DOI 10.15276/imms.v15.no1.95
УДК 004.056.53

Informatics and Mathematical Methods in Simulation
Vol.15 (2025), No. 1, pp. 95-106

РОЗРОБКА ЗАСТОСУНКУ ДЛЯ АВТОМАТИЗОВАНОГО ВИЯВЛЕННЯ КІБЕРСТАЛКІНГУ

О.В. Кузан, Н.І. Кушніренко, В.О. Назаров, В.В. Подуфалов

Національний університет «Одеська політехніка»
1, Шевченка пр., м.Одеса, 65044, Україна
Email: kushnirenko@op.edu.ua

На сьогодні, кіберсталкінг показав себе як надзвичайно швидкозростаюче явище, яке спричинило зміну парадигм переслідування. Воно полягає у нав'язливому, систематичному втручанні в онлайн-життя жертви з метою психологічного тиску, маніпуляцій, залякування, тощо. Незважаючи на масштаб проблеми, ефективних засобів автоматизованого виявлення таких загроз українською мовою на сьогодні бракує. Метою даної роботи є підвищення безпеки користувачів у цифровому середовищі шляхом розробки застосунку для виявлення кіберсталкерської поведінки в електронних листах, з використанням машинного навчання. В рамках дослідження проведено аналіз явища кіберсталкінгу, а також сучасних підходів до його виявлення, який дозволив визначити напрямки розробки та основні завдання роботи. Розроблений застосунок базується на використанні алгоритму SVM для побудови моделей, навчання яких здійснюється на спеціально сформованому датасеті. Створено дві моделі: бінарну – для класифікації листів на «кіберсталкінг» і «не кіберсталкінг» та багатокласову – для поділу за типами: залякування, сексуальний харасмент і звичайний кіберсталкінг. Цей інструмент також надає можливість зберігати виявлені повідомлення кіберсталкера як докази, задля подальшого звернення жертв до кіберполіції. Тестування підходу показало його високу ефективність: загальна точність бінарної моделі склала 99%, багатокласова досягла 89%. Отримані результати засвідчують значний потенціал застосунку у виявлення кіберпереслідувань і таким чином підвищення рівня цифрової безпеки користувачів в онлайн-комунікаціях, тож запропонований застосунок для виявлення кіберсталкінгу може бути успішно впроваджений для особистого використання. Результати даної роботи можуть бути використані при подальших дослідженнях та розробках у сфері кібербезпеки та боротьби з кіберсталкінгом.

Ключові слова: кіберсталкінг, електронна пошта, машинне навчання, класифікація тексту, кібербезпека.

Вступ. В наш час цифрові технології стали невід'ємною складовою повсякденного життя суспільства. Їх стрімкий розвиток відкрив нові горизонти для комунікації та обміну інформацією в інтернеті. Однак, водночас із позитивними змінами, цифрова епоха принесла людству низку нових викликів, зокрема стрімке зростання кіберзлочинності, тобто злочинності саме в кіберпросторі. Традиційні злочини, пройшовши через технічну адаптацію, створили своїх кібернащадків, які незважаючи на деякі схожі риси, містять в собі важливі відмінності: мають власні механізми реалізації та інакший вплив, тому їх не можна розглядати виключно як продовження традиційних злочинів. Одним з таких кіберзлочинів, який наразі набирає обертів у своєму поширенні є кіберсталкінг. Сам термін походить від розуміння традиційного сталкінгу і відноситься до переслідування, яке відбувається в онлайн-середовищі. На сьогодні, не існує єдиного визначення чи чітких критеріїв кіберсталкерської поведінки, проте якщо проаналізувати які види поведінки найчастіше включаються в концептуалізацію та які конкретні критерії необхідні для того, щоб класифікувати модель поведінки як кіберсталкінг, то можна зробити певне узагальнення, що це форма нав'язливої поведінки, яка відбувається в кіберпросторі за допомогою цифрових технологій. Вона характеризується такими елементами як: небажаність поведінки, повторюваність, систематизованість, спосіб

здійснення – використання цифрових технологій; руйнівний вплив [1]. До найпоширеніших компонентів кіберсталкінгу вносять: пошук і збір інформації про жертву з метою переслідування, погроз і залякування онлайн; неодноразову небажану розсилку електронних листів та інших типів повідомлень; крадіжка особистих даних; підписка жертви на послуги; купівля товарів і послуг від імені жертви; видавання себе за жертву в Інтернеті; надсилання або розміщення приватних матеріалів, дезінформації; а також втягнення інших користувачів Інтернету в переслідування або погрози на адресу жертви [2, 3]. Тобто можна сказати, що кіберсталкерами задіяні майже всі куточки кіберпростору; і також, доречно зазначити, що ці елементи завжди трансформуватимуться в міру розвитку технологій.

Кіберсталкер – це особа, яка використовує технології та кіберпростір як засоби для залякування, погроз, переслідування та нагнітання страху за допомогою тактики переслідування. Кіберсталкери не утворюють єдиної, чітко визначеної групи, оскільки їхня поведінка та мотиви значною мірою залежать від особистісних характеристик конкретної особи. Кіберсталкер може бути знайомий жертві, або навпаки, дуже часто абсолютно незнайомі люди діють онлайн на умовах анонімності [2]. До сьогодні, найпопулярнішою типологією вважається розробка Л. Макфарлейн та П. Бочого, які провели ґрунтовне дослідження феномену, в результаті виділивши чотири основні типи кіберсталкерів – стриманий, інтимний, колективний, мстивий [4].

Особливу увагу приділяють саме інтимним кіберсталкерам. Хоча будь-хто може бути інтимним кіберсталкером, який, наприклад, надсилає жертві нав'язливі повідомлення з інтимним підтекстом, немає нікого, хто б краще підходив для цього злочину окрім попередніх партнерів. Оскільки попередні партнери часто мають у своєму розпорядженні особисту інформацію своїх колишніх, переслідування стає легким завданням. З цим питанням тісно переплітається інше: відповідно до досліджень, більшість кіберсталкерів є чоловіки, а жертви зазвичай – жінки. Важливим є також те, що кіберсталкінг відноситься до основних видів насильницької поведінки якій сприяють технології, тобто кібернасильства. Особливої поширеності він набув при скоєнні підвиду кібернасильства – кіберсексуального насильства [5, 6]. Жінки-переслідувані зазвичай гостріше відчувають негативні наслідки кіберпереслідувань, але їхні повідомлення про це часто ігноруються або сприймаються менш серйозно, ніж аналогічні випадки, про які повідомляють жертви-чоловіки, що сприяє продовженню кіберсталкерської поведінки, бо зловмисники відчувають безкарність у своїх діях [7].

Говорячи про наслідки, то їх усвідомлення відіграє головну роль у розробці дієвих заходів для протидії цьому явищу, підвищенні рівня обізнаності користувачів. І незважаючи на розбіжності в способах вимірювання злочину, були зроблені чіткі висновки щодо його негативного впливу. Прийнято вважати, що кіберсталкінг може змінюватися від небезпечних електронних повідомлень до потенційно смертельної зустрічі між кривдником і жертвою. І чим довше триває кіберпереслідування та чим інтенсивнішим воно стає, тим серйознішими є наслідки для жертви. Поведінка кіберзлочинця завдає шкоду фінансовій, кар'єрній, емоційній, психологічній, соціальній та фізичній сферам життя переслідуваного [1, 2]. Ці наслідки можуть не поширюватися на всіх жертв кіберсталкінга, проте коли вони виникають, то вносять хаос у життя жертви. Незважаючи на це, кіберпереслідування в суспільстві часто розглядається не як карне правопорушення, а просто аморальний вчинок. Це становить вагому проблему, тому що доведено, що кіберсталкінг – це перш за все кіберзлочин [7]. За даними досліджень, найпоширенішим серед переслідуваних є неформальне звернення за допомогою, а останнім йде звернення до правоохоронних органів. Багато хто з потерпілих вирішує не повідомляти про випадки кіберсталкінгу, побоюючись несерйозного ставлення. Навіть якщо справа заведена, потерпілі не рідко стикаються з проблемами протягом розслідування [8].

Є питання і поза відношення правової системи до кіберсталкінга як злочину. Притягнути злочинців до відповідальності досить складно з таких причин як: процедурні виклики, міжнародне та анонімне переслідування. Початковий етап розслідування кіберпереслідувань визначається жертвами як найбільш розчаровуючий. Кожен випадок кіберсталкінга має бути зафіксований, щоб встановити модель поведінки, яка відповідає пороговому значенню – необхідно довести щонайменше два інциденти. Тобто потерпілий повинен вести записи про свою власну взаємодію, що як доведено може травмувати жертв [9]. На даний момент в Україні кіберсталкінг не розглядається як окреме правопорушення. На практиці для захисту осіб, які зазнають переслідування, застосовуються інші норми законодавства. З останніх новин, 2 жовтня 2024 року на офіційному порталі Верховної Ради зареєстрували новий законопроєкт №12088, де пропонується внесення змін до чинного Закону «про запобігання та протидію домашньому насильству», які передбачають введення визначення терміну кіберпереслідування: «кіберпереслідування – вид злочинного переслідування, що передбачає постійне спостереження за постраждалою особою без її згоди чи законного дозволу за допомогою інформаційних та комунікаційних технологій, що може здійснюватися шляхом обробки персональних даних жертви, зокрема в результаті крадіжки особистих даних або шпигування за жертвою через її соціальні мережі або платформи обміну повідомленнями, електронну пошту та телефон, крадіжку паролів або злом її пристроїв для доступу до приватного простору за допомогою встановлення додатків геолокації, у тому числі шпигунського програмного забезпечення, або через викрадення її пристроїв» [10]. Проте у створеному законопроєкті наразі не запропоновано, до якого саме покарання притягуватимуть винних в кіберсталкінгу. До того ж, відсутні чітко визначені механізми для виявлення, фіксування, підтвердження кіберсталкінгу, збору доказів і захисту жертв.

Аналітичний огляд літератури дав зрозуміти, що, зважаючи на помірну ефективність тактик управління кіберсталкінгом, мечем для виявлення, захисту та запобігання цьому злочину мають стати саме технології [11]. Незважаючи на популярність так званих інтервенційних підходів, виявлення є життєво важливим у протистоянні проти кіберсталкерів. Воно необхідне для подальших дій, спрямованих на ізоляцію правопорушення, попередження жертви, блокування небажаної комунікації, а також для підтримки цифрового криміналістичного розслідування. Можна виділити такі методи, які наразі використовуються для виявлення кіберпереслідувань, як: статистичні; інтелектуальний аналіз даних; машинне навчання. Останній підхід відіграє головну роль у виявленні кіберсталкінгу. Машинне навчання активно використовується як самостійна методологія або як частина гібридного підходу для ідентифікації злочинної поведінки. Основна перевага цих алгоритмів – здатність швидко та точно аналізувати величезні масиви інформації, адаптуючись до нових викликів та змін. Методи машинного навчання поділяються на контрольовані, неконтрольовані, напівконтрольовані та навчання з підкріпленням. Кожен із них має свої особливості та застосовується залежно від специфіки завдання. Серед подібних робіт, можна виділити використання алгоритмів випадкового лісу та логістичної регресії для автоматичного виявлення повідомлень кіберзалякування, а також гібридні підходи, як наприклад використання методу кластеризації на основі правил і нечіткої логіки для виявлення мови ворожнечі, які досягли 94,5 % f1-міри [12]. В даній роботі буде використовуватися саме контрольоване навчання, яке передбачає використання маркованих даних, де кожен вхідний приклад має відповідний вихідний результат. Алгоритм навчається на основі залежностей між цими змінними, що дозволяє йому прогнозувати майбутні результати. Такий підхід широко застосовується для завдань регресії та потрібної нам класифікації. Результати досліджень показали, що числове представлення TF-IDF досягає найкращих показників при роботі з алгоритмами класифікації [13]. TF визначає, як часто слово зустрічається в тексті і IDF відповідає за важливість терміна. Фінальна вага терміна у тексті визначається

як добуток TF та IDF, тобто рідкісні слова, які часто зустрічаються в окремих текстах, отримують вищу вагу, ніж ті, що є загальновживаними. Алгоритм машинного навчання SVM із методом TF-IDF показує найкращі результати навчання, особливо в завданні класифікації саме текстових даних [14]. Метод опорних векторів працює за принципом пошуку оптимальної гіперплощини, яка розділяє дані на категорії, його мета – знайти межу розділу між ними так, щоб максимізувати відстань між найближчими точками категорій. Він добре працює з невеликими датасетами і нерівномірними розподілами, що є безумовною перевагою щодо інших алгоритмів машинного навчання [15].

Тож, враховуючи подану вище інформацію, а також визнаючи значний рівень популяризації української мови в міжособистісному спілкуванні серед українців будь-якого віку на даний час, і додавши до цього відсутність розробок технічного формату у виявленні кіберсталкінгу в українському кіберпросторі, актуальним буде рішення про створення застосунку, який би виявляв кібersталкерську поведінку на українській мові. Вибір технології для обміну цифровими повідомленнями пав на електронну пошту, яка поки залишається одним із найпопулярніших засобів комунікації, як для офіційного листування, так і для особистого спілкування. До тепер, вона тримає одну з лідуючих позицій для здійснення кіберсталкінга, що робить її важливим джерелом для аналізу загроз цього типу. До того ж, електронні листи можуть використовуватися як юридичні докази в справах про кібersталкінг, тому їх важливо виявляти та зберігати.

Мета і задачі дослідження. Метою цієї роботи є підвищення безпеки користувачів в онлайн-середовищі, шляхом розробки застосунку для автоматизованого виявлення ознак кібersталкерської поведінки з використанням машинного навчання.

Для досягнення поставленої мети були визначені наступні задачі:

- відбір навчальних даних для датасета;
- оцінка ефективності моделей на тренувальних даних;
- розробка застосунку для автоматизованого виявлення кіберсталкінгу;
- тестування застосунку.

Основна частина. В даній роботі використовується набір сучасних технологій, які надають можливість реалізувати ідею автоматизованого виявлення кіберсталкінгу на основі машинного навчання. Основою розробки стала мова програмування Python, яка відзначається великою кількістю спеціалізованих бібліотек для машинного навчання, таких як scikit-learn. Основні модулі застосунку включають в себе:

- автентифікацію та доступ до електронної пошти;
- попередню обробку тексту;
- класифікацію загроз за допомогою машинного навчання;
- оцінку ефективності класифікації;
- фільтрацію по відправнику;
- генерацію звітів та інтеграцію з користувацьким інтерфейсом.

Тобто, основна робота застосунку виглядатиме так: буде виконуватися підключення до електронної пошти користувача, отримання листів із поштових папок Inbox та Spam, відбір релевантних листів на основі часових обмежень (враховуються повідомлення за останні 30 днів). Для отримання електронних листів застосунок використоватиме Google Gmail API, а автентифікація та авторизація проходитиме через OAuth 2.0 – сучасний протокол, який надає обмежений доступ до ресурсів користувача. Текст листів пройде нормалізацію за допомогою функції, яка очищає його від зайвих символів, видаляючи URL-адреси, електронні пошти, числа та пунктуацію, видаляє символи іншої мови, тощо. Наступним йде перетворення тексту у векторний вигляд, тобто використання TfidfVectorizer з бібліотеки scikit-learn, яка перетворює текст у числовий формат, використовуючи метод TF-IDF. Отримана матриця ознак розділиться на навчальні та тестові вибірки у співвідношенні 80% на 20%. Окремо створюватимуться два набори: один для класифікації повідомлень як кібersталкінгових або ні, а інший – для визначення їхньої підкатегорії. Після підготовки даних ініціалізуватимуться дві

моделі SVM із лінійним ядром та збалансованими вагами класів. Задля покращення точності моделей застосовуватиметься GridSearchCV (Grid Search with Cross-Validation) – метод автоматизованого підбору оптимальних гіперпараметрів для моделей машинного навчання з 5-кратною крос-валідацією. Ця технологія використовується для пошуку оптимального значення параметра c в моделі SVM. Вона контролює баланс між правильною класифікацією навчальних даних і запобіганням перенавчанню. Одразу як завершиться навчання моделей, оцінюватиметься їх точність класифікації за допомогою певних метрик ефективності і ROC-кривих, результати фіксуватимуться у логах. Фільтрація по відправнику означає аналіз частоти повідомлень від окремих відправників. Якщо відправник надсилає понад 2 загрозливі листи, повідомлення вважається підозрілим, бо кіберсталкінг, як було визначено – це явище систематичне.

Для того, щоб розробити ефективну інтелектуальну систему, яка виконуватиме завдання класифікації текстових даних, спочатку необхідно зібрати датасет, на якому моделі і будуть навчатися. Якість даних безпосередньо впливає на точність, узагальнюваність та стійкість моделі, тому набір даних повинен відповідати певним вимогам, таким як репрезентативність та збалансованість. Зібраний набір даних на початку складався з 5559 повідомлень, що є достатньою вибіркою для початкового навчання, хоча збільшення розміру набору даних в подальшому суттєво покращило результати. Серед повідомлень є як кіберсталкерські, так і звичайні тексти. Дані були зібрані більшою частиною синтетично, тобто повідомлення формувалися на основі аналізу існуючих загроз та шаблонів поведінки кіберсталкерів, щоб максимально точно відобразити їхні реальні прояви; а також використовувались реальні повідомлення з датасетів англійською мовою, що включали тексти: кібербулінгу, хейтспічу, з сексуальним підтекстом, спам/неспам повідомлення, переведені на українську. Як було вказано, було проведено аугментацію даних, щоб збільшити набір навчальних даних. Аугментація – це методика штучного розширення обсягу даних шляхом їх модифікації для покращення узагальнювальної здатності моделей та зменшення ризику перенавчання. В цій роботі було використано два методи – переклад на іншу мову і назад, а також міксування слів у повідомленні для зміни порядку слів у реченні. Це дозволило збільшити різноманітність текстів, при цьому не змінюючи їхній зміст, і створити більш стійку підборку повідомлень. Отже, в результаті було створено файл у форматі Excel з 13466 повідомлень, який є набором даних для навчання двох моделей. Файл містить 3 мітки: Text, Label, Types, – де Text є текстом повідомлення, Label є класифікацією кіберсталкінг/не-кіберсталкінг і Types – це тип повідомлення. За розподілом повідомлень у датасеті: 7771 належать до категорії «Cyberstalking», а 5695 – до «No-Cyberstalking». Серед кіберсталкінгу, виділено три основні типи повідомлень: «Сексуальний харасмент» (2221), «Залякування» (2885) і «Звичайний» (2665). На рисунку 1 можна побачити як виглядає файл, який являє собою датасет для навчання моделей в застосунку.

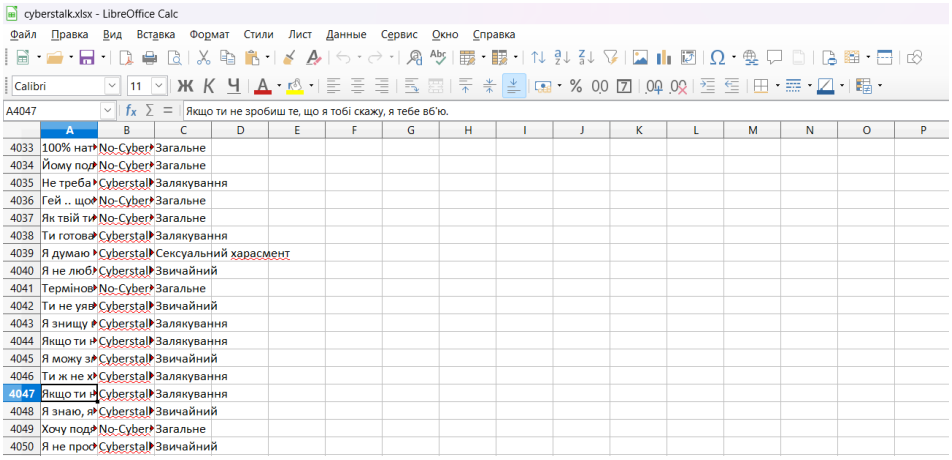


Рис.1. Зібраний набір даних

Об’єктивна оцінка якості побудованих моделей базуватиметься на низці загальновживаних у машинному навчанні метрик та підходів. Матриця помилок є одним із найважливіших інструментів. Вона представляє собою таблицю, яка відображає співвідношення між передбаченими та фактичними значеннями класів. Елементи цієї матриці включають кількість правильно класифікованих позитивних і негативних прикладів, а також кількість помилкових позитивних і негативних передбачень. На основі цієї матриці помилок розраховуються вже стандартні метрики ефективності моделі. Ті з них, які використовуватимуться, вказані нижче [17].

Влучність показує наскільки модель влучна у передбаченні позитивного класу:

$$\Pr ecision = \frac{TP}{TP + FP},$$

Повнота демонструє, яку частку з усіх реальних позитивних прикладів модель змогла правильно виявити:

$$\mathrm{Re} call = \frac{TP}{TP + FN},$$

F1-міра – гармонійне середнє між влучністю та повнотою, і використовується як зведений показник ефективності класифікації:

$$F - score = 2 * \frac{\Pr ecision * \mathrm{Re} call}{\Pr ecision + \mathrm{Re} call},$$

Загальна точність є загальним показником правильності роботи моделі:

$$Acurracy = \frac{TP + TN}{TP + TN + FP + FN}.$$

Додатково, будуть використовуватись середні значення показників (macro avg, weighted avg), задля врахування особливостей різного розподілу класів у тестовій вибірці. Macro avg – середнє арифметичне для всіх класів. Weighted avg – усереднене з урахуванням кількості прикладів кожного класу.

Тож, за допомогою описаних вище методів, проведемо оцінювання ефективності моделей із запропонованого підходу виявлення кіберсталкерської поведінки. Для основної бінарної моделі, яка виявляє чи є в повідомленні ознаки кібersталкінгу, клас 0 буде означати кібersталкінг, а клас 1 – не кібersталкінг. Що стосується багатокласової моделі визначення типів кібersталкінгу, то тут: клас 0 буде означати залякування, 1 – звичайний кібersталкінг, 2 – сексуальний харасмент. Першим етапом проходитиме оцінка бінарної моделі. Як видно з таблиці 2, модель допустила лише 22 помилки FN та 17 випадків FP.

Таблиця 2

Матриця помилок бінарної моделі

№	Клас	Прогноз 0	Прогноз 1
1	0	1523	22
2	1	17	1131

Таблиця 3 вказує, що модель має майже симетричні показники влучності та повноти, що є ознакою її стабільності. Значення F1-міри перебуває на рівні 0.98 – 0.99, а загальна точність становить 0.99. Високі значення macro avg та weighted avg говорять про відсутність критичного дисбалансу в розпізнаванні обох класів.

Таблиця 3

Оцінки ефективності бінарної моделі

№	Метрика	Клас 0	Клас 1	№	Метрика	Клас 0
1	Precision	0.99	0.98	5	Accuracy	0.99
2	Recall	0.99	0.99	6	Macro avg	0.99
3	F1-score	0.99	0.98	7	Weighted avg	0.99
4	Support	1545	1148			

В свою чергу, аналіз матриці помилок багатокласової моделі, таблиця 4, демонструє, що найбільша кількість правильних класифікацій спостерігається в класі «Залякування». Більшість помилок зосереджена між класами 0 і 2, та 1 і 2, це вказує на наявність перехресних ознак у змісті повідомлень. В таблиці 5 ми бачимо, що всі три класи мають показники precision і recall в діапазоні 0.88 – 0.91, тобто достатньо рівномірна якість класифікації, загальна точність моделі становить 0.89. Значення macro avg і weighted avg підтверджують, що модель справляється з усіма класами порівняно однаково добре, незважаючи на можливий дисбаланс у навчальній вибірці.

Таблиця 4

Матриця помилок багатокласової моделі

№	Клас	Прогноз 0	Прогноз 1	Прогноз 2
1	0	502	43	17
2	1	38	472	26
3	2	16	24	417

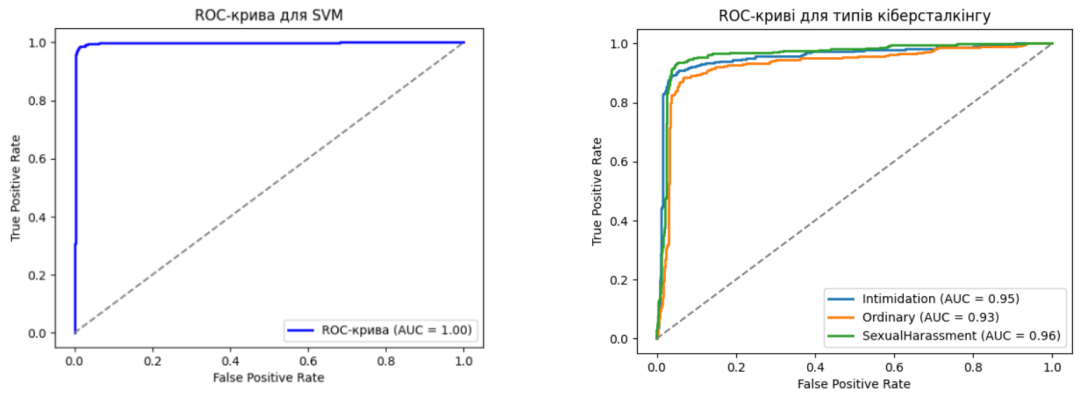
Таблиця 5

Оцінки ефективності багатокласової моделі

№	Метрика	Клас 0	Клас 1	Клас 2
1	Precision	0.90	0.88	0.91
2	Recall	0.89	0.88	0.91
3	F1-score	0.90	0.88	0.91
4	Support	562	536	457
5	Accuracy	0.89		
6	Macro avg	0.90		
7	Weighted avg	0.89		

Для візуалізації оцінки ефективності, побудуємо ROC-криві, які відображатимуть співвідношення між чутливістю та часткою хибно позитивних спрацювань моделей. Оптимальним є класифікатор, що наближається до верхнього лівого кута ROC-графіка. На обох графіках, криві мають AUC – площу під кривою – метрикою, яка кількісно виміряє здатність відокремлювати класи. Значення AUC наближається до одиниці для ідеального класифікатора.

Як можна побачити, на рисунку 2, в двох результатах, крива значно вище діагоналі випадкового вгадування, для бінарної моделі AUC = 1.00. Для багатокласової моделі AUC також зберігається на високому рівні: від 0.93 до 0.96, вказуючи таким чином на ефективність моделі навіть у складних сценаріях класифікації.



а) *б*)

Рис.2. Графіки ROC-кривих зі значеннями AUC:
а – бінарна модель; *б* – багатокласова модель

Для проведення тестування самого застосунку, на електронну пошту було спеціально надіслано 13 повідомлень, які містять в собі ознаки кіберсталкінгу, з яких 3 потрапили у спам і 10 у вхідні листи; 3 відправника надіслало більше 2 повідомлень, і в результаті вийшло 10 листів з кіберсталкінгом. Після запуску веб-застосунку у браузері відкривається перше представлення (рис. 3).

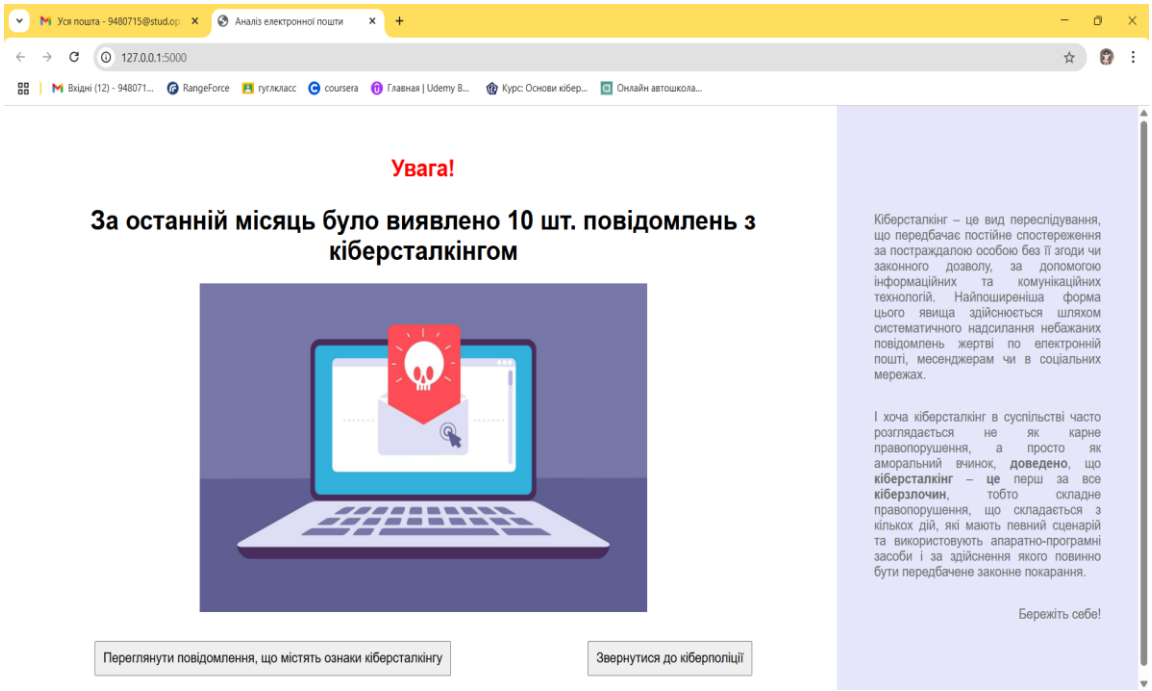


Рис.3. Головна сторінка веб-застосунку

Як можна побачити, програма виявила всі листи, які відповідають заданим фільтрам. З правого боку сторінки знаходиться невелика довідка для користувача, де визначається поняття кіберсталкінг і повідомляється користувачу, що він є карним кібезлочином.

Натискаємо кнопку перегляду повідомлень з кіберсталкінгом і таким чином переходимо до другого представлення (рис. 4).

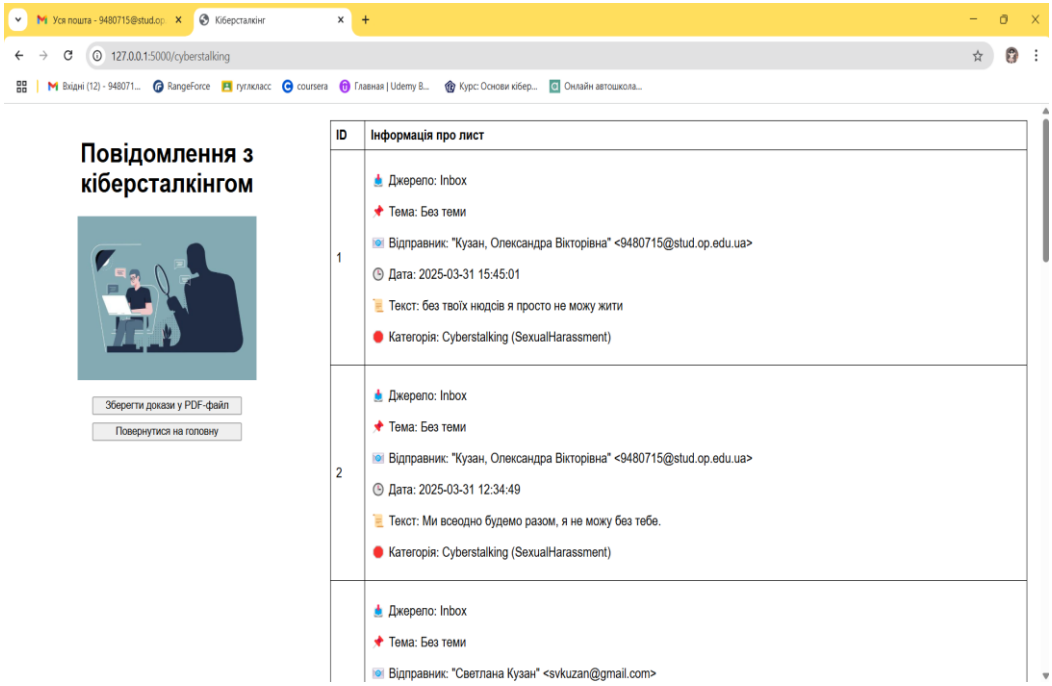
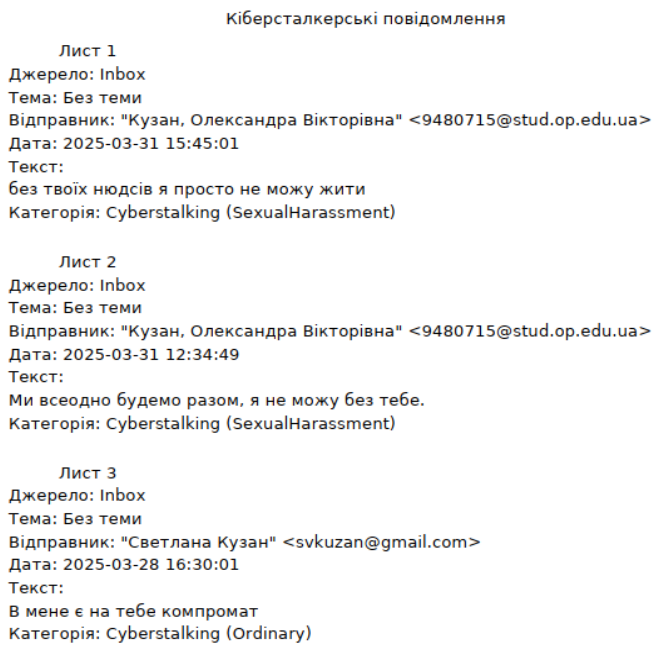


Рис.4. Друга сторінка веб-застосунку

Користувачу надаються відомості з виявлених листів, додається визначення типу кіберсталкерського повідомлення. Застосунок визначив типи повідомлень коректно, підібравши найбільш доречний варіант для кожного повідомлення. Сторінка містить в собі кнопку збереження доказів, натиснувши на яку, завантажується файл з інформацією, яку користувач бачить на екрані (рис. 5).



Кіберсталкерські повідомлення

Лист 1
Джерело: Inbox
Тема: Без теми
Відправник: "Кузан, Олександра Вікторівна" <9480715@stud.op.edu.ua>
Дата: 2025-03-31 15:45:01
Текст:
без твоїх нюдсів я просто не можу жити
Категорія: Cyberstalking (SexualHarassment)

Лист 2
Джерело: Inbox
Тема: Без теми
Відправник: "Кузан, Олександра Вікторівна" <9480715@stud.op.edu.ua>
Дата: 2025-03-31 12:34:49
Текст:
Ми всеодно будемо разом, я не можу без тебе.
Категорія: Cyberstalking (SexualHarassment)

Лист 3
Джерело: Inbox
Тема: Без теми
Відправник: "Светлана Кузан" <svkuzan@gmail.com>
Дата: 2025-03-28 16:30:01
Текст:
В мене є на тебе компромат
Категорія: Cyberstalking (Ordinary)

Рис.5. Завантажений пдф-файл

Повернувшись на першу сторінку, натискаємо на другу кнопку – звернення до кіберполіції, де користувачу надаються посилання для звернення по допомогу (рис. 6).

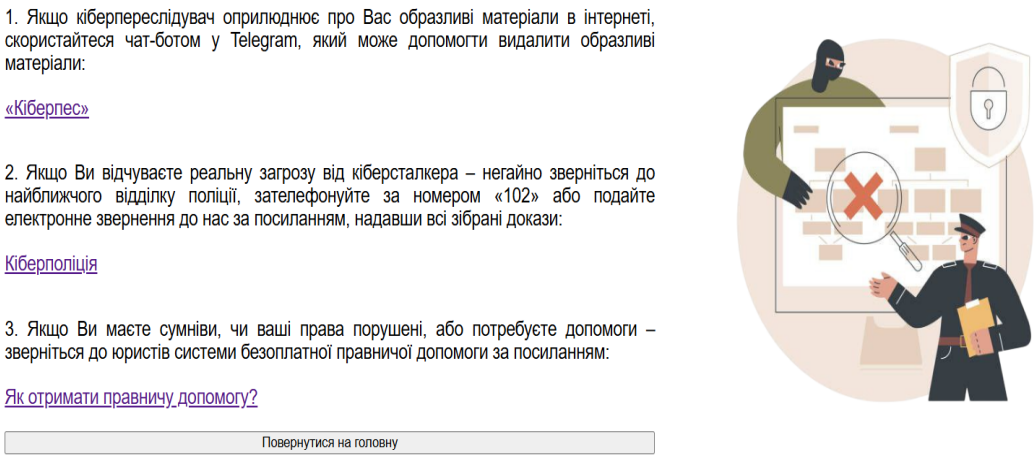


Рис.6. Третя сторінка веб-застосунку

Висновки. В даній роботі було проведено аналіз явища кіберсталкінгу, розглянуто його характеристики та небезпечні прояви, а також існуючі підходи до його виявлення і, на основі цього було розроблено застосунок, який дозволяє виявляти потенційні ознаки кіберсталкінгу в повідомленнях електронної пошти. В результаті роботи застосунку, користувач отримує автоматизований аналіз своєї електронної пошти на наявність кіберсталкінгових повідомлень та можливість вчасно вжити заходів для покарання зловмисника, без важких наслідків для себе. Ефективність застосунку забезпечує висока точність класифікації повідомлень за допомогою перевірених алгоритмів ML, які гарно показали себе у виявленні загроз: бінарна модель досягла загальної точності 99%, багатокласова модель має оцінку точності в 89% і впевнено розрізняє основні види кіберпереслідувань. Система працює з даними саме українською мовою, що є новим та актуальним для українського суспільства, може оновлюватися та навчатися на нових даних, а за допомогою веб-інтерфейсу користувач отримує чітке уявлення про виявлені загрози кіберсталкінгу. Отже, це дослідження має важливе теоретичне та практичне значення. Розроблений підхід чи сам застосунок може бути використаний для покращення безпеки користувачів в їх онлайн-комунікаціях.

Список літератури

1. Sheridan L. What is Cyberstalking? A Review of Measurements. *Journal of Interpersonal Violence.* 2021. URL: https://www.researchgate.net/publication/348467830_What_is_Cyberstalking_A_Review_of_Measurements.
2. Ahmed N. Cyberstalking: A content analysis of gender-based offenses committed online. 2019.
3. Goyal S., Kaur A. Impact of Cyberstalking on Women and Children: An Analysis. *International Journal of Research Publication and Reviews*. 2024. URL: https://www.researchgate.net/publication/385494443_Impact_of_Cyberstalking_on_Women_and_Children_An_Analysis.
4. Mcfarlane L., Bocij P. An exploration of predatory behaviour in cyberspace: Towards a typology of cyberstalkers. *First Monday.* 2003. URL: https://firstmonday.org/ojs/index.php/fm/article/view/1076.
5. Holyst B. Cyberstalking as a form of cyberharassment. *Ius novum.* 2015. URL: https://www.lazarski.pl/fileadmin/user_upload/dokumenty/czasopisma/ius-novum/2015/Ius_Novum_2_15_holyst.pdf.
6. Charan J. L. Cyber Stalkers and Cyber Bullies: Protecting Women in the Digital Age. *Cyber Crime & Cyber Securities in India*. 2023. URL: https://www.researchgate.net/publication/375120797_Cyber_Stalkers_and_Cyber_Bullies_Protecting_Women_in_the_Digital_Age.
7. Miftha A. The Social, Legal, and Technical Perspectives of Cyberstalking in India. 2024. URL: https://uobrep.openrepository.com/handle/10547/626190.

8. Kaur P., Dhir A., Tandon A., Alzeiby E. A. A systematic literature review on cyberstalking. An analysis of achievements and future promises. *Technological Forecasting & Social Change.* 2021. URL: https://www.researchgate.net/publication/347381968_A_systematic_literature_review_on_cyberstalking_An_analysis_of_past_achievements_and_future_promises.
9. O'shea B., Asquith N. L., Prichard J. Mapping Cyber-Enabled Crime: Understanding Police Investigations and Prosecutions of Cyberstalking. *International Journal for Crime, Justice and Social Democracy.* 2022. URL: https://www.crimejusticejournal.com/article/view/2096.
10. Проект Закону про внесення змін до Кримінального процесуального кодексу України та Закону України «Про запобігання та протидію домашньому насильству» щодо встановлення відповідальності за злочинне переслідування (сталкінг). Номер, дата реєстрації: 12088 від 02.10.2024. Верховна рада України. 2024. URL: https://itd.rada.gov.ua/billInfo/Bills/Card/44972.
11. Tokunaga R. S., Aune K. S. Cyber-Defense: A Taxonomy of Tactics for Managing Cyberstalking. *Journal of Interpersonal Violence.* 2015. URL: https://www.researchgate.net/publication/278793403_Cyber-Defense_A_Taxonomy_of_Tactics_for_Managing_Cyberstalking.
12. Gautam A. K., Bansal A. Automatic Cyberstalking Detection on Twitter in Real-Time using Hybrid Approach. *International Journal of Modern Education and Computer Science*. 2023. URL: https://www.mecs-press.org/ijmecs/ijmecs-v15-n1/IJMECS-V15-N1-5.pdf.
13. Gautam A. K., Bansal A. A Review on Cyberstalking Detection Using Machine Learning Techniques: Current Trends and Future Direction. *International Journal of Engineering Trends and Technology*. 2022. URL: https://ijettjournal.org/archive/ijett-v70i3p211.
14. Wen Z., Taketoshi Y., Xijin T. TFIDF, LSI and Multi-word in Information Retrieval and Text Categorization. *IEEE.* 2008. URL: http://meta-synthesis.amss.cas.cn/Publication/MSKS_Publications/Conference_papers/Paperlists/201410/P020141013596946632792.pdf.
15. Gautam A. K., Bansal A. Performance analysis of supervised machine learning techniques for cyberstalking detection in social media. *Journal of Theoretical and Applied Information Technology.* 2022. URL: https://scholar.google.fr/citations?view_op=view_citation&hl=ja&user=g0KKFIsAAAAJ&citation_for_view=g0KKFIsAAAAJ:QIV2ME_5wuYC.
16. Dakhaz M. A., Adnan M. A. Machine Learning Applications based on SVM Classification: A Review. *Qubahan Academic Journal.* URL: https://journal.qubahan.com/index.php/qaj/article/view/50/36.
17. Geeksforgeeks. Confusion Matrix in Machine Learning. URL: https://www.geeksforgeeks.org/confusion-matrix-machine-learning/.

О.В. Кузан, Н.І. Кушніренко, В.О. Назаров, В.В. Подуфалов

DEVELOPMENT OF AN APPLICATION FOR AUTOMATED DETECTION OF CYBERSTALKING

O.V. Kuzan, N.I. Kushnirenko, V.O. Nazarov, V.V. Podufalov

National Odesa Polytechnic University
1, Shevchenko Ave., Odesa, 65044, Ukraine
Email: kushnirenko@op.edu.ua

Today, cyberstalking has shown itself to be an extremely fast-growing phenomenon that has caused a paradigm shift in persecution. It consists of intrusive, systematic interference in the victim's online life for the purpose of psychological pressure, manipulation, intimidation, etc. Despite the scale of the problem, there are currently no effective means of automated detection of such threats in Ukrainian. The aim of this paper is to improve user safety in the digital environment by developing an application for detecting cyberstalker behavior in emails using machine learning. The study analyzed the phenomenon of cyberstalking, as well as modern approaches to its detection, which allowed us to determine the directions of development and the main tasks of the work. The developed application is based on the use of the SVM algorithm to build models, which are trained on a specially formed dataset. Two models have been created: a binary model for classifying emails into “cyberstalking” and “non-cyberstalking” and a multi-class model for dividing them by type: intimidation, sexual harassment, and ordinary cyberstalking. The tool also provides the ability to store detected cyberstalker messages as evidence for victims to report to the cyber police. Testing of the approach has shown its high efficiency: accuracy of the binary model was 99%, and the multi-class model reached 89%. The obtained results show the significant potential of the application in detecting cyberstalking and thus increasing the level of digital security of users in online communications, so the proposed application for detecting cyberstalking can be successfully implemented for personal use. The results of this work can be used in further research and development in the field of cybersecurity and cyberstalking.

Keywords: cyberstalking, email, machine learning, text classification, cybersecurity.