Міністерство освіти і науки України
Харківський національний університет імені В. Н. Каразіна
Навчально-науковий інститут комп'ютерних наук та штучного інтелекту

Кафедра кібербезпеки інформаційних систем, мереж і технологій

До захисту допущено
Кафедрою КІСМіТ протокол №____ від «___» грудня 2025 р.

завідувач кафедри _____ Марина ЄСІНА
(підпис) *(ім'я, прізвище)*
«___» грудня 2025 р.

Кваліфікаційна робота

здобувача другого (магістерського) рівня вищої освіти

Дослідження та аналіз механізмів захисту від соціально-інженерних атак та рекомендації із розробки методів їх виявлення
(назва роботи)

Спеціальність (спеціалізація) 125 «Кібербезпека та захист інформації»

Освітня програма «Безпека інформаційних і комунікаційних систем»

Виконавець _____ Катерина ІЛЬЧЕНКО
(підпис) *(ім'я, прізвище)*

Науковий керівник _____ Марина ЄСІНА
(підпис) *(ім'я, прізвище)*

Харків – 2025

РЕФЕРАТ

Кваліфікаційна робота на здобуття другого (магістерського) рівня вищої освіти має обсяг 72 сторінок. Робота містить 10 рисунків, 4 таблиці. Перелік посилань розміщено на 2 сторінках і містить 24 джерел.

Метою роботи є дослідження та критичний аналіз існуючих механізмів захисту від соціально-інженерних атак, розробка науково-обґрунтованих рекомендацій для вдосконалення методів їх виявлення та створення комплексної адаптивної моделі протидії, орієнтованої на підвищення рівня інформаційної безпеки персоналу АТ «Укрзалізниця».

Для досягнення мети застосовано методи прогнозування, аналізу поточної ситуації, математичного моделювання, а також теоретичні та практичні методи захисту інформації. В експериментальній частині використано обчислювальний апарат для реалізації методів машинного навчання: Naive Bayes, Random Forest, Support Vector Machine та Logistic Regression на спеціалізованому датасеті для класифікації фішингових повідомлень.

Наукова новизна полягає в уточненні поняття захисту від соціальної інженерії та розвитку методу підвищення рівня інформаційної безпеки шляхом захисту від СІА. Основними результатами є розробка Адаптивної моделі протидії соціальній інженерії та структури програми навчання для персоналу АТ «Укрзалізниця». Створено модель розслідування інцидентів, яка використовує соціальний граф організації для математичної оцінки ймовірності поширення загрози. Експериментально доведено, що класифікатор Random Forest демонструє найвищу ефективність у виявленні фішингових повідомлень.

Запропонована методика збільшення рівня захисту від соціальної інженерії може бути успішно застосована на підприємствах різних форм власності. Розроблений проєкт захисту для АТ «Укрзалізниця» рекомендовано

до реалізації для підвищення загального рівня захищеності критично важливих даних та покращення розуміння важливості інформаційної безпеки серед персоналу.

Робота має високу практичну значущість, оскільки пропонує інтегрований, багаторівневий підхід, що дозволяє закрити прогалину між технічними та організаційними засобами захисту. Встановлено, що ключем до успішної протидії є поєднання високоточних МН-детекторів з цілеспрямованим, адаптивним навчанням, яке враховує психологічні тригери СІА. Це забезпечує проактивний захист від загроз, які експлуатують людський фактор.

Перспективним напрямком є інтеграція розроблених класифікаторів з системами User and Entity Behavior Analytics для проактивного виявлення, а також використання навчання з підкріпленням для автоматизації процесів реагування на кібератаки та підвищення адаптивності систем захисту.

Ключові слова: СОЦІАЛЬНО-ІНЖЕНЕРНІ АТАКИ, ІНФОРМАЦІЙНА БЕЗПЕКА,КІБЕРБЕЗПЕКА, ФІШИНГ, АДАПТИВНА МОДЕЛЬ, УКРЗАЛІЗНИЦЯ, МАШИННЕ НАВЧАННЯ, RANDOM FOREST, ЗАХИСТ ІНФОРМАЦІЇ,ЮДСЬКИЙ ФАКТОР.

ABSTRACT

The Master's qualification thesis has a volume of approximately 72 pages. The work contains 10 illustrations, 4 tables. The list of references is placed on 2 pages and contains 24 sources.

The goal of the work is to research and critically analyze existing protection mechanisms against Social Engineering Attacks (SEA), develop scientifically grounded recommendations for improving their detection methods, and create a comprehensive adaptive countermeasure model aimed at increasing the information security level of JSC "Ukrzaliznytsia" personnel.

To achieve the goal, methods of forecasting, current situation analysis, mathematical modeling, as well as theoretical and practical methods of information protection were applied. In the experimental part, a computational apparatus was used to implement machine learning methods: Naive Bayes, Random Forest, Support Vector Machine, and Logistic Regression on a specialized dataset for classifying phishing messages.

The scientific novelty lies in the refinement of the concept of social engineering protection and the development of a method for increasing the level of information security through protection against SEA. The main results include the development of an Adaptive Social Engineering Countermeasure Model and a training program structure for the personnel of JSC "Ukrzaliznytsia". A unique incident investigation model utilizing the organizational social graph for mathematical assessment of threat spread was created. Experimental research proved that the Random Forest classifier demonstrates the highest effectiveness in detecting phishing messages.

The proposed methodology for increasing the level of protection against social engineering can be successfully applied at enterprises of various forms of ownership. The developed protection project for JSC "Ukrzaliznytsia" is recommended for

implementation to increase the overall security level of critical data and improve the understanding of information security importance among personnel.

The work has high practical significance as it offers an integrated, multi-layered approach that allows closing the gap between technical and organizational means of protection. It is established that the key to successful counteraction is the combination of highly accurate ML detectors with targeted, adaptive training that takes into account the psychological triggers of SEA. This ensures proactive defense against threats exploiting the human factor.

A promising direction is the integration of the developed classifiers with User and Entity Behavior Analytics systems for proactive detection, as well as the use of Reinforcement Learning to automate the processes of responding to cyberattacks and increase the adaptability of protection systems.

Keywords: SOCIAL ENGINEERING ATTACKS, INFORMATION SECURITY, CYBERSECURITY, PHISHING, ADAPTIVE MODEL, UKRZALIZNYTSIA, MACHINE LEARNING, RANDOM FOREST, INFORMATION PROTECTION, HUMAN FACTOR.

ЗМІСТ

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ

АТ - Акціонерне товариство

ЗІ - Захист інформації

ІТ - Інформаційні технології

МН – Машинне навчання

НЛП - Нейролінгвістичне програмування

ОТ - Операційні технології (Operational Technology)

ПЗ - Програмне забезпечення

СІ - Соціальна інженерія

СІА - Соціально-інженерні атаки

СКУД - Система контролю та управління доступом

УЗ - АТ «Укрзалізниця»

BEC - Business Email Compromise (Компрометація ділової електронної пошти)

DLP - Data Loss Prevention (Запобігання витоку даних)

DNS - Domain Name System (Система доменних імен)

EDR - Endpoint Detection and Response (Виявлення та реагування на кінцевих точках)

IAM - Identity and Access Management (Управління ідентифікацією та доступом)

IEC - International Electrotechnical Commission (Міжнародна електротехнічна комісія

IVR - Interactive Voice Response (Інтерактивна голосова відповідь)

LBFGS - Limited-memory Broyden–Fletcher–Goldfarb–Shanno (алгоритм оптимізації)

LLM - Large Language Models (Великі мовні моделі)

MFA - Multi-Factor Authentication (Багатофакторна автентифікація)

NLP - Natural Language Processing (Обробка природної мови)

OSINT - Open Source Intelligence (Розвідка на основі відкритих джерел)

PoLP - Principle of Least Privilege (Принцип найменших привілеїв)

RBL - Real-time Blackhole List (Чорні списки в реальному часі)

RL - Reinforcement Learning (Навчання з підкріпленням)

SCADA - Supervisory Control And Data Acquisition (Диспетчерське управління та збір даних)

SMO - Sequential Minimal Optimization (Послідовна мінімальна оптимізація)

SVM - Support Vector Machine (Метод опорних векторів)

TF-IDF - Term Frequency-Inverse Document Frequency (Частота терміну – обернена частота документа)

UEBA - User and Entity Behavior Analytics (Аналіз поведінки користувачів та сутностей)

ВСТУП

В епоху тотальної цифровізації та інформатизації суспільства, коли інформація стала ключовим активом будь-якої організації, питання її захисту виходить на перший план. Проте, як показує практика, найслабшою ланкою в будь-якій, навіть найнадійнішій, системі захисту залишається людина. Соціально-інженерні атаки, що експлуатують психологічні вразливості та маніпулюють людською поведінкою, демонструють стабільно високу ефективність, обходячи складні технічні засоби захисту.

Сучасні зловмисники активно застосовують найновітніші технології, включаючи інструменти на базі штучного інтелекту та великих мовних моделей, щоб створювати дуже персоналізовані та переконливі атаки, наприклад, цільовий фішинг або діпфейки. Це спричиняє значні фінансові та репутаційні втрати для компаній у всьому світі.

Об’єкт дослідження - дослідження та реалізація захисту від атак соціальної інженерії, а також соціальна інженерія в цілому.

Предмет дослідження - методологія знань і способи протидії таким атакам.

Завдання дослідження:

1. Вивчити стан актуальної проблеми в українській та зарубіжній науковій літературі та підвищити рівень захищеності від атак соціальної інженерії.

2. Розвинути та удосконалити теперішні методики збільшення рівня носіїв інформації.

3. Розробити проєкт захисту інформації для персоналу АТ «Укрзалізниця» з метою підвищення рівня захищеності даних.

Методи дослідження: прогнозування, аналіз, вивчення поточної ситуації – для вивчення теперішнього стану; теоретичні та практичні методи захисту інформації для прогнозування загроз безпеки інформації; математичне

моделювання, розрахунок вміння захищати інформацію – для визначення профілю надійності, цілісності та конфіденційності інформації.

Наукова новизна дослідження:

- уточнено поняття захисту від соціальної інженерії;
- набув розвитку метод підвищення рівня інформаційної безпеки методом захисту від атак соціальної інженерії.

Практичне значення дослідження:

- методика зі збільшення рівня захисту від соціальної інженерії, яка має можливість бути застосованою на підприємстві;
- проєкт захисту від атак АТ «Укрзалізниця» доцільно реалізувати у організації для збільшення захищеності інформації та покращення розуміння важливості інформації.

Експериментальна частина роботи присвячена дослідженню ефективності застосування методів машинного навчання для протидії атакам соціальної інженерії, зокрема фішингу. Розробити реалізацію та порівняльний аналіз підходів: класичного методу на основі правил та алгоритмів інтелектуального аналізу даних. На основі отриманих метрик точності та повноти виявлення загроз обґрунтувати вибір оптимальної моделі класифікації для автоматизованої системи захисту.

Результати роботи апробовані на XI Міжнародній науково-практичній конференції “GLOBAL TRENDS IN SCIENCE AND EDUCATION”, та на III Міжнародній науково-практичній конференції «Сучасні аспекти діджиталізації та інформатизації в програмній та комп’ютерній інженерії».

1 СОЦІАЛЬНА ІНЖЕНЕРІЯ В ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЯХ

1.1 Теоретичні основи та історія розвитку соціальної інженерії як загрози для інформаційної безпеки

У сучасному, глибоко комп'ютеризованому суспільстві, де дані – один із найцінніших ресурсів, ідея про інформаційну безпеку зазвичай зосереджена на технічних способах захисту: шифруванні, брандмауерах і антивірусних програмах. Однак дослідження випадків показують, що найслабшим місцем у будь-якій системі захисту залишається людський фактор. Саме цю вразливість намагається використати соціальна інженерія. У світі інформаційної безпеки соціальна інженерія – це набір психологічних прийомів, які спонукають людину до дій, що можуть поставити під загрозу безпеку. Це може бути розкриття конфіденційної інформації (паролів, фінансових даних), отримання доступу до систем без дозволу або переказ грошей. На відміну від атак, які використовують програмні або апаратні вразливості, соціальна інженерія користується людською довірливістю і психологічними слабкостями.

Теоретичні засади ефективності соціальної інженерії базуються на основних принципах соціальної психології. Злочинці навмисно або інтуїтивно використовують когнітивні упередження та евристики – автоматичні й спрощені шляхи, якими людський мозок швидко приймає рішення. Одним із найбільш відомих у цій галузі є класифікація психологічних принципів впливу, запропонована Робертом Чалдіні [1]. Ці принципи допомагають зрозуміти, чому люди часто довіряють та слідують певним новонаверненим. Наприклад, принцип авторитету базується на тому, що люди схильні виконувати вказівки тих, кого сприймають як авторитетних. В соціальній інженерії це зазвичай реалізується через імітацію дзвінка або листа від керівника або співробітника ІТ-служби. Ще один важливий принцип – прихильність, коли людям легше погодитися з проханнями тих, хто їм симпатичний. Зловмисники активно користуються відкритими джерелами

інформації , щоб знаходити дані про жертв у соцмережах і створювати фальшиві історії для здобуття довіри.

До інших психологічних важелів належать соціальний доказ, який спонукає людину діяти так, як завжди, наприклад: “Усі співробітники вашого відділу вже оновили свій пароль”. Також важливий принцип дефіциту, що створює штучне відчуття терміновості – “Ваш обліковий запис буде заблоковано через 24 години”. Не менш переконливим є й принцип взаємного обміну: пропозиція безкоштовної послуги або файлу викликає у жертви почуття обов’язку. Ще один важливий механізм – принцип послідовності. Зловмисники починають з невеликого, невинного запиту і поступово підвищують важливість своїх прохань.

Методи соціальної інженерії не є сталими – вони постійно змінюються та підлаштовуються під технологічний розвиток. Цю еволюцію можна умовно поділити на кілька головних етапів. Класична соціальна інженерія, пов'язана з такими іменами, як Кевін Митник, базувалася на телефонних дзвінках та фізичному проникненні до об'єктів [2]. Ці атаки вимагали високої майстерності, але були здебільшого точковими. З часом, у умовах масового Інтернету 1990-х і 2000-х років, з'явилися масштабні фішингові кампанії. Атаки стали менш витонченими, але значно масштабнішими – наприклад, "нігерійські листи" або підроблені банківські сайти. Це перетворилося на гру великого числа, відому як Spray and Pray – з низьким шансом успіху, але великими обсягами та прибутком за рахунок кількості.

Наступний етап, ера Web 2.0 та соціальних мереж у 2010-х роках, кардинально змінив ситуацію. Зростання популярності платформ, таких як Facebook і LinkedIn, дало зловмисникам безпрецедентний доступ до інструментів OSINT, що дозволило їм проводити більш персоналізовані атаки. З'явилися цільовий фішинг, спрямований на конкретних людей, та китобійний фішинг – Whaling, що цілиться на керівників вищої ланки. Рівень фейків і ступінь маніпуляцій значно зросли. Сучасний етап характеризується активним впровадженням штучного інтелекту. LLM використовуються для створення

ідеально точних фішингових повідомлень будь-якою мовою, що імітують стиль спілкування конкретної особи. Найризикованішою тенденцією стає deepfake phishing – використання синтезованого голосу або навіть відео керівника для дачі прямих інструкцій, наприклад, щодо переказу коштів.

Отже, підсумовуючи, соціальна інженерія перетворилася з мистецтва обману окремих людей у високотехнологічну, масштабовану і автоматизовану загрозу. Хоча методи та напрямки атак постійно оновлюються, основні психологічні принципи, на яких вони ґрунтуються, залишаються незмінними. Це створює суттєві виклики для систем захисту, які мають аналізувати не тільки технічні елементи, а й зрозуміти контекст і виявляти маніпулятивні моделі у людському спілкуванні.

1.2 Класифікація та аналіз життєвого циклу сучасних соціально-інженерних атак (CIA)

Проте, щоб створити ефективні методи виявлення, недостатньо аналізувати лише окремі типи атак. Важливо розуміти, що більшість сучасних систем інформаційної безпеки – це не одностадійні події, а багатоступінчасті процеси, які мають свій життєвий цикл. Саме розуміння цього циклу є ключовим для побудови проактивного, а не просто реактивного захисту, оскільки воно допомагає виявити атаку ще на ранніх її стадіях. Хоча в науці існують різні моделі: наприклад, Lockheed Martin Cyber Kill Chain [3], MITRE ATT&CK [4], загальний життєвий цикл кібератаки виглядає приблизно так. Все починається з розвідки, коли зловмисник пасивно і активно збирає якомога більше інформації про ціль: вивчає корпоративний сайт, профілі співробітників у LinkedIn, відкриті звіти, випадки витоку даних і навіть коментарі на форумах. Це допомагає йому ідентифікувати ієрархію компанії, ключових людей, проєкти та використовувані технології. Далі слідує етап озброєння та встановлення довіри. На основі зібраної інформації оформляється претекст, створюється інструмент – фішинговий сайт або шкідливий документ, а також підготовлюється психологічне озброєння – текст

листа чи сценарій дзвінка, який ретельно адаптується для активації психологічних триггерів, описаних раніше , наприклад, терміновість або авторитет. Наступний крок – доставка та використання. Тут іде безпосередня взаємодія з людиною – надсилання листа, телефонний дзвінок. Експлуатаціяв цьому контексті означає не технічний злом, а успішну психологічну маніпуляцію: момент, коли жертва натискає на посилання, відкриває файл, повідомляє пароль або авторизується для оплати. Фінальна стадія – виконання конкретних дій. Після успішного використання людської вразливості зловмисник досягає своєї кінцевої мети, яка може бути різною: від встановлення програм-вимагачів, тривалого шпигунства в мережі до викрадення баз даних або швидкої фінансової вигоди.

Отже, огляд сучасної літератури показує, що сучасні системи інформаційної безпеки є складними, багатовекторними та орієнтованими на процеси загрозами. Вони успішно поєднують технічні засоби доставки з глибоким психологічним впливом. Цей аналіз допомагає перейти до наступного кроку – критичного огляду існуючих механізмів, що мають протидіяти цим загрозам на різних етапах їхнього життя.

1.3 Огляд та порівняльний аналіз існуючих механізмів захисту

Боротьба зі складними багатоетапними загрозами, зокрема соціально-інженерними атаками, потребує багатошарового підходу до захисту, оскільки ні один окремий засіб або політика не здатні забезпечити повний захист. У науковій та технічній літературі, а також у практиці бізнесу, захисні механізми зазвичай поділяють на два основні типи: технічні засоби та організаційно-освітні методи (рис. 1.1). Аналізуючи кожен з них, ми можемо побачити їхні сильні сторони, а що важливіше для нашого дослідження – і їхні основні недоліки.

Рисунок 1.1 – Основні компоненти соціальної інженерії

Технічні засоби захисту – це перша лінія оборони, і вони здебільшого сфокусовані на точках входу та виходу цифрових даних, найчастіше – на електронній пошті. Найбільш базовим рівнем є шлюзи безпеки електронної пошти та спам-фільтри. Їхня функціональність історично базується на аналізі репутації відправника – IP-адреси, домени, перевірці наявності в чорних списках RBLs та сигнатурному аналізі. Хоча вони ефективно відсіюють масовий спам та відомі вірусні розсилки, їхня ефективність проти сучасних СІА є вкрай низькою. Цільовий фішинг зазвичай надходить з легітимних, але скомпрометованих облікових записів, не містить очевидних спам-тригерів і надсилається в малих обсягах, що робить його "невидимим" для таких фільтрів.

Наступний, більш просунутий рівень технічного захисту включає системи пісочниці Sandboxing та детонації посилань. Ідея їхньої роботи полягає в тому, що кожне вкладення або посилання в листі спочатку відкривається у віртуальному, ізольованому середовищі – так званій пісочниці. Якщо система виявляє ознаки шкідливої поведінки (наприклад, спробу шифрування файлів або з'єднання з відомим командним центром), вона одразу блокує доставку листа користувачу. Втім, зловмисники навчилися обходити і цей захист. Вони використовують сплячі шкідливі програми, які активуються лише через деякий час, коли перевірка в пісочниці вже давно завершилася. Крім того, зловмисники все частіше розміщують шкідливий

контент на легітимних, довірених сервісах – таких як Google Drive, Dropbox або Microsoft OneDrive, – туди пісочниця просто не має підстав блокувати доступ.

З усвідомленням того, що запобігти початковій компрометації не завжди можливо, були створені системи аналізу поведінки користувачів та сутностей. Над цим працювали не лише для виявлення вже скомпрометованих облікових записів, але й для кращого розуміння звичайних звичок користувачів: з яких IP-адрес зазвичай підключаються, у який час, до яких файлів звертаються, з ким найчастіше спілкуються. Якщо, наприклад, скомпрометований обліковий запис раптом почне масово завантажувати файли опівночі з незвичайної геолокації, система зафіксує це як потенційну проблему. Однак важливо розуміти, що UEBA здебільшого фіксує порушення, що вже трапилось, – злом – і не дуже допомагає запобігти початковому інциденту, наприклад, переходу за фішинговим посиланням. До того ж, через підвищену чутливість системи часто виникає багато помилкових спрацьовувань, що ускладнює роботу.

Нарешті, системи запобігання витоку даних Data Loss Prevention та DLP стають останнім захистом. Їхня мета – не допустити успішної атаки, а запобігти її завершенню. DLP-системи аналізують вихідний трафік: листи, веб-форми, месенджери на предмет чутливих даних, а саме номери кредитних карток, фрагменти коду, особисті дані, ключові слова "конфіденційно". Якщо зловмисник, контролюючи робочу станцію, намагається відправити архів з інформацією назовні, DLP повинен його зупинити. Проте є недолік: зловмисники можуть шифрувати викрадені дані або використовувати неконтрольовані канали – наприклад, хмарні сервіси, які не підлягають моніторингу – від чого DLP-система стає сліпою.

Інший важливий захист – це організаційно-освітні заходи. Вони спрямовані на найслабшу ланку – людину. Найпоширенішим методом є тренінги з обізнаності. У своєму класичному вигляді це щорічні лекції або відеокурси, що розповідають співробітникам про те, що таке фішинг і чому він шкідливий . Ефективність такого підходу, як показує практика, близька до

нульової. Інформація є занадто загальною, подається у відриві від контексту і швидко забувається. Співробітники сприймають це як формальну необхідність, а не як інструмент для реальної роботи.

Найефективнішим сучасним підходом є використання платформ для моделювання фішингових атак, таких як KnowBe4, Proofpoint, PhishER [5]. Їхня суть у тому, що служба безпеки компанії регулярно надсилає співробітникам реалістичні, але безпечні фішингові листи. Якщо працівник натисне на посилання, система не карає його, а одразу перенаправляє на коротке навчальне повідомлення: "Ви щойно натиснули на симуляцію фішингу. Ось ознаки, на які варто звернути увагу...". Такий підхід, заснований на швидкому зворотному зв'язку під час роботи, допомагає краще запам'ятовувати та розвивати здоровий скептицизм. Проте він не є чарівним рішенням, оскільки більшість залежить від того, щоб користувач був уважним у будь-який момент.

До організаційних методів також належать внутрішні політики та процедури. Наприклад, жорстка політика IT-відділ ніколи не запитає ваш пароль або процедура : "Усі фінансові перекази на суму понад Х тисяч гривень вимагають обов'язкового усного підтвердження по телефону з відомого номера". Це дуже корисні інструменти контролю, але їхня проблема полягає в тому, що іноді вони залишаються лише на папері, не досягають співробітників або ігноруються заради швидкості роботи та зручності бізнесу.

Отже, огляд свідчить, що існуючі механізми захисту є розсіяними і недостатніми. Технічні засоби не здатні зрозуміти психологічний контекст і намір повідомлення. Організаційні методи покладаються на людську пильність, яка не завжди достатня. Кожен інструмент працює самостійно, не обмінюючись інформацією з іншими. Цей розрив між засобами і є тою прогалиною знань, яку зловмисники використовують. Порівняльний аналіз механізмів захисту від багатоетапних загроз зображено в табл. 1.1. Мета цього дослідження – допомогти закрити цю прогалину і розглянути види загроз та порівняти їх між собою.

Таблиця 1.1 – Порівняльний аналіз механізмів захисту від багатоетапних загроз

Тип Засобу	Приклади (Механізми)	Основний Фокус (Перевага)	Критичний Недолік (Прогалина)
Технічні Засоби	Шлюзи безпеки/Спам-фільтри (ESGs)	Відсіювання масового спаму та відомих загроз за сигнатурами.	Неефективні проти цільового фішингу з легітимних або скомпрометованих облікових записів.
	"Пісочниці" та Детонація посилань	Ізольована перевірка файлів і посилань на шкідливу поведінку перед доставкою.	Обхід часом (сплячі програми) або розміщення шкідливого контенту на довірених хмарних сервісах (Google Drive, OneDrive).
	UEBA (Аналіз поведінки)	Виявлення скомпрометованих облікових записів через аномалії у звичках користувача (час, геолокація, обсяг даних).	Фіксує порушення, що вже трапилось, – злом; висока кількість помилкових спрацьовувань.
	DLP (Запобігання витоку)	Контроль вихідного трафіку для блокування передачі чутливих даних (останній рубіж).	Зловмисники можуть шифрувати дані або використовувати неконтрольовані канали (хмарні сервіси, месенджери).
Організаційно-освітні Методи	Тренінги з обізнаності	Формальне інформування співробітників про основи кібербезпеки (наприклад, "що таке фішинг").	Низька ефективність, інформація занадто загальна, швидко забувається і сприймається як формальність.
	Моделювання фішингових атак	Навчання на практиці через швидкий зворотний зв'язок і розвиток здорового скептицизму.	Залежить від постійної пильності користувача; не усуває, а лише знижує ризик людської помилки.
	Внутрішні політики та процедури	Встановлення чітких правил (наприклад, усне підтвердження фінансових переказів, заборона запиту пароля).	Часто залишаються на папері або ігноруються заради швидкості та зручності бізнесу.

продовження таб. 1.1

Додатковий критичний захист	IAM (управління ідентифікацією): багатофакторна автентифікація (mfa/2fa), принцип мінімальних привілеїв.	Посилення контролю доступу і запобігання використанню викрадених облікових даних для доступу.	MFA-бомбардування (втома користувача), компрометація ключів сесії після входу, політики мінімальних привілеїв часто занадто складні для впровадження.

2 ОСНОВИ ТА МЕТОДОЛОГІЯ АТАК СОЦІАЛЬНОЇ ІНЖЕНЕРІЇ

2.1 Принципи та теоретичні основи соціальної інженерії

Соціальна інженерія – це складна та багатогранна галузь, яка полягає у здобутті інформації або доступу без використання технологічних вразливостей. Це не просто набір хитрощів, а цілісний аспект зламу, що ґрунтується на глибокому розумінні людської психології. Вона виникла як окремий прикладний напрям психології, методи якого застосовуються навіть у підготовці шпигунів та секретних агентів.

Ключова методологічна основа соціальної інженерії – це експлуатація когнітивних упереджень, тобто систематичних помилок у мисленні, що впливають на прийняття рішень. Атака часто базується на початковому бажанні людей допомагати іншим або на їхній схильності довіряти авторитету та емоційному впливу.

Соціальна інженерія є найменш технічним, але водночас найбільш ефективним інструментом в арсеналі зловмисників. Вона використовує обман, вплив та переконання для незаконного отримання конфіденційних даних, але принципи маніпуляції можуть бути застосовані і в законних цілях, наприклад, для спонукання особи до виконання певних дій.

Поглиблений погляд на природу соціальної інженерії відкриває співвідношення між свідомістю та підсвідомістю. Ідеї, схожі до нейролінгвістичного програмування та гіпнозу, давно показують, що сила підсвідомих рішень є значною [6]. Сучасні дослідження доводять, що підсвідомі рішення іноді виникають ще до того, як ми усвідомлюємо це на рівні свідомості, за кілька секунд. Це дає технологічну основу для впливу на людей та змушує їх відкривати важливу інформацію.

Організації інвестують значні кошти у сучасні технології безпеки: брандмауери, системи виявлення вторгнень, біометричні пристрої

автентифікації. Вони навчають персонал, наймають охоронців, проте залишаються вразливими.

Чим більше рівнів технологій впроваджується, тим складнішою стає програмна мережа організації. Але часто зневажають внутрішні проблеми, пов'язані з людським фактором. Навіть найнадійніша мережа може бути обійдена через дії однієї людини. Зловмисники цілеспрямовано використовують соціальну інженерію, щоб обманути користувачів. Виходить, що безпека – це не лише технічна справа, а й питання людського мислення і управління. Щоб створити справжню систему захисту, потрібно не лише техніка, а й турбота про людей: постійне навчання працівників, виховання свідомого ставлення до безпеки та протидії соціальним інженерам.

Атака соціального інженера – це ретельно спланований процес, який проходить у три основні стадії підготовки: спершу розробляється детальний план дій, а потім слідує моральна підготовка та тренінг. Це найважливіший етап, під час якого опрацьовуються сценарії з урахуванням психологічної моделі потенційної жертви. Кожне слово та дія ретельно підбирається, щоб відповідати очікуваній реакції об'єкта і максимально підвищити шанси на успіх.

Загальна робота соціальних інженерів полягає у зборі інформації про потенційну жертву та її оточення перед тим, як встановлювати прямий контакт.

Соціальна інженерія, часто в комбінації з технічними знаннями про системи захисту, використовується для досягнення важливих стратегічних цілей:

- Збір інформації про потенційну жертву та її організацію.
- Отримання конфіденційних даних: це може статися під час тривалого спілкування, коли зловмисник поступово завойовує довіру, щоб отримати потрібну інформацію під різними приводами.
- Збір даних, необхідних для несанкціонованого доступу.

- Використання об'єкта для виконання потрібних дій: наприклад, спонукання жертви перейти за небезпечним посиланням або встановити шкідливу програму, щоб створити можливість для несанкціонованого доступу.

Атаки можна поділити на два типи: короткострокові, які швидко здійснюються і не вимагають багато ресурсів, але мають обмежений вплив на виконання складних або важливих дій, та довготривалі, що потребують часу на формування довіри й стосунків і дають змогу отримати доступ до важливої інформації. Техніки, які використовують соціальні інженери, завжди базуються на когнітивних упередженнях і застосовуються у різних комбінаціях для створення ефективної стратегії обману. Їхня головна мета – змусити людину вчинити необхідні дії.

Фішинг – це найпоширеніший вид Інтернет-шахрайства, метою якого є отримання доступу до конфіденційних даних користувачів : логінів, паролів, реквізитів банківських карток [7] .

Метод реалізується через масову розсилку електронних листів-спам від імені популярних брендів, банківських служб, платіжних систем чи соціальних мереж. Повідомлення містить посилання на підроблений сайт, часто використовуючи автоматичне перенаправлення користувачів. На підробленому сайті застосовуються психологічні методи, щоб переконати користувача ввести облікові дані. Фішери активно націлені на клієнтів банків, електронних платіжних систем та особливо соціальних мереж. Експерти оцінюють, що понад 70% шахрайських атак у соціальних мережах є успішними через низьку пильність користувачів.

Машинне навчання відіграє критично важливу роль у кіберзахисті, слугуючи високоадаптивним і масштабованим інструментом для протидії атакам соціальної інженерії, які є однією з найпідступніших і найнебезпечніших загроз. Сучасні зловмисники дедалі частіше використовують генеративний штучний інтелект для створення надзвичайно переконливих фішингових повідомлень, що робить традиційні методи

захисту, які ґрунтуються на простих правилах і списках заборон, неефективними. Саме тут машинне навчання демонструє свою перевагу, оскільки воно здатне виявляти тонкі, динамічні патерни і аномалії, які лежать в основі цих маніпуляцій.

Найпоширенішою формою соціальної інженерії є фішинг через електронну пошту, і тут застосування маш. навчання є найбільш потужним. Моделі обробки природної мови , зокрема ті, що базуються на глибокому навчанні, як-от трансформери, аналізують вміст повідомлення далеко за межами простих ключових слів. Вони оцінюють тональність тексту, шукаючи ознаки терміновості, залякування або надмірної привабливості, які є психологічними тригерами для жертви. Крім того, вони можуть виявляти синтаксичні та граматичні аномалії або нетипові звороти, що не відповідають звичайному стилю спілкування організації, яку нібито представляє відправник.

Машинне навчання також аналізує технічні метадані електронного листа: невідповідності між іменем відправника та фактичною адресою домену, незвичні маршрути надсилання, підозріле кодування чи наявність посилань на URL-адреси, які використовують омогліфи для маскування справжньої адреси.

Соціальна інженерія часто завершується компрометацією облікового запису, після чого зловмисник починає діяти в мережі. У цьому випадку МН використовується для аналізу поведінки користувачів та об'єктів [8] . Моделі створюють базовий профіль нормальної поведінки для кожного працівника – з яких пристроїв, у який час, до яких систем він зазвичай отримує доступ, і яка його звичайна активність наприклад, обсяги завантажених файлів. Будь-яке значне відхилення від цього профілю – спроба доступу до конфіденційних даних у неробочий час, вхід з незвичної геолокації, або надто швидке введення облікових даних після отримання підозрілого листа – розглядається як аномалія, що може свідчити про успішну атаку соціальної інженерії або використання викрадених даних.

Крім виявлення, МН допомагає створити адаптивні системи захисту. Наприклад, системи, що використовують навчання з підкріпленням , можуть тренуватися в симульованому середовищі кібератак, щоб автоматично визначати оптимальну стратегію реагування. Якщо система виявляє новий вектор атаки або варіацію фішингу, вона може миттєво оновити свої правила класифікації, не вимагаючи ручного втручання. Таким чином, МН не лише реагує на відомі загрози, але й проактивно адаптується до нових, що є необхідним в умовах, коли ворожі сили також активно використовують ШІ.

Розповсюджувачі шкідливого програмного забезпечення (наприклад, троянів) використовують ті самі психологічні прийоми, що й маркетологи, експлуатуючи людські слабкості: бажання переглянути цікавий чи сенсаційний контент ; інтерес до дефіцитного товару або можливостей швидкого збагачення : фінансові піраміди, супер-ідеї для бізнесу.

Жертва, відкриваючи прикріплений до листа файл, добровільно встановлює на комп’ютер шкідливе ПЗ, надаючи зловмиснику доступ до інформації. Для маскування можуть застосовуватися прийоми, як-от подвійне розширення файлу ,наприклад, xxx.Jpg.Exe, що дозволяє обійти деякі засоби захисту поштових серверів.

Ще однією технікою впливу може бути спуфінг: зловмисник видає себе за відомого або довіреного користувача : керівника, колегу, технічну підтримку, і надсилає повідомлення електронною поштою або через миттєві месенджери.

В Інтернеті система доменних імен діє як ефективний каталог, пов'язуючи домени з відповідними сервісами. Це як зателефонувати другу та підключитися до його конкретного номера. Коли користувачі вводять домен, вони безперешкодно перенаправляються на відповідний веб-сайт.

Однак хакери опанували стратегію, відому як DNS-спуфінг. Ця прихована маніпуляція змінює записи DNS, приводячи нічого не підозрюючих користувачів на шахрайські веб-сайти, відстежуючи їхню активність або розгортаючи шкідливе програмне забезпечення. Цей трюк дуже цікавий для

зловмисників через свою невидимість – жертви часто абсолютно не помічають вторгнення.

Популярним методом DNS-спуфінгу є отруєння кешу DNS. Це передбачає пошкодження записів DNS у кеші резольвера [9] . У результаті сервер відображає користувачам оманливу IP-адресу, спрямовуючи їх на небезпечне факсиміле цільового веб-сайту. На рис. 2.1, що ілюструє цю техніку, соціальний інженер діє як знайомий користувач, розраховуючи на зниження пильності одержувача. На рис. 2.2, зображено різновиди спуфінгу які також часто використовуються зловмисниками для отримання інформації, та використання її у власних цілях.

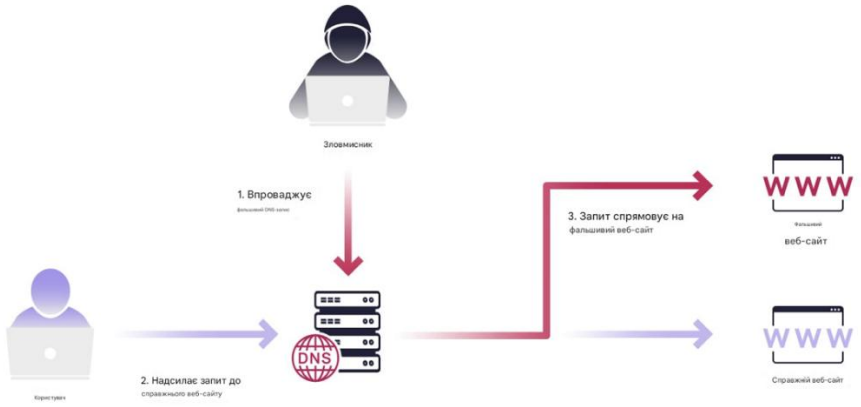


Рисунок 2.1 – Схема роботи DNS спуфінгу

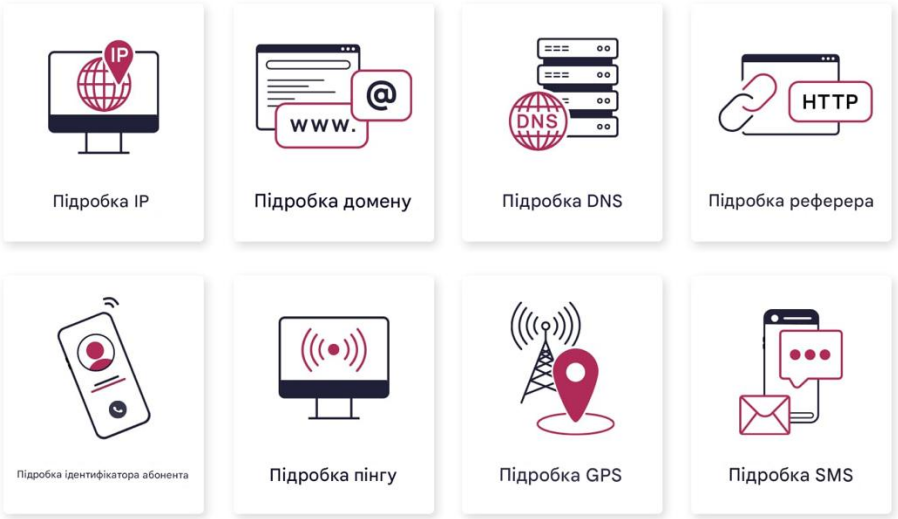


Рисунок 2.2 – Різновиди спуфінгу

Використання IVR-систем: Зловмисники можуть використовувати інтерактивні голосові системи, оскільки користувачі сприймають їх як звичайну телефонну послугу і не пов’язують із потенційними програмними загрозами.

Чат та миттєві повідомлення: Можливість використання псевдонімів значно розширюють можливості для атаки, оскільки користувачі більш відкриті у неформальному спілкуванні.

Успіх атаки залежить від рівня підготовки соціального інженера та статусу жертви. Чим нижчий рівень повноважень жертви, тим вища ймовірність успішного отримання доступу порівняно з атаками на начальника чи адміністратора. Більш детальна систематизація аспектів соціальної інженерії та технік впливу представлена у таб. 2.1.

Таблиця 2.1 – Систематизація аспектів соціальної інженерії та технік впливу

Категорія аспекту	Підкатегорія	Ключові засади та приклади	Основний вектор впливу
Методологічна основа	Психологічні фундаменти	Експлуатація когнітивних упереджень (бажання допомогти, підкорення авторитету). Вплив на підсвідоме прийняття рішень.	Обхід свідомого контролю
	Етапи планування атаки	Детальний збір інформації, розробка сценарію та моральна підготовка з урахуванням психології жертви.	Підвищення шансів на успіх
Стратегічні цілі	Отримання доступу/даних	Збір конфіденційної інформації (паролі, реквізити) та даних для несанкціонованого доступу.	Компрометація систем
	Спонукання до дії	Змусити жертву перейти за посиланням	Створення вектору атаки

продовження таб. 2.1

		або встановити шкідливе ПЗ.	
Техніки реалізації	Фішинг	Масова розсилка від імені довірених організацій із посиланням на шахрайський сайт.	Експлуатація довіри
	Шкідливе ПЗ (Трояни)	Використання людських слабкостей (інтерес, жадібність), щоб жертва добровільно відкрила інфікований файл.	Приховане встановлення
	Імітація (Spoofing)	Видача себе за керівника або технічну підтримку через Email/IM для зниження пильності.	Вплив авторитетом
	Неформальні канали	Використання чатів, IM та IVR-систем для підвищення відкритості та зниження пильності користувачів.	Використання зручності
Головна вразливість	Людський фактор	Навіть найнадійніша мережа може бути обійдена через помилки користувачів, недоліки у навчанні чи зловживання повноваженнями.	Критична точка відмови

Для забезпечення інформаційної безпеки використовуються міжнародні стандарти:

ДСТУ ISO/IEC 27004:2018: Визначає методи захисту та системи управління інформаційною безпекою, а також необхідність постійного контролю та управління заходами безпеки, визначеними ISO/IEC 27001 [10].

ДСТУ ISO/CEI 27005:2015: Надає керівництво щодо управління ризиками інформаційної безпеки. Цей стандарт наголошує, що підхід до управління ризиками розробляється самою організацією, але він повинен підтримувати вимоги системи управління безпекою [11]. Детальний аналіз стандартів зображено у таблиці 2.2.

Таблиця 2.2 – Аналіз стандартів ISO/IEC для управління інформаційною безпекою (у контексті людського фактору)

Найменування стандарту	Основне призначення	Ключове положення
ДСТУ ISO/IEC 27001:2022 (Фундаментальний)	Вимоги до створення, впровадження та підтримки системи управління інформаційною безпекою.	Це основний стандарт, який вимагає від організації систематичного підходу до безпеки, включаючи управління ризиками, пов'язаними з людським фактором [12].
ДСТУ ISO/IEC 27002:2022 (Критично важливий)	Настанова та звід практик для впровадження заходів (контролів) інформаційної безпеки.	Містить розділ "Контролі, пов’язані з персоналом" (наприклад, п. 5.10 "Дозволене використання активів") та "Обізнаність та навчання" (п. 6.3), що є прямою відповіддю на "організаційно-освітні методи" [13].
ДСТУ ISO/IEC 27004:2018	Моніторинг, вимірювання, аналіз та оцінювання СУІБ.	Дозволяє виміряти ефективність впроваджених заходів, наприклад, наскільки успішними є тренінги з обізнаності (їхня ефективність "близька до нульової" без перевірки).
ДСТУ ISO/ IEC 27005:2015	Настанова щодо управління ризиками інформаційної безпеки.	Надає методологію для ідентифікації та оцінки ризиків, де соціальна інженерія та людські помилки мають бути ідентифіковані як високопріоритетні загрози.

Особливу увагу слід приділяти джерелам загроз, що виникають внаслідок людської діяльності, які включають:

- Недоліки у навчанні та кваліфікації персоналу.
- Зловживання наданими повноваженнями.
- Помилки користувачів при роботі з конфіденційною інформацією.

Комплексне навчання забезпечує інформаційну безпеку, охоплюючи політику безпеки, техніки протидії соціальним інженерам та правильне поводження з даними. Крім того, якість паролів, що генеруються вручну, є низькою, що підкреслює необхідність автоматизованих засобів та постійного контролю за їхньою складністю та періодичною зміною.

2.2 Процес аналізу та реагування на інциденти соціальної інженерії

Багато наукових праць присвячено аналізу кіберінцидентів, що виникають через суто технічні прогалини в безпеці інформаційних систем. Проте, незважаючи на таку технічну орієнтацію, окремі методології з цих робіт виявляються придатними і для розслідування атак, заснованих на соціальній інженерії.

Наприклад, існують дослідження, що пропонують аналізувати цифрові криміналістичні події. Цей підхід, корисний на ранніх стадіях передбачення для реконструкції дій зловмисника, фокусується на автоматизації вивчення низькорівневих «цифрових спотворень» з метою виявлення аномальної поведінки, що вимагає детального аналізу.

Інші дослідження підкреслюють надзвичайний обсяг та різноманітність персональних даних, які можна зібрати із соціальних мереж, та вказують на пов'язані з цим ризики порушення приватності. Демонструється, наприклад, що аналіз «лайків» користувача дозволяє з високою точністю прогнозувати його стать, расову приналежність, релігійні та політичні погляди, певні риси характеру, рівень інтелекту та навіть схильність до шкідливих звичок. У тому ж контексті наводяться цікаві спостереження: студенти з високими показниками екстраверсії, як правило, були членами багатьох груп у Facebook, але мали обмежене коло друзів (пояснюється перевагою миттєвих контактів), тоді як особи з високим рівнем невротизму частіше публікували текстові записи на своїй сторінці, уникаючи поширення особистих фотографій [14].

Автори ще однієї праці концентруються на вивченні факторів, що стимулюють взаємодію користувачів у Facebook. Ці фактори поділяються на

ендогенні (зовнішні, як-от активність SMM-менеджерів чи маркетингові інвестиції) та екзогенні (внутрішні, пов'язані зі структурою самої мережі). Для нашого аналізу особливий інтерес становлять саме екзогенні фактори, тобто динаміка соціальних зв’язків у соціальному графі. Дослідження підтверджує дві гіпотези: по-перше, існує пряма залежність між щільністю соціального графа та активністю взаємодії користувачів; по-друге, кластеризація (виділення щільніших підгруп у мережі) також позитивно корелює з рівнем взаємодії [15]. Цей висновок є цінним для моделювання та оцінки поширення багатоетапних атак.

Соціальну інженерію також можна розглядати як різновид інформаційно-психологічного впливу, що змінює сприйняття реальності людиною, особливо її поведінкові реакції [16]. При формуванні профілів вразливості користувачів та профілів компетенцій зловмисників може бути корисною детальна класифікація видів, засобів та тактик інформаційного впливу. Перспективним напрямком розвитку моделей, що використовуються, є також розгляд інформаційно-психологічної зброї, зокрема механізмів поширення дезінформації через соціальні мережі.

Питаннями формалізованого моделювання зміни свідомості під дією зовнішніх чинників займалися автори іншого дослідження [17]. Запропоновані ними моделі можуть бути використані в майбутньому для симуляції соціоінженерних атак, особливо в умовах обмежених ресурсів у зловмисника. Для моделювання такого впливу також актуальними є результати праці, де піднімається проблема виявлення схильності індивідів до інформаційного впливу залежно від їхньої ролі та місця в соціальній структурі. Крім того, при розробці рекомендаційних систем для запобігання неправомірним діям користувачів, корисними можуть бути висновки дослідження.

Незважаючи на існування різноманітних превентивних заходів, організації часто стикаються з фактом вже реалізованої соціоінженерної атаки. У таких ситуаціях фахівцям з безпеки доступна лише інформація про те, що певні критичні активи системи були скомпрометовані, але невідомо, яким

чином. Процес розслідування подібних інцидентів є надзвичайно трудомістким, оскільки вимагає комплексного аналізу численних компонентів інформаційної системи. Це породжує актуальну задачу розробки методологій для повної або часткової автоматизації цього процесу.

Модель розслідування інцидентів соціальної інженерії – це злагоджена система, яка об'єднує технічний, соціальний та психологічний аналіз для точного відтворення сценаріїв атак і визначення того, як вони поширюються в інформаційній системі. Вона формується шляхом побудови соціального графу користувачів організації та оцінки ймовірності успіху атак, що можуть передаватися від одного співробітника до іншого.

Соціальний граф $G = (U, E)$ – орієнтований зважений граф, де

$U = \{U_1, U_2, \dots, U_n\}$ – множина користувачів;

$E = \{(U_i, U_j, p_{ij}, l_{ij})\}$ – множина орієнтованих ребер із вагами, де p_{ij} – ймовірність успіху соціоінженерної атаки від користувача U_i до U_j, а l_{ij} – інтенсивність або інший параметр зв’язку.

Для кожного користувача задана оцінка вразливості до атаки p_i, що відповідає ймовірності успішного впливу на нього.

Відновлення сценарію атаки: Нехай відомо, що критичний документ D_i було атаковано. Визначаємо множину користувачів, які мають доступ до D_i:

$$S_0 = \{U_j : (U_j, D_i) \in A\},$$

де A – відношення доступу користувачів до документів.

Відсіваємо користувачів із низьким рівнем вразливості, використовуючи порогове значення t_U:

$$S^{(1)} = \{U_k \in S_0 : p_k \geq t_U\}.$$

Ця множина містить користувачів, які потенційно піддалися прямій атаці.

Для багатоходової атаки розглядаємо можливі траєкторії T поширення атаки по графу:

$$T = (U_{i_1}, U_{i_2}, \dots, U_{i_T}).$$

Ймовірність реалізації траєкторії T:

$$p_T = p_{i_1} \times \prod_{k=1}^{T-1} p_{i_k, i_{k+1}},$$

де $p_{i_k, i_{k+1}}$ – ймовірність переходу атаки від користувача U_{i_k} до $U_{i_{k+1}}$.

Вводиться поріг t_E для відсіву траєкторій із низькою ймовірністю:

$$S^{(m)} = \{T : p_T \geq t_E\}.$$

Для випадку, коли події незалежні, модель застосовує класичне множення ймовірностей [18]. Якщо залежність подій виявляється істотною, застосовується апарат алгебраїчних байєсівських мереж для корекції розрахунків.

Побудова матриць ймовірностей: Визначимо матрицю ймовірностей розповсюдження атаки P, де $P = [p_{ij}]; p_{ij} =$ ймовірність переходу атаки від U_i до $U_j, i, j = 1..n$.

Для відновлення сценарію атаки використовується ітеративний процес, в якому обчислюють вектор ймовірностей успіху атаки моделюванням поширення по соціальному графу:

$$\mathrm{p}^{(k+1)} = \mathrm{p}^{(k)} \cdot P.$$

Початковий вектор $\mathrm{p}^{(0)}$ містить ймовірності прямої вразливості користувачів.

Врахування психологічних аспектів: Модель інтегрує характеристику користувачів c_i, що включає індивідуальні психологічні риси : екстраверсія, невротизм, інтелект та фактори сприйнятливості:

$$f(c_i) = w_1 \cdot \text{екстраверсія}_i + w_2 \cdot \text{невротизм}_i + \cdots.$$

Цей параметр впливає на корекцію оцінок p_i та p_{ij}, що підвищує точність моделювання.

Автоматизація розслідування: Модель використовує алгоритми сканування соціальних мереж і журналів інформаційної системи для збору вхідних даних:

Збір даних: $D = \{c_i, p_i, p_{ij}, A\}$ – психологічні профілі, оцінки вразливості, ймовірності переходів, доступ користувачів до документів.

Пошук траєкторій із $p_T \geq t_E$.

Візуалізація результатів для подальшого аналізу експертами:

1) Оцінка ймовірності успіху прямої атаки на користувача U_j:

   $p_j \geq t_U, t_U \in [0,1]$.

2) Ймовірність рядової траєкторії для багатоходової атаки:

   $$p_T = p_{i_1} \cdot \prod_{k=1}^{T-1} p_{i_k, i_{k+1}} \geq t_E.$$

3) Множина користувачів потенційно залучених в атаку:

   $S^{(m)} = \{U_k : p_k \geq t_U, \exists T : U_k \in T, p_T \geq t_E\}$.

4) Процес ітеративного поширення атаки по графу:

   $\mathrm{p}^{(k+1)} = \mathrm{p}^{(k)} \cdot P, \mathrm{p}^{(0)} = [p_1, p_2, \dots, p_n]$.

Практичне застосування моделі розслідування соціоінженерної атаки до інформаційної системи полягає у використанні алгоритмів аналізу доступів, соціальних зв'язків та ймовірностей компрометації, що дозволяє реконструювати шлях поширення атаки в межах організації.

Для прикладу розглянемо систему, яка містить кілька співробітників й документи на рис. 2.3.

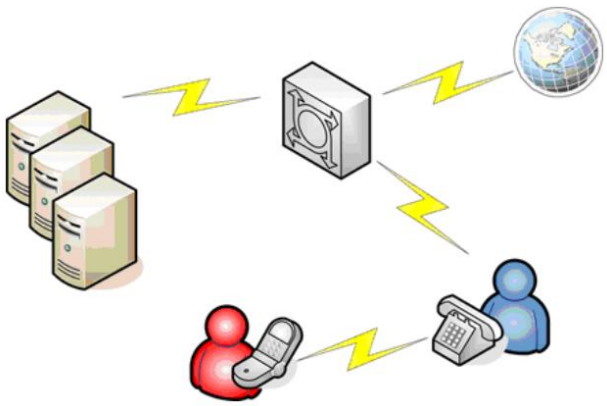


Рисунок 2.3 – Схема нападу по ланцюжку соціальних зв'язків

Після виявлення факту успішного проникнення до документа D_2, визначається початковий перелік користувачів, які мають до нього доступ – це $\{U_3, U_5\}$. Для подальшого аналізу застосовується порогове значення вразливості користувача $t_U = 0{,}1$, що дозволяє відсіяти малоймовірних кандидатів. Далі розглядається множина користувачів із відповідними оцінками прямої атаки; наприклад, якщо $p_3 = 0{,}4$, а $p_5 = 0{,}23$, то обидва користувачі задовольняють порогові умови.

Відповідно до представленої моделі, розслідування не обмежується одноходовим впливом, а також передбачає аналіз сценарію поширення атаки по ланцюжку соціальних зв'язків. Для багатоходового аналізу встановлюється поріг ймовірності реалізації траєкторії $t_E = 0{,}1$. Далі будується перелік всіх можливих шляхів, де обов'язково враховується ймовірність переходу атаки від одного користувача до іншого. Формально для двоходової атаки траєкторія через користувача U_2 до U_3 (який має доступ до D_2) розраховується як:

$$p_T = p_2 \cdot p_{2,3} = 0{,}92 \cdot 0{,}6 = 0{,}552.$$

А для впливу через U_5 отримаємо інший варіант:

$$p_T = p_5 \cdot p_{5,3} = 0{,}23 \cdot 0{,}47 = 0{,}1081.$$

Порівняння з порогом дозволяє залишити лише ті траєкторії, де ймовірність не менша за t_E. Всі ланцюжки атаки, що не задовольняють порогові умови (наприклад, якщо $p_T < 0{,}1$), відсіюються. Так траєкторія через U_3 до U_5:

$$p_T = p_3 \cdot p_{3,5} = 0{,}4 \cdot 0{,}12 = 0{,}048.$$

Є недостатньо імовірною для подальшого аналізу, тому виключається з розгляду.

Графічна візуалізація цього процесу відображає найбільш ймовірні сценарії атаки, виділяючи відповідних користувачів і документи. Такий підхід дає змогу експертам безпеки зосередити зусилля на аналізі найімовірніших ланцюгів впливу, здійснити цілеспрямовані перевірки та оперативно реагувати на наслідки інциденту.

Дана модель дозволяє не лише формалізувати та кількісно оцінити ймовірність розповсюдження соціоінженерних атак в інформаційній системі організації, а й автоматизувати процес розслідування інцидентів. Вона інтегрує соціальні, психологічні та технічні параметри, що робить її ефективною у відновленні сценаріїв атаки, ранньому виявленні загроз і підвищує якість інформаційної безпеки.

2.3 Висновки до другого розділу

У другому розділі докладно розглядається нормативно-методична база, яка регулює питання захисту від соціоінженерних атак і встановлює основні вимоги до інформаційної безпеки організації. Особливу увагу звернуто на те, що джерела загроз часто пов'язані саме з людським фактором – поведінкою співробітників, управлінськими рішеннями та внутрішніми процесами, що можуть створювати слабкі місця.

Початковий етап розробки включає систематичний огляд найбільш поширених методів, прийомів і типів атак, серед яких соціальна інженерія виділяється через свою здатність використовувати психологічний вплив і маніпуляції для отримання конфіденційної інформації або доступу до важливої інфраструктури. Аналізуються найбільш поширені сценарії: фішинг, скімінг, соціальний вплив з використанням електронних листів, телефонних дзвінків або взаємодії у соціальних мережах.

Важливим елементом дослідження є огляд міжнародних та національних стандартів, таких як ДСТУ ISO/IEC 27032:2016 [19], які прямо визначають вимоги до протидії атакам соціальної інженерії на рівні компаній та установ. Стандарти містять технічні й організаційні рекомендації щодо багаторівневого підходу, що включає:

- застосування систем контролю доступу, брандмауерів, шифрування та обмеження прав користувачів;
- впровадження політик безпеки, регулярне навчання персоналу й тренування на основі реальних сценаріїв для формування кіберкультури;

- постійний моніторинг аномалій та швидке реагування на підозрілі дії завдяки інтеграції сучасних технологій із залученням штучного інтелекту для аналізу патернів атак.

На основі висновків першого розділу поставлені задачі для подальших етапів – це побудова методики нейтралізації соціоінженерних технік, розробка практичних рекомендацій протидії. До них входять впровадження ефективних політик захисту даних, організація навчальних програм для співробітників, розробка процедур раннього виявлення та реагування на інциденти, а також впровадження превентивних заходів для мінімізації ризику людських помилок і невірних дій.

Для обґрунтування ефективності запропонованого підходу необхідно проаналізувати практичний і економічний ефект – адже йдеться не лише про технічну безпеку, а й про мінімізацію втрат, скорочення відновних витрат, підвищення стійкості бізнес-процесів. Це включає розрахунки зниження ризику, оцінку впливу на фінансові показники, а також формування звітності щодо запобігання випадкам порушення конфіденційності чи втрати даних.

3 АДАПТИВНА МОДЕЛЬ ПРОТИДІЇ СОЦІАЛЬНІЙ ІНЖЕНЕРІЇ ДЛЯ АТ «УКРЗАЛІЗНИЦЯ»

Цей розділ аналізує побудову ефективної моделі захисту організації від атак, спрямованих на персонал. Ключова проблема полягає в тому, що жодні, навіть найсучасніші, технічні заходи захисту (такі як антивіруси, брандмауери чи системи виявлення вторгнень) не здатні самі по собі гарантувати повну безпеку.

Це пояснюється тим, що соціальні інженери експлуатують не вразливості програмного коду, а «людський фактор» – психологічні слабкості, довірливість, неуважність та бажання допомогти. Якщо персонал, від рядового співробітника до керівника, не навчений розпізнавати ці маніпуляції, витік критичної інформації або несанкціоноване проникнення в корпоративну мережу є лише питанням часу.

У цьому контексті найефективнішим підходом до протидії соціальній інженерії є постійна, системна та гнучка робота з підвищення обізнаності людей. Це допоможе їм краще розпізнавати потенційні загрози та вживати заходів для їх запобігання. Адже кожна людина – це унікальна особистість, яка заслуговує на увагу та підтримку. Для реального підвищення рівня безпеки організації необхідно впровадити циклічну модель навчання. Вона має виходити за межі одноразових лекцій і включати:

1) Постійний моніторинг рівня знань працівників.
2) Регулярні тестування та опитування.
3) Контрольовані симуляції атак (наприклад, внутрішній тестовий фішинг або вішинг). Такі симуляції є критично важливими, оскільки виявляють реальний рівень підготовки співробітників у безпечних, але максимально наближених до реальних умовах.

Найважливіший момент, на який слід звернути увагу при навчанні користувачів: навчання – це не захід, а безперервний процес. Він повинен

періодично повторюватися та оновлюватися, щоб відповідати новим тактикам, які використовують зловмисники.

Комплексна програма навчання повинна поєднувати різні формати для максимального охоплення та засвоєння матеріалу:

- Теоретичні заняття: вивчення основ кібергігієни, огляд поширених методів маніпуляції : фішинг, вішинг, претекстинг тощо та аналіз психологічних прийомів, які використовують злочинці.
- Практикуми: колективний розбір реальних прикладів фішингових листів, аналіз записів дзвінків шахраїв, відпрацювання алгоритмів реагування.
- Онлайн-семінари : дозволяють оперативно оновлювати знання персоналу про нові виявлені загрози та інформувати про поточні атаки.
- Рольові ігри та симуляції: моделювання реалістичних сценаріїв атаки наприклад, дзвінок з технічної підтримки або лист від керівника, де співробітник повинен самостійно прийняти правильне рішення в умовах обмеженого часу.

3.1 Концептуальні основи створення механізмів захисту від атак соціальної інженерії

Для розробки дієвої методики необхідно проаналізувати, як саме соціальна інженерія використовується проти української критичної інфраструктури, зокрема об'єктів транспорту.

АТ «Укрзалізниця» неодноразово ставала ціллю потужних кібератак. Наприклад, у березні 2025 року компанія зазнала багаторівневої та систематичної атаки, яка призвела до масового збою в роботі онлайн-сервісів продажу квитків - сайт та мобільний застосунок. Хоча публічно не розголошувався точний вектор початкового проникнення, переважна більшість таких складних, багаторівневих атак починається саме з попереднього етапу соціальної інженерії.

Проблема полягає в тому, що зловмисникам достатньо одного успішного фішингового листа або телефонного дзвінка - вішинг, щоб

отримати початкові облікові дані. Здобувши доступ до мережі, вони можуть місяцями вивчати інфраструктуру, підвищувати свої привілеї і готувати руйнівну атаку, націлену на критичні системи – чи то сервери продажу квитків, чи, що небезпечніше, системи управління рухом.

Щоб зрозуміти ризики для УЗ, варто розглянути підтверджені випадки, де соціальна інженерія була ключовою:

- Атака на Київстар (грудень 2023): Ця атака, що мала катастрофічні наслідки, офіційно почалася з фішингу. Один зі співробітників відкрив шкідливий лист, що дозволило хакерам отримати початковий доступ. Згодом вони змогли заволодіти обліковими даними адміністраторів і зруйнувати значну частину цифрової інфраструктури компанії. Для Укрзалізниці це пряма аналогія: один необережний клік адміністратора чи бухгалтера може призвести до паралічу систем [20].
- Атака на Прикарпаттяобленерго (грудень 2015): Перший у світі підтверджений випадок блекауту, спричиненого кібератакою. Атака почалася з цільових листів співробітникам [21]. Вони містили шкідливий документ Word. Після його відкриття хакери отримали контроль над системами SCADA і вимкнули електропостачання для сотень тисяч людей. Для УЗ це найстрашніший сценарій – можливість зовнішнього втручання в системи диспетчеризації та управління рухом поїздів.

Отже, жодні технічні заходи захисту інформації в УЗ не дадуть 100% гарантії, доки існує людський фактор. Соціальні інженери використовують не вразливості ПЗ, а психологічні слабкості персоналу. У цьому контексті безперервна та цільова робота з персоналом є єдиним ефективним способом боротьби.

З метою підвищення безпеки організації слід постійно проводити спеціалізоване навчання, моніторити рівень знань, проводити контрольовані симуляції атак , щоб виявляти рівень підготовки працівників в реальних умовах.

Найважливішим моментом, на який слід звернути увагу, є те, що навчання – це циклічний процес, який необхідно періодично повторювати та оновлювати відповідно до нових тактик зловмисників.

Форми навчання можуть включати:

- Теоретичні інтерактивні модулі;
- Практикуми (наприклад, розбір фішингових листів);
- Онлайн-семінари з експертами;
- Рольові ігри та симуляції атак.

Для того, щоб створити методологію навчання, необхідно зрозуміти, чому люди вразливі до атак. Маніпуляція як метод впливу досліджується соціологами десятиліттями. Роберт Чалдіні у своїй праці «Психологія впливу» визначив шість основних принципів, які соціальні інженери найчастіше та найуспішніше використовують [1] .

Розглянемо 4 ключові методи в контексті «Укрзалізниці».

Авторитет: Люди схильні підкорятися особам, яких сприймають як авторитетних чи наділених владою.

Приклад атаки на УЗ: Зловмисник телефонує начальнику станції, представляючись куратором з СБУ або аудитором з Міністерства інфраструктури, та, посилаючись на термінову перевірку, вимагає надати службові дані про переміщення вантажів або доступ до робочої станції.

Прихильність / Симпатія: Люди охочіше виконують прохання тих, хто їм подобається, або тих, з ким вони мають щось спільне.

Приклад атаки на УЗ: Шахрай знаходить у соцмережах провідника, дізнається про його захоплення наприклад, риболовля, і втирається в довіру під виглядом однодумця. Через деякий час він просить просто відкрити файл з розкладом змагань на службовому планшеті, який насправді є шкідливим ПЗ.

Зобов'язання та послідовність: Люди прагнуть виконувати взяті на себе обіцянки або бути послідовними у своїх діях.

Приклад атаки на УЗ: Новому касиру дзвонить співробітник ІТ-підтримки та просить підтвердити згоду з новою політикою безпеки що є

логічним для новачка. Після отримання усної згоди: “Так, я згоден дотримуватись правил”, шахрай просить продиктувати тимчасовий пароль для остаточної реєстрації в системі, і співробітник, вже погодившись, з більшою ймовірністю це зробить.

Соціальний доказ: Люди схильні вважати правильними ті дії, які виконують інші люди.

Приклад атаки на УЗ: Фішинговий лист надсилається на весь відділ бухгалтерії з текстом: “Всі вже заповнили нову форму для нарахування премій, залишились тільки ви. Ось посилання...” .Страх відстати від колективу або бути єдиним, хто не виконав завдання , змушує людину поспішити і клікнути, не перевіряючи посилання.

3.2 Структура програми навчання для АТ «Укрзалізниця»

Розробка заходів протидії має починатися зі створення Міжвідомчої робочої групи з кіберобізнаності. До неї повинні входити представники департаментів IT, кібербезпеки, безпеки руху, пасажирських перевезень та HR.

Обов'язки цієї групи:

- Забезпечення підтримки політики та процедур безпеки на всіх рівнях.
- Допомога в розробці та актуалізації навчальних матеріалів.
- Програма інформаційної безпеки повинна мати диференційовані вимоги для окремих груп співробітників, оскільки ризики для них різні.

Цільові групи для тренінгів в УЗ:

- Керівний склад: Фокус на стратегічних ризиках, прийнятті рішень під час атаки Київстар , репутаційних та фінансових втратах.
- Диспетчерський та операційний персонал (управління рухом): Критична група. Навчання має бути зосереджене на захисті систем управління. Сценарії: спроби вішингу з вимогою змінити маршрут, проігнорувати попередження системи, надати віддалений доступ «для оновлення».

- IT-персонал та адміністратори: Поглиблене технічне навчання щодо векторів атак, крадіжки облікових даних, методів підвищення привілеїв (вектор атаки на «Київстар»).
- Персонал «першої лінії» (касири, провідники, начальники поїздів): Фокус на захисті платіжних та квиткових систем, безпеці службових планшетів, протидії фізичній соціальній інженерії (спроби пасажирів отримати доступ до службових приміщень).
- Адміністративний та офісний персонал (бухгалтерія, HR, закупівлі): Основний фокус на фішингу (кейс «Прикарпаттяобленерго»), безпеці документообігу.
- Обслуговуючий персонал та охорона: Навчання протидії фізичній соціальній інженерії (прохід слідом , забув перепустку, підкидання USB-накопичувачів).

Технічні засоби навчання повинні включати:

- Демонстрацію соціальної інженерії через рольові ігри та симуляції.
- Огляд медіазвітів щодо останніх атак на УЗ та інші компанії в Україні.
- Обговорення шляхів запобігання втраті даних (наприклад, чіткий протокол: "Якщо вам дзвонять з IT і просять пароль – покладіть слухавку і зателефонуйте на внутрішній номер IT-підтримки самі").

Організація повинна не тільки встановити політику безпеки, але й заохочувати її дотримання. Необхідно переконатись, що всі співробітники розуміють причину прийняття певних правил (наприклад, "чому не можна заряджати особистий телефон від робочого комп'ютера диспетчера").

Для закріплення навичок необхідне посилення. Воно може бути двох типів:

Негативне : Це чітка відповідальність за грубі порушення політик безпеки наприклад, запис пароля від критичної системи на стікері. Це питання слід інтегрувати у щорічну атестацію співробітника, що демонструє серйозність відповідальності за безпеку.

Позитивне : Цей метод значно ефективніший. Він спонукає персонал дбати про безпеку.

Приклад: Замість того, щоб шукати порушників, публічно заохочуйте тих, хто вчасно і правильно повідомив про фішингову спробу.

Гейміфікація: Введення нагороди наприклад, додатковий день відпустки або брендована продукція для відділу, який показав найкращий результат у квартальній симуляції фішингової атаки тобто 0% співробітників клікнули на посилання.

Головною метою розробки системи навчання є зосередження уваги на тому, що:

- Співробітники УЗ можуть бути атаковані будь-коли;
- Працівники повинні вивчити і зрозуміти свою особисту роль у захисті інформації компанії;
- Навчання техніці безпеки повинно бути важливішим за формальне дотримання встановлених правил.

Організація може вважати, що мета досягнута, якщо всі співробітники звикнуть до думки, що інформаційна безпека є невід'ємною частиною їхньої відповідальності. Співробітники УЗ повинні чітко розуміти, що атаки соціальних інженерів є реальними, і що втрата інформації загрожує не тільки компанії, але й кожному працівникові особисто, їхній роботі, добробуту та безпеці руху в країні.

3.3 Аналіз технічних засобів захисту від соціальної інженерії

Технічний захист може включати засоби запобігання отриманню інформації та засоби запобігання використанню отриманої інформації. Загалом, засоби забезпечення захисту даних з точки зору запобігання навмисним діям, залежно від способу реалізації, можна розділити на групи.

Технічні засоби. Це різні типи пристроїв (механічні, електромеханічні, електронні тощо), які вирішують проблему захисту даних за допомогою апаратних засобів. Вони або запобігають фізичному вторгненню, або, якщо

сталося втручання, доступу до інформації, включаючи її маскування. Першу частину проблеми вирішують замки, решітки вікон, охоронна сигналізація тощо. По-друге, генератори шуму, мережеві фільтри, скануючі радіостанції та багато інших пристроїв, які «блокують» потенційні канали витоку інформації або дозволяють їх виявлення. Переваги технічних засобів пов'язані з їх надійністю, незалежністю від суб'єктивних факторів, високою стійкістю до модифікацій. Слабким сторонам бракує гнучкості, відносно великого обсягу та ваги, високих витрат.

До програмних засобів належать програми для ідентифікації користувачів, контролю доступу, шифрування інформації, видалення залишкової (робочої) інформації, такої як тимчасові файли, контроль тесту системи безпеки тощо. Переваги програмного забезпечення – універсальність, гнучкість, надійність, простота установки, можливість модифікувати та розвивати. Недоліками є обмежена функціональність мережі, використання деяких ресурсів файлового сервера та робочих станцій, висока чутливість до випадкових або навмисних змін, можлива залежність від типів комп’ютерів (їх обладнання).

Організаційні засоби бувають організаційно-технічними (підготовка приміщень за допомогою комп’ютерів, прокладка кабельної системи з урахуванням вимог щодо обмеження доступу до неї тощо) та організаційно-правовими (закони та нормативні акти, національні норми, встановлені керівництвом певної компанії). Переваги організаційних інструментів полягають у тому, що вони дозволяють вирішити безліч різних проблем, їх легко реалізувати, швидко реагувати на небажані дії в мережі та мати необмежені можливості налаштування та розробки. Недоліки – сильна залежність від суб’єктивних факторів, включаючи загальну організацію праці в певному підрозділі.

3.4 Актуальні проблеми соціальної інженерії в організації АТ «УЗ»

Важливо чітко розуміти, що технічні засоби самі по собі не запобігають соціальній інженерії, оскільки не можуть вплинути на психологічну маніпуляцію. Натомість вони слугують критично важливим другим ешелоном оборони. Їхня головна мета – мінімізувати ризики та заблокувати атаку вже на технічному рівні, навіть, якщо співробітник внаслідок маніпуляції припустився помилки. Технічний захист в «Укрзалізниці» має одночасно запобігати отриманню інформації (наприклад, блокувати фішингові листи) та запобігати використанню вже отриманої інформації (наприклад, блокувати запуск шкідливого ПЗ або робити вкрадений пароль марним). Ці засоби можна умовно розділити на три ключові групи.

Перша група – це технічні (апаратні) засоби. Вони являють собою фізичні пристрої, що створюють бар'єри для зловмисника. До таких засобів належать системи контролю та управління доступом, що включають біометричні сканери, карткові пропуски та турнікети. Вони необхідні для обмеження доступу у диспетчерські центри, серверні кімнати, вузли зв'язку та ремонтні депо. Це є прямою протидією фізичній соціальній інженерії, такій як спроби проникнення під виглядом кур'єра чи нового співробітника. Крім того, сюди відносяться апаратні міжмережеві екрани , які критично важливі для фізичної сегментації мережі. Вони повинні на апаратному рівні ізолювати корпоративну IT-мережу , де працюють бухгалтери і HR, від операційно-технологічної OT-мережі, яка керує стрілками, сигналами та тяговими підстанціями. Охоронна сигналізація та відеонагляд доповнюють цей захист, дозволяючи виявляти несанкціоновану фізичну присутність. Переваги апаратних засобів полягають у їхній надійності та незалежності від суб'єктивних факторів, але їхніми слабкими сторонами є низька гнучкість та висока вартість впровадження, особливо в масштабах «Укрзалізниці».

Друга група – це програмні засоби, які є основним інструментом блокування безпосередніх наслідків кібератак, що сталися через людську помилку. До них належать системи ідентифікації та управління доступом .

Найважливішим програмним засобом у цій категорії є багатофакторна автентифікація . Навіть, якщо зловмисник виманить у співробітника пароль, він не зможе увійти в систему без другого фактору (наприклад, коду із застосунку). Це доповнюється принципом найменших привілеїв , який програмно обмежує права доступу. Наприклад, касир не повинен мати жодних прав у системі управління рухом. Також ключове значення має захист поштових шлюзів, що автоматично аналізує листи на ознаки фішингу, перевіряє підозрілі посилання у пісочниці та блокує підробку адрес нібито від керівництва УЗ. Якщо ж співробітник все ж таки запустить шкідливий файл, сучасні системи Endpoint Detection and Response виявлять аномальну поведінку як-от шифрування файлів та автоматично ізолюють робочу станцію від мережі. Нарешті, системи запобігання витоку даних не дозволяють співробітнику, навмисно чи випадково, відправити назовні файли з грифом «службова таємниця» чи персональні дані пасажирів. Переваги програмного забезпечення – це універсальність та гнучкість, а недоліки – залежність від правильних налаштувань та своєчасних оновлень.

Третя група – організаційні засоби. Вони створюють «правила гри» та регламенти, які об'єднують технічні та людські аспекти безпеки. Вони бувають організаційно-технічними, наприклад, політики безпеки, що регламентують прокладання кабельних мереж на станціях або підготовку приміщень диспетчерських центрів. Інший підвид – організаційно-правові засоби. Це основа всієї системи, що включає як національні норми (наприклад, Закон України «Про основні засади забезпечення кібербезпеки», оскільки УЗ є об'єктом критичної інфраструктури), так і внутрішні політики та регламенти [22]. Саме вони є основою для навчання персоналу та юридичною базою для застосування санкцій за порушення. Переваги організаційних інструментів у тому, що вони дозволяють вирішити безліч нетехнічних проблем, але їхній головний недолік – сильна залежність від суб'єктивних факторів та загальної культури дотримання правил у компанії.

Замість одного застарілого продукту, для «Укрзалізниці» актуальним є комплексний набір промислового рівня. Він повинен включати захист персональних даних та контроль приватності, не дозволяючи співробітникам переглядати дані пасажирів, що не стосуються їхніх обов'язків. Віртуальна клавіатура може бути корисна для захисту від шпигунського ПЗ на терміналах у касах. Контроль USB-пристроїв є обов'язковим для централізованого блокування неавторизованих флешок, що нейтралізує атаку з підкиданням носіїв. Шифрування файлів та дисків на ноутбуках керівництва та мобільних співробітників захистить дані у випадку крадіжки. Нарешті, гарантоване знищення файлів необхідне для безпечної утилізації старих робочих станцій, щоб конфіденційні дані неможливо було відновити.

3.5 Методика протидії соціальній інженерії

Для ефективного впровадження захисних заходів в АТ «Укрзалізниця» ключовим є розробка чітких протоколів реагування на інциденти соціальної інженерії. Кожен інцидент надає безцінну інформацію для поточного аудиту безпеки та вдосконалення моделі захисту. Після отримання інформації про спробу або факт атаки, ІТ-служба та служба безпеки повинні негайно провести цільовий аудит. Управління інцидентами вимагає використання єдиного протоколу, який фіксує щонайменше таку інформацію, як жертва атаки наприклад, «Диспетчер станції 'Х'», дату та час інциденту, напрям (вектор) атаки (фішинг, вішинг тощо), детальний опис атаки та використану легенду, результат атаки (успішна/неуспішна), наслідки атаки (фінансові/операційні збитки) та рекомендації щодо запобігання. Систематичне ведення такого журналу дозволяє визначати стійкі схеми атак та покращувати захист.

При повному аналізі організації такого масштабу, як УЗ, виявляються специфічні фактори, які роблять її вразливою. По-перше, масштаб та розпорошеність структури (тисячі об'єктів та величезний штат) дозволяють соціальному інженеру легко видавати себе за колегу з іншого регіону або аудитора з Києва. По-друге, публічна інформація про персонал (наприклад,

графіки руху) дозволяє зловмиснику посилатися на реальних співробітників, місцезнаходження яких він знає, що підвищує довіру. По-третє, доступність внутрішньої структури, наприклад, телефонних довідників чи назв відділів, допомагає зловмиснику впевнено маніпулювати іменами під час атаки. По-четверте, прогалини в політиках та навчанні призводять до того, що персонал не знає, як поводитись у критичних ситуаціях. Нарешті, неналежне поводження з фізичними носіями : загублені документи, службові записки у сміттєвих контейнерах створює ризик витоку комерційної таємниці.

Для захисту такої складної організації слід використовувати інтегровані багаторівневі системи безпеки. Фізична безпека є першим бар'єром, що обмежує доступ до будівель, диспетчерських та серверних, а також контролює утилізацію документів зі сміттєвих контейнерів. Безпека даних є наступним рівнем і стосується чітких принципів роботи як з паперовими, так і з електронними носіями на всіх рівнях. Безпека програмного забезпечення також є критичною і має включати визначення чітких політик щодо того, які програми можна використовувати, особливо на робочих місцях диспетчерів чи касирів (політика «білого списку»). Це доповнюється загальною безпекою комп’ютерних систем, що охоплює захист серверів та клієнтських систем.

Окремим вектором атаки є комунікаційні системи, зокрема офісні та диспетчерські АТС. Існує три основних типи атак на них: отримання доступу для «вільного» використання міжміського зв'язку, отримання доступу до системи комунікацій для прослуховування або переадресації службових розмов, або порушення роботи системи. Термін для таких атак – фрікінг (phreaking). Найпростіший підхід – це моделювання соціальним інженером ролі телефонного інженера або ІТ-спеціаліста провайдера зв'язку. Така атака особливо ефективна, коли зловмисник телефонує на внутрішні номери, недоступні для широкої громадськості, створюючи в жертви ілюзію, що той, хто телефонує, вже є своїм. Якщо працівник особливо відповідальний, як-от диспетчер отримує дзвінок з невідомого номера з проханням надати технічний доступ і не може перевірити особу абонента, він повинен негайно перервати

розмову та терміново зв’язатися з відповідальним співробітником Служби Безпеки УЗ за відомим внутрішнім номером.

Нарешті, методика повинна включати збалансовану політику мотивації та відповідальності. Слід нагадати працівникам, що у разі доведеного розголошення комерційної чи службової таємниці вони можуть бути притягнуті до відповідальності, включно зі звільненням за відповідними статтями КЗпП України [23]. Водночас, негативна мотивація не повинна переважати. Ключовим елементом успіху є позитивне заохочення. Співробітників, які стали об'єктом атаки, але вчасно розпізнали її та повідомили службу безпеки, слід публічно заохочувати. Такі заохочення призводять до того, що всі працівники будуть більш обережні та не боятимуться повідомляти про свої підозри.

3.6 Навчання на базі комплексу атак «важливі документи - інформаційна система - персонал - зловмисник»

У межах цієї роботи було розглянуто моделі імітації соціоінженерних атак на користувачів інформаційної системи, зокрема, на соціальному графі зв’язків між працівниками компанії. Такі моделі надають кількісну оцінку стійкості користувачів до впливу зловмисника та визначають потенційні шляхи поширення атаки всередині корпоративного середовища. Проте, стандартні підходи часто ігнорують рівень обізнаності чи досвіду працівника у питаннях кібербезпеки.

Для підвищення точності оцінки ризиків, у даному дослідженні пропонується розширити модель: окрім структури зв’язків враховувати «профіль документів» та індивідуальні характеристики користувача (від доступів до важливої інформації до властивостей поведінки). Модель користувача доцільно розглядати як структуру:

$$U = \langle C, Z, PV, L \rangle,$$

C – критичні документи у доступі користувача,

Z – зони перебування (наприклад, серверні, офісні приміщення),

PV – профіль вразливості персоналу,

L – матриця стосунків з іншими співробітниками.

Зловмисник подається як:

$M = \langle R, S_0, U_0, P, G \rangle$,

R – ресурси (час, гроші, особисті властивості),

S_0 – наявні знання про систему,

U_0 – користувачі, до яких є початковий доступ,

P – компетенції у застосуванні технік атак,

G – ціль атаки (захоплення документу, доступу тощо).

Ймовірність успішного впливу оцінюється функціонально через ступінь володіння певною атакою та схильність користувача до неї:

$p_{ij} = f(D_j(K_i), T_i)$,

де $D_j(K_i)$ – рівень майстерності зловмисника у техніці K_i, T_i – гранична компетентність, а сама функція може бути деталізована як об'єкт дослідження.

Узагальнену оцінку ймовірності прориву отримують через серію взаємодій між ресурсом чи дією та конкретними вразливостями працівника:

$p = f(D_j(K_i), S_q(V_l, K_i))$.

Тут S_q – чутливість користувача q до техніки K_i, а також враховуються різні ресурси зловмисника, які можуть бути використані гнучко (від хабаря до психологічного тиску).

У контексті «Укрзалізниці» це особливо важливо, адже зловмисники активно використовували фішингові та вішингові підходи (дзвінки, підроблені листи), а співробітники часто не ідентифікували обман і виконували інструкції зловмисника, вважаючи його частиною IT-команди або керівництва. Форма впливу від типового фішингу до підроблених телефонних дзвінків (вішинг) добре працювала саме через відсутність глибокого розуміння методики атак та рідкісне проведення тренінгів. У листах маскувались справжні імена, стилістика, термінові нагадування, а самі повідомлення підштовхували

перейти за підробленим посиланням або відкрити вкладення, що несло шкідливий код.

Для запобігання таким загрозам рекомендовано такі системні заходи:

1) Регулярні навчальні кампанії та інтерактивні семінари: тренінги, симуляція «живих» атак, розбір нових тактик фішингу та вішингу з демонстрацією відомих прикладів.

2) Персоналізований аудит рівня кіберобізнаності, моделювання індивідуальної зони ризику і цілеспрямована робота з найвразливішими категоріями співробітників.

3) Впровадження політик суворого контролю листування, електронних підписів і підготовка користувачів до критичного аналізу будь-яких підозрілих листів чи дзвінків (наприклад, ідентифікація невідповідності між доменом і відправником, некоректними підписами, посиланням на зовнішній ресурс зі схожою адресою).

Також важливо використовувати гнучку систему мотивації: співробітники, що демонструють підвищений рівень кіберобізнаності, заохочуються, а ігнорування заходів безпеки тягне за собою дисциплінарну відповідальність.

Особливої ролі набуває й структурований внутрішній аудит: всі звернення, обробку запитів до технічної підтримки та важливі операції слід регулярно аналізувати, вести й перевіряти журнали дій. Це мінімізує зловживання правами і дозволяє швидко реагувати у випадку підозрілого інциденту.

Системна робота над підвищенням кіберусвідомленості можлива лише в умовах, коли компанія гарантує безперервність освітньої програми, підтримує зворотній зв’язок, проводить рольові ігри і стимулює прогрес співробітників через відкриту програму мотивації. За умови імплементації таких стратегій ризик успішних соціоінженерних атак на великі інфраструктурні підприємства, як «Укрзалізниця», знижується у рази.

У реаліях «Укрзалізниці» важливо враховувати, що атаки соціальної інженерії можуть розгортатися не лише у офісі компанії, а й безпосередньо під час поїздки – як щодо працівників (провідників, операторів чи інженерів), так і щодо пасажирів, які використовують цифрові сервіси перевізника.

Сценарії розвитку загроз під час поїздки можуть мати кілька проявів:

1) Комунікація через підроблені дзвінки: зловмисники можуть телефонувати персоналу під виглядом диспетчера чи керівника змін, наголошуючи на аварійній ситуації (наприклад, загрозі безпеці руху) та випрошувати терміновий доступ до інформаційних систем або паролі. Створюючи відчуття невідкладності у дорозі, знижують рівень критичного мислення.

2) Соціальна інженерія у вагонах: у великих поїздах із Wi-Fi або із залученням мобільних пристроїв, працівники або пасажири можуть отримувати фішингові листи від імені служби підтримки з вимогою поновити дані авторизації або під'єднатись до підробленого Wi-Fi, що дозволяє зловмисникам перехопити паролі.

3) Вішингові атаки на пасажирів під час поїздки: популярною стає схема, коли пасажира або співробітника під час поїздки просять по телефону чи месенджері підтвердити банківські дані, посилаючись на «системні перевірки» або шахрайство з їхнім квитком. Людина, перебуваючи далеко від домашнього середовища, менш схильна відразу звернутись до справжньої служби підтримки і є більш вразливою.

4) Маніпулятивні впливи в кабіні локомотива чи технічних приміщеннях: зловмисник може звертатись у вигляді контролера, аварійного диспетчера чи іншого службовця, намагатися отримати коди чи доступ до систем для «термінових ремонтних робіт».

5) Підробка внутрішньої документації та інструкцій: персонал може отримати у месенджерах чи е-мейлах підроблений документ або наказ із вимогою виконати позапланову дію (відкрити доступ, змінити маршрутизацію

тощо), що особливо актуально для співробітників на маршруті, які часто діють під тиском часу та у відриві від офісу.

Всі ці сценарії демонструють, наскільки важливо закладати у політики безпеки організації не лише базові процедури для офісних працівників, а й враховувати моделі поведінки персоналу та пасажирів під час поїздки – коли емоційна, фізична чи інформативна вразливість значно зростає.

Рекомендовано:

1) Регулярно проводити тренінги для працівників, які подорожують, з акцентом на кібергігієну та розпізнавання підозрілих комунікацій у рейсі.

2) Вводити тренувальні симуляції атак саме під час поїздок, коли рівень уваги працівників природно нижчий.

3) Забезпечити чіткі канали термінового зв'язку з офіційними представниками компанії у разі отримання будь-яких підозрілих запитів під час відрядження або чергування в дорозі.

4) Встановити політику негайного інформування служби безпеки при будь-яких спробах отримати конфіденційну інформацію зі сторонніх або сумнівних джерел.

Завдяки застосуванню таких заходів «Укрзалізниця» здатна не лише мінімізувати ризики класичних офісних атак, а й адаптувати захист до специфічних умов роботи, пов'язаних із поїздками, мобільністю персоналу та новими цифровими сервісами для пасажирів. Це робить модель протидії універсальною та сучасною, адаптованою до нових реалій залізничного бізнесу.

3.7 Висновки до третього розділу

У третьому розділі було розроблено комплексну методику боротьби з методами соціальної інженерії та проаналізовано адаптивну модель, яка може захистити організацію від відповідних атак. Ключовою метою методології є розробка цілісної політики та процедур безпеки, спрямованих на захист як окремих працівників, так і організації в цілому. Для розробки цієї методики

було враховано теоретичні аспекти психології впливу, що лежать в основі маніпуляцій, та детально проаналізовано технічні засоби захисту, які слугують критичним другим ешелоном оборони.

Ґрунтовний аналіз проблем соціальної інженерії та її практичних проявів було проведено на прикладі специфіки АТ «Укрзалізниця». Було проаналізовано специфічні вразливості цієї критичної інфраструктурної компанії, що дозволило визначити ключові вимоги до політики безпеки, а також класифікувати типи та рівні ризиків для різних категорій персоналу – від операційного (диспетчери, провідники) до адміністративного апарату.

На основі виявлених вразливостей підготовлено перелік конкретних контрзаходів. Під час розробки методології було детально розкрито етапи її впровадження в організації, а також запропоновано методичні вказівки, необхідні заходи та процедури щодо дотримання правил безпеки для ефективної протидії методам соціальної інженерії.

4 ЕКСПЕРИМЕНТАЛЬНЕ ДОСЛІДЖЕННЯ

Машинне навчання відіграє критично важливу роль у кіберзахисті від атак соціальної інженерії, особливо тих, що використовують цифрові канали, як-от фішинг [8] . Оскільки соціальна інженерія спрямована на людські вразливості, МН використовується для виявлення підозрілих цифрових патернів і аномалій, які вказують на спробу атаки.

4.1 Опис датасету

Для проведення експериментального дослідження було використано датасет Phishing Email Dataset з платформи Kaggle, який містить реальні зразки електронних листів для класифікації фішингових повідомлень.

Датасет містить 18634 електронні листи, розподілені наступним чином:

- Легітимні листи: 11 322 (60,8%);
- Фішингові листи: 7 312 (39,2%).

Такий розподіл відображає реальні умови, де фішингові листи становлять меншість від загального обсягу кореспонденції. Співвідношення класів 1,55:1 не є критичним дисбалансом і дозволяє ефективно навчати моделі без додаткових методів балансування.

Змістовно фішингові листи в датасеті представляють типові атаки з використанням соціальної інженерії: заклики до термінових дій, імітацію офіційної кореспонденції від банків, пропозиції виграшів та запити на верифікацію облікових записів. Легітимні листи включають ділову кореспонденцію, технічні повідомлення, академічні листи та інформаційні розсилки.

Для забезпечення коректної роботи алгоритмів класифікації та уніфікації вхідних даних було проведено етап попередньої обробки тексту . У ході цього процесу текст було приведено до нижнього регістру, URL-адреси та email-адреси замінено на відповідні токени, а також видалено спеціальні

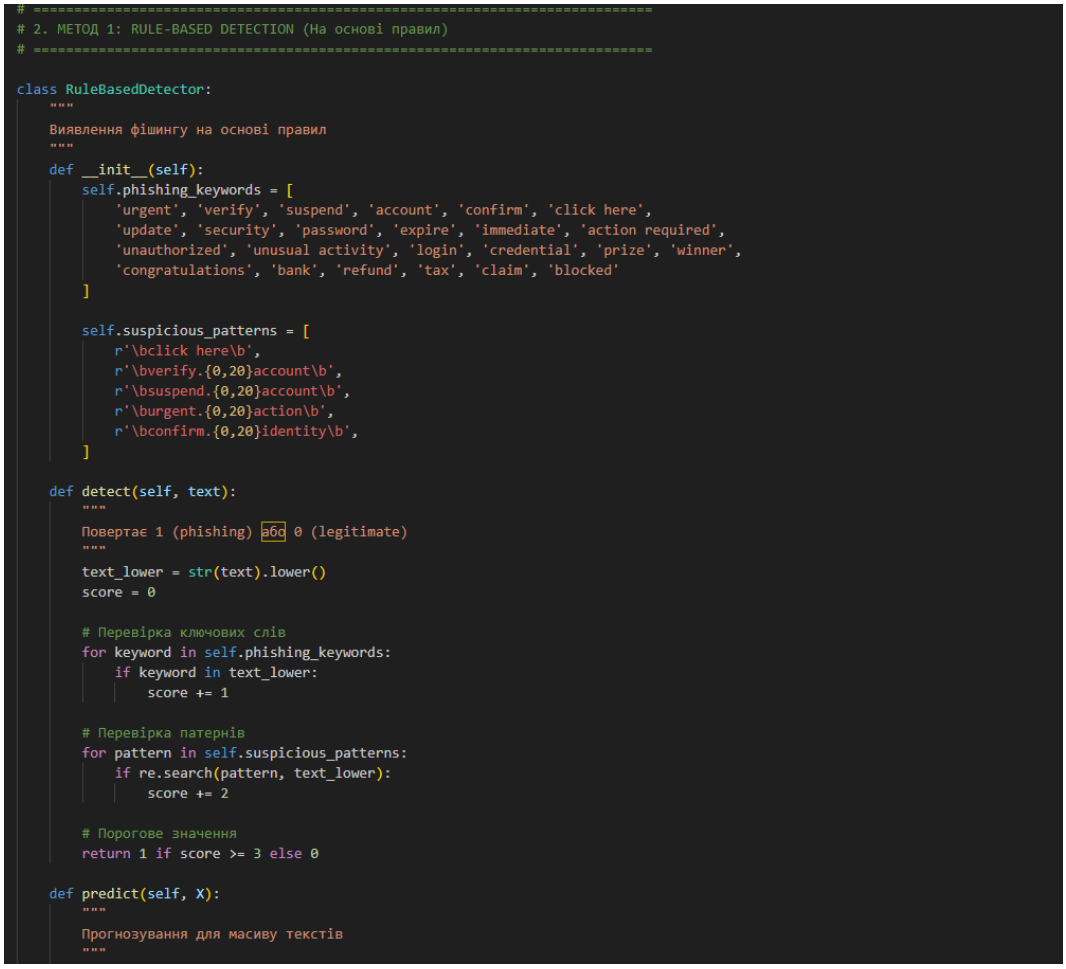
символи та зайві пробіли. Фрагмент даних після попередньої обробки зображено на рис. 4.1.

Рисунок 4.1 – Скріншот оброблених фішингових листів в датасеті

4.2. Реалізація методів

У рамках експериментального дослідження було реалізовано та протестовано п'ять різних підходів до виявлення фішингових атак: один класичний метод на основі правил та чотири алгоритми машинного навчання з різними принципами роботи. Такий вибір дозволяє порівняти традиційний експертний підхід із сучасними методами автоматичного навчання на даних.

Rule-Based Detection : Метод представляє традиційний підхід до виявлення фішингу, який базується на експертних знаннях про характерні ознаки зловмисних повідомлень. Цей підхід не потребує етапу навчання та працює виключно на попередньо визначених правилах і патернах. Далі буде представлено скріншоти реалізації методів , реалізація класичного методу на основі правил зображено на рис.4.2.



```
# ============================================================================
# 2. МЕТОД 1: RULE-BASED DETECTION (На основі правил)
# ============================================================================

class RuleBasedDetector:
    """
    Виявлення фішингу на основі правил
    """
    def __init__(self):
        self.phishing_keywords = [
            'urgent', 'verify', 'suspend', 'account', 'confirm', 'click here',
            'update', 'security', 'password', 'expire', 'immediate', 'action required',
            'unauthorized', 'unusual activity', 'login', 'credential', 'prize', 'winner',
            'congratulations', 'bank', 'refund', 'tax', 'claim', 'blocked'
        ]

        self.suspicious_patterns = [
            r'\bclick here\b',
            r'\bverify.{0,20}account\b',
            r'\bsuspend.{0,20}account\b',
            r'\burgent.{0,20}action\b',
            r'\bconfirm.{0,20}identity\b',
        ]

    def detect(self, text):
        """
        Повертає 1 (phishing) або 0 (legitimate)
        """
        text_lower = str(text).lower()
        score = 0

        # Перевірка ключових слів
        for keyword in self.phishing_keywords:
            if keyword in text_lower:
                score += 1

        # Перевірка патернів
        for pattern in self.suspicious_patterns:
            if re.search(pattern, text_lower):
                score += 2

        # Порогове значення
        return 1 if score >= 3 else 0

    def predict(self, X):
        """
        Прогнозування для масиву текстів
        """
```

Рисунок 4.2 – Скріншот реалізації класичного методу на основі правил

Архітектура детектора: Реалізований детектор складається з двох основних компонентів.

Перший компонент – це словник підозрілих ключових слів, який містить 24 терміни, згруповані за категоріями загроз. До категорії термінів терміновості входять слова "urgent", "immediate", "action required", які створюють психологічний тиск на жертву. Категорія термінів верифікації включає "verify", "confirm", "update", "security", "password" – типові запити в фішингових атаках. Терміни загроз представлені словами "suspend", "blocked", "unauthorized", "expire", які залякують користувача можливими наслідками. Фінансова категорія містить "bank", "refund", "prize", "winner", "claim" – характерні для шахрайських схем.

Другий компонент – це набір регулярних виразів для виявлення складних патернів у тексті. Патерн \bclick here\b виявляє прямі заклики до кліків на посилання. Вираз \bverify.{0,20}account\b знаходить запити верифікації облікового запису незалежно від формулювання в межах 20 символів між ключовими словами. Аналогічно працюють патерни \bsuspend:{0,20}account\b для загроз блокування та \burgent.{0,20}action\b для створення відчуття терміновості.

Алгоритм класифікації: Процес класифікації відбувається в декілька етапів. Спочатку вхідний текст приводиться до нижнього регістру для уніфікації. Потім ініціалізується змінна score зі значенням нуль. Детектор послідовно перевіряє наявність кожного ключового слова зі словника – при знаходженні збігу score збільшується на одиницю. Далі застосовуються регулярні вирази для пошуку складних патернів – кожен знайдений патерн додає два бали до загального score через їхню вищу значущість.

Фінальне рішення приймається на основі порогового значення: якщо сумарний score досягає трьох або більше балів, лист класифікується як фішинг, інакше вважається легітимним. Це порогове значення було емпірично підібрано для балансу між виявленням загроз та мінімізацією помилкових спрацювань.

Головна перевага методу – повна прозорість та інтерпретованість рішень. Експерт завжди може пояснити, чому конкретний лист було класифіковано певним чином. Метод не потребує етапу навчання та великих обчислювальних ресурсів, що робить його швидким у роботі. Однак основне обмеження – це низька адаптивність. Зловмисники легко обходять прості правила, використовуючи синоніми, перефразування або нестандартну лексику. Метод не враховує контекст та семантичні зв'язки між словами, що призводить до великої кількості помилок.

Naive Bayes : Ймовірнісний метод класифікації, який базується на теоремі Байєса з припущенням умовної незалежності ознак. Незважаючи на

спрощене припущення про незалежність слів у тексті, метод демонструє високу ефективність у задачах класифікації текстів.

Метод обчислює ймовірність належності документа до класу на основі формули Байєса: P(клас|текст) = P(текст|клас) × P(клас) / P(текст). Оскільки P(текст) однакова для всіх класів, фактично порівнюються значення P(текст| клас) × P(клас) для кожного класу. Припущення Naive Bayes полягає в тому, що всі слова в документі є незалежними один від одного при умові знання класу, що дозволяє розкласти ймовірність: P(текст|клас) = P($слово_1$| клас) × P($слово_2$|клас) × ... × P($слово_n$|клас).

Процес навчання та класифікації: Перший етап – це векторизація тексту за допомогою методу TF-IDF (Term Frequency – Inverse Document Frequency). Цей підхід перетворює кожен документ на вектор чисел, де кожна компонента відповідає вазі конкретного слова або n-грами. TF-IDF враховує не лише частоту слова в документі, але й його рідкість у всьому корпусі – рідкісні слова отримують більшу вагу як більш інформативні. У нашій реалізації використовується 3000 найважливіших n-грам, включаючи як окремі слова (uni-грами), так і пари слів (bi-грами). Це дозволяє врахувати не лише окремі терміни, але й деякі стійкі словосполучення.

На етапі навчання для кожного класу (фішинг/легітимний) обчислюються умовні ймовірності P(слово|клас) для всіх слів у словнику. Також оцінюються апріорні ймовірності класів P(фішинг) та P(легітимний) на основі їхньої частоти в навчальній вибірці. Використовується згладжування Лапласа для уникнення нульових ймовірностей для слів, які не зустрічалися в навчанні.

При класифікації нового листа спочатку виконується його векторизація тим самим TF-IDF перетворенням. Потім для кожного класу обчислюється логарифм апостеріорної ймовірності (використання логарифмів запобігає числовому переповненню): log P(клас|текст) = log P(клас)+Σ log P(слово|клас).

Вибирається клас з максимальним значенням цього виразу.

Naive Bayes особливо ефективний для текстових даних завдяки простоті обчислень та стійкості до високої розмірності ознак. Метод швидко навчається навіть на великих датасетах і добре працює з малими обсягами навчальних даних. Однак припущення про незалежність ознак часто порушується в реальних текстах – значення слова може залежати від контексту, що обмежує точність методу.

Random Forest : Ансамблевий метод машинного навчання, який будує множину дерев рішень та агрегує їхні прогнози для отримання фінального результату [24] . Цей підхід відноситься до методів навчання ансамблів, що значно покращує якість класифікації порівняно з одиночними моделями. Приклад реалізації зображено на рис. 4.3.

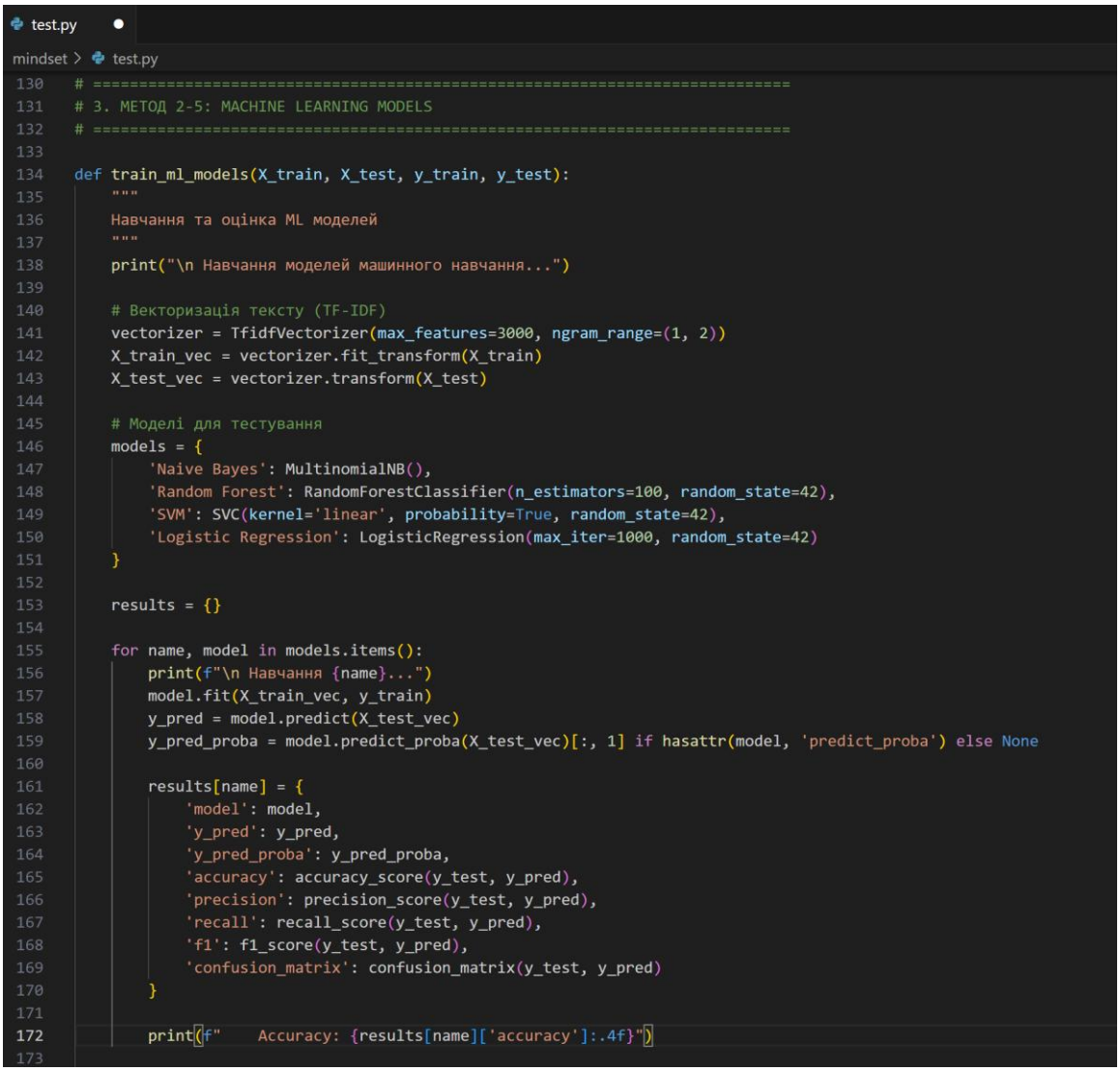


Рисунок 4.3 – Скріншот реалізації алгоритмів машинного навчання

Основна ідея полягає в тому, що комбінація багатьох слабких класифікаторів може створити сильний класифікатор з високою точністю. Кожне окреме дерево в лісі може робити помилки, але при агрегації прогнозів від сотні дерев помилки компенсують одна одну, що призводить до стабільного та точного результату.

У реалізованій моделі використовується 100 дерев рішень. Кількість дерев є компромісом між точністю та обчислювальними витратами – більша кількість зазвичай покращує результат, але збільшує час роботи. Критерієм розбиття вузлів служить індекс Джині, який вимірює нечистоту вузла – наскільки добре дані у вузлі розділені за класами. Максимальна глибина дерев не обмежена, що дозволяє моделі вивчати складні залежності в даних.

Навчання відбувається в два етапи рандомізації, які забезпечують різноманітність дерев у лісі. Перша рандомізація – це метод bootstrap aggregating. Для кожного дерева створюється випадкова вибірка з поверненням з навчального набору. Це означає, що деякі зразки можуть повторюватися, а інші взагалі не потрапити у вибірку. Таким чином кожне дерево навчається на трохи різних даних.

Друга рандомізація відбувається при побудові кожного дерева. У кожному вузлі при виборі ознаки для розбиття розглядається не весь набір ознак, а випадкова підмножина. Зазвичай використовується квадратний корінь від загальної кількості ознак. Для нашого випадку з 3000 TF-IDF ознаками це приблизно 55 ознак на кожному розбитті. Це запобігає домінуванню найсильніших ознак та створює більш різноманітні дерева.

Кожне дерево будується рекурсивно: починаючи з кореневого вузла, що містить всі дані, алгоритм вибирає оптимальну ознаку та порогове значення для розбиття, які максимально зменшують індекс Джині. Процес продовжується до повного розділення даних або досягнення критерію зупинки.

При класифікації нового листа він проходить через усі 100 дерев у лісі. Кожне дерево незалежно приймає рішення, проходячи від кореня до листка

відповідно до значень ознак. У результаті отримується 100 індивідуальних прогнозів. Фінальне рішення приймається методом голосування більшості (мажоритарне правило) – вибирається клас, за який проголосувало більше дерев.

Random Forest демонструє високу стійкість до перенавчання завдяки усередненню прогнозів багатьох дерев. Метод ефективно працює з великою кількістю ознак та автоматично оцінює їхню важливість [24]. Рандомізація на двох рівнях забезпечує різноманітність моделей та знижує кореляцію помилок між деревами. Метод стійкий до шуму в даних та викидів, що робить його надійним у реальних умовах.

Support Vector Machine : Метод бінарної класифікації, який шукає оптимальну гіперплощину для розділення класів у багатовимірному просторі ознак. Метод базується на принципі максимізації відстані між класами, що забезпечує хорошу узагальнюючу здатність.

У випадку лінійно роздільних даних SVM шукає таку гіперплощину, яка розділяє класи з максимальним зазором (margin). Зазор – це відстань від гіперплощини до найближчих точок кожного класу. Точки, які лежать на границях зазору, називаються опорними векторами – саме вони визначають положення гіперплощини. Інші точки не впливають на модель, що робить SVM стійким до викидів усередині класів.

Для лінійно роздільних даних задача формулюється як оптимізаційна проблема: знайти вектор ваг w та зміщення b, які максимізують зазор $2/||w||$ при обмеженнях $y_i(w \cdot x_i+b) \geq 1$ для всіх навчальних зразків. Тут y_i – мітка класу (+1 або -1), x_i – вектор ознак. Це еквівалентно мінімізації $||w||^2/2$ при тих самих обмеженнях.

Для реальних даних, які не є ідеально роздільними, використовується м'який зазор (soft margin) з параметром регуляризації C. Це дозволяє деяким точкам порушувати умову розділення, але зі штрафом. Параметр C контролює баланс між максимізацією зазору та мінімізацією помилок класифікації на навчанні.

У реалізації використовується лінійне ядро (linear kernel), яке будує лінійну розділяючу гіперплощину в оригінальному просторі ознак. Вибір лінійного ядра обумовлений високою розмірністю TF-IDF представлення (3000 ознак) – в такому просторі дані часто стають лінійно роздільними, і складні нелінійні ядра не дають значного виграшу, але збільшують обчислювальну складність.

Параметр регуляризації C встановлено у значення 1.0 за замовчуванням, що забезпечує баланс між складністю моделі та помилкою на навчанні. Увімкнено обчислення ймовірностей класів через калібрацію Платта, що дозволяє отримувати не лише бінарні прогнози, але й оцінки впевненості моделі.

Навчання SVM зводиться до розв'язання задачі квадратичної оптимізації. Використовується метод Sequential Minimal Optimization , який ефективно знаходить опорні вектори. Алгоритм ітеративно оптимізує пари множників Лагранжа, що відповідають обмеженням задачі, поки не досягне глобального оптимуму. Після навчання модель зберігає лише опорні вектори, а інші точки відкидаються, що робить прогнозування ефективним.

Для класифікації обчислюється значення функції рішення $f(x)=w \cdot x+b$, де w та b знайдені під час навчання. Знак цієї функції визначає клас: якщо $f(x)>0$, зразок належить до позитивного класу (фішинг), інакше – до негативного (легітимний). Відстань $|f(x)|$ від точки до гіперплощини характеризує впевненість класифікації.

SVM ефективний у високорозмірних просторах, що критично важливо для текстових даних з тисячами ознак. Метод має добру математичну основу та гарантує знаходження глобального оптимуму. Використання лише опорних векторів робить модель компактною. Однак навчання може бути повільним на великих датасетах, і метод чутливий до вибору параметрів.

Logistic Regression: Лінійна модель для класифікації, яка оцінює ймовірність належності зразка до класу через логістичну (сигмоїдну) функцію.

Незважаючи на назву "регресія", метод призначений саме для задач класифікації.

Модель припускає, що логарифм відношення ймовірностей (log-odds) є лінійною функцією ознак: $\log(P(y=1|x) / P(y=0|x)) = w \cdot x + b$. Звідси виводиться формула для ймовірності: $P(y=1|x) = 1 / (1 + \exp(-(w \cdot x + b)))$. Функція $\exp(-(w \cdot x + b))$ є експонентою, а весь вираз – це сигмоїдна функція, яка відображає будь-яке дійсне число в інтервал [0, 1], що інтерпретується як ймовірність.

Для навчання використовується логарифмічна функція втрат (log loss або binary cross-entropy): $L(w,b) = -\Sigma[y_i \log(P(y=1|x_i)) + (1-y_i) \log(1-P(y=1|x_i))]$. Ця функція штрафує модель за невпевнені правильні прогнози та впевнені неправильні прогнози. Мета навчання – знайти параметри w та b, які мінімізують цю функцію на навчальних даних.

У реалізації використовується алгоритм оптимізації Limitedmemory Broyden–Fletcher–Goldfarb–Shanno – це метод квазі-ньютонівської оптимізації другого порядку. LBFGS ефективніший за звичайний градієнтний спуск, оскільки використовує апроксимацію гессіана (матриці других похідних) для вибору напрямку та величини кроку. Метод потребує обмеженої пам'яті, що робить його придатним для великих задач.

Встановлено максимальну кількість ітерацій 1000, чого достатньо для збіжності на датасеті. Алгоритм зупиняється раніше, якщо досягається збіжність за критерієм малої зміни функції втрат між ітераціями.

Для запобігання перенавчанню використовується L2-регуляризація (ridge регуляризація), яка додає до функції втрат штраф за великі значення ваг: $L\_total = L(w,b) + \lambda||w||^2$. Параметр λ контролює силу регуляризації – більші значення призводять до простіших моделей з меншими вагами. Регуляризація особливо важлива у високорозмірних задачах, де існує ризик перенавчання.

Для нового зразка спочатку обчислюється лінійна комбінація ознак $z=w \cdot x+b$, потім застосовується сигмоїдна функція для отримання ймовірності $P(y=1|x) = 1/(1+\exp(-z))$. За замовчуванням використовується поріг 0.5: якщо ймовірність більша за 0.5, зразок класифікується як позитивний клас (фішинг),

інакше – як негативний (легітимний). Поріг може налаштовуватися залежно від балансу між false positives та false negatives.

Logistic Regression має високу інтерпретованість – ваги моделі показують внесок кожної ознаки в рішення, причому позитивні ваги збільшують ймовірність фішингу, а негативні – зменшують. Модель швидко навчається та працює, потребує мало пам'яті. Вихідні ймовірності добре калібровані та можуть використовуватися для ранжування. Метод стабільний та легко масштабується на великі датасети. Основне обмеження – лінійність моделі, яка може бути недостатньою для складних нелінійних залежностей у даних. На рис. 4.4 зображено теплову карту метрик.

Рисунок 4.4 – Теплова карта метрик ефективності

4.3 Результати експериментів

Тестування проводилось на всьому датасеті (18634 зразки) для отримання об'єктивної оцінки ефективності методів. Результати представлено у табл. 4.1.

Таблиця 4.1 – Оцінка ефективності методів

Метод	Accuracy	Precision	Recall	F1-Score
Naive Bayes	0.9507	0.9351	0.9397	0.9374
Random Forest	0.9892	0.9732	0.9999	0.9864
SVM	0.9777	0.9589	0.9854	0.9719
Logistic Regression	0.9692	0.9516	0.9710	0.9612
Rule-Based	0.6719	0.7850	0.2257	0.3505

Для всебічної оцінки якості розроблених моделей було використано чотири ключові метрики: Accuracy, Precision, Recall та F1-Score, кожна з яких висвітлює різні аспекти надійності системи. Загальна ефективність методу визначається показником Accuracy, який відображає частку правильно класифікованих листів серед усієї вибірки. У цьому контексті беззаперечним лідером став алгоритм Random Forest, який досяг точності 98,92%, що свідчить про вкрай низький рівень помилок — лише 1,08%. Важливим аспектом аналізу є метрика Precision, яка показує, наскільки можна довіряти вердикту системи, коли вона позначає лист як загрозу. Для Random Forest цей показник склав 97,32%, що означає, що із 7512 листів, маркованих як фішинг, лише 201 виявився легітимним повідомленням (False Positives).

Однак найбільш критичною метрикою для систем кібербезпеки є Recall, що демонструє здатність моделі виявляти всі наявні загрози. Random Forest продемонстрував тут майже ідеальний результат у 99,99%, пропустивши лише одну атаку із 7311 реальних випадків фішингу (True Positives). Саме цей показник є вирішальним, оскільки кожен пропущений лист може призвести до успішної кібератаки. Для балансування між точністю та повнотою використовується F1-Score, де Random Forest також показав найкращий результат — 0.9864. Статистика цього методу підтверджує його надійність: окрім високого рівня виявлення загроз, він правильно ідентифікував 11121 легітимний лист (True Negatives), припускаючись мінімальної кількості хибних спрацьовувань.

На противагу лідеру, метод Rule-Based (на основі правил) продемонстрував найгірші результати, які можна охарактеризувати як небезпечні для реального застосування. Система змогла виявити лише 1650 фішингових листів (True Positives), що становить всього 22,6% від загальної кількості загроз. Відповідно, кількість пропущених атак (False Negatives) сягнула катастрофічної цифри у 5662 листи, тобто 77,4% фішингу залишилися непоміченими, що робить цей метод непридатним для використання як основного засобу захисту.

Проміжні позиції між лідером та аутсайдером зайняли інші методи машинного навчання. Метод опорних векторів (SVM) показав другий найкращий результат за F1-Score (0.9719), пропустивши 107 фішингових атак (1,46%). Логістична регресія виявилася менш ефективною, не помітивши 212 загроз (2,90%). Найслабшим серед ML-методів виявився наївний байєсівський класифікатор (Naive Bayes), який пропустив 441 фішинговий лист (6,03%), що значно поступається показникам Random Forest та SVM і свідчить про недостатню здатність моделі враховувати складні залежності в даних.

4.4.Аналіз помилок

Причини помилок Rule-Based методу: Метод на основі правил показав найгірші результати (F1-Score: 0.3505), пропустивши 77,4% фішингових листів. Основні причини наступні.

Обхід простих правил зловмисниками. Сучасні фішери уникають явних ключових слів типу "urgent", "verify", використовуючи синоніми або контекстуальні формулювання. Наприклад, замість "Click here to verify" використовують "Please review your account details at your earliest convenience".

Відсутність контекстного аналізу. Метод не розуміє семантику тексту. Легітимний лист від банку може містити слова "account", "verify", "update", але в іншому контексті. Правила не можуть це розрізнити.

Висока кількість false negatives. З 7312 фішингів виявлено лише 1650. Це означає, що 5662 атаки пройшли через фільтр – неприйнятний рівень для системи захисту.

Загальний графік порівнянь зображено нижче на рис. 4.5.

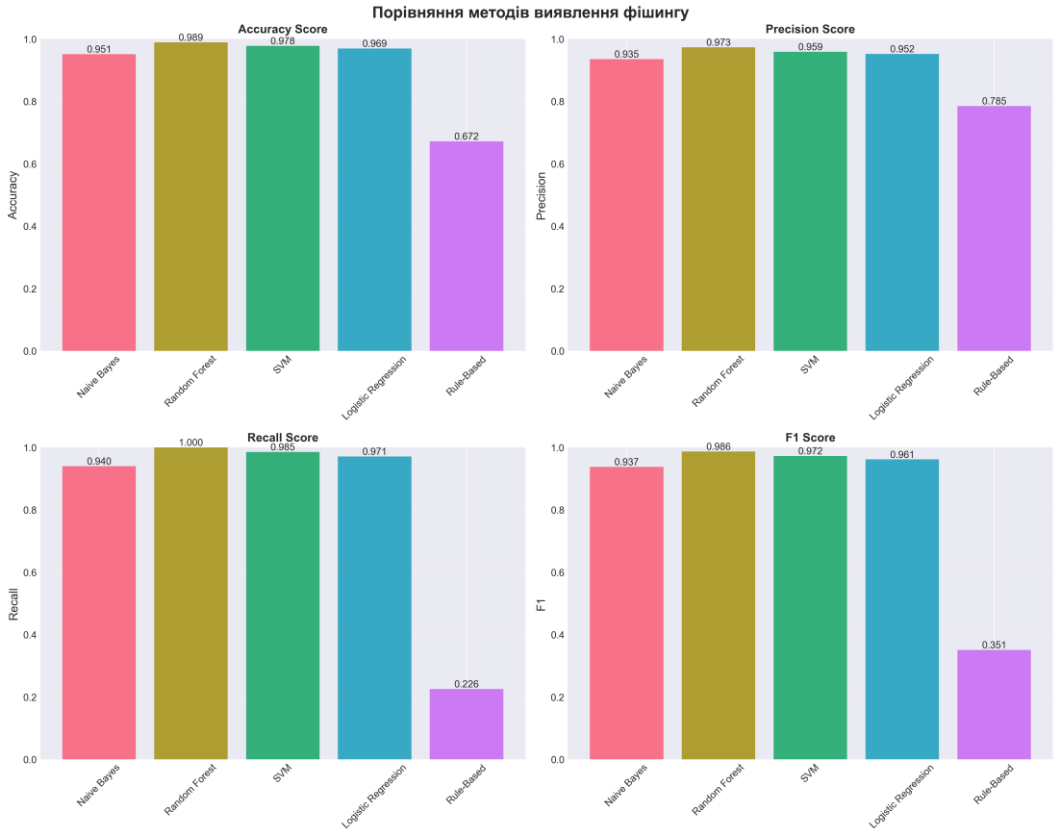


Рисунок 4.5 – Порівняння методів виявлення фішингу

Далі розглянемо причини помилок методів машинного навчання.

False Positives (легітимні позначені як фішинг): Random Forest помилково класифікував 201 легітимний лист. Аналіз показав, що це офіційні листи з банків, платіжних систем, які містять фрази на кшталт "підтвердити транзакцію", "оновити дані картки". Модель навчилася асоціювати такі фрази з фішингом, оскільки вони дійсно часто зустрічаються в атаках.

SVM помилково позначив 309 легітимних листів (2,73%) , що зображено на матриці помилок – рис.4.6. Причина – схожість лексичного складу з фішингом в окремих випадках.

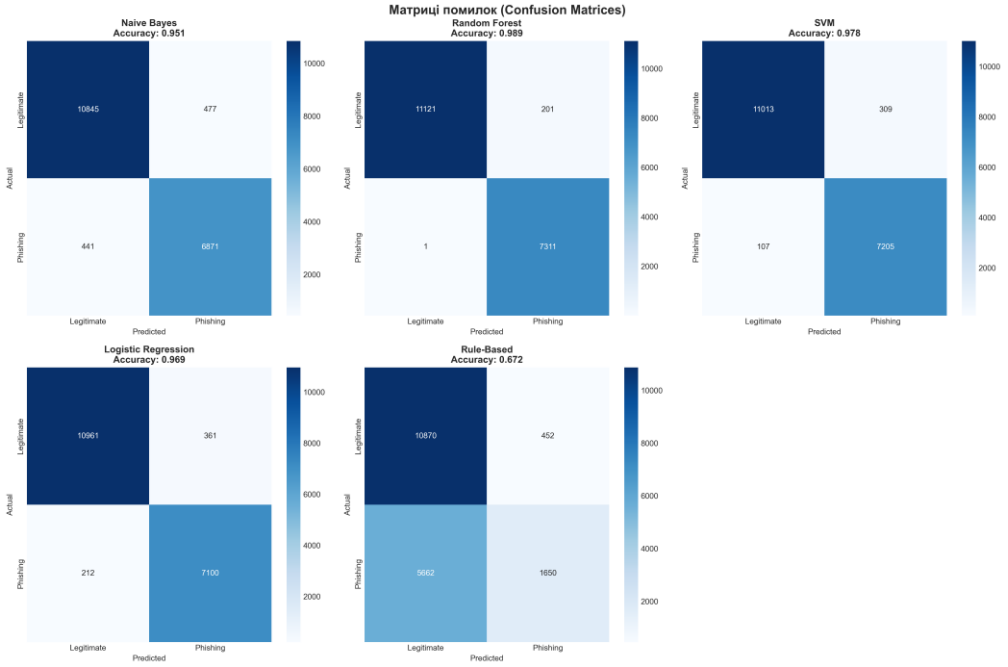


Рисунок 4.6 – Матриці помилок

Ось перероблений та розширений текст у формі цілісної аналітичної записки без використання списків.

Детальний аналіз помилок другого роду (False Negatives) демонструє суттєву різницю в ефективності обраних моделей. Беззаперечним лідером у цьому дослідженні виявився алгоритм Random Forest, який пропустив лише один фішинговий лист. Цей винятковий випадок став наслідком специфічної, витонченої атаки, де зловмисники використали нестандартну лексику та стилістику, що не зустрічалися у навчальній вибірці, тим самим обійшовши загальні правила класифікації. Натомість метод опорних векторів (SVM) показав дещо слабший результат, не ідентифікувавши 107 фішингових повідомлень, що становить 1,46% від загальної кількості. Основною вразливістю цієї моделі стали складні атаки, замасковані під звичайне ділове листування з використанням нейтральної мови та мінімумом очевидних підозрілих термінів. Найбільш критичні результати продемонстрував наївний байєсівський класифікатор (Naive Bayes), який пропустив 441 загрозу, що склало 6,03%. Така значна похибка пояснюється фундаментальним обмеженням алгоритму — припущенням про повну незалежність ознак.

Оскільки модель не враховує семантичні зв'язки між словами, вона втрачає контекст, який часто є вирішальним для виявлення маніпуляцій у реальних текстах.

Для суттєвого покращення результатів детекції необхідно застосувати комплексний підхід, модернізуючи як методи на основі правил, так і алгоритми машинного навчання. У контексті Rule-Based методу першочерговим завданням є значне розширення словника індикаторів до понад ста термінів, що дозволить охопити ширший спектр загроз. Однак простого пошуку за ключовими словами недостатньо, тому слід впровадити контекстний аналіз для перевірки оточення підозрілих слів, а також використовувати регулярні вирази, які допоможуть виявляти спроби обфускації (наприклад, написання "v3r1fy" замість "verify"). Крім того, аналіз має виходити за межі тіла листа і включати перевірку метаданих, зокрема заголовків та репутації відправника.

Що стосується методів машинного навчання, то шлях до вдосконалення лежить через збагачення даних та ускладнення архітектури. Критично важливим є розширення датасету зразками нестандартних та новітніх атак, що дозволить моделям навчатися на складних кейсах. Паралельно необхідно провести поглиблений Feature Engineering, додавши такі ознаки, як довжина URL, кількість гіперпосилань, наявність вкладень та вік домену. Для подолання недоліків окремих алгоритмів доцільно використовувати ансамблеві методи, наприклад, комбінацію Random Forest та SVM, що дозволить знизити рівень False Negatives. Найбільш перспективним напрямком є інтеграція методів Deep Learning, таких як моделі BERT або GPT, які здатні розуміти глибокий контекст і семантику тексту краще за класичні алгоритми.

Запорукою стабільної роботи системи є регулярне, щомісячне оновлення моделей на свіжих даних, оскільки фішингові техніки постійно еволюціонують. Зрештою, налаштування системи повинно базуватися на балансі метрик відповідно до бізнес-цілей. Якщо пріоритетом є безпека

критичної інфраструктури, банків чи державних установ, необхідно оптимізувати метрику Recall, навіть якщо це призведе до збільшення кількості хибних спрацьовувань. У випадках, коли критично важливо не блокувати легітимну комунікацію користувачів, фокус варто змістити на максимізацію Precision.

4.5. Висновки до четвертого розділу

Експериментальні дослідження показали значну перевагу методів машинного навчання порівняно з класичним підходом на основі правил у завданнях виявлення фішингових атак. Random Forest продемонстрував найкращі результати серед усіх методів: точність 98,92%, F1-Score 0.9864, та критично важливий показник Recall 99,99% (лише 1 фішингова атака була пропущена з 7312). Це робить метод оптимальним вибором для систем захисту електронної пошти, де пропуск атаки може призвести до серйозних наслідків. Ансамблевий характер методу забезпечує стійкість до перенавчання та здатність виявляти складні патерни.

Другий найкращий результат показав SVM F1-Score 0.9719, пропустивши 1,46% фішингів. Метод ефективний у високорозмірних просторах та може використовуватися як альтернатива Random Forest або в ансамблі з ним для підвищення надійності.

Logistic Regression та Naive Bayes показали прийнятні, але гірші результати (F1-Score 0.9612 та 0.9374 відповідно Перевага в тому, що вони швидкі та прості, що може бути корисним у системах з обмеженими ресурсами. Rule-Based метод виявився неефективним (F1-Score 0.3505), пропустивши 77,4% фішингів. Сучасні атаки легко обходять прості правила, що робить цей підхід непридатним як самостійне рішення. Проте метод може використовуватися як додатковий фільтр швидкої перевірки перед застосуванням ML моделей.

Аналіз помилок показав, що основна проблема Rule-Based методу – відсутність адаптивності та контекстного розуміння. ML методи краще

впораються з варіативністю мови та складними патернами, але потребують регулярного оновлення для виявлення нових типів атак.

Для максимальної ефективності рекомендується використовувати Random Forest або гібридний підхід: комбінацію Random Forest + SVM + аналіз метаданих листа - репутація відправника, структура посилань. Це забезпечить Recall >99% при збереженні високої точності.

ВИСНОВКИ

У кваліфікаційній роботі магістра на тему «Дослідження та аналіз механізмів захисту від соціально-інженерних атак та рекомендації із розробки методів їх виявлення», я досягла всіх поставлених перед собою цілей. Це дослідження не лише підтвердило, що людський фактор залишається найслабшою ланкою в кіберзахисті, але й дозволило мені створити практичний інструментарій для протидії цій загрозі.

Виконавши усі завдання, отримано наступні науково-практичні результати.

Я детально проаналізувала теоретичну базу та історію розвитку соціальної інженерії, включаючи її сучасні прояви з використанням штучного інтелекту для створення переконливих атак. На основі цього я обґрунтувала необхідність відходу від формальних методів навчання та технічних фільтрів на користь цілісної, людино-орієнтованої стратегії захисту.

Була розроблена методика боротьби із соціальною інженерією. Ця методика базується на інтеграції психологічного розуміння маніпуляцій із технологічним захистом. Її головна мета – створити ефективну політику безпеки та процедури реагування, які захищають як інфраструктуру, так і кожного окремого працівника.

Я дослідила специфіку застосування цієї методології на прикладі критичної інфраструктури – АТ «Укрзалізниця». Визначила ключові вразливості, класифікувала рівні ризиків для різних груп персоналу (від диспетчерів до офісних працівників) та підготувала конкретний перелік контрзаходів, адаптованих під реальні залізничні операції.

Для підвищення ефективності реагування на інциденти було створено унікальну модель розслідування. Ця модель не просто фіксує факт атаки, а використовує соціальний граф організації для математичної оцінки

ймовірності поширення загрози по ланцюжку контактів, дозволяючи оперативно ідентифікувати джерело та зупинити ланцюгову реакцію.

Мною були деталізовані всі етапи впровадження розробленої системи захисту, включаючи створення адаптивної програми навчання через практичні симуляції та рольові ігри.

Економічна доцільність методики була підтверджена розрахунками. Витрати на впровадження системи є виправданими, оскільки мінімізується потенційна шкода від успішних атак на критичні вузли, що в кінцевому підсумку захищає компанію від мільйонних фінансових і репутаційних втрат.

На завершальному етапі дослідження проведено експериментальну перевірку ефективності методів машинного навчання для виявлення фішингових атак. У ході роботи було реалізовано та порівняно п’ять різних підходів, включаючи Rule-Based метод та алгоритми Naive Bayes, Random Forest, SVM і Logistic Regression. Результатом цього етапу стало обґрунтування вибору оптимальної моделі класифікації на основі аналізу показників точності та повноти.

Таким чином, моя дипломна робота пропонує практично орієнтовану, гнучку і науково обґрунтовану стратегію, яка може бути успішно застосована у великих організаціях для побудови стійкої культури кібербезпеки.

ПЕРЕЛІК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Чалдіні Р. Б. (Cialdini R.B.) Класифікація психологічних принципів впливу. New York : HarperCollins Publishers, 2007. 330-336 с.

2. Митник К. Д. (Mitnick K.D.) Класична соціальна інженерія. Методичний посібник. NYC: Wiley Books, 2008. 273 с.

3. Lockheed Martin Cyber Kill Chain. Модель життєвого циклу кібернападу [Електронний ресурс]. Режим доступу: https://www.lockheedmartin.com/en-us/capabilities/cyber/cyber-kill-chain.html

4. MITRE ATT&CK. Модель життєвого циклу кібернападу [Електронний ресурс]. Режим доступу: https://attack.mitre.org/

5. KnowBe4, Proofpoint, PhishER. Платформи для моделювання фішингових атак [Електронний ресурс]. Режим доступу: https://www.g2.com/categories/security-awareness-training

6. Концепції про нейролінгвістичне програмування (НЛП) та гіпноз. Дослідження сили підсвідомих рішень [Електронний ресурс]. Режим доступу: https://www.frontiersin.org/articles/10.3389/fpsyg.2014.00940/full

7.Що таке фішинг? [Електронний ресурс]. Режим доступу: https://www.microsoft.com/uk-ua/security/business/security-101/what-is-phishing

8. Machine Learning, ML [Електронний ресурс]. Режим доступу: https://www.it.ua/knowledge-base/technology-innovation/machine-learning

9. DNS Spoofing: Як Захистити Систему? [Електронний ресурс]. Режим доступу: https://cyberset.com.ua/beginners/network/dns-spoofing-yak-zahistiti-systemu/

10. ДСТУ ISO/IEC 27004:2018. Інформаційні технології. Методи та засоби забезпечення безпеки. Системи управління інформаційною безпекою. Моніторинг, вимірювання, аналіз і оцінювання. Київ : ДП «УкрНДНЦ», 2018 [Електронний ресурс]. Режим доступу: https://online.budstandart.com/ua/catalog/doc-page.html?id_doc=78518

11. ДСТУ ISO/CEI 27005:2015. Інформаційні технології. Методи та засоби забезпечення безпеки. Управління ризиками інформаційної безпеки. Київ : ДП «УкрНДНЦ», 2015 [Електронний ресурс]. Режим доступу: https://online.budstandart.com/ua/catalog/doc-page.html?id_doc=66912

12. ДСТУ ISO/IEC 27001:2022. Інформаційна безпека, кібербезпека та захист приватності. Системи управління інформаційною безпекою. Вимоги [Електронний ресурс]. Режим доступу: https://online.budstandart.com/ua/catalog/doc-page.html?id_doc=104398

13. ДСТУ ISO/IEC 27002:2022. Інформаційна безпека, кібербезпека та захист приватності. Засоби контролювання безпеки [Електронний ресурс]. Режим доступу:https://online.budstandart.com/ua/catalog/doc-page.html?id_doc=106958

14. Дослідження про соціальні мережі. Аналіз "лайків", прогнозування рис характеру, вплив екстраверсії та невротизму [Електронний ресурс]. Режим доступу: https://www.pnas.org/doi/10.1073/pnas.1717621115

15. Дослідження про соціальний граф та кластеризацію. Аналіз факторів, що стимулюють взаємодію користувачів [Електронний ресурс]. Режим доступу: https://arxiv.org/abs/1803.07686

16. Дослідження про інформаційно-психологічний вплив. Маніпуляція та зміна сприйняття реальності [Електронний ресурс]. Режим доступу: https://www.tandfonline.com/doi/full/10.1080/01900692.2018.1524339

17. Дослідження про моделювання зміни свідомості під дією зовнішніх чинників [Електронний ресурс]. Режим доступу: https://www.nature.com/articles/s41598-021-99884-2

18. Ф. Е. ГЕЧЕ ТЕОРІЯ ЙМОВІРНОСТЕЙ І МАТЕМАТИЧНА СТАТИСТИКА [Навчальний посібник] Режим доступу: https://fpk.in.ua/images/biblioteka/2fmb_finansy/Teoriya--ymovirnostey.pdf

19. ДСТУ ISO/IEC 27032:2016. Інформаційні технології. Методи та засоби забезпечення безпеки. Настанова щодо кібербезпеки [Електронний

ресурс]. Режим доступу: https://online.budstandart.com/ua/catalog/doc-page.html?id_doc=69128

20. Кейс атаки на АТ «Київстар» (грудень 2023). Аналіз початку атаки з фішингу [Електронний ресурс]. Режим доступу: https://dou.ua/lenta/news/kyivstar-cyber-attack-restoration/

21. Кейс атаки на «Прикарпаттяобленерго» (грудень 2015). Аналіз початку атаки з цільового фішингу [Електронний ресурс]. Режим доступу: https://www.bbc.com/ukrainian/society/2016/01/160106_cyber_attacks_electricity_ukraine_vc

22. Кодекс законів про працю України (КЗпП України) [Електронний ресурс]. Режим доступу: https://zakon.rada.gov.ua/laws/show/322-08

23. Закон України «Про основні засади забезпечення кібербезпеки» [Електронний ресурс]. Режим доступу: https://zakon.rada.gov.ua/laws/show/2163-19

24. Random Forests [Електронний ресурс]. Режим доступу: https://www.it.ua/knowledge-base/technology-innovation/machine-learning