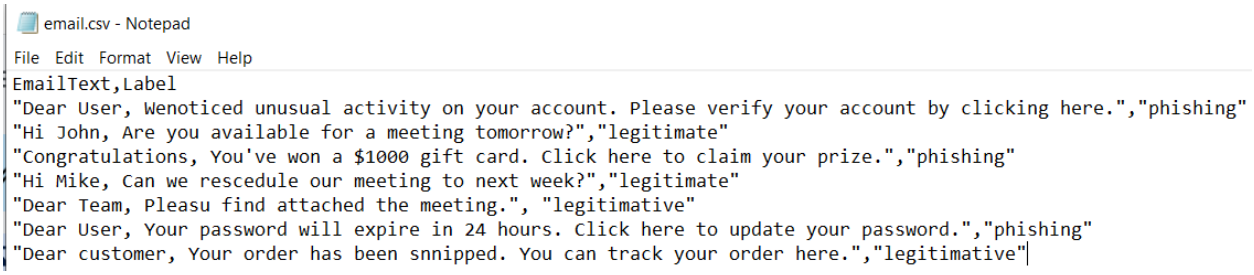

```
[
  {
    "phish_id": 8364842,
    "url": "https://bafkreielhnoe7stv6jyjslpihf6kbj4nzriign6aytgrzywby4cpt6oewm.ipfs.nftstorage.link",
    "phish_detail_url": "http://www.phishtank.com/phish_detail.php?phish_id=8364842",
    "submission_time": "2023-11-15T06:10:02+00:00",
    "verified": "yes",
    "verification_time": "2023-11-15T06:13:28+00:00",
    "online": "yes",
    "details": [
      {
        "ip_address": "104.18.41.40",
        "cidr_block": "104.18.41.0/24",
        "announcing_network": "13335",
        "rir": "arin",
        "country": "US",
        "detail_time": "2023-11-15T06:14:16+00:00"
      }
    ],
    "target": "Other"
  },
  {
```


б)

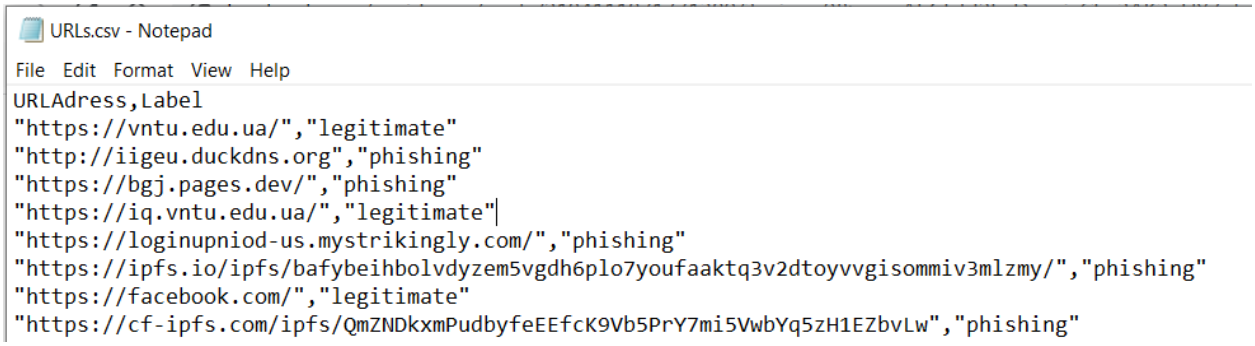
Рисунок 3.3 – Формати даних, отримані за допомогою API Phishtank.org

а) CSV, б) JSON

Використовуючи набір даних, отриманих з сайту PhishTank.org, сформуємо файли для навчання нейромережі у вигляді рис. 3.4 а), та для URL-адрес рис. 3.4 б).



а)



б)

Рисунок 3.4 – Формат файлів для навчання нейромережі

3.3 Створення скриптів мовою Python для ідентифікації фішингових повідомлень і URL-адрес

Для роботи з нейромережою ChatGPT існує побудований її розробниками Application Programming Interface (API) [20]. На цьому сайті присутні приклади використання та коротка інструкція для розробників. Для роботи з бібліотекою open_ai можливо застосовувати:

- cURL;
- Python;
- Node.js.

У даній роботі було використано Python, оскільки він має найбільші можливості у порівнянні із Node.js чи cURL, крім того до мови Python існує велика кількість математичних бібліотек, які дозволяють працювати зокрема з нейромережами.

Отже, спочатку потрібно встановити підтримку мови програмування Python та бібліотек, що потрібні для роботи з нейромережами. Крім бібліотеки Open AI також було використано бібліотеку Pandas [21] та бібліотеку SKLearn [22]. Для встановлення бібліотек було використано package installer for Python. Для встановлення бібліотеки потрібно ввести таку команди:

pip install pandas
pip install openai
pip install sklearn

Для завантаження даних з файлу використовується бібліотека pandas. Для її використання потрібно її імпортувати і скрипті таким чином:

import pandas as pd,

а також підключимо модулі потрібні для роботи з нейромережею

from sklearn.model_selection import train_test_split
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.naive_bayes import MultinomialNB
from sklearn.metrics import accuracy_score, confusion_matrix

та здійснити зчитування даних з файлу, що містяться у форматі csv рис. 3.4 а, у змінну data

data = pd.read_csv('emails.csv')

Після зчитування даних, їх потрібно розподілити у набори даних та визначити кількість токенів у кожному наборі за допомогою CountVecrorizer із бібліотеки SKLearn. Ці дії відбуваються наступним чином:

vectorizer = *CountVectorizer()*
counts = *vectorizer.fit_transform(data['EmailText'].values)*

З data вибирається перший рядок даних, що має назву «EmailText» та за допомогою функції fit_transform вивчається словниковий запас і повертається матриця термів документу.

Далі потрібно розділити дані на навчальні та тестові набори за допомогою функції train_test_split таким чином

X_train, X_test, y_train, y_test = train_test_split(counts, data['Label'], test_size=0.2, random_state=42)

Ця функція приймає на вхід 4 параметри: масив підготовлених даних (у цьому випадку опрацьовані повідомлення з електронної пошти), набір даних, що вказує тип повідомлення(“phishing” or “legitimate”), test_size визначає пропорцію набору даних, що включаються у тестовий розподіл і random_state – керує перетасуванням, застосованим до даних перед застосуванням розділення.

Для класифікації даних і навчання моделі було використано багаточленний наївний класифікатор Байєса (Multinomial Naive Bayes) — це версія наївного класифікатора Байєса, яка часто використовується для класифікації текстової інформації, зокрема для фільтрації спаму в електронній пошті чи аналізу тональності текстів. В нашому випадку він також підходить, оскільки

використовується робота з текстовими повідомленнями, в яких присутні певні тональності.

Основна ідея наївного класифікатора Байєса полягає у використанні теореми Байєса для розрахунку ймовірностей класів для даного набору ознак. "Наївний" у назві походить від припущення, що всі ознаки незалежні одна від одної, навіть якщо це не завжди відповідає реальній ситуації. Це робить алгоритм простим та ефективним для багатьох завдань.

У випадку багаточленного наївного класифікатора Байєса, вважається, що дані є мультиноміально розподіленими, що часто відповідає ситуаціям, коли маємо лічильні дані, наприклад, кількість входжень кожного слова у текстовому документі.

Використовуючи бібліотеку SKLearn створимо модель нейронної мережі та передаємо дані для навчання, які були отримано раніше:

```
model = MultinomialNB()
model.fit(X_train, y_train)
```

Крім MultinomialNB можна використовувати BernoulliNB або ComplementNB. BernoulliNB як і MultinomialNB, цей класифікатор підходить для дискретних даних. Різниця полягає в тому, що в той час як MultinomialNB працює з підрахунком входжень, BernoulliNB розроблений для двійкових/булевих функцій. Водночас, наївний байєсівський класифікатор Complement був розроблений для виправлення «суворих припущень», зроблених стандартним мультиноміальним наївним байєсовським класифікатором. Він особливо підходить для незбалансованих наборів даних.

Після навчання моделі отримаємо прогнози на основі тестових даних, як показано нижче

predictions = model.predict(X_test)

Для оцінювання точності роботи нейромережі використаємо функцію accuracy_score, якій на вхід передаємо «правильні» відповіді та прогнози отримані на попередньому кроці і виведемо значення у консоль:

print("Accuracy: ", accuracy_score(y_test, predictions))

Також потрібно побудувати матрицю невідповідності(confusion matrix). Ця матриця дозволяє візуалізувати роботу алгоритму. Кожен її рядок містить зразки прогнозованого класу, а стовпчики прогнозованого класу [23, 24]. Її можливо представити у вигляді як показано на рис. 3.5.

		Прогнозований клас	
	Загальна кількість = P + N	Predicted Positive (PP)	Predicted Negative (PN)
Справжній клас	Positive (P)	True positive (TP)	False negative (FN)
	Negative (N)	False positive (FP)	True negative (TN)

Рисунок 3.5 – Матриця невідповідності в узагальненому вигляді

Для нашого випадку матрицю невідповідності можна отримати таким чином:

print("Confusion Matrix: \n", confusion_matrix(y_test, predictions))

Після аналізу показника точності та матриці невідповідності, можна переходити до перевірки справжніх повідомлень для визначення їх типів. При цьому нейромережа може точно визначати тип повідомлення лише у випадку, коли її точність має значення більше 92%.

3.4 Тестування створених скриптів та аналіз результатів

3.4.1 Тестування скрипта для виявлення фішингових URL-адрес. Для навчання будемо використовувати дані, що містять 24250 фішингових та 24500 легітимних URL-адрес, що в цілому утворює 48750 записів для навчання. Дані отримано зі скрипта, що було проаналізовано у попередньому підрозділі із використанням 3-х схожих класифікаторів:

1) MultinomialNB;
2) BernoulliNB;
3) ComplementNB.

Також варто зазначити, що дані для аналізу було отримано для всіх 3-х класифікаторів з різною тестовою вибіркою(test_size): 0.2, 0.3, 0.4, 0.5. У результаті для кожного значення тестової вибірки і вибраного класифікатора було отримано матрицю невідповідності за допомогою функції confusion_matrix(y_test, predictions) та параметри точності класифікації. Розглянемо більш детально позначення, які використовуються для представлення отриманих даних:

True positive – правильно класифікована легітимна URL-адреса;

False positive – неправильно класифікована легітимна URL-адреса;

False negative – неправильно класифікована фішингова URL-адреса;

True negative – правильно класифікована фішингова URL-адреса;

Precision score legitimate – точність класифікації легітимних URL-адрес;

Precision score phising – точність класифікації фішингових URL-адрес;

Total score – загальна точність класифікації.

Для випадку класифікатора MultinomialNB отримані дані наведено в табл. 3.1. Отримані дані доцільно представити у графічному вигляді, як показано на рис. 3.6.

Таблиця 3.1 – Результати тестової вибірки для MultinomialNB класифікатора

Test size	True positive	False positive	False negative	True negative	Precision score legitimate	Precision score phishing	Total score
0.2	4250	150	300	5050	0,966	0,944	0,955
0.3	6637	152	647	7189	0,978	0,917	0,948
0.4	8752	351	752	9648	0,962	0,928	0,945
0.5	11750	447	903	11305	0,964	0,926	0,945

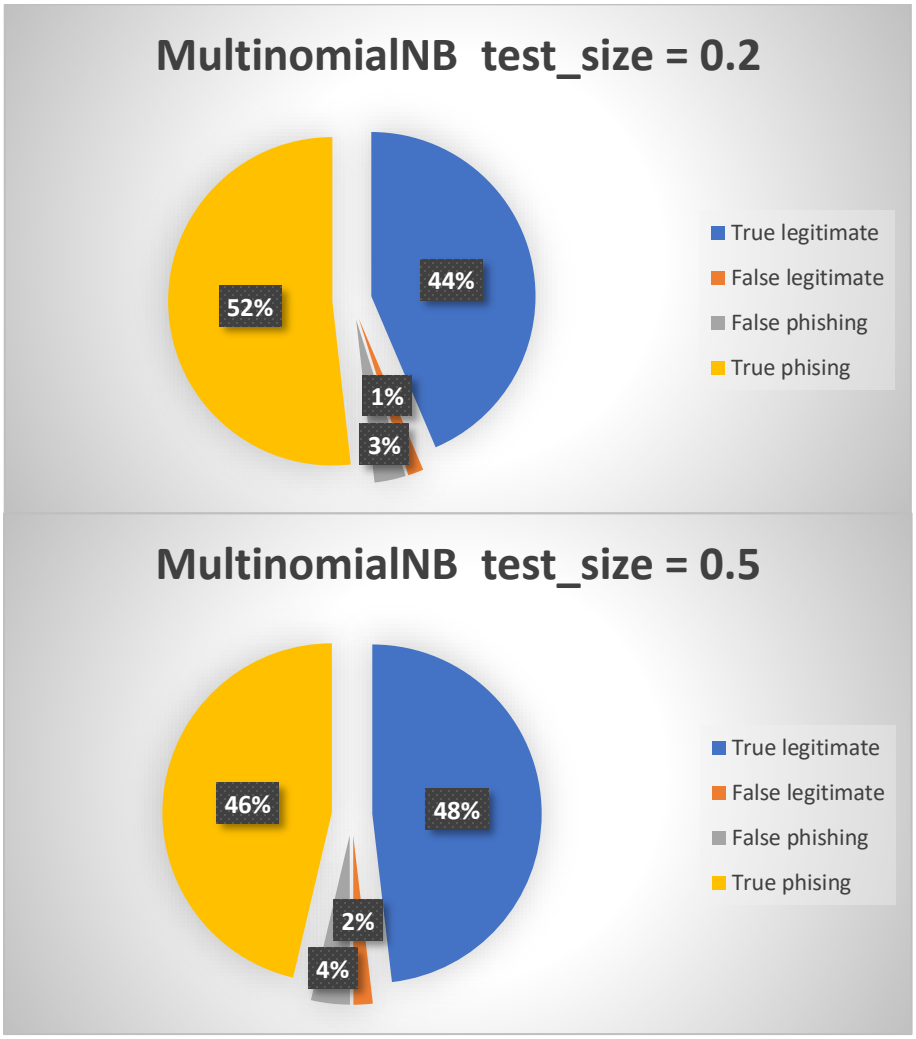


Рисунок 3.6 – Графічне представлення матриці невідповідності для різних значень тестової вибірки

З наведених графіків зрозуміло, що найкращі параметри точності отримано з тестовою вибіркою test_size = 0.2. При такому значенні, цікавим є те, що кількість правильно класифікованих фішингових адрес перевищило значення 50%, ще означає, що під час «перемішування» даних для навчання до тестової вибірки потрапило більше одного типу значень ніж іншого.

Дані у випадку використання BernoulliNB класифікатора показано у табл. 3.2.

Таблиця 3.2 –Результати тестової вибірки для BernoulliNB класифікатора

Test size	True positive	False positive	False negative	True negative	Precision score legitimate	Precision score phishing	Total score
0.2	4601	251	0	4898	1,0	0,948	0,974
0.3	6742	495	97	7291	0,987	0,932	0,959
0.4	9298	552	201	9449	0,979	0,944	0,962
0.5	11405	852	151	11998	0,987	0,933	0,959

На рис. 3.7 графічно представлено матрицю невідповідності для таблиці, наведеної вище.

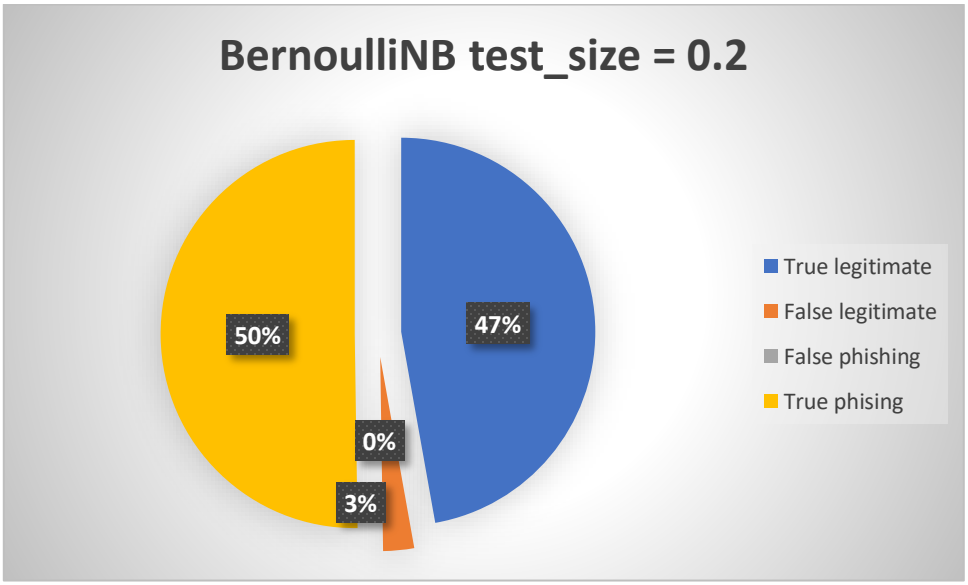


Рисунок 3.7 – Графічне представлення матриці невідповідності для BernoulliNB класифікатора

Таким чином, на відміну від попереднього класифікатора BernoulliNB краще визначає допустимі URL-адреси ніж фішингові, хоч і значення точності Precision score legitimate не набагато більше. Також було отримано дані і для випадку використання ComplementNB класифікатора, що поміщено у табл. 3.3.

Таблиця 3.3 –Результати тестової вибірки ComplementNB класифікатора

Test size	True positive	False positive	False negative	True negative	Precision score legitimate	Precision score phishing	Total score
0.2	4896	103	149	4602	0,969	0,979	0,974
0.3	7051	199	402	7005	0,946	0,973	0,959
0.4	9598	251	647	9003	0,933	0,975	0,954
0.5	11789	449	796	11340	0,934	0,963	0,949

На рис. 3.8 подано графічну інтерпретацію матриці невідповідності для ComplementNB класифікатора.

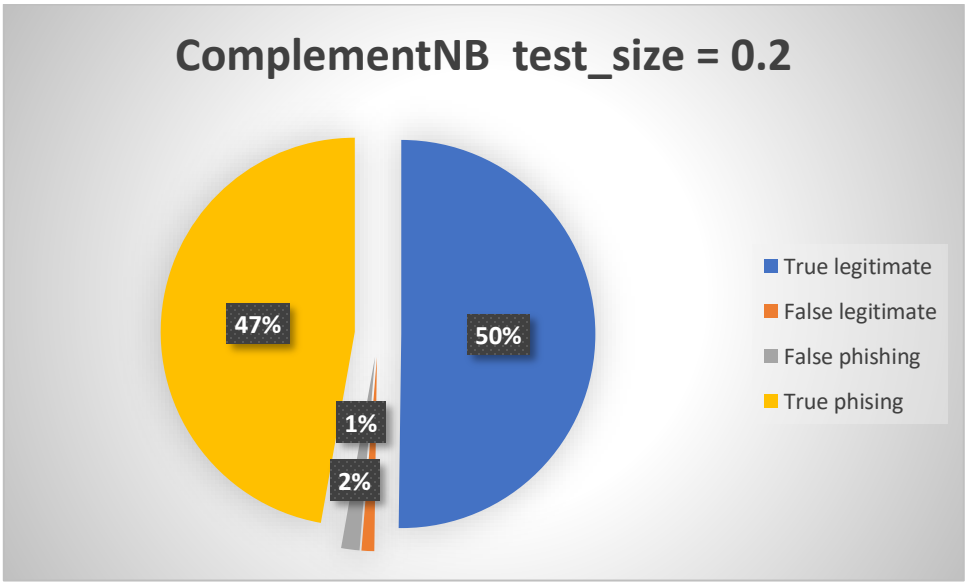


Рисунок 3.8 – Графічне представлення матриці невідповідності для ComplementNB класифікатора

Результати дослідження класифікаторів подано у табл. 3.4, яка дозволяє більш наочно порівняти їх ефективність роботи.

Таблиця 3.4 – Порівняння класифікаторів із тестовою вибіркою 0.2

Класифікатор	True positive	False positive	False negative	True negative	Precision score legitimate	Precision score phishing	Total score
MultinomialNB	4250	150	300	5050	0,966	0,944	0,955
BernoulliNB	4601	251	0	4898	1,0	0,948	0,974
ComplementNB	4896	103	149	4602	0,969	0,979	0,974

Аналізуючи значення наведені у попередній таблиці, можна зробити висновок, що BernoulliNB та ComplementNB працюють практично однаково і їх загальна точність ідентична. Водночас, ComplementNB має краще значення виявлення фішингових адрес, а BernoulliNB – легітимних.

3.4.2. Тестування скрипта для виявлення фішингових текстових повідомлень. Для навчання було отримано 10000 повідомлень допустимих та фішингових. Для класифікації повідомлень було використано BernoulliNB як такого, що показав одні з найкращих результатів. Результати роботи скрипта було занесено до табл. 3.5

Таблиця 3.5 –Результати тестової вибірки для BernoulliNB класифікатора

Test size	True positive	False positive	False negative	True negative	Precision score legitimate	Precision score phishing	Total score
0.2	946	104	79	871	0,901	0,917	0,909

Графічно інтерпретувати результати можна у вигляді як показано на рис. 3.9. Таким чином, у цьому розділі було здійснено розробку скриптів для ідентифікації фішингових атак, які з використанням засобів штучного інтелекту здатні класифікувати ці повідомлення на безпечні та фішингові, за рахунок аналізу текстового повідомлення та присутньої в ньому URL-адреси із точністю більше 90%. Це дозволяє отримати додатковий бар'єр безпеки та зменшити вірогідність втрати цінних даних чи певних активів.

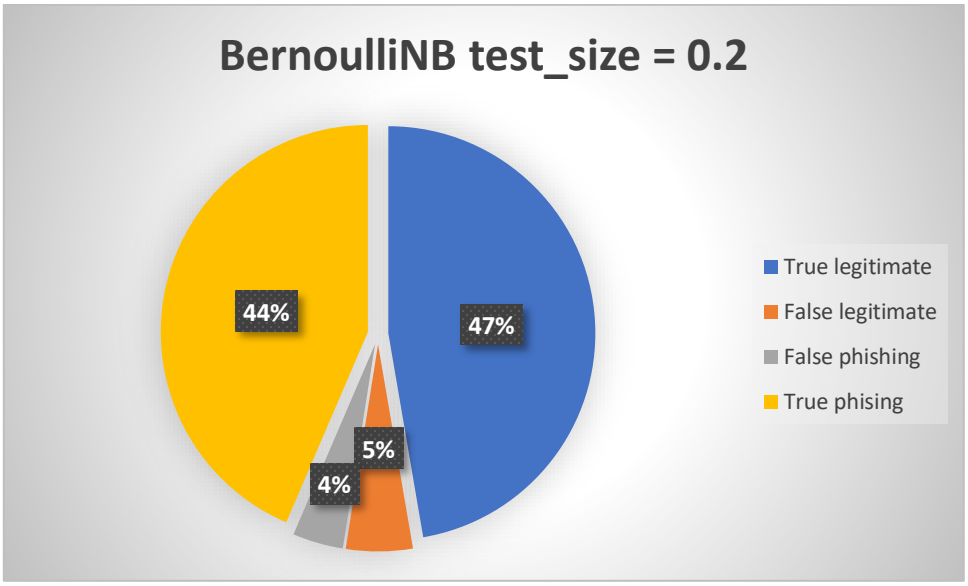


Рисунок 3.9 – Графічне представлення матриці невідповідності для BernoulliNB класифікатора

Водночас, варто зазначити, що такий підхід не може гарантувати 100% результат, оскільки останнім бар'єром у захисті залишається саме користувач і саме його дії призводять до відповідних втрат. Таким чином, крім використання описаного підходу у цій роботі, потрібно здійснювати постійне навчання користувачів оскільки з року в рік фішингові атаки стають все більш витонченими і час від часу досягають своєї мети.

4 ЕКОНОМІЧНА ЧАСТИНА

4.1 Комерційний та технологічний аудит науково-технічної розробки

Метою даного розділу є проведення технологічного аудиту, в даному випадку нового програмного модулю для виявлення фішингових атак у електронній пошті. Метою розробки програмного модулю для ідентифікації фішингових атак є захист користувацьких даних від витоку та блокування шкідливого контенту.

За аналог можна вибрати Email Security Defender за ціною 5403 грн.

Для проведення комерційного та технологічного аудиту залучено 3-х незалежних експертів з кібербезпеки:

1) Войтович О. П. – доцент кафедри захисту інформації;

2) Каплун В. А. – старший викладач кафедри захисту інформації;

3) Лукічов В. В. – доцент кафедри захисту інформації.

Оцінювання науково-технічного рівня розробки та її комерційного потенціалу рекомендується здійснювати із застосуванням п’ятибальної системи оцінювання за 12-ма критеріями, у відповідності із табл. 4.1.

Таблиця 4.1 – Рекомендовані критерії оцінювання комерційного потенціалу розробки та їх можлива бальна оцінка

Бали (за 5-ти бальною шкалою)					
Кри-терій	0	1	2	3	4
Технічна здійсненність концепції					
1	Достовірність концепції не підтверджена	Концепція підтверджена експертними висновками	Концепція підтверджена розрахунками	Концепція перевірена на практиці	Перевірено роботоздат-ність продукту в реальних умовах

Продовження табл. 4.1

Ринкові переваги					
2	Багато аналогів на малому ринку	Ринкові п Мало аналогів на малому ринку	Кілька аналогів на великому ринку	Один аналог на великому ринку	Продукт не має аналогів на великому ринку
3	Ціна продукту значно вища за ціни аналогів	Ціна продукту дещо вища за ціни аналогів	Ціна продукту приблизно дорівнює цінам аналогів	Ціна продукту дещо нижче за ціни аналогів	Ціна продукту значно нижче за ціни аналогів
4	Технічні та споживчі властивості продукту значно гірші, ніж в аналогів	Технічні та споживчі властивості продукту трохи гірші, ніж в аналогів	Технічні та споживчі властивості продукту на рівні аналогів	Технічні та споживчі властивості продукту трохи кращі, ніж в аналогів	Технічні та споживчі властивості продукту значно кращі, ніж в аналогів
5	Експлуатаційні витрати значно вищі, ніж в аналогів	Експлуатаційні витрати дещо вищі, ніж в аналогів	Експлуатаційні витрати на рівні експлуатаційних витрат аналогів	Експлуатаційні витрати трохи нижчі, ніж в аналогів	Експлуатаційні витрати значно нижчі, ніж в аналогів
Ринкові перспективи					
6	Ринок малий і не має позитивної динаміки	Ринок малий, але має позитивну динаміку	Середній ринок з позитивною динамікою	Великий стабільний ринок	Великий ринок з позитивною динамікою
7	Активна конкуренція великих компаній на ринку	Активна конкуренція	Помірна конкуренція	Незначна конкуренція	Конкурентів немає
Практик на здійсненність					
8	Відсутні фахівці як з технічної, так і з комерційної реалізації ідеї	Необхідно наймати фахівців або витрачати значні кошти та час на навчання наявних фахівців	Необхідне незначне навчання фахівців та збільшення їх штату	Необхідне незначне навчання фахівців	Є фахівці з питань як з технічної, так із комерційної реалізації ідеї
9	Потрібні значні фінансові ресурси, які відсутні. Джерела фінансування ідеї відсутні	Потрібні незначні фінансові ресурси. Джерела фінансування відсутні	Потрібні значні фінансові ресурси. Джерела фінансування є	Потрібні незначні фінансові ресурси. Джерела фінансування є	Не потребує додаткового фінансування
10	Необхідна розробка нових матеріалів	Потрібні матеріали, що використовуються у військово-промисловому комплексі	Потрібні дорогі матеріали	Потрібні досяжні та дешеві матеріали	Всі матеріали для реалізації ідеї відомі та давно використовуються у виробництві

Продовження табл. 4.1

11	Термін реалізації ідеї більший за 10 років	Термін реалізації ідеї більший за 5 років. Термін окупності інвестицій більше 10-ти років	Термін реалізації ідеї від 3-х до 5-ти років. Термін окупності інвестицій більше 5-ти років	Термін реалізації ідеї менше 3-х років. Термін окупності інвестицій від 3-х до 5-ти років	Термін реалізації ідеї менше 3-х років. Термін окупності інвестицій менше 3-х років
12	Необхідна розробка регламентних документів та отримання великої кількості дозвільних документів на виробництво та реалізацію продукту	Необхідно отримання великої кількості дозвільних документів на виробництво та реалізацію продукту, що вимагає значних коштів та часу	Процедура отримання дозвільних документів для виробництва та реалізації продукту вимагає незначних коштів та часу	Необхідно тільки повідомлення відповідним органам про виробництво та реалізацію продукту	Відсутні будь-які регламентні обмеження на виробництво та реалізацію продукту

Усі дані по кожному параметру занесено в таблиці 4.2

Таблиця 4.2 – Результати оцінювання комерційного потенціалу розробки

Критерії оцінювання	ПІБ експертів		
	Войтович О. П.	Каплун В. А.	Лукічов В. В.
	Бали		
Технічна здійсненність концепції	3	3	4
Наявність аналогів на ринку	3	3	4
Цінова політика	4	3	3
Технічні та споживчі властивості виробу	3	4	4
Експлуатаційні витрати	4	3	3
Ринок збуту	3	4	4
Конкурентоспроможність	4	3	3
Фахівці з технічної і комерційної реалізації	3	4	3
Фінансування	4	4	3
Матеріально-технічна база	4	3	3
Термін реалізації ідеї	3	4	3
Супровідна документація	3	3	4
Сума	41	41	41
Середньоарифметична сума балів	(41+41+41) / 3 = 41		

За даними таблиці 4.2 можна зробити висновок щодо рівня комерційного потенціалу даної розробки. Для цього доцільно скористатись рекомендаціями, наведеними в таблиці 4.3.

Таблиця 4.3 - Рівні комерційного потенціалу розробки

Середньоарифметична сума балів СБ , розрахована на основі висновків експертів	Рівень комерційного потенціалу розробки
0-10	Низький
11-20	Нижче середнього
21-30	Середній
31-40	Вище середнього
41-48	Високий

Як видно з таблиці, рівень комерційного потенціалу розроблюваного нового програмного продукту є високим, що досягається за рахунок того, що програмний продукт відрізняється від існуючих тим, що дана технологія є програмним засобом для виявлення фейкової інформації у соцмережах, який систематизує загальні підходи до опису і формалізації фейків, на основі яких визначено основні способи поширення фейкового контенту в соціальних мережах.

4.2 Прогнозування витрат на виконання науково-дослідної (дослідно-конструкторської) роботи

4.2.1 Витрати на оплату праці. Основна заробітна плата розробників, яка розраховується за формулою:

$$З_о = \frac{M}{T_р} \cdot t, \tag{4.1}$$

де М – місячний посадовий оклад конкретного розробника (дослідника), грн.;

$Т_р$ – число робочих днів в місяці, 23 днів;

t – число днів роботи розробника (дослідника).

Результати розрахунків зведемо до таблиці 4.4.

Таблиця 4.4 – Основна заробітна плата розробників

Найменування посади	Місячний посадовий оклад, грн.	Оплата за робочий день, грн.	Число днів роботи	Витрати на заробітну плату, грн.
Керівник проекту	12500	543,48	5	2717,4
Програміст	15000	652,17	30	19565,1
Всього				22282,5

Так як в даному випадку розробляється програмний продукт, то розробник виступає одночасно і основним робітником, і тестувальником розроблюваного програмного продукту.

4.2.2 Додаткова заробітна плата розробників, які приймали участь в розробці обладнання. Додаткова заробітна плата прийнято розраховувати як 12 % від основної заробітної плати розробників та робітників:

$$З_{дод} = З_о \cdot Н_{дод} / 100\ \% \qquad (4.2)$$

$З_{дод} = (22282,5 \cdot 12\ \% / 100\ \%) = 2673,9$ (грн.)

4.2.3 Відрахування на соціальні заходи. Згідно діючого законодавства нарахування на заробітну плату складають 22 % від суми основної та додаткової заробітної плати.

$$Н_з = (З_о + З_д) \cdot 22\ \% / 100\% \qquad (4.3)$$

$Н_з$ = (22282,5 + 2673,9) · 22 % / 100 % = 24956,4 (грн.)

4.2.4 Сировина і матеріали. Витрати на матеріали (М) у вартісному вираженні розраховуються окремо для кожного виду матеріалів за формулою:

$$М = \sum_{j=1}^{n} H_j \cdot Ц_j \cdot K_j - \sum_{j=1}^{n} В_j \cdot Ц_{вj}.$$

Використаємо формулу наведену вище та занесемо дані у табл. 4.2.

Таблиця 4.2 – Витрати на матеріали

Найменування матеріалу, марка, тип, сорт	Ціна за 1 кг, грн	Норма витрат, кг	Величина відходів, кг	Ціна відходів, грн/кг	Вартість витраченого матеріалу, грн
Бумага А4, 80 гр/м^2	70	2,5	0,5	3	199,75
Тонер для друку BROTHER TN-1075	2875	0,04	-	-	132,25
Всього					332

4.2.5 Розрахунок витрат на комплектуючі. Оскільки для розроблювального пристрою не потрібно витрачати матеріали та комплектуючи, то витрати на матеріали і комплектуючі дорівнюють нулю.

4.2.6 Програмне забезпечення для наукових(експериментальних) робіт. Для створення комплексу використовувалися лише безкоштовне програмне забезпечення, тому витрати такого типу відсутні.

4.2.7 Амортизація обладнання, яке використовувалось для проведення розробки. Амортизація обладнання, що використовувалось для розробки в спрощеному вигляді амортизація обладнання, що використовувалась для розробки розраховується за формулою:

$$A = \frac{Ц}{Тв} \cdot \frac{t_{вик}}{12} \text{ [грн.].} \quad (4.4)$$

де Ц – балансова вартість обладнання, грн.;

Т – термін корисного використання обладнання згідно податкового законодавства, років

$t_{вик}$ – термін використання під час розробки, місяців

Розрахуємо, для прикладу, амортизаційні витрати на комп'ютер балансова вартість якого становить 25000 грн., термін його корисного використання згідно податкового законодавства – 2 роки, а термін його фактичного використання – 1 міс.

$$A_{обл} = \frac{25000}{2} \times \frac{1}{12} = 1041{,}66 \; грн.$$

Аналогічно визначаємо амортизаційні витрати на інше обладнання та приміщення. Розрахунки заносимо до таблиці 4.2. Для розрахунку амортизації нематеріальних ресурсів використовується формула:

$$A_{н.р.} = Ц_{н.р.} * Н_а * \frac{t_{вик.}}{12} \quad (4.5)$$

Але, так як вартість ліцензійної ОС та спеціалізованих ліцензійних нематеріальних ресурсів менше 20000 грн, то даний нематеріальний актив (Microsoft Windows 10) не амортизується, а його вартість включається у вартість розробки повністю, $В_{нем.ак.}$ = 1100 грн.

Таблиця 4.5 – Амортизаційні відрахування матеріальних і нематеріальних ресурсів для розробників

Найменування обладнання	Балансова вартість, грн.	Строк корисного використання, років	Термін використання обладнання, місяців	Амортиза-ційні відрахуван-ня, грн.
Комп'ютер та комп'ютерна периферія (Ноутбук ASUS)	25000	2	1	1041,66
Офісне обладнення (меблі)	20000	4	1	416,66
Приміщення	300000	20	1	1250,00
Всього				2708,32

4.2.8 Витрати на електроенергію. Тарифи на електроенергію для непобутових споживачів (промислових підприємств) відрізняються від тарифів на електроенергію для населення. При цьому тарифи на розподіл електроенергії у різних постачальників (енергорозподільних компаній), будуть різними. Крім того, розмір тарифу залежить від класу напруги (1-й або 2-й клас). Тарифи на розподіл електроенергії для всіх енергорозподільних компаній встановлює Національна комісія з регулювання енергетики і комунальних послуг (НКРЕКП). Витрати на силову електроенергію розраховуються за формулою:

$$В_е = В \cdot П \cdot Ф \cdot К_п, \quad (4.6)$$

де В – вартість 1 кВт-години електроенергії для 1 класу підприємства, В = 7,5 грн./кВт;

П – встановлена потужність обладнання, кВт. П = 0,4 кВт;

Ф – фактична кількість годин роботи обладнання, годин.

$К_п$ – коефіцієнт використання потужності, $К_п$ = 0,9.

$$В_е = 0,9 \cdot 0,4 \cdot 8 \cdot 30 \cdot 7,5 = 648 \text{ (грн.)}$$

4.2.9 Інші витрати та загальновиробничі витрати. До статті «Інші витрати» належать витрати, які не знайшли відображення у зазначених статтях витрат і можуть бути віднесені безпосередньо на собівартість досліджень за прямими ознаками. Витрати за статтею «Інші витрати» розраховуються як 50…100% від суми основної заробітної плати дослідників:

$$I_в = (З_о + З_р) \cdot \frac{Н_{ів}}{100\%}, \quad (4.7)$$

де $Н_{ів}$ – норма нарахування за статтею «Інші витрати».

$I_в$ = 22282,5 * 75% / 100% = 16711,88 (грн.)

До статті «Накладні (загальновиробничі) витрати» належать: витрати, пов'язані з управлінням організацією; витрати на винахідництво та раціоналізацію; витрати на підготовку (перепідготовку) та навчання кадрів; витрати, пов'язані з набором робочої сили; витрати на оплату послуг банків; витрати, пов'язані з освоєнням виробництва продукції; витрати на науково-технічну інформацію та рекламу та ін. Витрати за статтею «Накладні (загальновиробничі) витрати» розраховуються як 100…150% від суми основної заробітної плати дослідників:

$$Н_{нзв} = (З_о + З_р) \cdot \frac{Н_{нзв}}{100\%}, \quad (4.8)$$

де $Н_{нзв}$ – норма нарахування за статтею «Накладні (загальновиробничі) витрати».

$Н_{нзв}$ = 22282,5 * 150 % / 100 % = 33423,75 (грн.)

4.2.10 Витрати на проведення науково-дослідної роботи.

Сума всіх попередніх статей витрат дає загальні витрати на проведення науково-дослідної роботи:

$В_{заг}$ = 22282,5+2673,9+24956,4+332+2708,32+1100+648+16711,88+

+33423,75 = 104836,75 грн.

4.2.11 Розрахунок загальних витрат на науково-дослідну (науково-технічну) роботу та оформлення її результатів.

Загальні витрати на завершення науково-дослідної (науково-технічної) роботи та оформлення її результатів розраховуються *ЗВ*, визначається за формулою:

$$ЗВ = \frac{В_{заг}}{\eta} \text{ (грн)}, \qquad (4.9)$$

де η – коефіцієнт, який характеризує етап (стадію) виконання науково-дослідної роботи.

Так, якщо науково-технічна розробка знаходиться на стадії: науково-дослідних робіт, то η=0,1; технічного проектування, то η=0,2; розробки конструкторської документації, то η=0,3; розробки технологій, то η=0,4; розробки дослідного зразка, то η=0,5; розробки промислового зразка, то η=0,7; впровадження, то η=0,9. Оберемо η = 0,5, так як розробка, на даний момент, знаходиться на стадії дослідного зразка:

$ЗВ$ = 104836,75 / 0,5 = 209673,5 грн.

4.3 Розрахунок економічної ефективності науково-технічної розробки за її можливої комерціалізації потенційним інвестором

В ринкових умовах узагальнювальним позитивним результатом, що його може отримати потенційний інвестор від можливого впровадження результатів тієї чи іншої науково-технічної розробки, є збільшення у потенційного інвестора величини чистого прибутку. Саме зростання чистого прибутку забезпечить потенційному інвестору надходження додаткових коштів, дозволить покращити фінансові результати його діяльності, підвищить конкурентоспроможність та може позитивно вплинути на ухвалення рішення щодо комерціалізації цієї розробки.

Для того, щоб розрахувати можливе зростання чистого прибутку у потенційного інвестора від можливого впровадження науково-технічної розробки необхідно:

а) вказати, з якого часу можуть бути впроваджені результати науково-технічної розробки;

б) зазначити, протягом скількох років після впровадження цієї науково-технічної розробки очікуються основні позитивні результати для потенційного інвестора (наприклад, протягом 3-х років після її впровадження);

в) кількісно оцінити величину існуючого та майбутнього попиту на цю або аналогічні чи подібні науково-технічні розробки та назвати основних суб'єктів (зацікавлених осіб) цього попиту;

г) визначити ціну реалізації на ринку науково-технічних розробок з аналогічними чи подібними функціями.

При розрахунку економічної ефективності потрібно обов'язково враховувати зміну вартості грошей у часі, оскільки від вкладення інвестицій до отримання прибутку минає чимало часу. При оцінюванні ефективності інноваційних проектів передбачається розрахунок таких важливих показників:

- абсолютного економічного ефекту (чистого дисконтованого доходу);

- внутрішньої економічної дохідності (внутрішньої норми дохідності);
- терміну окупності (дисконтованого терміну окупності).

Аналізуючи напрямки проведення науково-технічних розробок, розрахунок економічної ефективності науково-технічної розробки за її можливої комерціалізації потенційним інвестором можна об'єднати, враховуючи визначені ситуації з відповідними умовами.

4.3.1 Розробка чи суттєве вдосконалення програмного засобу (програмного забезпечення, програмного продукту) для використання масовим споживачем.

В цьому випадку майбутній економічний ефект буде формуватися на основі таких даних:

$$\Delta П_i = (\pm \Delta Ц_0 \cdot N + Ц_0 \cdot \Delta N)_i \cdot \lambda \cdot \rho \cdot (1 - \frac{\vartheta}{100}), \qquad (4.10)$$

де $\pm\Delta Ц_o$ – зміна вартості програмного продукту (зростання чи зниження) від впровадження результатів науково-технічної розробки в аналізовані періоди часу;

N – кількість споживачів які використовували аналогічний продукт у році до впровадження результатів нової науково-технічної розробки;

$Ц_o$ – основний оціночний показник, який визначає діяльність підприємства у даному році після впровадження результатів наукової розробки, $Ц_o = Ц_б \pm \Delta Ц_o$;

$Ц_б$ – вартість програмного продукту у році до впровадження результатів розробки;

ΔN – збільшення кількості споживачів продукту, в аналізовані періоди часу, від покращення його певних характеристик;

λ – коефіцієнт, який враховує сплату податку на додану вартість. Ставка податку на додану вартість дорівнює 20%, а коефіцієнт $\lambda = 0{,}8333$.

p – коефіцієнт, який враховує рентабельність продукту;

ϑ – ставка податку на прибуток, у 2022 році ϑ =18%.

Припустимо, що при прогнозованій ціні 1500 грн. за одиницю виробу, термін збільшення прибутку складе 3 роки. Після завершення розробки і її вдосконалення, можна буде підняти її ціну на *500* грн. Кількість одиниць реалізованої продукції також збільшиться: протягом першого року – на 5000 шт., протягом другого року – на 4000 шт., протягом третього року на 3000 шт. До моменту впровадження результатів наукової розробки реалізації продукту не було:

$\Delta П_1 = (0*500 + (1500 + 500)*5000)* 0{,}8333* 0{,}25) * (1 - 0{,}18) = 1701500$ грн.

$\Delta П_2 = (0*500 + (1500 + 500)*(5000+4000)* 0{,}8333* 0{,}25) * (1 - 0{,}18) = 3062700$ грн.

$\Delta П_3 = (0*500 + (1500 + 500)*(5000+4000+3000)* 0{,}8333* 0{,}25) * (1 - 0{,}18) = 4083600$ грн.

$$\Delta П_{заг} = \sum_{i=1}^{3} \Delta П_i = 1701500 + 3062700 + 4083600 = 8847800 \text{ грн.}$$

Отже, комерційний ефект від реалізації результатів розробки за три роки складе 8847800 грн.

4.3.2 Розрахунок ефективності вкладених інвестицій та періоду їх окупності.

Розраховуємо приведену вартість збільшення всіх чистих прибутків *ПП*, що їх може отримати потенційний інвестор від можливого впровадження та комерціалізації науково-технічної розробки:

$$ПП = \sum_{1}^{T} \frac{\Delta П_i}{(1+\tau)^t}, \qquad (4.11)$$

де $\Delta П_i$ – збільшення чистого прибутку у кожному із років, протягом яких виявляються результати виконаної та впровадженої науково-дослідної (науково-технічної) роботи, грн;

Т – період часу, протягом якою виявляються результати впровадженої науково-дослідної (науково-технічної) роботи, роки;

τ – ставка дисконтування, за яку можна взяти щорічний прогнозований рівень інфляції в країні, $\tau = 0,05...0,15$;

t – період часу (в роках).

Збільшення прибутку отримаємо починаючи з першого року:

ПП= $(1701500/(1+0,1)^1)+(3062700/(1+0,1)^2)+(4083600/(1+0,1)^3)$ = 1546818,18 + 2531157,02 + 3068069,12 = 7146044,32 грн.

Далі розраховують величину початкових інвестицій *PV*, які потенційний інвестор має вкласти для впровадження і комерціалізації науково-технічної розробки. Для цього можна використати формулу:

$$PV = k_{інв} * ЗВ, \quad (4.12)$$

де $k_{інв}$ – коефіцієнт, що враховує витрати інвестора на впровадження науково-технічної розробки та її комерціалізацію. Це можуть бути витрати на підготовку приміщень, розробку технологій, навчання персоналу, маркетингові заходи тощо; зазвичай $k_{інв}=2...5$, але може бути і більшим;

ЗВ – загальні витрати на проведення науково-технічної розробки та оформлення її результатів, грн.

$$PV = 2 * 209673,5 = 419347 \text{ грн.}$$

Тоді абсолютний економічний ефект $E_{абс}$ або чистий приведений дохід *(NPV, Net Present Value)* для потенційного інвестора від можливого впровадження та комерціалізації науково-технічної розробки становитиме:

$$E_{абс} = ПП - PV, \tag{4.13}$$

$E_{абс}$ = 7146044,32 – 419347 = 6726697,32 грн.

Оскільки $E_{абс} > 0$ то вкладання коштів на виконання та впровадження результатів даної науково-дослідної (науково-технічної) роботи може бути доцільним.

Для остаточного прийняття рішення з цього питання необхідно розра-хувати внутрішню економічну дохідність або показник внутрішньої но-рми дохідності *(IRR, Internal Rate of Return)* вкладених інвестицій та порівняти її з так званою бар'єрною ставкою дисконтування, яка визначає ту мінімальну внутрішню економічну дохідність, нижче якої інвестиції в будь-яку науково-технічну розробку вкладати буде економічно недоцільно.

Розрахуємо відносну (щорічну) ефективність вкладених в наукову розробку інвестицій $E_в$. Для цього використаємо формулу:

$$E_в = \sqrt[T_ж]{1 + \frac{E_{абс}}{PV}} - 1, \tag{4.14}$$

$T_ж$ – життєвий цикл наукової розробки, роки.

$$E_в = \sqrt[3]{1 + 6726697{,}32/419347} - 1 = 1{,}57$$

Визначимо мінімальну ставку дисконтування, яка у загальному вигляді визначається за формулою:

$$\tau = d + f, \tag{4.15}$$

де d – середньозважена ставка за депозитними операціями в комерційних банках; в 2023 році в Україні d = (0,09…0,14);

f – показник, що характеризує ризикованість вкладень; зазвичай, величина f = (0,05...0,5).

$$\tau_{\min} = 0,14 + 0,05 = 0,19.$$

Так як $E_в > \tau_{min}$, то інвестор може бути зацікавлений у фінансуванні даної наукової розробки.

Розрахуємо термін окупності вкладених у реалізацію наукового проекту інвестицій за формулою:

$$T_{ок} = \frac{1}{E_в}, \tag{4.16}$$

$T_{ок}$ = 1 / 1,57 = 0,63 р.

Оскільки $T_{ок}$ < 3-х років, а саме термін окупності рівний 0,63 роки, то фінансування даної наукової розробки є доцільним.

Висновки до розділу: економічна частина даної роботи містить розрахунок витрат на розробку нового програмного продукту, сума яких складає 104836,75 гривень. Було спрогнозовано орієнтовану величину витрат по кожній з статей витрат. Також розраховано чистий прибуток, який може отримати виробник від реалізації нового технічного рішення, розраховано період окупності витрат для інвестора та економічний ефект при використанні даної розробки. В результаті аналізу розрахунків можна зробити висновок, що розроблений програмний продукт за ціною дешевший за аналог і є висококонкурентоспроможним. Період окупності складе близько 0,63 роки.

ВИСНОВКИ

Здійснено аналіз кіберзагроз за останній рік із використанням даних від Держспецзв'язку. Аналіз показав, що за останній рік кількість фішингових атак на державні та приватні підприємства зросла у 2-4 рази і такий від атак є досить небезпечним, оскільки зазвичай призводить до втрати цінної інформації.

Виконаний аналіз літературних джерел щодо існуючих методів і стратегій фішингових атак показав, що існує велика кількість таких підходів і вони використовуються різноманітні особливості та вразливості комп'ютерних мереж для проведення атак. Крім того, напрямок фішингових атак є досить актуальним незважаючи на те, що він не новий і з кожним роком з'являється все більше і більше підходів та засобів для боротьби з фішинговими атаками.

Визначено, що бурхливий розвиток штучного інтелекту, зокрема, великих мовних моделей несуть такі ризики: 1) покращення якості написання фішингових текстів; 2) зниження технічного порогу входження для зловмисників, оскільки великі мовні моделі можуть створювати сайти за побажанням користувача та поліпшувати їх код, які в свою чергу можуть бути використані для викрадення даних користувачів. 3) прискорення певних фаз фішингових атак та збільшення потенційної шкоди. Таким чином, зловмисники отримали додатковий інструмент для своєї діяльності, який здатен навчатися та покращуватися при взаємодії з користувачами.

Подальшого розвитку отримав метод виявлення фішингових атак із використанням штучного інтелекту, який дозволяє підвищити ефективність виявлення фішингових текстових повідомлень та URL-адрес за рахунок використання фішингових онлайн сервісів і штучного інтелекту.

Здійснено експериментальні дослідження запропонованих засобів ідентифікації фішингових атак, підготовлено дані для навчання штучного інтелекту: для розпізнавання URL-адрес – 48750, а для розпізнавання фішингових

текстових повідомлень – 10000. Підготовлені дані та відлагоджені скрипти дозволили здійснювати визначення фішингових URL-адрес із вірогідністю 95-97%, а текстових повідомлень – 89-90%.

Слід відзначити, що не можна повністю покладатися на будь-які засоби боротьби із фішинговими атаками, оскільки жоден з них не гарантує 100% захисту від таких атак. Водночас, для підвищення захисту потрібно використовувати усі можливі засоби в тому числі і навчання користувачів, оскільки саме вони є останнім бар’єром захисту і більшості випадків можуть призвести до втрати цінних даних.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Моніторинг дотримання цифрових прав. Сайт громадської організації «Платформа прав людини» URL: https://www.ppl.org.ua/ monitoring/monitoring-cifrovix-prav (дата звернення: 08.11.2023).
2. Сайт Держаної служби спеціального зв'язку та захисту інформації України URL: https://cip.gov.ua/ua (дата звернення: 08.11.2023).
3. Шахрайські та фішингові сайти. Аналіз, тренди та рекомендації для клієнтів, 2022/2023 URL: https://www.ema.com.ua/citizens/cyber-safety-school/ shahrajski-ta-fishingovi-sajti-analiz-trendi-ta-rekomendacii-dlja-kliientiv-2022-2023/ (дата звернення: 08.11.2023).
4. Merwe, A. v. d., Marianne, L., and Marek, D. (2005). "Characteristics and responsibilities involved in a Phishing attack, in WISICT '05: proceedings of the 4th international symposium on information and communication technologies. Trinity College Dublin, 249–254 p.
5. Phishing Attacks: A Recent Comprehensive Study and a New Anatomy. URL: https://www.frontiersin.org/articles/10.3389/fcomp.2021.563060/full (дата звернення: 08.11.2023).
6. Що таке фішинг? Microsoft official site. URL: https://www.microsoft.com/ uk-ua/security/business/security-101/what-is-phishin(дата звернення: 08.11.2023).
7. APPLICATION IDENTIFICATION FOR PHISHING DETECTION: Pat. US2023344866 (A1); H04L51/21; H04L61/5007; app. Numb. US202217729723 20220426H04L9/40; prior. numb: US202217729723 20220426.
8. Method of Detect an Email Phishing Attempt or Fraudulent Email Within an Email Domain: Pat. US2023291767 (A1); G06Q10/107; H04L9/40; app. numb. US202318117456 20230305; prior. numb. US202318117456 20230305.
9. WEBPAGE PHISHING AUTO-DETECTION: Pat. US2023344868 (A1); H04L9/40; app. numb. US202318309240 20230428; prior. numb. US202318309240 20230428.
10. Кравченко, В., Руденко, О., Доманов, І. і Казначей, С. (2022) «АНАЛІЗ ФІШИНГ-АТАК. ДОСЛІДЖЕННЯ МЕТОДІВ ЗАПОБІГАННЯ ТА ЗАХИСТУ», Збірник наукових праць Державного науково-дослідного інституту випробувань і сертифікації озброєння та військової техніки, 11(1), с. 85-95. doi: 10.37701/dndivsovt.11.2022.10.
11. Advanced protection to safeguard your inboxes. Cisco official site. URL: https://www.cisco.com/site/us/en/products/security/secure-email/index.html (дата звернення: 08.11.2023)
12. Максимальна довжина URL-адреси в Internet Explorer – 2083 символи. Microsoft Support Site. URL: https://support.microsoft.com/uk-ua/topic/%D0%BC%D0%B0%D0%BA%D1%81%D0%B8%D0%BC%D0%B0

%D0%BB%D1%8C%D0%BD%D0%B0-%D0%B4%D0%BE%D0%B2
%D0%B6%D0%B8%D0%BD%D0%B0-url-%D0%B0%D0%B4%D1%80
%D0%B5%D1%81%D0%B8-%D0%B2-internet-explorer-2083-
%D1%81%D0%B8%D0%BC%D0%B2%D0%BE%D0%BB%D0%B8-
174e7c8a-6666-f4e0-6fd6-908b53c12246 (дата звернення: 08.11.2023).

13. Detection of Phishing Attacks: A Machine Learning Approach. [Електронний ресурс] / Ram Basnet, Srinivas Mukkamala, Andrew H. Sung // 85 Researchgate, 2008. Режим доступу до ресурсу: https://www.researchgate.net/publication/226420039_Detection_of_Phishing_Attacks_A_Machine_Learning_Approach.
14. Benavides-Astudillo, E.; Fuertes, W.; Sanchez-Gordon, S.; Nuñez-Agurto, D.; Rodríguez-Galán, G. A Phishing-Attack-Detection Model Using Natural Language Processing and Deep Learning. Appl. Sci. 2023, 13, 5275. https://doi.org/10.3390/app13095275.
15. What Is ChatGPT Doing … and Why Does It Work? URL: https://writings.stephenwolfram.com/2023/02/what-is-chatgpt-doing-and-why-does-it-work/(дата звернення: 08.11.2023).
16. Гарнага В. А. Neural networks in phishing attacks. Молодь в науці: дослідження, проблеми, перспективи 2023: матеріали Всеукраїнської науково-практичної інтернет-конференції / Він. нац. техн. ун-т., Вінниця, 2023.
17. Nikhil Ketkar. Deep Learning with Python: A Hands-on Introduction. Apres, 2017, 226 p.
18. Aurélien Géron. Hands-On Machine Learning with ScikitLearn and TensorFlow. O'Reilly Media, 2017, 751 p.
19. Аллен Б.Дауни. Think DSP. Цифровая обработка сигналов на Python, 2017, 160 с.
20. Welcome to the OpenAI developer platform. Open AI API reference. URL https://platform.openai.com/docs/overview(дата звернення: 15.11.2023).
21. Pandas. Pandas official site. URL: https://pandas.pydata.org/ (дата звернення 17.11.2023)
22. Scikit-learn Machine Learning in Python. SKLearn site. URL https://scikit-learn.org/stable/index.html# (дата звернення 17.11.2023)
23. Сперкач М. О., Юзьвак Д. Ю. Розв'язання задачі класифікації текстів методами обробки природньої мови та машинного навчання // Науковий огляд №4 (57), 2019. 11 с.
24. Stehman, Stephen V. (1997). Selecting and interpreting measures of thematic classification accuracy. Remote Sensing of Environment 62 (1): 77–89 p. doi:10.1016/S0034-4257(97)00083-7.

ДОДАТКИ

ТОВАРИСТВО З ОБМЕЖЕНОЮ ВІДПОВІДАЛЬНІСТЮ
«ДЕВЕЛОПМЕНТ БІЛДІНГ ГРУП»
04070, м. Київ, вул. Фролівська, буд. 1/6
Код ЭДРПОУ №30519952 тел. 097 417 29 24

№12
від 8 грудня 2023 р.

АКТ
про впровадження результатів магістерської роботи
Гарнаги Володимира Анатолійовича

Комісія у складі: голова комісії – директор Олександр Кучеренко, розглянувши матеріали магістерської роботи Володимира Гарнаги на тему «Метод та засіб ідентифікації фішингових атак на основі штучного інтелекту», склала цей акт про те, що у ТОВ «ДЕВЕЛОПМЕНТ БІЛДІНГ ГРУП» впроваджено результати цієї роботи, а саме засіб ідентифікації фішингових атак.

Впроваджені результати дозволяють покращити захист даних користувача шляхом блокування або маркування шкідливих повідомлень.

Голова комісії:
Директор **Олександр КУЧЕРЕНКО**

Додаток А

ПРОТОКОЛ ПЕРЕВІРКИ МАГІСТЕРСЬКОЇ КВАЛІФІКАЦІЙНОЇ РОБОТИ НА НАЯВНІСТЬ ТЕКСТОВИХ ЗАПОЗИЧЕНЬ

Назва роботи: Метод та засіб ідентифікації фішингових атак на основі штучного інтелекту

Автор роботи: Гарнага Володимир Анатолійович

Тип роботи: магістерська кваліфікаційна робота

Підрозділ кафедра захисту інформації ФІТКІ
(кафедра, факультет)

Показники звіту подібності Unicheck

Оригінальність – 96,41 %. Схожість – 3,59 %.

Аналіз звіту подібності (відмітити потрібне):

√ 1. Запозичення, виявлені у роботі, оформлені коректно і не містять ознак плагіату.

□ 2. Виявлені у роботі запозичення не мають ознак плагіату, але їх надмірна кількість викликає сумніви щодо цінності роботи і відсутності самостійності її виконання автором. Роботу направити на розгляд експертної комісії кафедри.

□ 3. Виявлені у роботі запозичення є недобросовісними і мають ознаки плагіату та/або в ній містяться навмисні спотворення тексту, що вказують на спроби приховування недобросовісних запозичень.

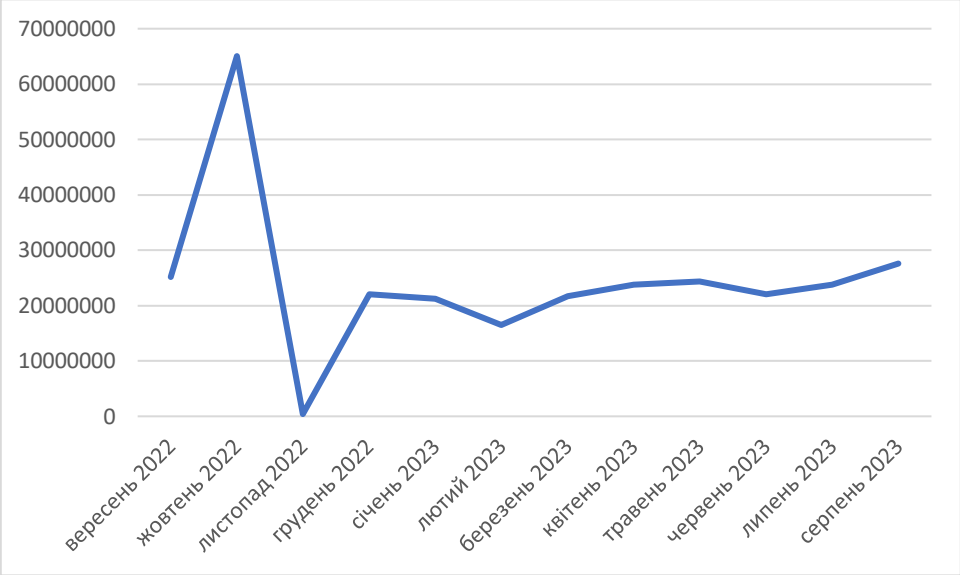
Особа, відповідальна за перевірку _____ (підпис) Валентина КАПЛУН

Ознайомлені з повним звітом подібності, який був згенерований системою Unicheck щодо роботи.

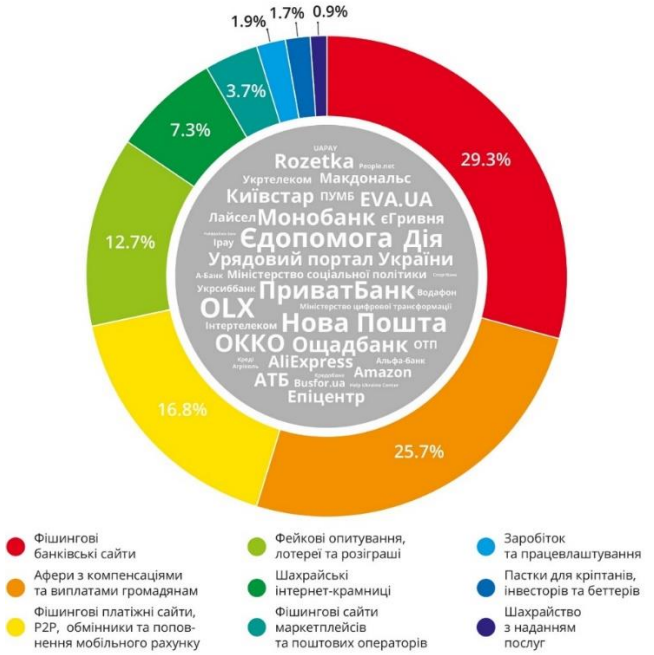
Автор роботи _____ (підпис) Володимир ГАРНАГА

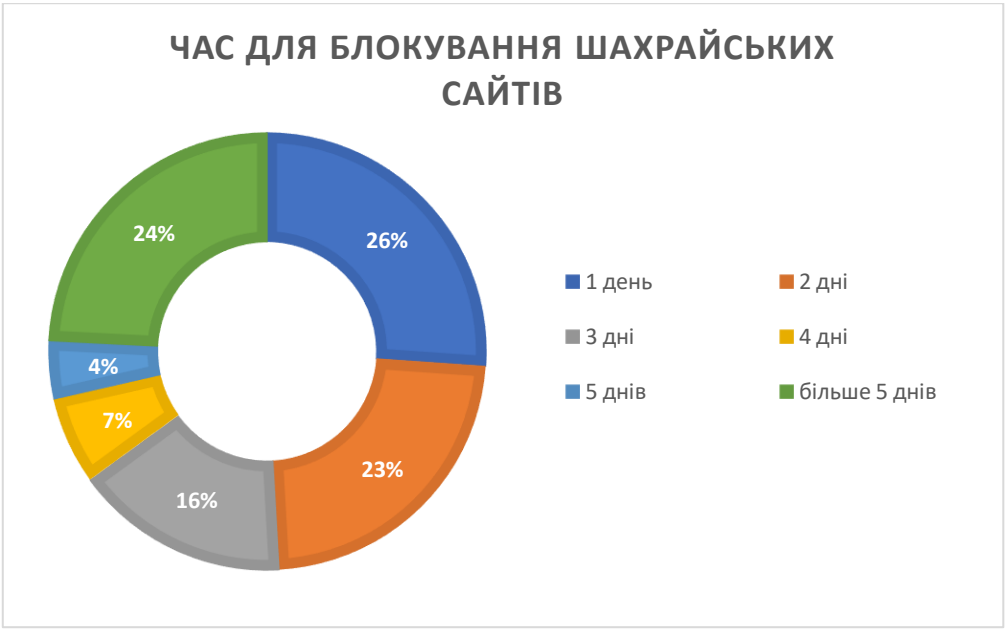
Керівник роботи _____ (підпис) Олеся ВОЙТОВИЧ

Кількість кіберподій та фішингових атак у кіберпросторі України

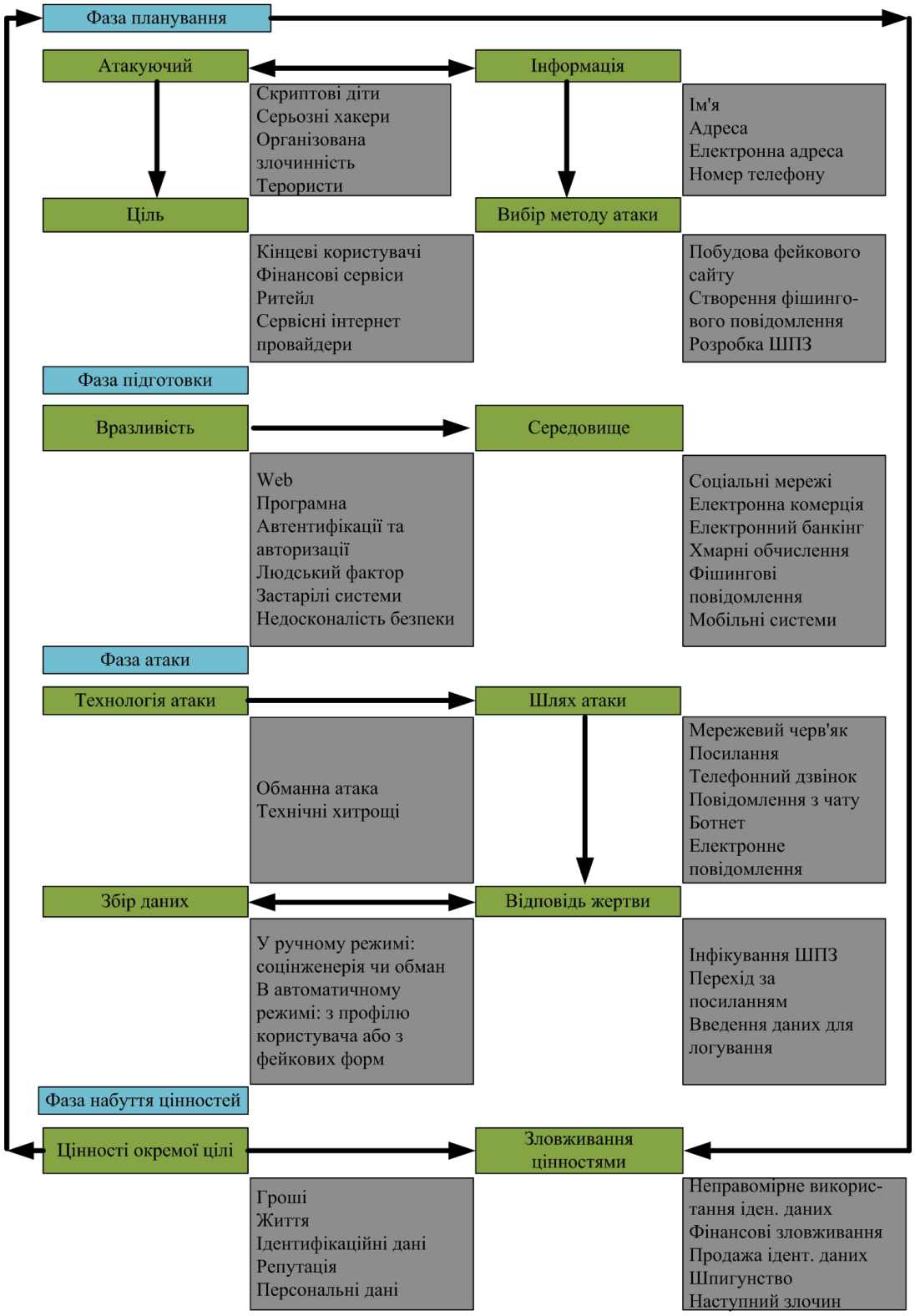


Бренди, що використовуються під час фішингу та швидкість блокування шкідливих сайтів у 2022 році

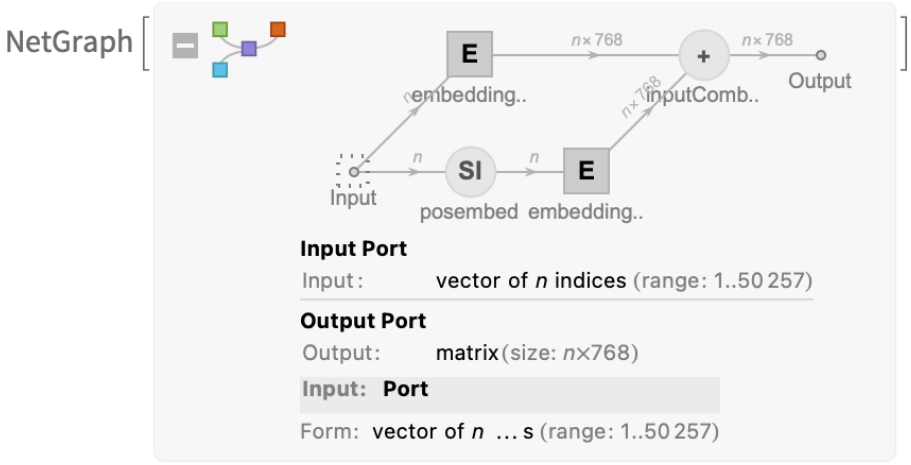


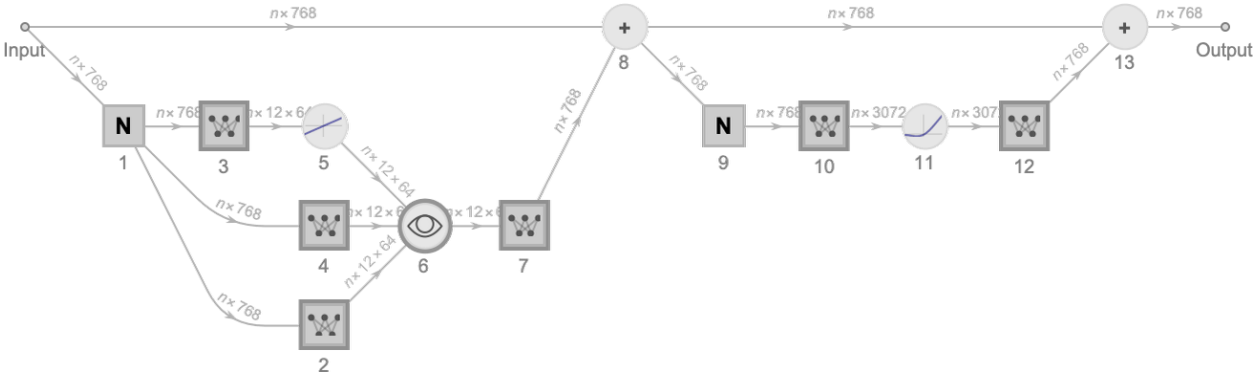


Фази та етапи фішингової атаки

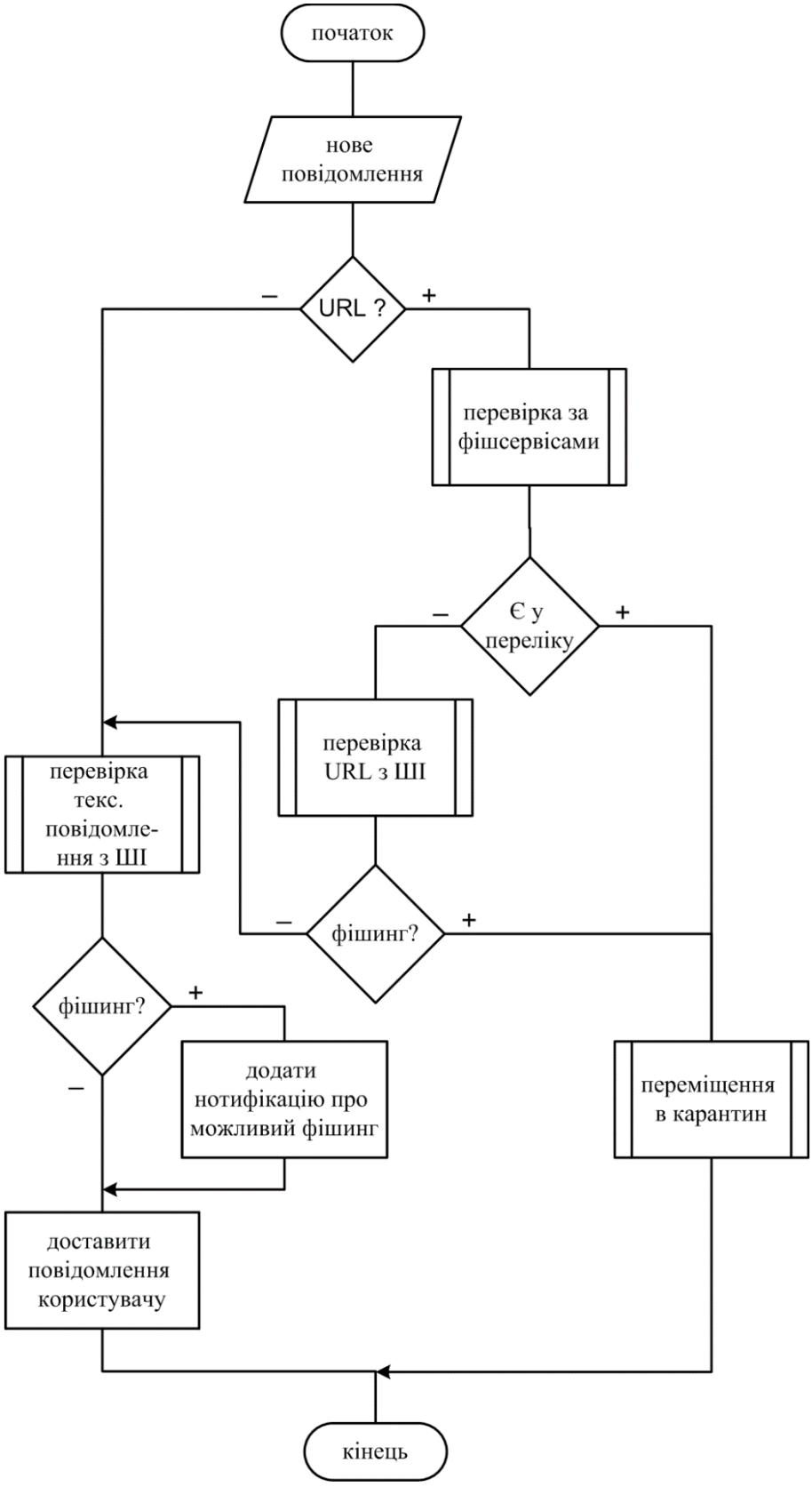


Складові великих мовних моделей

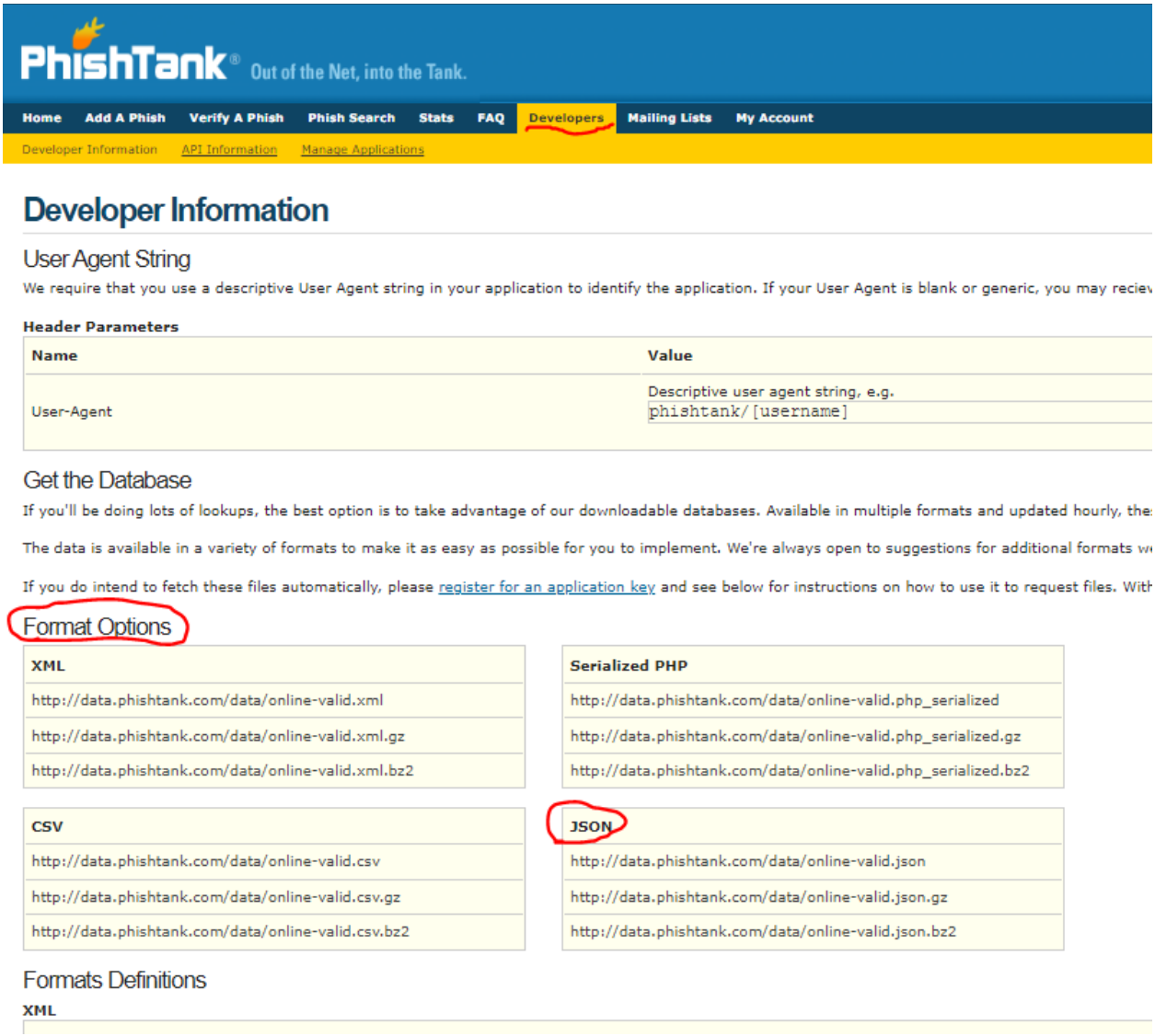




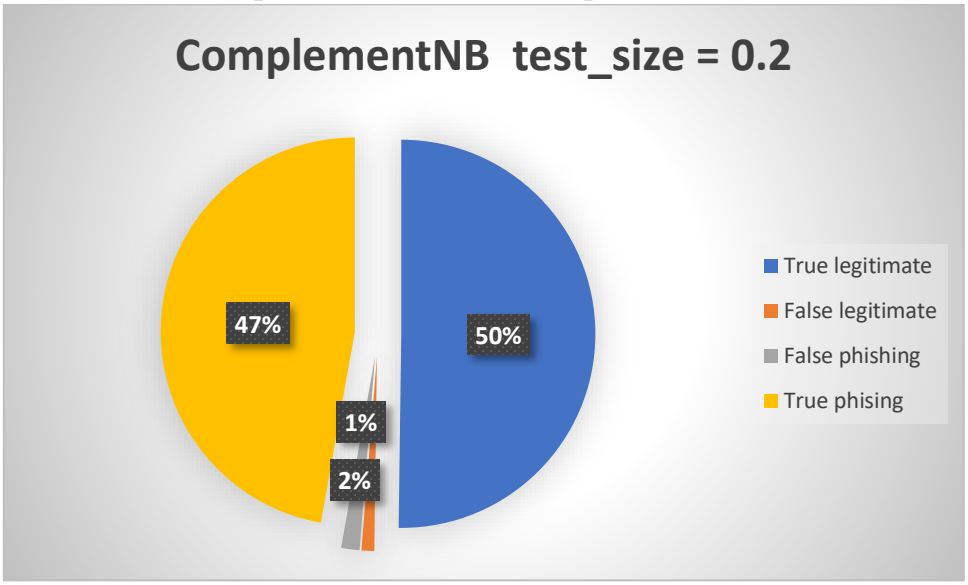
Алгоритм методу ідентифікації фішингових атак



API для аналізу активних шкідливих сайтів



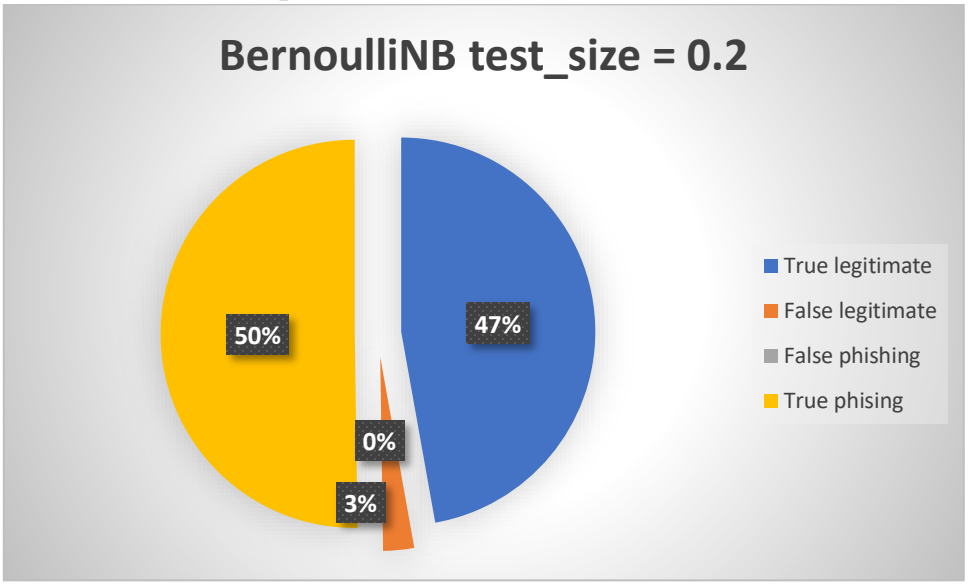
Отримані дані для ComplementNB



Таблиця 1 –Результати тестової вибірки ComplementNB класифікатора

Test size	True positive	False positive	False negative	True negative	Precision score legitimate	Precision score phishing	Total score
0.2	4896	103	149	4602	0,969	0,979	0,974
0.3	7051	199	402	7005	0,946	0,973	0,959
0.4	9598	251	647	9003	0,933	0,975	0,954
0.5	11789	449	796	11340	0,934	0,963	0,949

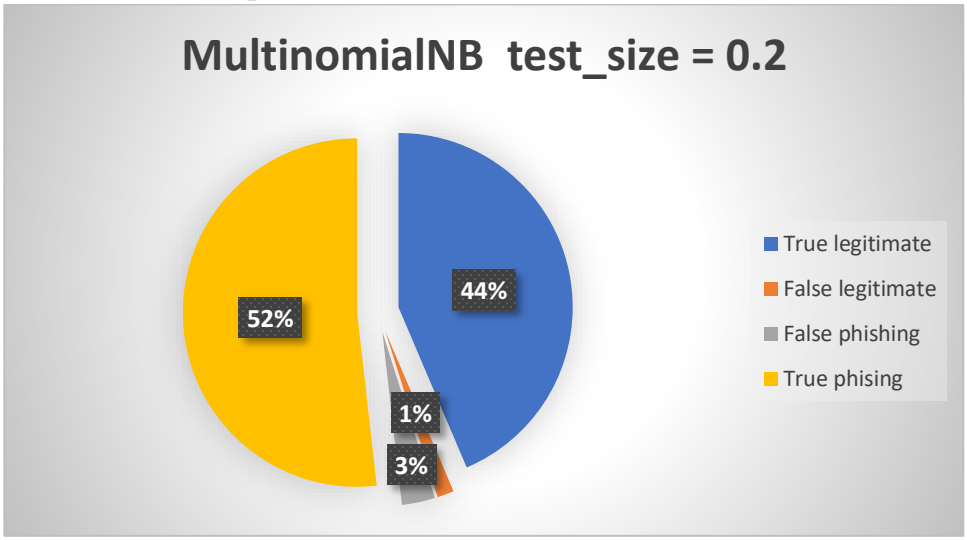
Отримані дані для BernoulliNB



Таблиця 2 –Результати тестової вибірки для BernoulliNB класифікатора

Test size	True positive	False positive	False negative	True negative	Precision score legitimate	Precision score phishing	Total score
0.2	4601	251	0	4898	1,0	0,948	0,974
0.3	6742	495	97	7291	0,987	0,932	0,959
0.4	9298	552	201	9449	0,979	0,944	0,962
0.5	11405	852	151	11998	0,987	0,933	0,959

Отримані дані для MultinominalNB



Таблиця 3 – Результати тестової вибірки для MultinomialNB класифікатора

Test size	True positive	False positive	False negative	True negative	Precision score legitimate	Precision score phishing	Total score
0.2	4250	150	300	5050	0,966	0,944	0,955
0.3	6637	152	647	7189	0,978	0,917	0,948
0.4	8752	351	752	9648	0,962	0,928	0,945
0.5	11750	447	903	11305	0,964	0,926	0,945

Порівняльний аналіз

Таблиця 4 – Порівняння класифікаторів із тестовою вибіркою 0.2

Класифікатор	True positive	False positive	False negative	True negative	Precision score legitimate	Precision score phishing	Total score
MultinomialNB	4250	150	300	5050	0,966	0,944	0,955
BernoulliNB	4601	251	0	4898	1,0	0,948	0,974
ComplementNB	4896	103	149	4602	0,969	0,979	0,974