Національний університет кораблебудування
ім.адм. Макарова

Кваліфікаційна бакалаврська робота

на тему «**Розроблення інформаційної системи виявлення фішингових сайтів та підозрілих листів на основі ML**»

Здобувач Роман НЕМЧЕНКО
Керівник Ірина БАКОВСЬКА

Миколаїв, 2026

Недоліки існуючих рішень

1. Залежність від чорних списків.
2. Використання схожих доменів та піддоменів.
3. Недостатній аналіз контенту вебсторінки
4. Недостатня адаптивність до змін у фішингових атаках.
5. Проблема соціальної інженерії.
6. Використання підроблених адрес відправника
7. Складність аналізу вкладень
8. Використання HTML та прихованих посилань
9. Високий рівень помилкових спрацювань
10. Низький рівень виявлення прихованих атак
11. Відсутність комплексного підходу
12. Відсутність автоматичного навчання та оновлення

Постановка задачі

Об'єктом дослідження є процеси виявлення та класифікації фішингових вебсайтів і підозрілих електронних повідомлень.

Предметом дослідження є методи та алгоритми машинного навчання, що використовуються для аналізу вебресурсів і електронних листів з метою виявлення ознак фішингу.

Метою даної роботи є підвищення ефективності виявленння виявлення фішингових сайтів та підозрілих листів на основі методів машинного навчання для підвищення ефективності захисту користувачів від кіберзагроз.

Концепція системи

№	Концепція системи	Опис	Переваги	Недоліки
1	Blacklist / репутаційні бази	Перевірка URL, доменів і адрес відправників за чорними списками та репутаційними сервісами	- висока швидкість перевірки- проста реалізація- мінімальні ресурси	- не виявляє нові загрози (zero-day)- залежність від оновлення баз- обмежена функціональність (без аналізу тексту листа)
2	Rule-based система (евристики)	Аналіз URL та листів за правилами: ключові слова, символи, структура URL, вкладення, кількість посилань	- швидка робота- зрозуміла логіка перевірки- не потребує навчальних даних	- потребує постійного оновлення правил- багато хибних спрацювань- слабка адаптація до нових атак
3	**ML-система (машинне навчання)**	**Використання моделей ML для класифікації сайтів і листів за набором ознак (URL, текст, метадані)**	**- висока точність- здатність виявляти нові атаки- можливість самонавчання- комплексний аналіз (URL + текст)**	**- потрібні навчальні дані- складніша реалізація- потребує ресурсів для навчання моделей**

Математичне забезпечення

Одним із базових методів класифікації є логістична регресія, яка обчислює ймовірність того, що об’єкт належить до класу фішингових. Функція ймовірності задається сигмоїдною функцією

$$P(y = 1|X) = \sigma(z) = \frac{1}{1 + e^{-z}}$$

де

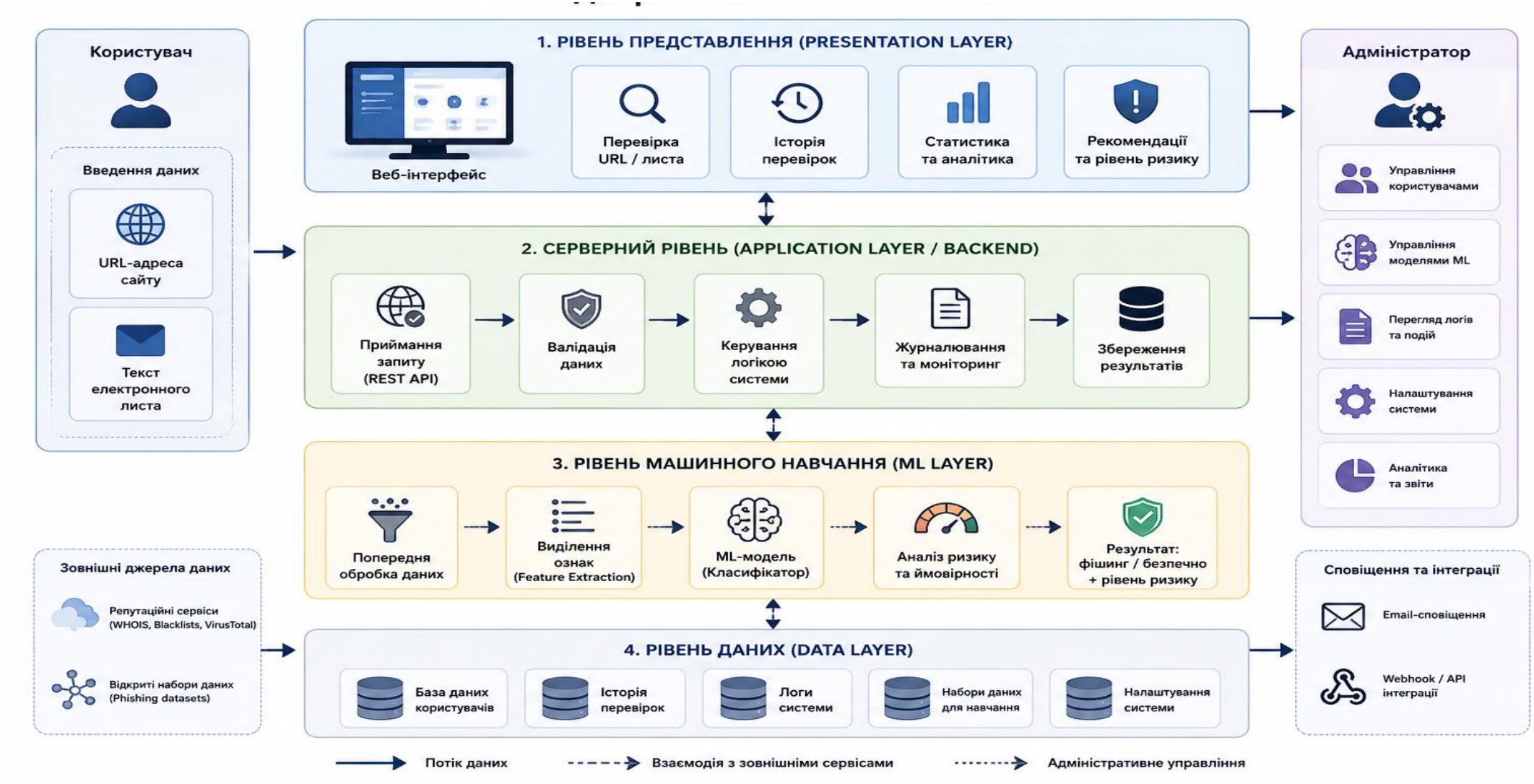
$$z = w_0 + w_1x_1 + w_2x_2 + ... + w_nx_n$$

w_0 — зміщення (bias), w_i — вагові коефіцієнти ознак.

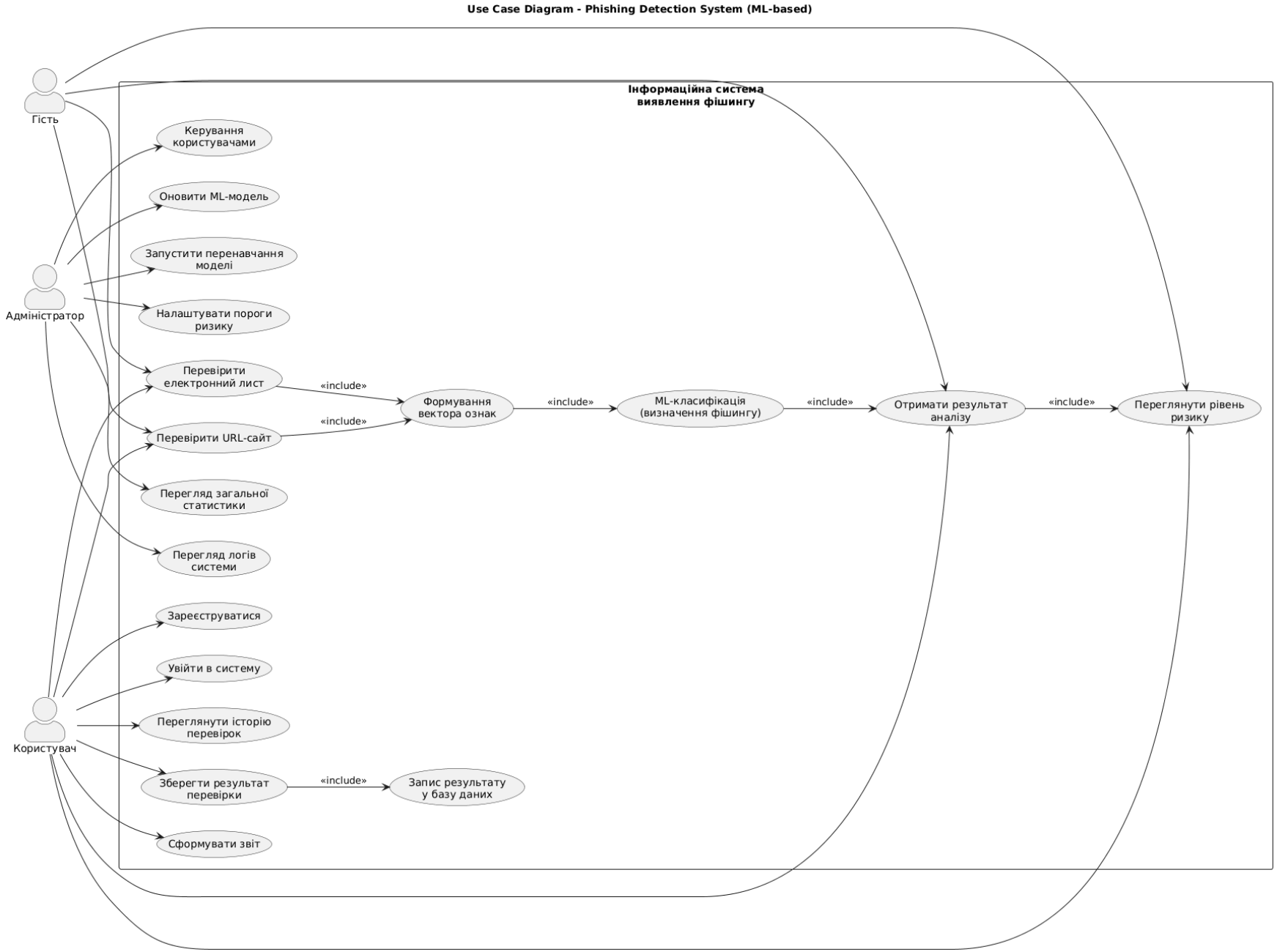
Рішення приймається за правилом:

$$\hat{y} = \begin{cases} 1, & P(y = 1|X) \geq 0.5 \\ 0, & P(y = 1|X) < 0.5 \end{cases}$$

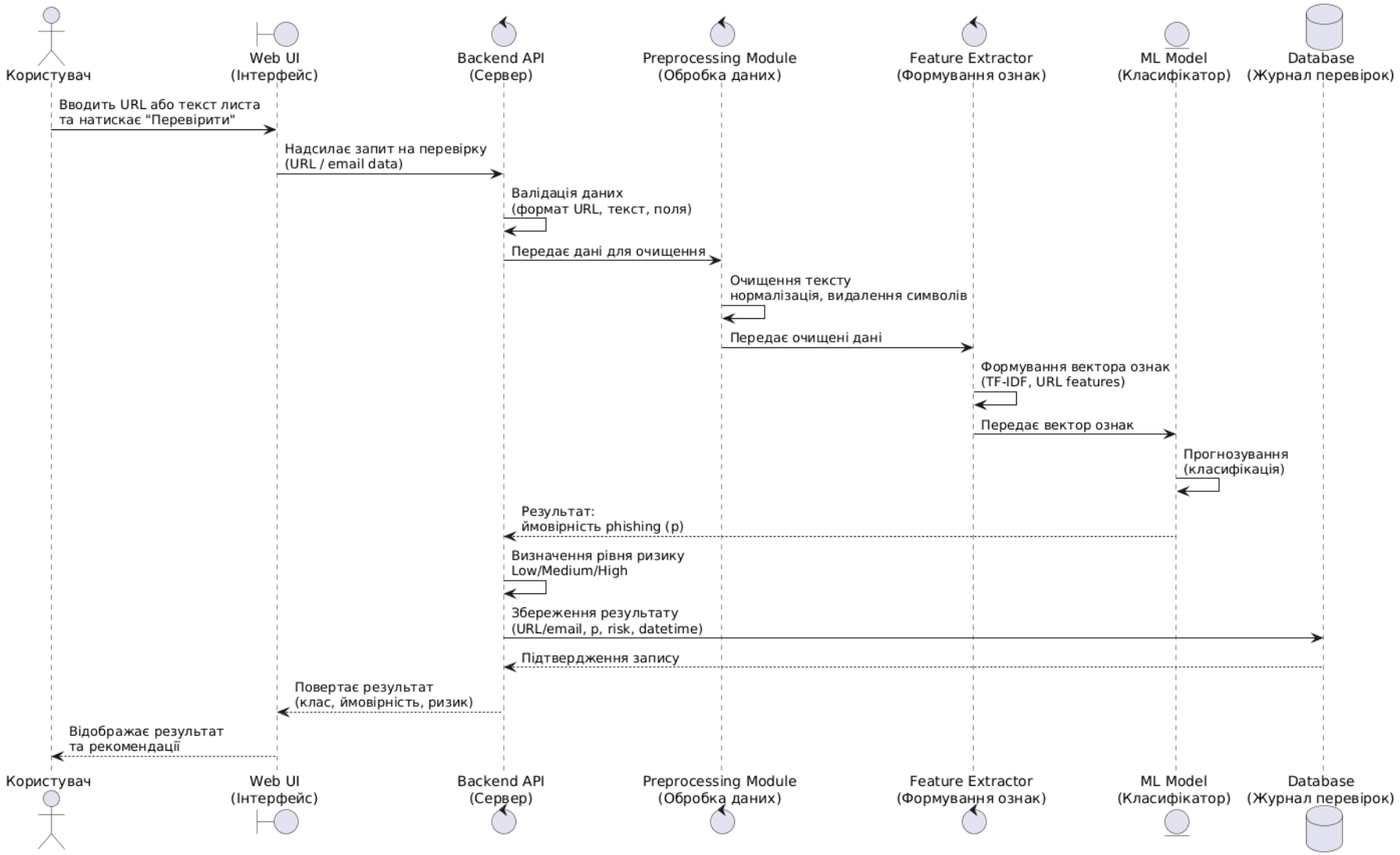
Архітектура системи



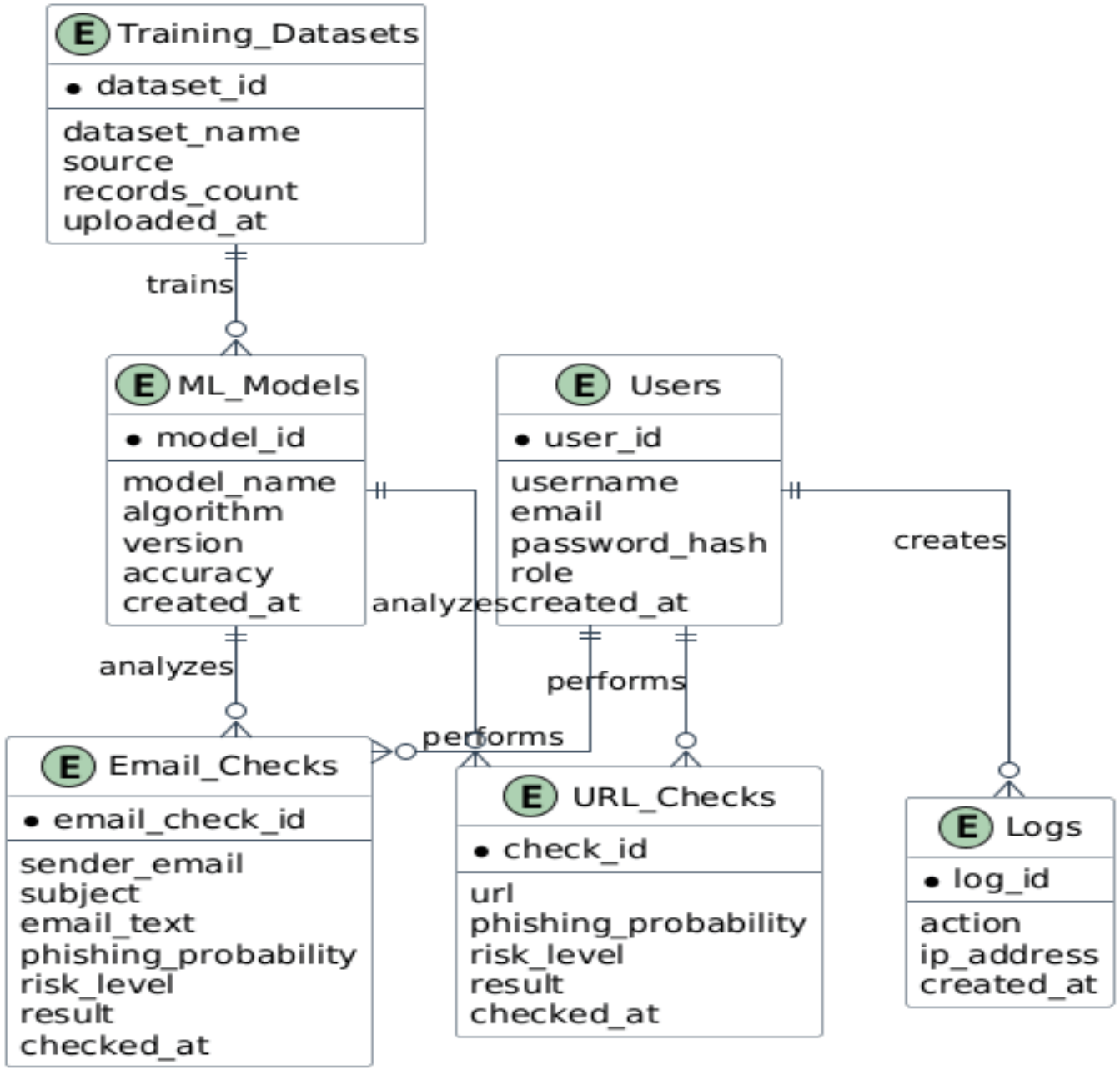
USE CASE


Use Case Diagram - Phishing Detection System (ML-based)
Інформаційна система виявлення фішингу
Гість
Адміністратор
Користувач
Керування користувачами
Оновити ML-модель
Запустити перенавчання моделі
Налаштувати пороги ризику
Перевірити електронний лист
Перевірити URL-сайт
Перегляд загальної статистики
Перегляд логів системи
Зареєструватися
Увійти в систему
Переглянути історію перевірок
Зберегти результат перевірки
Сформувати звіт
Формування вектора ознак
ML-класифікація (визначення фішингу)
Отримати результат аналізу
Переглянути рівень ризику
Запис результату у базу даних
«include»
«include»
«include»
«include»
«include»
«include»

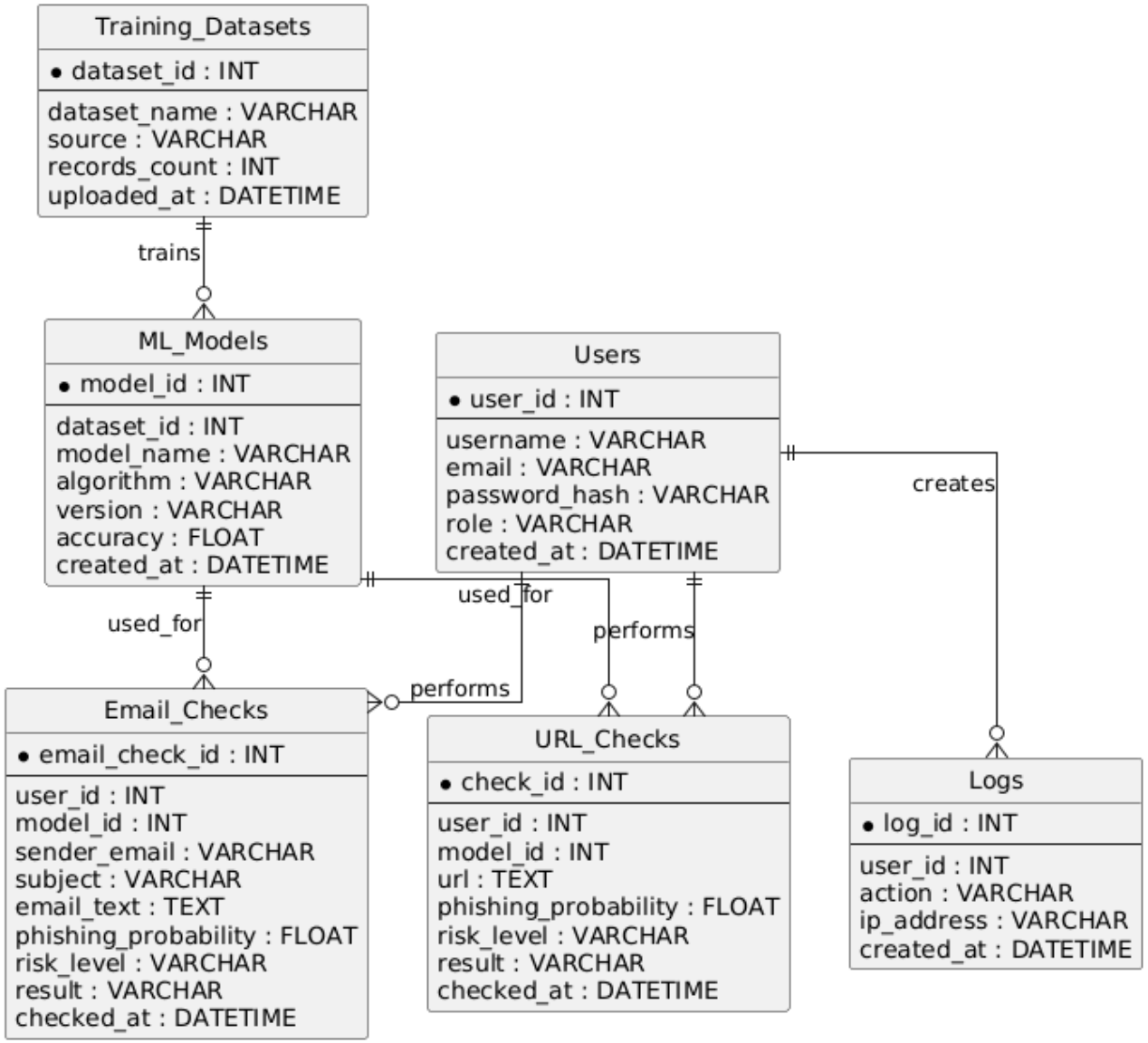
Діаграма послідовності



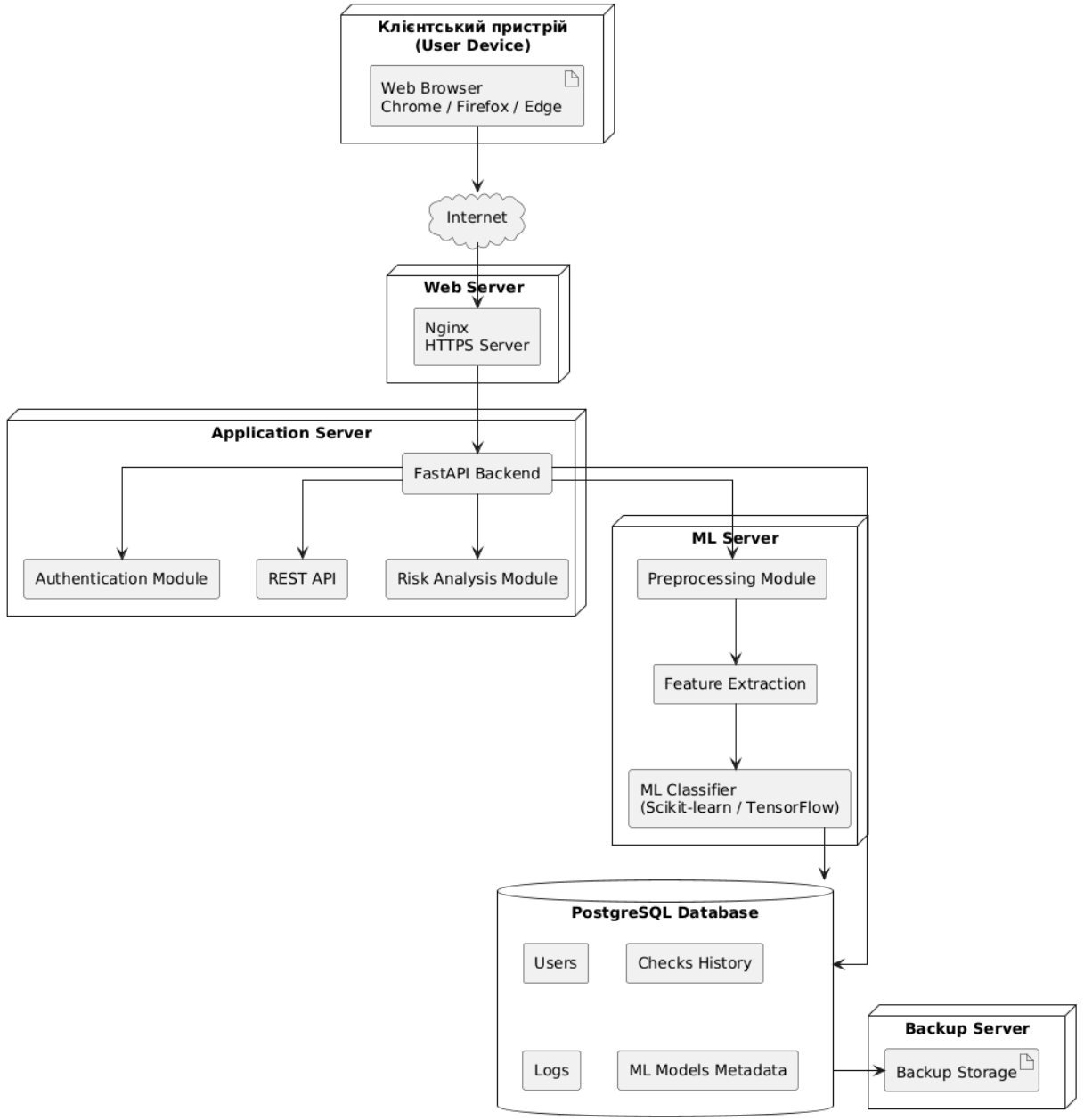
ER- Діаграма


Training_Datasets
dataset_id
dataset_name
source
records_count
uploaded_at
trains
ML_Models
model_id
model_name
algorithm
version
accuracy
created_at
Users
user_id
username
email
password_hash
role
created_at
creates
analyzes
analyzes
performs
performs
Email_Checks
email_check_id
sender_email
subject
email_text
phishing_probability
risk_level
result
checked_at
URL_Checks
check_id
url
phishing_probability
risk_level
result
checked_at
Logs
log_id
action
ip_address
created_at

Логічна модель даних



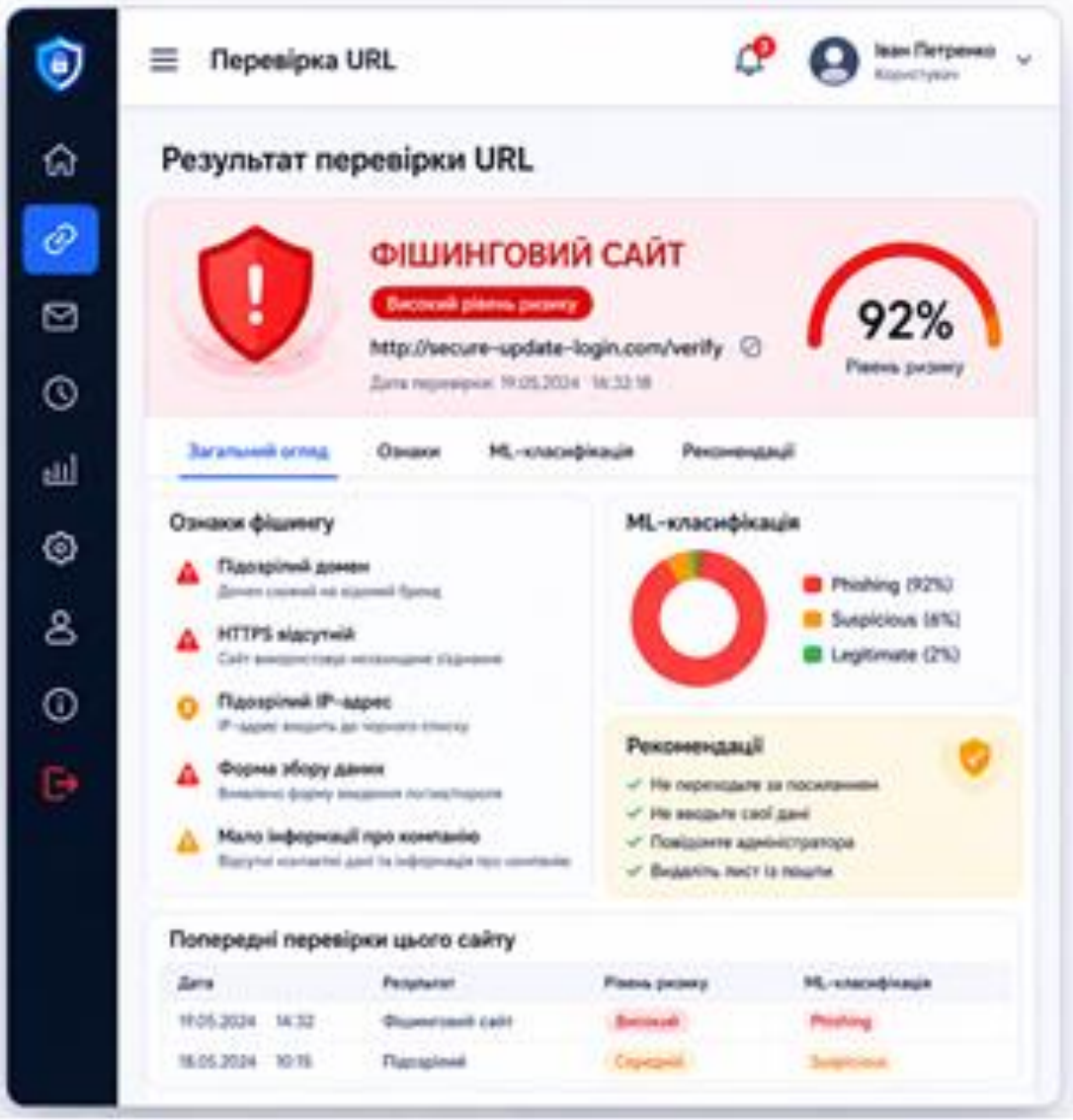
Діаграма розгортання



Екранні форми розробки

Екранна форма 1 – Головна панель системи

Екранна форма 2 – Результат перевірки URL



Висновки

1. Розроблено інформаційну систему виявлення фішингових сайтів та підозрілих електронних листів на основі ML.
2. Проведено аналіз предметної області та сучасних методів кіберзахисту.
3. Обґрунтовано використання методів машинного навчання для виявлення загроз.
4. Спроєктовано архітектуру системи, базу даних та UML-діаграми.
5. Визначено вимоги до програмного й апаратного забезпечення та модель розгортання.
6. Розроблено заходи з охорони праці та безпечної експлуатації системи.
7. Запропоноване рішення підвищує швидкість виявлення загроз і рівень інформаційної безпеки.

Дякую за увагу!