https://doi.org/10.31891/2307-5732-2025-357-56
УДК 04.056.53:004.89

ВІЖЕВСЬКИЙ ПЕТРО
Хмельницький національний університет
https://orcid.org/0009-0009-4851-0839
e-mail: vizhevskyipv@khmnu.edu.ua
КРАВЧИК ЮРІЙ
Хмельницький національний університет
https://orcid.org/0000-0002-2780-5605
e-mail: yurii_kravchyk@khmnu.edu.ua

МОДЕЛЬ СИСТЕМИ ЗАПОБІГАННЯ ВИТОКУ ДАНИХ З ЕВОЛЮЦІЙНОЮ АДАПТАЦІЄЮ НА ОСНОВІ ГЕНЕТИЧНИХ АЛГОРИТМІВ

У роботі запропоновано модель системи запобігання витоку даних (DLP) з еволюційною адаптацією, що інтегрує три компоненти у чотирирівневій архітектурі: класифікацію документів на основі генетичного алгоритму з продукційними IF-THEN правилами, двовіконний детектор дрейфу з архівом політик теплого старту, та поведінкове профілювання користувачів з експоненційним забуванням. Хромосомне кодування правил класифікації забезпечує повну інтерпретованість рішень, а багатокритеріальна функція пристосованості балансує точність зі складністю набору правил. На основі експериментальної оцінки на наборах даних SMS Spam Collection (5574 повідомлення) та Synthetic PII (2000 документів) встановлено, що генетичний класифікатор досягає F1-міри 0,985 на задачі виявлення персональних даних, поступаючись ансамблевим методам (F1 = 1,000) лише на 1,5 відсоткових пункти. Механізм еволюційної адаптації забезпечує приріст prequential accuracy на 13,9% порівняно зі статичною моделлю, а детектор дрейфу коректно виявляє всі точки раптового дрейфу (Recall = 1,00). Запропонована модель придатна для розгортання у регульованих галузях, де відповідність стандартам GDPR та HIPAA вимагає обґрунтування рішень про блокування інформаційних потоків. Підхід може бути розширений інтеграцією трансформерних моделей для вилучення ознак, адаптований до багатокласової класифікації та впроваджений у розподілені корпоративні середовища із застосуванням федеративного навчання.

***Ключові слова**: запобігання витоку даних; генетичні алгоритми; еволюційна адаптація; дрейф концепції; поведінкове профілювання; класифікація документів.*

VIZHEVSKYI PETRO
KRAVCHYK YURII
Khmenlnystskyi National University

A MODEL OF A DATA LEAKAGE PREVENTION SYSTEM WITH EVOLUTIONARY ADAPTATION BASED ON GENETIC ALGORITHMS

The paper proposes a data leakage prevention (DLP) system model with evolutionary adaptation that integrates three components within a four-tier architecture: document classification based on a genetic algorithm with production IF-THEN rules, a dual-window concept drift detector combining model uncertainty monitoring and the Kolmogorov–Smirnov test with a warm-start policy archive, and user behavioral profiling with exponential forgetting. Content analysis is based on a classifier that synthesizes a set of rules in the IF-THEN form using a genetic algorithm. Each rule sets conditions on specific features of the document and evaluates the degree of belonging to a certain confidentiality class. Chromosomal encoding of classification rules ensures full interpretability, while a multi-objective fitness function balances accuracy with rule set complexity. Experimental evaluation on the SMS Spam Collection (5574 messages) and Synthetic PII (2000 documents) datasets shows that the genetic classifier achieves an F1-score of 0.985 on the personal data detection task, falling short of ensemble methods (F1 = 1.000) by only 1.5 percentage points. The evolutionary adaptation mechanism provides a 13.9% improvement in prequential accuracy compared to a static model, and the drift detector correctly identifies all abrupt drift points (Recall = 1.00). The proposed model is suitable for deployment in regulated industries where GDPR and HIPAA compliance requires justification of decisions to block information flows. The approach can be extended through integration of transformer models for feature extraction, adapted to multi-class classification, and deployed in distributed corporate environments using federated learning.

***Keywords:** data leakage prevention; genetic algorithms; evolutionary adaptation; concept drift; behavioral profiling; document classification.*

Стаття надійшла до редакції / Received 01.09.2025
Прийнята до друку / Accepted 24.09.2025

Постановка проблеми у загальному вигляді та її зв'язок із важливими науковими чи практичними завданнями

Впровадження гібридних моделей роботи та активне використання хмарних сервісів змінили вигляд корпоративних інформаційних потоків. Тепер, замість виключно внутрішньої мережі, дані циркулюють між робочими та особистими комп'ютерами, мобільними пристроями, хмарними платформами та інфраструктурою партнерів. За даними Verizon Data Breach Investigations Report 2025, близько третини всіх зафіксованих інцидентів витоку пов'язані з діяльністю постачальників, підрядників та партнерів, що мають доступ до ресурсів організації [1].

Згідно з IBM Cost of Data Breach Report 2024, середня вартість одного інциденту сягнула 4,88 мільйона доларів, причому майже 40% випадків витоків даних лишаються невиявленими протягом понад 200 днів [2]. Обсяг завданих збитків прямо пропорційний тому як багато часу пройшло від витоку до його виявлення.

Інсайдерські інциденти становлять до 35% усіх випадків компрометації даних [3]. Інсайдери можуть діяти умисно або ж бути недостатню обізнаними із політиками безпеки; в кожному з випадків необхідно використовувати окремі стратегії виявлення. Переважаючий підхід в існуючих DLP-системах зосереджений на статичному аналізі контенту та порівнянні з попередньо визначеними правилами й сигнатурами. Більшість розгорнутих DLP рішень не здатні розрізняти зловмисне та випадкове розкриття інформації, що генерує високу кількість хибнопозитивних тривог [4]. Використання DNS-тунелювання, стеганографії та шифровані канали передачі ускладнюють виявлення реальних загроз.

Додатковим ускладненням є дрейф концепції, Оскільки поведінка користувачів з часом змінюється то моделі, навчені на історичних вибірках, поступово деградують. Жодна статична модель не здатна гарантувати стабільну якість класифікації за нестаціонарних умов [5].

Таким чином, існує потреба в моделі DLP-системи, яка б поєднувала автоматичну адаптацію до змін у характері загроз, інтерпретованість рішень для верифікації експертами та інтеграцію контентного і поведінкового аналізу.

Аналіз досліджень та публікацій

В дослідженнях проблематики виявлення витоків даних акцент поступово зміщується від сигнатурних методів до підходів на основі машинного навчання та поведінкового аналізу.

Ефективний підхід для захисту data-in-use запропонований в роботі [6], Efficient DLP-visor – це тонкий гіпервізор, що перехоплює системні виклики в операційних системах Windows і гарантує, що конфіденційна інформація не може покинути визначений набір каталогів. Але це не вирішує задачу класифікації документів та адаптації до змін.

У роботі [7] запропоновано метод виявлення інсайдерських витоків на основі one-hot кодування, синтетичного генерування прикладів меншості (SMOTE) та комбінації класифікаторів машинного навчання. Цей підхід дозволяє боротися з дисбалансом класів, характерним для реальних журналів подій DLP-систем, проте результуючі моделі є непрозорими для аналітиків.

Система виявлення інсайдерських загроз BRITD, яка базується на ритмах поведінки користувача з урахуванням часової динаміки та індивідуальної адаптації продемонстрована в роботі [8]. Система моделює добові та тижневі цикли активності, що підвищує точність виявлення аномалій пов'язаних з нехарактерною поведінкою користувачів, однак не передбачає механізму адаптації до довгострокових змін у поведінці користувачів.

Проблему дрейфу концепції у потокових даних систематизовано в оглядових роботах [9, 10]. Найкращі результати демонструють performance-aware детектори, що безпосередньо відстежують деградацію якості прогнозів. При рекурентному дрейфі, коли раніше спостережені зсуви повертаються через певний час, збереження архівних моделей суттєво прискорює реадаптацію. В роботі [11] запропоновано ансамблевий алгоритм ROSE для онлайн-навчання на незбалансованих потокових даних з дрейфом, який динамічно регулює склад ансамблю та стратегію балансування, але внутрішня структура ROSE не передбачає інтерпретованості рішень.

Актуальні прикладні аспекти генетичних алгоритмів (ГА) висвітлено у оглядах [12-14]. Сучасні ГА перейшли від класичних схем Голланда до гібридних варіантів, хоча проблеми балансування між дослідженням та використанням залишається [12]. Вибір операторів селекції, кросоверу та мутації залежно від характеру задачі впливає на збіжність алгоритму [13]. Загалом еволюційні алгоритми демонструють ефективність для широкого спектру задач оптимізації, включаючи планування, маршрутизацію та налаштування параметрів складних систем [14].

В роботі [15] розроблено модель глибокого навчання для запобігання витоку інформації з цифрових документів, на основі MLP-архітектури в комбінації з n-грамним TF-IDF дескриптором для класифікації чутливого контенту. Модель забезпечує високу точність виявлення конфіденційних документів, проте не передбачає адаптації до змін у даних з часом.

Аналіз літератури засвідчує, що існуючі рішення зосереджуються або на статичній класифікації контенту, або на поведінковому аналізі, або на адаптації до дрейфу – але не пропонують усі три інструменти в єдиній системі. Окрім того, підходи на основі машинного навчання не передбачають інтерпретованості рішень, що ускладнює їх застосування в галузях, де рішення про блокування інформаційних потоків потребують обґрунтування.

Формулювання цілей статті

Метою роботи є: розробка формальної моделі системи запобігання витоку даних з еволюційною адаптацією, яка не деградуватиме за умов дрейфу концепції та дозволяє аналітику безпеки верифікувати логіку прийняття рішень згідно GDPR та HIPAA.

Виклад основного матеріалу

Модель DLP-системи з еволюційною адаптацією побудовано за модульним принципом, вона складається з чотирьох функціональних рівнів, кожен з яких відповідає за окремий аспект. Визначимо систему формально як кортеж:

$$S = \langle L_C, L_A, L_R, L_E, \Phi \rangle, \quad (1)$$

де L_C – рівень збору даних, L_A – рівень аналізу, L_R – рівень реагування, L_E – рівень еволюційної адаптації, Ф – множина функцій взаємодії між компонентами. Рівень збору даних агрегує

інформацію з мережевого трафіку, кінцевих точок, хмарних платформ та систем автентифікації. Рівень аналізу виконує класифікацію документів за ступенем конфіденційності та оцінює відхилення поведінки користувачів від профілю. Рівень реагування реалізує захисні дії – блокування, карантин, сповіщення – на підставі комбінованого скору ризику. Рівень еволюційної адаптації функціонує як фоновий процес, що безперервно оптимізує параметри аналітичних моделей та політик безпеки засобами генетичного алгоритму. Узагальнена архітектура запропонованої DLP-системи зображена на рисунку 1.

Рівень збору визначимо як множину спеціалізованих колекторів, кожен з яких відповідає за окремий канал надходження даних:

$$L_C = \{C_{net}, C_{end}, C_{cloud}, C_{auth}, C_{email}\}. \quad (2)$$

Мережевий колектор C_{net} здійснює DPI на рівні застосунків аналізуючи HTTP/HTTPS-трафік, поштові протоколи та файлові передачі. Колектор кінцевих точок C_{end} працює безпосередньо на кінцеивх ком'ютерах і серверах, фіксуючи файлові операції, активність буфера обміну, підключення зовнішніх носіїв та друк документів. Хмарний колектор C_{cloud} інтегрується з API інтерфейсами SaaS платформ для моніторингу операцій зі спільними документами, завантажень та наданням доступу. Колектори автентифікації C_{auth} та електронної пошти C_{email} відстежують події входу й вміст повідомлень.

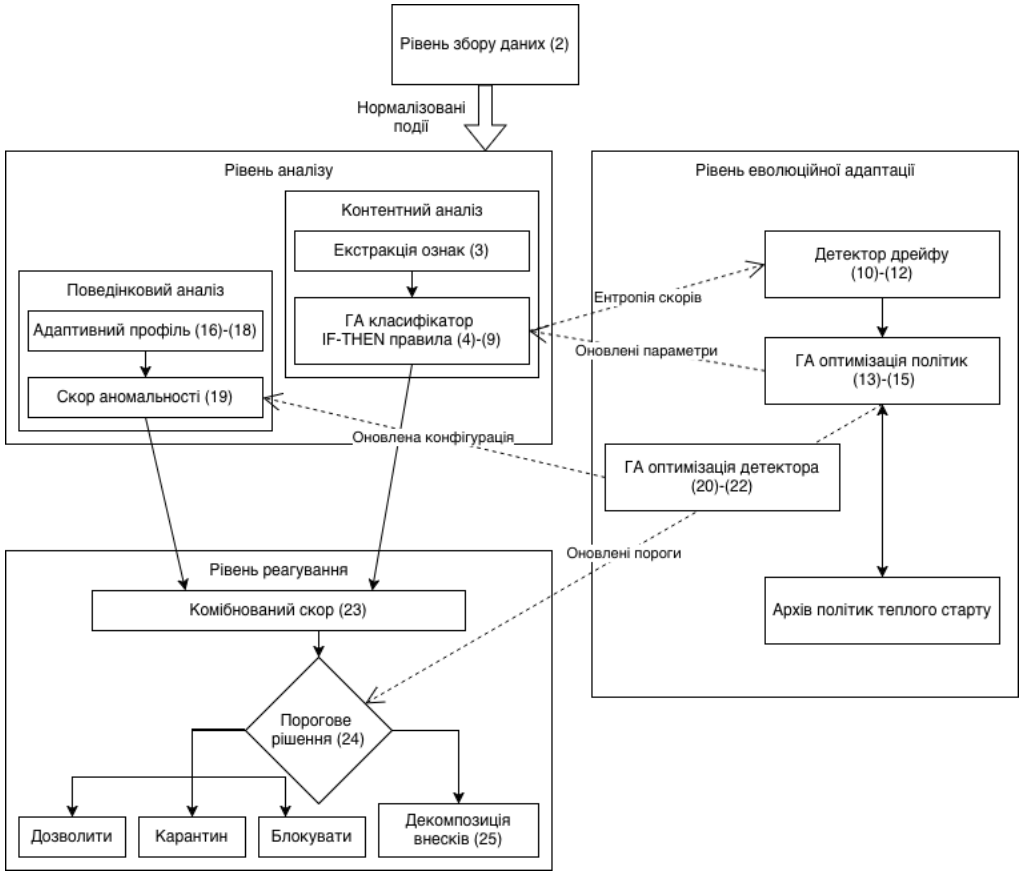

Рис. 1. Чотирирівнева архітектура DLP-системи з еволюційною адаптацією

Складений вектор ознак документа обчислюємо як:

$$F = F_{lex} \oplus F_{pattern} \oplus F_{meta}, \quad (3)$$

де F_{lex} – лексичні ознаки, побудовані на основі TF-IDF векторизації з підтримкою n-грам, $F_{pattern}$ – патернові ознаки, що включають кількість виявлених PII-сутностей різних типів (номери телефонів, електронні адреси, ідентифікаційні коди), співвідношення цифр до букв, ентропію тексту та індикатор наявності конфіденційних маркерів, F_{meta} – метадані документа. Оператор $\oplus$ позначає конкатенацію векторів.

Контентний аналіз базується на класифікаторі, що синтезує набір правил у формі IF-THEN за допомогою генетичного алгоритму. Кожне правило задає умови на конкретні ознаки документа та визначає ступінь належності до певного класу конфіденційності. Такий підхід забезпечує повну інтерпретованість та підтримку коригування правил вручну.

Елементарну умову правила визначаємо як інтервальне обмеження на значення ознаки:

$$C_f: x_f \in [a, b]. \quad (4)$$

Правило складаємо з кон'юнкції елементарних умов, до якої приєднано цільовий клас та вагу для механізму голосування:

$$r = (C_1 \wedge C_2 \wedge \ldots \wedge C_n) \rightarrow (k, w), \quad (5)$$

Хромосомою кодуємо повний набір правил разом із бінарними прапорцями активності, що дозволяє генетичному алгоритму не лише оптимізувати параметри окремих правил, а й вмикати та вимикати правила:

$$G = \{(r_1, a_1), (r_2, a_2), \ldots, (r_K, a_K)\}, \quad a_i \in \{0,1\}. \quad (6)$$

Якість набору правил оцінюємо функцією пристосованості, що враховує точність класифікації та штрафує надмірну складність:

$$F(G) = F_1(G) - \lambda \cdot X(G), \quad (7)$$

де $F_1(G)$ – F1-міра на валідаційній вибірці, $X(G)$ – нормована складність набору правил, λ – коефіцієнт регуляризації, що контролює баланс між точністю та компактністю.

Складність визначаємо як зважену суму кількості активних правил та умов у них:

$$X(G) = \frac{1}{K}\sum_{i=1}^{K} a_i \cdot (1 + \lambda_{cond} \cdot |C_i|). \quad (8)$$

де $|C_i|$ – кількість умов у правилі r_i, λ_{cond} – вага складності умов. Штраф за складність запобігає перенавчанню та сприяє генерації компактного набору правил.

Ініціалізація популяції поєднує два підходи: частина хромосом генерується випадково для забезпечення різноманітності, решта створюється евристично на основі статистик класів:

$$C_f = (f, \mu_f - \sigma_f, \mu_f + \sigma_f). \quad (9)$$

де μ_f та σ_f – середнє та стандартне відхилення значень ознаки f для документів відповідного класу. Евристичні хромосоми забезпечують «теплий старт» еволюції, оскільки початкові правила вже частково відповідають розподілу даних.

Еволюційний процес використовує турнірну селекцію з розміром турніру $k = 3$, одноточковий кросовер із ймовірністю $p_c = 0{,}8$ та мутацію порогових значень із ймовірністю $p_m = 0{,}02$. Елітизм гарантує збереження n_{elite} найкращих хромосом у наступному поколінні.

Для виявлення дрейфу розроблено двовіконний статистичний детектор, що поєднує два незалежних джерела сигналу. Перше джерело відстежує невизначеність моделі: для кожної події e_t з ризиковим скором s_t обчислюємо ентропію розподілу скорів:

$$H(e_t) = -s_t \log s_t - (1 - s_t)\log(1 - s_t) \quad (10). \quad (10)$$

Детектор підтримує два ковзних вікна – W_{ref}, що відповідає стабільному періоду, та поточне W_{cur}. Сигнал дрейфу за невизначеністю формуємо на основі порівняння середніх значень ентропії у вікнах:

$$I_{drift}^{H} = \mathbb{1}\left[\frac{\bar{H}_{cur} - \bar{H}_{ref}}{\sigma_{ref}} > \kappa\right], \quad (11)$$

де $\bar{H}_{ref}, \bar{H}_{cur}$ – середні значення ентропії у відповідних вікнах, σ_{ref} – стандартне відхилення у референтному вікні, κ – поріг чутливості. Друге джерело сигналу моніторить зміни у розподілі ознак через двовибірковий тест Колмогорова-Смирнова для кожної ознаки f_j у парі вікон. Фінальний бінарний сигнал дрейфу формуємо як диз'юнкцію обох джерел:

$$I_{drift} = I_{drift}^{H} \vee I_{drift}^{KS}, \quad (12)$$

що забезпечує виявлення як деградації впевненості моделі, так і зміни розподілу ознак. Для запобігання хибним спрацюванням через випадкові флуктуації застосовується механізм підтвердження: адаптація активується лише після k послідовних детекцій дрейфу (типово $k = 3$), що є необхідною ціною за стабільність системи. Після виявлення дрейфу система ініціює еволюційну оптимізацію політик. Політику DLP-системи кодуємо як хромосому з параметрами реагування та навчання:

$$P = (\tau_q, \tau_b, W, \rho, \eta, \mu_0, \mu_1), \quad (13)$$

де τ_q – поріг карантину, τ_b – поріг блокування, $W = (w_1, \ldots, w_5)$ – ваги складових цільової функції, ρ – параметр затухання, η – швидкість навчання, μ_0 та μ_1 – базова та додаткова інтенсивність мутації відповідно.

Якість політики оцінимо функцією пристосованості:

$$F(P) = \sum_{j=1}^{5} w_j \cdot J_j(P), \quad (14)$$

де J_1 – частка хибнонегативних спрацювань, J_2 – частка хибнопозитивних спрацювань, J_3 – затримка обробки подій, J_4 – складність політик, J_5 – стабільність потоку сповіщень. Цей набір критеріїв відображає компроміс: жорсткі політики знижують ризик витоків, але генерують надмірну кількість хибних тривог, які збільшуватимуть навантаження на аналітиків.

При виявленні дрейфу інтенсивність мутації автоматично збільшується, що розширює область

пошуку в просторі параметрів:

$$\mu = \mu_0 + \mu_1 \cdot I_{drift}. \quad (15)$$

Підвищена мутація зберігається протягом перехідного періоду, достатнього для адаптації популяції до нових умов.

Окремим механізмом є архів політик із теплим стартом. Система зберігає політики, що продемонстрували високу ефективність за різних умов, разом із описом цих умов (розподіл типів подій, рівень активності, характеристики трафіку). При виявленні дрейфу архів перевіряється на наявність політики, ефективної за аналогічних обставин. Якщо така знайдена, вона включається до початкової популяції, що значно прискорює адаптацію до повторюваних патернів [10].

Поведінковий детектор аномалій будує індивідуальний профіль для кожного користувача на основі наступних ознак: часові патерни активності (розподіл подій протягом доби та тижня), обсяги і типи оброблюваних даних, мережева активність та характеристики контенту.

Оскільки поведінка користувачів природно змінюється у зв'язку з новими проєктами, обов'язками, інструменти, – профіль має адаптуватися, надаючи більшу вагу нещодавнім спостереженням. Для цього застосовуємо механізм експоненційного забування. Адаптивне середнє значення i-ї ознаки для користувача u оновлюємо за рекурентною формулою:

$$\mu_i^u(k) = \rho \cdot \mu_i^u(k-1) + (1-\rho) \cdot x_i^u(k), \quad (16)$$

де $\rho \in (0,1)$ – коефіцієнт забування, що визначає швидкість відкидання минулих спостережень, $x_i^u(k)$ – поточне значення ознаки.

Адаптивну дисперсію оновлюємо аналогічно:

$$(\sigma_i^u(k))^2 = \rho \cdot (\sigma_i^u(k-1))^2 + (1-\rho) \cdot (x_i^u(k) - \mu_i^u(k))^2. \quad (17)$$

Для нормалізованої оцінки відхилення використовуємо z-score:

$$z_i^u(k) = \frac{x_i^u(k) - \mu_i^u(k)}{\sigma_i^u(k) + \varepsilon}. \quad (18)$$

де $\varepsilon > 0$ – мала константа для забезпечення чисельної стійкості

Загальну оцінку аномальності визначимо як зважену суму перетворених z-score за всіма активними ознаками:

$$S^u(k) = \sum_{i=1}^m b_i \cdot w_i \cdot \varphi(z_i^u(k)), \quad (19)$$

де $b_i \in \{0,1\}$ – індикатор включення i-ї ознаки, $w_i \geq 0$ – вага ознаки, $\varphi(z) = \max(0, z)$ – функція активації, що враховує лише позитивні відхилення (поведінка «більша за норму»).

Конфігурацію детектора (маска активних ознак, ваги та поріг) кодуємо як окрему хромосому, що оптимізується генетичним алгоритмом:

$$G_d = (B, W, \tau, \rho), \quad (20)$$

де $B = (b_1, \dots, b_m)$ – бінарна маска, $W = (w_1, \dots, w_m)$ – вектор ваг, τ – поріг тривоги, ρ – коефіцієнт забування.

Функція пристосованості детектора враховує якість виявлення, хибнопозитивну частку та стабільність потоку тривог:

$$F(G_d) = F_1(G_d) - \alpha \cdot FPR(G_d) - \beta \cdot CV_{alert}(G_d) - \gamma \cdot C(G_d) \quad (21), \quad (21)$$

де FPR – частка хибнопозитивних спрацювань, CV_{alert} – коефіцієнт варіації потоку тривог, α, β, γ – коефіцієнти штрафів. Штраф за волатильність потоку тривог CV_{alert} стимулює стабільне навантаження на аналітиків, запобігаючи ситуаціям, коли тривоги надходять нерівномірними сплесками.

Волатильність визначається як:

$$V_{alert} = \frac{\sigma_A}{\bar{A} + \varepsilon}, \quad (22)$$

де σ_A – стандартне відхилення кількості тривог за часове вікно, $\bar{A}$ – середня кількість тривог.

Результати контентного та поведінкового аналізу об'єднуються у комбіновану оцінку ризику:

$$R_{combined} = \lambda \cdot S_{content} + (1-\lambda) \cdot S_{behavior}, \quad (23)$$

де $\lambda \in (0,1)$ – коефіцієнт балансу. Значення λ оптимізується еволюційним контуром спільно з іншими параметрами політики, що забезпечує автоматичне калібрування відносного внеску обох компонентів.

Рівень реагування обирає конкретну дію відповідно до порогових значеннь, які є частиною хромосоми політики (13):

$$A(R) = \begin{cases} A_{allow}, & R < \tau_q \\ A_{quarantine}, & \tau_q \leq R < \tau_b \\ A_{block}, & R \geq \tau_b \end{cases}. \quad (24)$$

Для кожної згенерованої тривоги система обчислює відносний внесок окремих ознак:

$$CN_i^u(k) = \frac{b_i \cdot w_i \cdot \varphi(z_i^u(k))}{S^u(k)+\varepsilon}. \quad (25)$$

Ця декомпозиція надає пояснення причин спрацювання детектора: які саме аспекти поведінки користувача відхилилися від норми та наскільки значним є кожне відхилення. Для DLP-систем, де рішення про блокування інформаційного потоку може зупинити критичний бізнес-процес, пояснюваність є операційною необхідністю.

Для експериментальної перевірки класифікаційної складової моделі використано два набори даних із різними характеристиками.

SMS Spam Collection містить 5574 текстові повідомлення з бінарною розміткою (спам / не спам). Цей набір характеризується суттєвим дисбалансом класів (приблизно 87% нормальних повідомлень проти 13% спаму) та коротким текстовим контентом, що ускладнює побудову інформативних лексичних ознак. Хоча SMS-повідомлення відрізняються від корпоративних документів за форматом, задача бінарної класифікації текстів за наявності дисбалансу класів є типовою для DLP-систем.

Synthetic PII складається з 2000 синтетичних документів, що імітують корпоративну кореспонденцію з персональними даними (номери телефонів, електронні адреси, ідентифікаційні коди) та без них. Цей набір моделює типову задачу DLP: виявлення документів, що містять чутливу персональну інформацію.

Для обох наборів сформовано вектори ознак відповідно до формули (3): лексичні ознаки на основі TF-IDF (до 500 найінформативніших термів), патернові ознаки (кількість PII-сутностей за типами, ентропія, співвідношення цифр до букв) та метадані. Оцінка якості виконувалася за F1-мірою на тестовій вибірці (30% даних) з використанням стратифікованого розбиття.

Таблиця 1

Порівняння якості класифікації

Метод	F1 (SMS Spam)	F1 (Synthetic PII)
GA-класифікатор (запропонований)	0,785	0,985
Random Forest	0,926	1,000
Gradient Boosting	0,926	1,000
SVM	0,857	0,970
Naive Bayes	0,809	0,941

У таблиці 1 наведено порівняння запропонованого GA-класифікатора з чотирма методами-еталонами. На наборі Synthetic PII генетичний класифікатор демонструє F1-міру 0,985, що лише на 1,5 відсоткових пункти поступається ансамблевим методам (Random Forest та Gradient Boosting), які досягають ідеальної класифікації. На більш складному наборі SMS Spam, де дисбаланс класів та короткі тексти створюють додаткові труднощі, розрив збільшується: GA-класифікатор отримує 0,785 порівняно з 0,926 у ансамблевих методів. Naive Bayes (0,809) та SVM (0,857) займають проміжну позицію.

Зниження якості GA-класифікатора на незбалансованих даних пояснюється обмеженнями інтервальних IF-THEN правил: вони оперують прямокутними областями у просторі ознак, тоді як ансамблеві методи апроксимують довільні нелінійні межі. Проте для задачі виявлення PII, де ключові ознаки мають виражений пороговий характер (наявність або відсутність патерну персональних даних), IF-THEN правила виявляються практично такими ж ефективними, як і складніші моделі.

Для оцінки механізму адаптації змодельовано потік даних з трьома точками раптового дрейфу, де розподіл ознак змінюється стрибкоподібно. Порівняно дві конфігурації: статичну модель (навчену на початковому вікні без подальшого оновлення) та адаптивну модель із запропонованим двовіконним детектором дрейфу та еволюційною оптимізацією.

Таблиця 2

Результати адаптації до дрейфу концепції

Конфігурація	Prequential accuracy	Стандартне відхилення
Без адаптації	0,604	0,169
З адаптацією (запропонований детектор + GA)	0,688	0,151

Адаптивна модель забезпечує приріст prequential accuracy на 13,9% (з 0,604 до 0,688) та одночасно знижує стандартне відхилення з 0,169 до 0,151, що свідчить про стабільнішу якість класифікації в умовах змін. Запропонований детектор коректно виявив усі три точки дрейфу (Recall = 1,00) при двох хибних спрацюваннях (Precision = 0,60, F1 = 0,75). Механізм підтвердження з $k = 3$ послідовними детекціями суттєво зменшив кількість хибних спрацювань порівняно з одноразовим порогом. Зайві цикли перенавчання не погіршували якість класифікації, оскільки еволюційний алгоритм за відсутності реального дрейфу зберігав наявну конфігурацію.

Окрім якості класифікації, для практичного розгортання DLP-системи критичними є обчислювальні витрати та здатність до інтерпретації рішень. У таблиці 3 наведено порівняння за цими критеріями на наборі Synthetic PII.

Таблиця 3

Обчислювальна складність

Метод	Час навчання (с)	F1 (Synthetic PII)
GA-класифікатор	65,7	0,985
Random Forest	0,06	1,000
Gradient Boosting	0,25	1,000
SVM	0,03	0,970
Naive Bayes	0,01	0,941

Час навчання генетичного класифікатора (65,7 секунди) суттєво перевищує час навчання конкурентів, що зумовлено ітеративною природою еволюційного пошуку. Це обмеження є допустимим із двох причин. По-перше, навчання виконується у фоновому режимі й не впливає на латентність обробки подій у реальному часі. По-друге, після навчання застосування IF-THEN правил для класифікації нового документа є миттєвим.

Головна перевага GA-класифікатора – повна інтерпретованість. Результатом навчання є набір правил, що людина може прочитати: наприклад, «ЯКЩО кількість PII-сутностей > 2 ТА ентропія тексту < 4,5, ТО документ конфіденційний». У регульованих галузях, де відповідність стандартам GDPR та HIPAA вимагає пояснення кожного рішення про блокування інформаційного потоку, ця властивість має стратегічне значення [4].

Висновки з даного дослідження
і перспективи подальших розвідок у даному напрямі

Розроблено формальну модель системи запобігання витоку даних з еволюційною адаптацією, що інтегрує три складові в єдиній архітектурі: класифікацію документів на основі генетичного алгоритму з IF-THEN правилами, еволюційну адаптацію політик безпеки із детекцією дрейфу концепції та поведінкове профілювання користувачів. Комбінована оцінка ризику та механізм пояснення забезпечують прозору і узгоджену роботу контентного і поведінкового аналізу.

На задачі виявлення PII експериментальна оцінка генетичного класифікатору досягає F1-міри 0,985, поступаючись ансамблевим методам лише на 1,5 відсоткових пункти при інтерпретованості правил. Механізм адаптації до дрейфу забезпечує приріст prequential accuracy на 13,9% порівняно зі статичною моделлю. Для організацій, що працюють відповідно до вимог GDPR та HIPAA, пояснюваність і адаптивність є критичними перевагами, що компенсують незначний програш у точності.

В рамках подальших досліджень перспективним є інтеграція нейромереж для виокремлення ознак із текстових документів, що дозволить GA-класифікатору працювати з компактними представленнями замість високорозмірних TF-IDF векторів. Також перспективним є дослідження застосованості моделі у випадку розширення бінарної класифікації до багатокласової з кількома рівнями конфіденційності (публічні, внутрішні, конфіденційні, секретні). Важливо також виконати тестування моделі на реальних корпоративних наборах даних, щоб оцінити ефективність в умовах реальних масштабів та різноманітності.

Література

1. Verizon. 2025 Data Breach Investigations Report [Електронний ресурс] / Verizon Business. – 2025. – Режим доступу: https://www.verizon.com/business/resources/reports/dbir/

2. IBM Security. Cost of a Data Breach Report 2024 [Електронний ресурс] / IBM Corporation. – 2024. – Режим доступу: https://www.ibm.com/reports/data-breach

3. Herrera Montano I. Survey of Techniques on Data Leakage Protection and Methods to address the Insider threat / I. Herrera Montano, J.J. García Aranda, J. Ramos Diaz, S. Molina Cardín, I. de la Torre Díez, J.J.P.C. Rodrigues // Cluster Computing. – 2022. – Vol. 25. – P. 4289–4302. – DOI: https://doi.org/10.1007/s10586-022-03668-2.

4. Domnik J. On Data Leakage Prevention Maturity: Adapting the C2M2 Framework / J. Domnik, A. Holland // Journal of Cybersecurity and Privacy. – 2024. – Vol. 4, No. 2. – P. 167–195. – DOI: https://doi.org/10.3390/jcp4020009.

5. Arora S. A systematic review on detection and adaptation of concept drift in streaming data using machine learning techniques / S. Arora, R. Rani, N. Saxena // WIREs Data Mining and Knowledge Discovery. – 2024. – Vol. 14, No. 4. – Article e1536. – DOI: https://doi.org/10.1002/widm.1536.

6. Kiperberg M. Efficient DLP-visor: An efficient hypervisor-based DLP / M. Kiperberg, G. Amit, A. Yeshooroon, N.J. Zaidenberg // 2021 IEEE/ACM 21st International Symposium on Cluster, Cloud and Internet Computing (CCGrid). – 2021. – P. 344–355. – DOI: https://doi.org/10.1109/CCGrid51090.2021.00044.

7. Al-Shehari T. An insider data leakage detection using one-hot encoding, synthetic minority oversampling and machine learning techniques / T. Al-Shehari, R.A. Alsowail // Entropy. – 2021. – Vol. 23,

No. 10. – Article 1258. – DOI: https://doi.org/10.3390/e23101258.

8. Song S. BRITD: behavior rhythm insider threat detection with time awareness and user adaptation / S. Song, N. Gao, Y. Zhang, C. Ma // Cybersecurity. – 2024. – Vol. 7, Article 2. – DOI: https://doi.org/10.1186/s42400-023-00190-9.

9. Bayram F. From concept drift to model degradation: An overview on performance-aware drift detectors / F. Bayram, B.S. Ahmed, A. Kassler // Knowledge-Based Systems. – 2022. – Vol. 245. – Article 108632. – DOI: https://doi.org/10.1016/j.knosys.2022.108632.

10. Suárez-Cetrulo A.L. A survey on machine learning for recurring concept drifting data streams / A.L. Suárez-Cetrulo, D. Quintana, A. Cervantes // Expert Systems with Applications. – 2023. – Vol. 213. – Article 118934. – DOI: https://doi.org/10.1016/j.eswa.2022.118934.

11. Cano A. ROSE: Robust online self-adjusting ensemble for continual learning on imbalanced drifting data streams / A. Cano, B. Krawczyk // Machine Learning. – 2022. – Vol. 111. – P. 2561–2599. – DOI: https://doi.org/10.1007/s10994-022-06168-x.

12. Katoch S. A review on genetic algorithm: Past, present, and future / S. Katoch, S.S. Chauhan, V. Kumar // Multimedia Tools and Applications. – 2021. – Vol. 80. – P. 8091–8126. – DOI: https://doi.org/10.1007/s11042-020-10139-6.

13. Alhijawi B. Genetic algorithms: Theory, genetic operators, solutions, and applications / B. Alhijawi, A. Awajan // Evolutionary Intelligence. – 2024. – Vol. 17. – P. 1245–1256. – DOI: https://doi.org/10.1007/s12065-023-00822-6.

14. Slowik A. Evolutionary algorithms and their applications to engineering problems / A. Slowik, H. Kwasnicka // Neural Computing and Applications. – 2020. – Vol. 32. – P. 12363–12379. – DOI: https://doi.org/10.1007/s00521-020-04832-8.

15. Guha A. A Deep Learning Model for Information Loss Prevention From Multi-Page Digital Documents / A. Guha, D. Samanta, A. Banerjee, D. Agarwal // IEEE Access. – 2021. – Vol. 9. – P. 80451–80465. – DOI: https://doi.org/10.1109/access.2021.3084841.

References

1. Verizon. 2025 Data Breach Investigations Report [Electronic resource] / Verizon Business. – 2025. – Access mode : https://www.verizon.com/business/resources/reports/dbir/

2. IBM Security. Cost of a Data Breach Report 2024 [Electronic resource] / IBM Corporation. – 2024. – Access mode : https://www.ibm.com/reports/data-breach

3. Herrera Montano I. Survey of Techniques on Data Leakage Protection and Methods to address the Insider threat / I. Herrera Montano, J.J. García Aranda, J. Ramos Diaz, S. Molina Cardín, I. de la Torre Díez, J.J.P.C. Rodrigues // Cluster Computing. – 2022. – Vol. 25. – P. 4289–4302. – DOI: https://doi.org/10.1007/s10586-022-03668-2.

4. Domnik J. On Data Leakage Prevention Maturity: Adapting the C2M2 Framework / J. Domnik, A. Holland // Journal of Cybersecurity and Privacy. – 2024. – Vol. 4, No. 2. – P. 167–195. – DOI: https://doi.org/10.3390/jcp4020009.

5. Arora S. A systematic review on detection and adaptation of concept drift in streaming data using machine learning techniques / S. Arora, R. Rani, N. Saxena // WIREs Data Mining and Knowledge Discovery. – 2024. – Vol. 14, No. 4. – Article e1536. – DOI: https://doi.org/10.1002/widm.1536.

6. Kiperberg M. Efficient DLP-visor: An efficient hypervisor-based DLP / M. Kiperberg, G. Amit, A. Yeshooroon, N.J. Zaidenberg // 2021 IEEE/ACM 21st International Symposium on Cluster, Cloud and Internet Computing (CCGrid). – 2021. – P. 344–355. – DOI: https://doi.org/10.1109/CCGrid51090.2021.00044.

7. Al-Shehari T. An insider data leakage detection using one-hot encoding, synthetic minority oversampling and machine learning techniques / T. Al-Shehari, R.A. Alsowail // Entropy. – 2021. – Vol. 23, No. 10. – Article 1258. – DOI: https://doi.org/10.3390/e23101258.

8. Song S. BRITD: behavior rhythm insider threat detection with time awareness and user adaptation / S. Song, N. Gao, Y. Zhang, C. Ma // Cybersecurity. – 2024. – Vol. 7, Article 2. – DOI: https://doi.org/10.1186/s42400-023-00190-9.

9. Bayram F. From concept drift to model degradation: An overview on performance-aware drift detectors / F. Bayram, B.S. Ahmed, A. Kassler // Knowledge-Based Systems. – 2022. – Vol. 245. – Article 108632. – DOI: https://doi.org/10.1016/j.knosys.2022.108632.

10. Suárez-Cetrulo A.L. A survey on machine learning for recurring concept drifting data streams / A.L. Suárez-Cetrulo, D. Quintana, A. Cervantes // Expert Systems with Applications. – 2023. – Vol. 213. – Article 118934. – DOI: https://doi.org/10.1016/j.eswa.2022.118934.

11. Cano A. ROSE: Robust online self-adjusting ensemble for continual learning on imbalanced drifting data streams / A. Cano, B. Krawczyk // Machine Learning. – 2022. – Vol. 111. – P. 2561–2599. – DOI: https://doi.org/10.1007/s10994-022-06168-x.

12. Katoch S. A review on genetic algorithm: Past, present, and future / S. Katoch, S.S. Chauhan, V. Kumar // Multimedia Tools and Applications. – 2021. – Vol. 80. – P. 8091–8126. – DOI: https://doi.org/10.1007/s11042-020-10139-6.

13. Alhijawi B. Genetic algorithms: Theory, genetic operators, solutions, and applications / B. Alhijawi, A. Awajan // Evolutionary Intelligence. – 2024. – Vol. 17. – P. 1245–1256. – DOI: https://doi.org/10.1007/s12065-023-00822-6.

14. Slowik A. Evolutionary algorithms and their applications to engineering problems / A. Slowik, H. Kwasnicka // Neural Computing and Applications. – 2020. – Vol. 32. – P. 12363–12379. – DOI: https://doi.org/10.1007/s00521-020-04832-8.

15. Guha A. A Deep Learning Model for Information Loss Prevention From Multi-Page Digital Documents / A. Guha, D. Samanta, A. Banerjee, D. Agarwal // IEEE Access. – 2021. – Vol. 9. – P. 80451–80465. – DOI: https://doi.org/10.1109/access.2021.3084841.