Міністерство освіти і науки України
Національний технічний університет
«Дніпровська політехніка»

Інститут електроенергетики
Факультет інформаційних технологій
Кафедра безпеки інформації та телекомунікацій

ПОЯСНЮВАЛЬНА ЗАПИСКА
кваліфікаційної роботи ступеня бакалавра

студента *Павлова Сергія Олексійовича*

академічної групи **125м-21-1**

спеціальності *125 Кібербезпека*

спеціалізації[1]

за освітньо-професійною програмою *Кібербезпека*

на тему *Методика захисту від ботів на основі аналізу повідомлень*

Керівники	Прізвище, ініціали	Оцінка за шкалою		Підпис
		рейтинговою	інституційною	
кваліфікаційної роботи	к.т.н Сафаров О.О.			
розділів:				
спеціальний	к.т.н Сафаров О.О.			
економічний	к.е.н. Пілова Д.П.			

Рецензент				

Нормоконтролер				

Дніпро
2022

ЗАТВЕРДЖЕНО:
завідувач кафедри
безпеки інформації та телекомунікацій
_____д.т.н., проф. Корнієнко В.І.

«____»_____2022 року

ЗАВДАННЯ
на кваліфікаційну роботу ступеня бакалавра

студенту ***Павлову Сергію Олексійовичу*** ***академічної групи*** ***125м-21-1***
(прізвище та ініціали) (шифр)

спеціальності ***125 Кібербезпека***

спеціалізації[1]

за освітньо-професійною програмою ***Кібербезпека***

на тему ***Методика захисту від ботів на основі аналізу повідомлень***

Затверджую наказом ректора НТУ «Дніпровська політехніка» від _____ №_____

Розділ	**Зміст**	**Термін виконання**
Розділ 1	*Аналіз загальної інформації про ботів, визначення предмету досліджень*	28.10.2022
Розділ 2	*Аналіз існуючих рішень виявлення ботів, перелік запропонованих методів детекції і захисту*	15.11.2022
Розділ 3	*Економічне обґрунтування доцільності впровадження запропонованих рішень кваліфікаційної роботи*	20.12.2022

Завдання видано _____ _____
(підпис керівника) (прізвище, ініціали)

Дата видачі завдання: _____

Дата подання до екзаменаційної комісії: _____

Прийнято до виконання _____ Павлов С.О._____
(підпис студента) (прізвище, ініціали)

РЕФЕРАТ

Пояснювальна записка: 59 с., 6 рис., 1 табл., 4 додатків, 61 джерело.

Об’єкт дослідження: додатки в яких отримується інформація.

Мета роботи: дослідити методи захисту від соціальних ботів.

Методи дослідження: спостереження, аналіз, класифікація.

У першому розділі викладені основні особливості предмету аналізу, визначені причини для проведення аналізу. Була проаналізована нормативно-правова база у сфері дезінформації в Україні.

У спеціальній частині проаналізовані існуючі методики детекції соціальних ботів і програмні методики захисту від соціальних ботів, викладено ефективність існуючих алгоритмів.

У економічному розділі визначений соціально-економічний вплив від діяльності шкідливих соціальних ботів, розраховано витрати на розробку системи інформаційної безпект детекції соціальних ботів.

Результати аналізу можуть бути застосовані для реалізації програмного забезпечення з пошуку зловмисних соціальних ботів для їх подальшої ліквідації.

ІНФОРМАЦІЙНА БЕЗПЕКА, СОЦІАЛЬНІ МЕРЕЖІ, СОЦІАЛЬНІ БОТИ, TWITTER, ЕКОНОМІЧНА ДОЦІЛЬНІСТЬ, МЕТОДИ ЗАХИСТУ, АЛГОРИТМ

ABSTRACT

Explanatory note: 59 p., 6 figures, 1 tables, 4 supplements, 61 sources.

Object of study: information and news applications.

The purpose of the qualification work: explore social bot protection methods.

Methods that were used: observation, analysis, classification.

In the first part of the study the main features of the subject of the analysis were outlined, the reasons for the analysis were determined. The regulatory and legal framework in the field of disinformation in Ukraine was analyzed.

In the second part of the study existing social bot detection methods and software protection methods against social bots were analyzed. Were outlined the effectiveness of existing algorithms.

In the economic part socio-economic impact of malicious social bots were identified. The costs of developing social bot detection software were calculated. Indicators of economic efficiency by the implementation security system were calculated and analyzed.

The results of the analysis can be used to implement searching software for malicious social bots that must be further eliminated.

INFORMATION SECURITY, SOCIAL NETWORK, SOCIAL BOTS, TWITTER, ECONOMIC FEASIBILITY, SECURITY METHODS, ALGORYTHM

СПИСОК УМОВНИХ СКОРОЧЕНЬ

ЗМІ – засоби масової інформації;

ПЗ – програмне забезпечення;

ПК – персональний комп'ютер;

РФ – Російська федерація;

ТЗ – технічне завдання;

DDoS атака – атака типу denial of service;

HTTP – Hypertext transfer protocol;

SVM – Support Vector Machines;

URL – Uniform Resource Locator.

ЗМІСТ

С.

ВСТУП 8

РОЗДІЛ 1. СТАН ПИТАННЯ. ПОСТАНОВКА ЗАДАЧІ 9

1.1.Загальні відомості 9

1.2.Визначення предмету аналізу 13

1.3.Підстави проведення аналізу 15

1.4.Проблемні аспекти виявлення і ідентифікації соціальних ботів 18

1.5.Висновок і постановка задачі 20

РОЗДІЛ 2. СПЕЦІАЛЬНА ЧАСТИНА 21

2.1.Сучасні методи аналізу даних з метою виявлення ботів 21

2.1.1.Natural Language Processing 22

2.1.2.Алгоритми нейронних мереж 24

2.1.3.Статистична класифікація, алгоритми класифікації 27

2.1.4.Семантичний аналіз тексту 29

2.2.Існуючі програмні рішення для визначення ботів у соціальних мережах 31

2.3.Порівняльна характеристика методів захисту 32

2.4.Висновки 34

РОЗДІЛ 3. ЕКОНОМІЧНИЙ РОЗДІЛ. 35

3.1 Розрахунок (фінансових) капітальних витрат 35

3.1.1 Визначення трудомісткості розробки та опрацювання ПЗ 35

3.1.2 Розрахунок витрат на створення програмного продукту 37

3.1.3 Капітальні (фіксовані) витрати на створення системи інформаційної безпеки 39

3.2 Розрахунок експлуатаційних витрат 40

3.3 Оцінка величини збитку 43

3.4 Загальний ефект від впровадження системи інформаційної безпеки 45

3.5 Визначення та аналіз показників економічно ефективності системи 45

3.6 Висновки до економічного розділу 47

ВИСНОВКИ 48
ПЕРЕЛІК ПОСИЛАНЬ 49
ДОДАТОК А. Відомість матеріалів кваліфікаційної роботи
ДОДАТОК Б. Перелік документів на оптичному носії
ДОДАТОК В. Відгук керівника економічної частини
ДОДАТОК Г. Відгук керівника кваліфікаційної роботи

ВСТУП

Сучасний світ не можливо представити без використання систем передачі інформації. Тенденції населення до збільшення кількості споживання інформаційного контенту, такого як новини, блоги, соціальні медіа, показують необхідність захисту і фільтрації такої інформації.

У зв'язку з останніми військовими подіями Україна перебуває у кризовому стані, а саме тому доцільними постає визначення переліку причин цього стану. Перші виявлення інформаційно-психологічних впливу з боку країни агресора були зафіксовані з 2013 року і досліджені. Актуальною задачею стає зниження подібного впливу на населення України і підвищення обізнаності від ефектів такого роду.

Одним із найвплиловіших інструментів донесення інформації є додатки і сайти з швидким доступом до інформації – соціальні мережі, месенджери, додатки з новинами і навіть сайти відеохостингу. Кількість інформації, що передається таким шляхом, збільшується з кожним днем, а користувачі цих додатків не встигають забезпечити надійність джерела прочитаної інформації і самої інформації. Саме тому джерела з великим рівнем довіри межують з джерелами сфальсифікованої і неправдивої інформації. Під час використання таких ресурсів для отримання актуальних новин нерідко можна зустріти використання агресивної пропаганди і поширення великої кількості містифікації з використанням автоматизованих програм – ботів.

Соціально-психологічні результати подібного впливу з боку країни агресора вже були зафіксовані з 2013 року, а повний вплив на населення важко визначити через його нову і приховану природу нових методів гібридної війни.

Було вирішено створити систему інформаційної безпеки, детекції і захисту від соціальних ботів, реалізація якої зменшить соціально-психологічного і дезінформаційний вплив на людей.

РОЗДІЛ 1. СТАН ПИТАННЯ. ПОСТАНОВКА ЗАДАЧІ

1.1. Загальні відомості

Одним із визначних феноменів людства є соціальні взаємозв'язки. Люди завжди збиралися у групи і використовували кількісну перевагу у різноманітних сферах життя – будь-то збиральництво, полювання, землеробство, будування, тощо. Однак з часом на ці сфери життєдіяльності значною мірою почав впливати технологічний прогрес, що можна спостерігати при зміні епох. На кінці 20 століття одним із найголовніших факторів, що вплинули на функціонування всіх сфер життя людства, стала інформаційна революція. Поширення використання мережі «Інтернет» призвело до великого збільшення обміну інформацією, а також до легшого і швидшого доступу до будь-якої інформації.

Прискорення доступу до інформації і збільшення її кількості зменшує ймовірність отримання достовірних знань. Великий обсяг знань призводить до надлишковості зайвої інформації, яка не несе під собою ніякої користі для користувачів. Такий феномен як дезінформація відомий людство здавна і, як відомо з історії, має великий негативний вплив. Одним із прикладів такого впливу можна привести під час компанії антивакцинації, яка була створена у соціальній мережі «Twitter» ботами і тролями - користувачами, які підривали міжнародний консенсус по цій темі [1]

Однак, найбільшого впливу дезінформація набуває під час кризових ситуацій, коли необхідно вирішувати життєво важливі питання. До таких кризових ситуацій можна віднести епідемії, епізоотії, природні катастрофи, військові конфлікти, тощо. Викривлення інформації, маніпулювання народною думкою і інформаційно-психологічні операції, як форми нелетального ведення війни, стали складовою частиною військової справи [2].

Сучасні війни відбуваються не лише не площині збройного протиборства, а і у різноманітних площинах людської діяльності: «…війну натепер варто розглядати як складне суспільно–політичне явище, що включає сукупність різних форм

боротьби (політичної, економічної, збройної, інформаційної, психологічної й ін.), що ведуть між собою держави або коаліції держав» [2].

Термін «гібридна війна» набув своєї популярності після подій анексії Криму 2014 року. [3] .Дискусії на тему визначення термінології «гибридної війни» мають велике значення на методи реагування і захисту, які будуть застосовуватися на рівні держав. Гібридний підхід до визначення війни у науковому журналі [4] має свої недоліки: розмиття чіткої зони міжнародних відносин від військового втручання у так звану «сіру зону конфлікту». Однак такі методи впливу під час асиметричних війн, коли слабкість ворогуючої сторони протиставляється перевагою іншої сторони, і визначають успішність військових стратегій.

Гібридна війна включає у собі заперечення факту акту агресії, використання економічного, диверсійного і технологічного тиску, використання внутрішньополітичних конфліктів. А тому після четвертої інформаційної революції можна виділити окрему складову гібридної війни – війну інформаційну.

Інформаційна війна – інтенсивна боротьба у інформаційному просторі, яка ставить своєю головною метою досягнення інформаційної, психологічної, ідеологічної переваги, нанесення пошкодження інформаційним системам, процесам і ресурсам, об’єктам критичної інфраструктури і комунікацій (інформаційно-технічна, сетецентрична та кібервійна), підриву політичних і соціальних систем, а також масової психологічної обробки особового складу військ і населення(інформаційно-психологічна війна).

Одним із методів впливу на суспільну думку є соціальні мережі. Неможливо повністю проаналізувати весь обсяг впливу, який генерується і обговорюється у бесідах соціальних мереж, однак найпомітніші події і явища є предметом аналізу соціально-психологічного впливу на думки користувачів(людей).

Дійсно, останні актуальні новини люди отримують за допомогою інформаційних агентств, програм з новинами, газет, радіостанцій, тощо. Однак останніми роками спостерігається тенденція отримання подібної інформації за допомогою соціальних мереж, месенджерів і додатків для дзвінків та обміну повідомленнями.

Схильність людей до отримання швидкої і фільтрованої інформації призвела до активного використання алгоритмів у фільтрації показаного контенту з боку соціальних мереж і додатків-агрегаторів, а також до збільшення додатків для новин. Дослідження [5] показує, що люди схильні довіряти і обирати фільтрацію алгоритмами більше, ніж редакторам або журналістам, в той час як для молодого покоління ці результати були ще більшими.

Аналіз активності користувачів у дослідженні [6] під час російського вторгнення 24 лютого показує причетність російських державних медіа та інші домени до сплеску поширення недостовірної інформації і закликає до перегляду впливу соціальних медіа на людей. На рисунку 1.1 показано, що #putin використовується набагато ширше, ніж #zelenskyy. Щоб краще зрозуміти кількість твітів, які містять #putin, ми також побудували графік кількості твітів, які використовують #biden для нинішнього президента Сполучених Штатів Джо Байдена, і кількості твітів, які використовують #trump, посилаючись на колишнього президента США, Дональда Трампа.

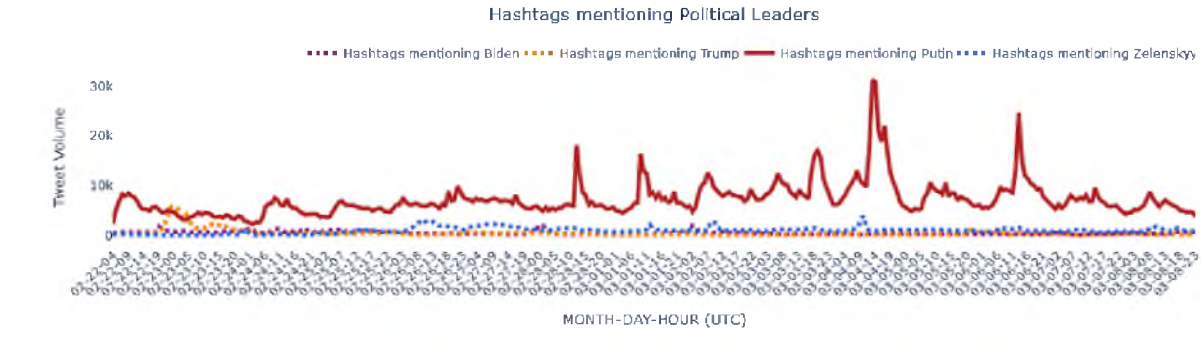

Рисунок 1.1 – Загальна кількість твітів, які використовують хештеги #biden, #trump, #putin і #zelenskyy

З інформації про щорічне опитування USAID-Internews [7] можна сказати що українці все більш схильни вибирати соціальні мережі як джерела новин. Відзначається продовження існуючої тенденції відходу від традиційних ЗМІ та збільшення обсягів споживання новин з Інтернету і соціальних мереж/месенджерів. Telegram – головний месенджер для отримання новин. Головним чином для отримання коротких новин, можливість бути в курсі подій. Набирають

популярності міські/локальні телеграм-канали, новини з яких можна безпосередньо перевірити самостійно, або новини з яких можуть мати вплив на життя в місті.

Такі соціальні додатки і месенджери як Telegram, YouTube, Facebook, Viber, Instagram, TikTok, Twitter респонденти дослідження «Київського міжнародного інституту соціології» [8] використовували за останні два місяці (березень-квітень 2022 року) для отримання новин.

Однак соціальні мережі окрім своїх прямих можливостей – спілкування і поширення думок серед користувачів, можуть бути використані як інструмент під час політичних, військових, терористичних або інших протистоянь. Так, під час пандемії COVID-19 боти генерували значну кількість негативних діалогів між користувачами і мали значний вплив на інформаційні потоки.

Дослідження впливу ботів на агресивне онлайн спілкування протягом пандемії COVID-19 [9] показало, що «Незважаючи на розбіжності в контексті, ми наголошуємо на двох головних ідеях, отриманих із цього аналізу. По-перше, боти, особливо ненависні боти, мають вплив і контролюють основні аспекти потоку інформації в розмові про пандемію. По-друге, люди відіграють значну роль у поширенні мови ненависті; таким чином, визнання органічної участі в онлайн-токсичності буде життєво важливим для розгляду шляхів її пом'якшення.»

Слід зауважити, що діяльність ботів може маніпулювати соціальною думкою і впроваджувати розбрат поміж користувачів. Окрім питання визначення приналежності користувача до ботів, постає проблема масової підтримки необхідної сторони з боку ботів. Так, наприклад, у статті [10] було помічено можливість існування чорного маркету ботів для політичного викривлення інформації, а також те, що боти можуть бути однією з причин початку великої кількості політичних дебатів серед користувачів соціальної мережі Twitter.

Подібні висновки щодо значного впливу на предметний дискурс та прискорення поширення хибної інформації було розглянуто дослідженнями [11, 12, 14, 15].

Дслідження соціальних мереж [13] показало, що звичайна соціальна інформація має неоднорідні рівні інформаційного навантаження та уваги, а отже

краща і якісніша інформація не мають значно більших шансів стати популярними порівняно з низькоякісною інформацією. Якщо розглядати ці висновки з боку містифікації та фейкових новин, тоді стає зрузумілим чому неякісна інформація поширюється так само швидко, як і надійна інформація в соціальних мережах.

1.2. Визначення предмету аналізу

Бот – програмне забезпечення, яке здатне виконувати автоматизовані задачі. До таких задач відносять моніторинг трафіку, виконання скриптів, пересилання повідомлень, імітація діяльності людини, букінг, спостереження за станом системи, тощо. Хоча боти і можуть імітувати діяльність людей, однак на відміну від людей, вони виконують свої задачі зі значною швидкістю. Більше ніж половина трафіку у мережі Інтернет генерується ботами.

Слід зазначити, що хоча боти і виконують автоматизовану роботу, проте сам термін «бот» за останні роки набув багато визначень, які мають значні відмінності і особливості у класифікації. Можна виділити наступні види ботів:

– Чат-боти (chatbots) – програми, що здатні вести діалог з людиною. Одним із найвідоміших представників є чат-бот “Eliza” – програма обробки природньої мови (Natural Language Processing) від дослідницького проекту Масачусетського Інституту Технологій розроблена в 1966 року [16]. Використовуючи «відповідність шаблонів» і методологію підстановки, програма дає стандартні відповіді, які змушували перших користувачів відчувати, що вони розмовляють з кимось, хто розуміє їхнє введення. Іншими представниками чат-ботів є віртуальні асистенти, такі як Amazon’s Alexa [17], Apple’s Siri [18] and Google Assistant [19]. Як зазначається у науковій статті Трофименко О. Г. [20] – «Однією з важливих можливостей чат-ботів є можливість збирати інформацію про користувачів, відслідковувати їхні дії, а потім за потреби проаналізувати їхні звички.»

– Соціальні боти – роботи, що здатні імітувати діяльність людей, вести з користувачами дискусії у соціальних мережах Були розглянуті наукові статті [13 – 15], що доказують важливість вивчення і необхідності протидії соціальним ботам.

– Боти для покупок, шоп-боти (Shopbots) – програми створені для автоматизації покупок, порівняння цін, отримання повідомлення про знижки.

– Інформаційні боти, ноу-боти (knowbots) – окремі програми, які можуть самостійно вишукувати інформацію у потрібних базах даних в мережі «Інтернет» від імені користувача. Після виконання пошуку цей бот надсилає результат користувачу.

– Боти-павуки або боти шукачі (Spiders or crawlers)- веб-боти, які отримують доступ до сайтів і збирають контент для індексації пошукових систем.

– Веб-скрапери, скрапери-пошукачі (web scraping crawlers) – веб-боти, що використовують для збору даних і отримання певного контенту з веб-сторінок, а також боти пошукових систем.

– Боти спостерігачі (monitoring bots) – боти, що використовують для спостерігання за станом здоров'я систем чи веб-сайтів.

– Боти для транзакцій (transactional bots) – тип ботів, який спрощує операції людей, які виконуються по телефону, наприклад блокування викраденої картки або підтвердження години роботи банку.

– Шкідливі боти (malicious bots) – боти, що використовуються зловмисниками для виконання протиправних дій і кіберзлочинів.

Ботнет – це група робочих машин, які були зламані/скомпрометовані та з'єднані за допомогою мережі «Інтернет». Кожна окрема робоча машина у такій системі називається ботом і здатна виконувати будь-які протиправні дії. Стати ботом може будь-який пристрій, який має вихід до мережі «Інтернет» - комп'ютер, ноутбук, смартфон або пристрій з так званої «Мережі пристроїв» (Internet of Things).

Перш за все слід відрізняти ботів дії яких направлені на користь – чатботи, боти помічники (Knowbots), боти для моніторингу, тощо, від ботів які використовують зловмисники для правопорушної активності: атака типу відмова сервісу (DoS, DDoS), спуфінг трафіку, розсилання зловмисного програмного забезпечення, спаму, пропаганди, та інше. Якщо не приділяти захисту від ботів достатньої уваги, компанії можуть зазнати значних втрат через відмову в

нормальному обслуговуванні від своїх сервісів, або зазнати репутаційних втрат. Звичайні користувачі будуть звертати увагу на ті новини або інформацію, яку підтримує велика кількість ботів, що створює ефект підтримки більшості і упередженість до певної сторони конфлікту.

На відміну від веб-ботів, які проводять автоматизовану роботу з веб-трафіком, боти в соціальних мережах можуть імітувати людську поведінку з ціллю отримання морально-психологічної переваги, створення початку дискусій на визначену тему або посівання розбрату серед інших користувачів.

Соціальні боти можуть використовувати один акаунт з людиною, щоб замаскувати свою активність. У науковій статті [21] таких користувачів називають «cyborg users» та зазначають, що через свою поведінкову унікальність вони мають більші можливості для поширення неправдивих, фейкових новин і інформації.

У науковій статті [22] було ідентифіковано приналежність ботів до створення політичних маніпуляцій, неправдивих новин, теорій змови, маніпуляції на фондовому ринку, маніпуляцією думок на тему публічного захисту здоров'я, пропаганди. Окреме дослідження ботів на тему вакцинації [23] показало, що «…боти і тролі соціальної мережі «Twitter» мають значний вплив на онлайн спілкування щодо вакцинації».

1.3. Підстави проведення аналізу

Зловмисники можуть використовувати ботів, щоб завдати наступні протиправні дії:

1) DDoS атаки, що спрямовані на посилання великої кількості запитів до сервісу, що призводить до відказу у його доступності і неможливості звернення до цього сервісу. Цей тип атак з використанням ботнету один із найтиповіших для серверів і сайтів.

2) Шпигунське ПЗ (spyware), яке спрямоване на порушення конфіденційності жертви і отримання особистої інформації про користувача – логіни, паролі, інформація про кредитні картки тощо. Якщо уражений комп'ютер був всередині

корпоративної мережі, зловмисники мають змогу до отримання корпоративної і конфіденційної інформації.

3) Спам поштою (E-mail spam), що зазвичай містить рекламу, спірні посилання на сайти з зловмисним ПЗ, або саме зловмисне ПЗ, яке може бути сховане всередині файлів.

4) Шахрайство з спливаючими вікнами (click fraud), яке може з'являтися від певних подій, наприклад під час перегляду веб-сторінок. Небезпечними ці вікна є тому, що вони є докучливими і створюють веб-трафік для персонального вигоди або комерційного прибутку.

5) Рекламне шахрайство (ad fraud) – використання ботів для просування популярності (отримання більшої кількості послідовників, підписників, переглядів тощо), для збільшення кількості кліків, які отримує реклама, що дозволяє сайтам отримувати більші виплати від рекламодавців.

6) Атаки типу підмішування облікових даних (credential stuffing attacks) використовують ботів для входу в облікові записи користувачів за допомогою вкрадених паролів, як, наприклад, під час атаки на General Motors у 2022 році.

7) Ботнет може використовуватися для видобутку біткоїнів (bitcoin mining), що несе отримання прибутку для оператора ботнету.

8) Саморозповсюдження (self-spreading) – можлива атака, яка проводиться з метою отримання більшої кількості інфікованих пристроїв у ботнет. Зазвичай ця функція автоматизована у ботнетах.

Згідно з чинним законодавством України однією з причин захисту користувачів є порушення прав і законів на отримання достовірної і повної інформації.

Закон України «Про інформацію» [24]: «Стаття 2. Основні принципи інформаційних відносин

1. Основними принципами інформаційних відносин є:

гарантованість права на інформацію;

відкритість, доступність інформації, свобода обміну інформацією;

достовірність і повнота інформації;

свобода вираження поглядів і переконань;

правомірність одержання, використання, поширення, зберігання та захисту інформації;

захищеність особи від втручання в її особисте та сімейне життя.»

Закон України «Про інформацію» [24]: «Стаття 28. Неприпустимість зловживання правом на інформацію

1. Інформація не може бути використана для закликів до повалення конституційного ладу, порушення територіальної цілісності України, пропаганди війни, насильства, жорстокості, розпалювання міжетнічної, расової, релігійної ворожнечі, вчинення терористичних актів, посягання на права і свободи людини».

Закон України «Про основні засади забезпечення кібербезпеки України» [25]: «Стаття 1. Визначення термінів.

6) кіберзагроза - наявні та потенційно можливі явища і чинники, що створюють небезпеку життєво важливим національним інтересам України у кіберпросторі, справляють негативний вплив на стан кібербезпеки держави, кібербезпеку та кіберзахист її об'єктів;

Одним із найбільш виражених прикладів гібридної війни для українського суспільства були анексія Криму у 2014 році і так звана «Спеціальна військова операція» з боку РФ. Обидві події несли у собі інформаційну складову, яку більш правильно називати «не інформаційну», так як використані міри в інформаційному просторі були направлені на викривлення реальної інформації і дезінформацію.

Як зазначається у дослідженнях [26-27] на тимчасово окупованих територіях України було відмічено широкий спектр технологій маніпулювання масовою свідомістю: інформаційна блокада, використання «лідерів думок», методи упереджального удару, метод зворотного зв'язку – проведення штучно інсценованих масових акцій, метод емоційного резонансу, психологічного шоку, рейтингування застосовувалося для виправдання агресивних дій. Внаслідок містифікації і активної міфологізації, фільтрації «необхідного» контенту і нав'язування російської ідеології у інформаційному просторі була нанесена

психологічно-смислова шкода та створена дезорганізація частини українського суспільства.

Окреме дослідження соціальної мережі Facebook [28] підтвердило факт використання російською військовою розвідкою «фейкових» акаунтів для ідентифікації публічних персон з України і сусідніх держав. «Деякі з них видавали себе за громадських журналістів і намагалися зв'язатися з політиками, журналістами та іншими громадськими діячами.»

Стаття від журналу «Washingtone post» [29] показала, що боти здатні збільшувати активність кандидата на виборах тапокращувати його упізнання, що у свою чергу впливає на їх результати. «Подібним чином у 2008 році зі збільшенням суми на рекламу, на яку більше витратився Барак Обама ніж Джона, зросла й віра виборців у те, що обрання республіканця від Арізони означає третій термін Буша – висновок який збільшив ймовірність голосування за Обаму».

1.4. Проблемні аспекти виявлення і ідентифікації соціальних ботів

Після того як було визначено потенційну небезпеку від ботів необхідно проаналізувати існуючі методи виявлення і протидії зловмисним ботам.

Захист відбувається від сторони соціальної мережі з боку можливості використання скарг від звичайних користувачів, а також від впливу деяких подій на репутацію соцмережі(президентські вибори, політичний скандал). Соціальна платформа Twitter має окрему власну команду, яка забезпечує цілісність платформи, та окрему команду криміналістів, які визначають приналежність акаунту:

- до такого, яким управляє людина, але він поводить себе як бот;
- до такого, яким управляє автоматизоване програмне забезпечення або бот;
- до такого, яким управляє людина. [30]

Під час розгляду особливостей, які відрізняють ботів від людей, у науковій статті [31] виділяють наступні характеристикиі:

– Невідповідність у профілі між іменем і статтю, відсутність зображення профілю чи місцезнаходження.

– Співвідношення соціальних звязків. Боти, як правило, стежать за іншими ботами, а люди – за іншими людьми.

– Дія корельованими у часі способами, що створює підпис, який можна розпізнати.

– Більшість розглянутих досліджень використовують Twitter як джерело даних. Twitter API дозволяє дослідникам надзвичайно легко збирати всі необхідні функції, однак це створює упередженість щодо дослідження інших платформ.

Можна виділити два основних підходи до виявлення ботів соціальних мереж: на основі обробки матеріалів, що публікуються користувачами і на основі обробки кількісно-якісних характеристик самих акаунтів. Можна виділити такі недоліки подібних методів: відсутність достатньо повного набору даних для перевірки якості виявлення ботів соціальних мереж; мовна обмеженість застосованих методів.

Наукова стаття [32] пропонує наступну таксономію систем виявлення соціальних ботів:

– на основі інформації з соціальних мереж;

– на основі краудсорсингу (crowd-sourcing) і використання людського інтелекту;

– методи машинного навчання на основі ідентифікації певних атрибутів, що розкривають ботів серед людей.

Використання людського інтелекту і краудсорсингу має свої переваги і недоліки. Так як краудсорсинг передбачає отримання роботи або інформації від великої групи людей, які можуть виконувати завдання платно і добровільно, то даний метод має брак достатньої довіри до «середньостатистичного» працівника. Такі команди все ще потребують своїх експертів для точного виявлення підроблених облікових записів. У результаті, щоб надійно побудувати правду про визначених ботів, великі соціальні мережі, такі як Facebook і Twitter, змушені наймати команди експертів-аналітиків. Подібні рішення можуть бути нерентабельним для платформи з великою базою користувачів, а також викликати

проблему порушення конфіденційності інформації про користувачів, яка передається працівникам.

Аналіз дослідження [33] виявив певні специфічні поведінкові групи облікових записів і виділив три окремі групи ботів: спамери, рекламодавці та облікові записи, які публікують вміст із підключених програм (наприклад новини, інформація з веб-сайтів).

1.5. Висновок і постановка задачі

Було вирішено проаналізувати існуючі методи захисту від зловмисних соціальних ботів, особливості викривлення інформації ботами у соціальних мережах і скласти перелік методів, які можна використати, щоб запобігти інформаційно-психологічній шкоді для користувачів.

РОЗДІЛ 2. СПЕЦІАЛЬНА ЧАСТИНА

2.1. Сучасні методи аналізу даних з метою виявлення ботів

Одним із вирішальних питань під час дослідження ботів є правильне розділення користувачів на тих, які контролюються людиною, і тих, що контролює бот. Приналежність акаунту до автоматизованого можна вимірювати за допомогою низки факторів: активність акаунту, співвідношення кількості підписок і підписників, кількість повідомлень/твітів (публікацій у соціальній мережі Twitter), кількість хештегів(hashtag), кількість відповідей на повідомлення/твіти інших користувачів, час життя акаунту, приватність акаунту, співвідношення активності акаунту до активності інших користувачів, тощо.

Як зазначається у науковій статті [34]: для взаємодії соціальним ботам необхідне технічна інфраструктура:

– профіль у соціальній мережі;

– технічна підготовка акаунту для автоматизації через інтерфейс програмного застосунку(API), або відповідні механізми взаємодії з веб-сайтом, або зовнішній інтерфейс(front-end);

– поведінковий алгоритм.

Наукова стаття [35] показало, що окрім того, що боти були задіяні у політичних програмах, було визначено наступні особливості проаналізованого ботнету: маскування своєї автоматизованої особистості, мімікрія під поведінку звичайних користувачів, використання хештегів для просування звичайним користувачам і реверберацийне поширення вибраних постів і повідомлень.

Дослідження і аналіз даних від соціальних ботів ведеться:

– за допомогою Natural Language Processing (NLP) – методів обробки природної мови;

– за допомогою різноманітних алгоритмів нейронних мереж: випадковий ліс, метод опорних векторів (Support Vector Machines), метод k-найближчих сусідів (k-nearest neighbors);

– за допомогою математичних або статистичних методів, алгоритмів класифікації - Наївний баєсів класифікатор, асоціативні правила (Rule Learner), дерево ухвалення рішень, класифікатор J48, алгоритм JRip (Repeated Incremental Pruning to Produce Error Reduction).

– за допомогою семантичного аналізу тексту повідомлень;

– а також за допомогою краудсорсингу і використання людського інтелекту.

Слід зазначити можливість визначення поведінки звичайних ботів за рахунок високої синхонності дій. Як зазначається у дослідженні [36]: «Ключовим спостереженням є те, що люди не можуть бути синхронними з високою точністю протягом тривалого часу; таким чином, облікові записи користувачів з великим рівнем синхронизації, швидше за все, є ботами.»

Розглянуті методи на основі графів, такі як у роботах [37-38], загалом перевершують традиційні підходи, які не розглядають сферу соціальної мережі Twitter як графи та мережі відношень. Такі результати демонструють важливість моделювання топологічної структури Twitter для виявлення ботів. Дослідники наукової статті [37] пропонують використання перших фреймворків для виявлення ботів з урахуванням гетерогенності відношень між користувачами.

2.1.1. Natural Language Processing

Під час розгляду існуючих робіт, головними питаннями, які вирішують за допомогою NLP постають: пошук закономірностей між текстами; отримання семантичної інформації з текстів; визначення зв’язків між словами всередині тексту.

Семантичний аналіз тексту – алгоритм автоматичного розпізнавання тексту, який виділяє семантичні відношення між словами і формує семантичне представлення цих зв’язків у вигляді графів. У лінгвістиці семантика вивчає смислове навантаження слів, фраз і пропозицій на різних рівнях.

Семантична мережа — графічна система позначень для подання знань у шаблонах пов'язаних вузлів і дуг. Семантичній мережа — це орієнтований граф, вершини якого — поняття, а дуги — відношення між ними. Приклад орієнтованого графа і відношень між його вершинами зображено на рисунку 2.1.

Вершини графа об'єднані відношеннями, які розглядаються як гіпероніми і гіпоніми.

Гіперонім— слово з ширшим значенням, яке виражає загальне, родове поняття, назва класу (множини) предметів (властивостей, ознак)».

Гіпонім — слово з вужчим значенням, яке називає предмет (властивість, ознака) як елемент класу (множини). Якщо розглядати гіпоніми одного рівня, тоді по відношенню один до одного вони – еквоніми.

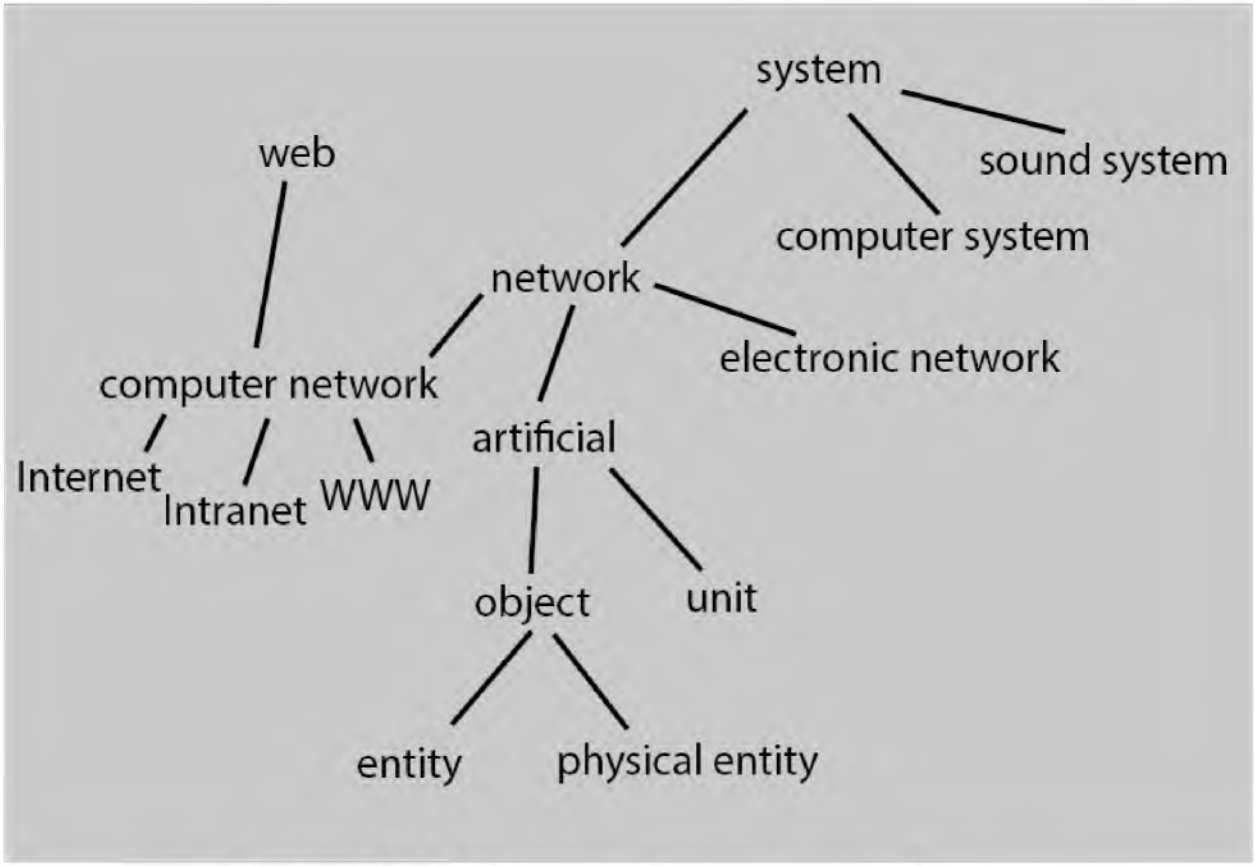

Рисунок 2.1 – Приклад орієнтованого графа

Семантичний аналіз використовується як частина комплексного аналізу акаунтів у соціальних мережах. Дослідження [39] використовує Наївний Байєсовий класифікатор для відбору необхідних ознак – текстові (співвідношення хештегів,

посилань, посилань на користувачів, ретвітів, текстових емоцій, емодзі, звуконаслідувань, абревіатур, алітерацій), на основі вмісту (співвідношення довжини твітів, однакових твітів, лексичне багатство), лексичні(лексикон людино-машинний, жіночо-чоловічий, лексикон настрою/почуттів людино-машини, лексикон настроїв/почуттів жіночо-чоловічий), та констатує, що використання мовних особливостей твітів і лексична інформація покращує показники зібраних базових особливостей.

2.1.2. Алгоритми нейронних мереж

Після згадування про перше створення обчислювальної моделі на основі нейронної активності у статті [40] було створено достатню базу штучного інтелекту для використання її у наукових і професійних цілях. Під час вивчення великих масивів даних перевагу у дослідженнях набувають методи машинного навчання завдяки їх обчислювальній швидкості, можливість навчання на вибраних вибірках і можливістю налаштування чутливості входів у моделях.

Під час використання нейронних мереж існують два види їх навчання: машинне навчання і глибоке навчання. Головною різницею для цих видів є вид представлення оброблюваних даних. Алгоритми машинного навчання як правило потребують структурованих даних, які заздалегідь визначені за деякими параметрами. Після навчання нейронної мережі на таких вибірках вони використовуються для отримання нових даних, хоча іноді виникає потреба у повторному навчанні або перенавчанні моделі. Алгоритми глибокого навчання у свою чергу використовують штучні нейронні мережі, які не потребують класифікацію даних, а з використанням багаторівневих шарів створюють власні ієрархії, на яких відбувається навчання.

Одним із підходів під час аналізу текстів користувачів у соціальних мережах, який використовувався у дослідженні [22], було використання архітектури рекурентної нейронної мережі (Recurrent Neural Network) з контекстною

довготривалою короткочасною пам'яттю (Contextual Long Short-Term Memory) для широкомасштабного акустичного моделювання [41]. Сехматичне зображення архітектури блоку LSTMP RNN зображено на рисунку 2.2.

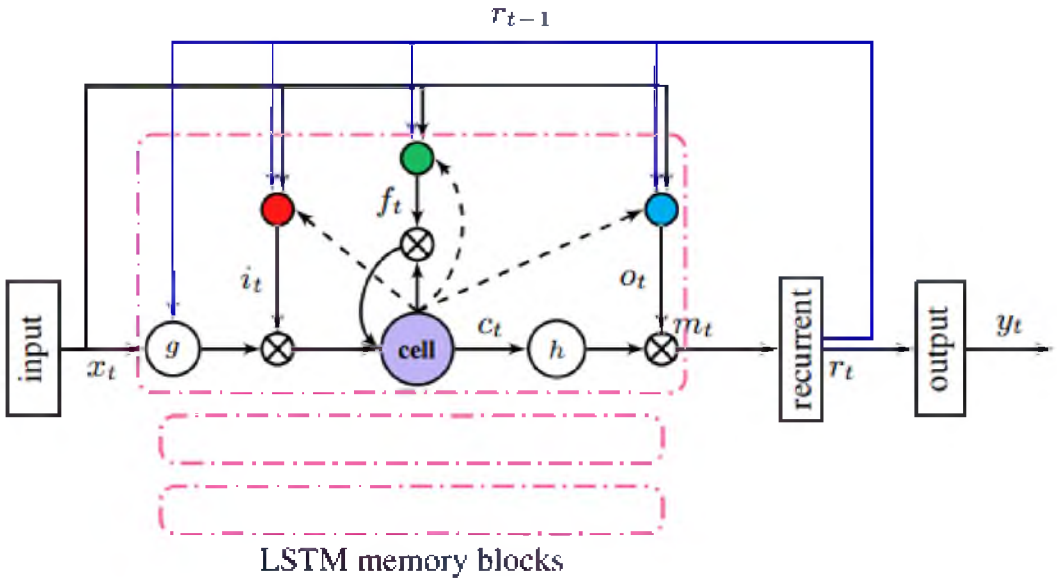

Рисунок 2.2 – Архітектура блоку LSTMP RNN

За допомогою методів застосунку програмного інтерфейсу (Application Program Interface) збирається інформація про акаунти користувачів соціальних мереж, а також інформація і зміст написаних повідомлень/твітів для подальшої обробки і аналізу. З усіх розглянутих архітектур, які можна побачита на рисунку 2.3, у роботі була використана контекстна архітектура LSTM з глибоким навчанням, яка використовувала текст публікацій і інформацію про метадані цих публікацій. Вибрана контекстна архітектура показала високий результат детекції ботів – 96% оцінки AUC (Area Under the Curve – площа під кривою ROC-кривої).

«Дерево ухвалення рішень» – метод класифікації, який має структуру подібну дереву і складається з «листя» і «гілок». На ребрах «гілки» дерева ухвалення рішення записані атрибути, від яких залежить цільова функція. В вузлах «листи» записані значення цільової функції або клас. Дерево ухвалення рішень не

вимагає складної підготовки даних, добре працює для числових і категоріальних змінних.

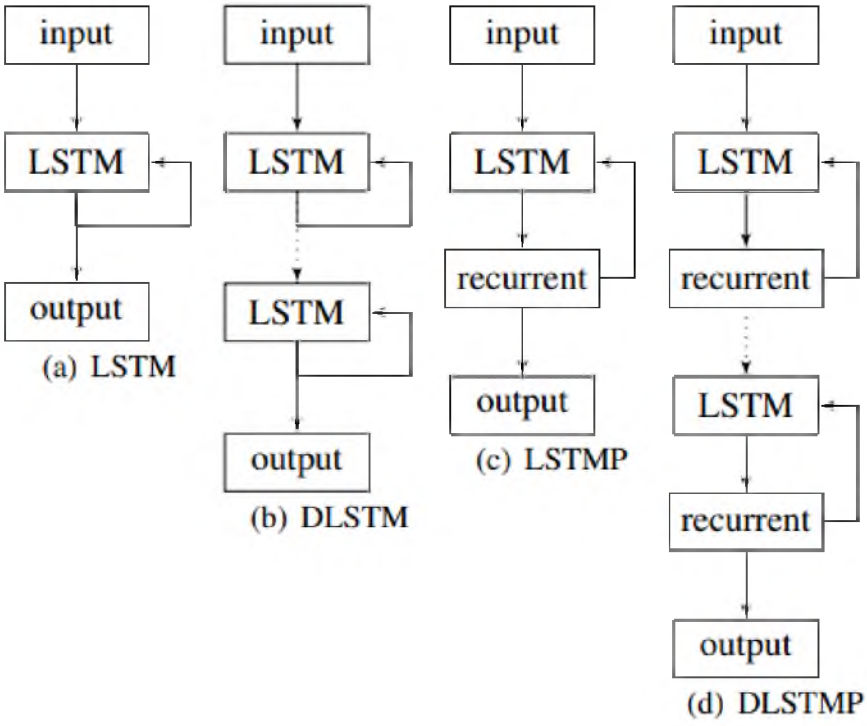

Рисунок 2.3 – Різноманітні архітектури LSTM RNN: Convensional LSTM, LSTM Deep, LSTM with Recurrent Projection Layer, Deep LSTMP.

«Випадковий ліс» (Random forest) – метод навчання класифікації, який працює за допомогою множини «дерев ухвалення рішень». Під час вирішення завдання класифікації рішення приймається голосуванням з більшості серед виданих відповідей.

1. Генерується випадкова підвибірка навчальної вибірки з повтореннями – по ній будується дерево (для кожного дерева своя підвибірка).

2. Для побудови кожного розщеплення у дереві проглядаються m (квадратний корінь із загальної кількості ознак) випадкових ознак (для кожного нового розщеплення свої випадкові ознаки).

3. Вибираються найкращі ознаки та розщеплення по ньому. Дерево будується до вичерпання вибірки (поки у листі не залишаться представники лише одного класу), якщо не задані обмеження (висота дерева, кількість об'єктів у листі та кількість об'єктів у підвиборці, при якому проводиться розщеплення).

2.1.3. Статистична класифікація, алгоритми класифікації

Алгоритми математичного розпізнавання образів використовуються для класифікації і виділення необхідних параметрів, які причетні до поведінки автоматизованої програми.

Кожен вектор ознак образів складається з компонентів-ознак, які характеризують невизначеність, яка у свою чергу може мати ймовірнісний характер. Багатовимірна ймовірнісна величина складається з векторів ознак образу, появу яких можна описати за допомогою закону розподілу ймовірностей, наприклад у формі густини розподілу ймовірностей. Вид і параметри функції щільності визначаються середовищем, в якому працює система розпізнавання образів.

Так як задачу розпізнання семантичного вмісту твітів або публікацій від ботів можна привести до задачі фільтрації вмісту. Автори статті [42] використали Байєсівський і Наївний Байєсівський класифікатори для фільтрування вмісту поштових повідомлень, схему яких можна побачити на рисунку 2.4. Байєсівський класифікатор передбачає до якого класу відноситься об’єкт за ймовірністю, наприклад ймовірність того, що даний зразок належить до певного класу. Наївний Байєсівський алгоритм припускає, що значення вплив атрибута на даний клас не залежить від значень інших атрибутів. Байєсовські класифікатори є популярними алгоритмами класифікації завдяки своїй простоті, обчислювальній ефективності та дуже хорошій продуктивності для реальних проблем. Іншою важливою перевагою також є те, що моделі Байєса швидко навчаються та оцінюються, і вони мають високу точність у багатьох областях.

Наукова стаття [43] використовує три ознаки на основі графів: кількість друзів, кількість підписників і співвідношення підписників. Очевидно, що якщо кількість підписників відносно невелика порівняно з кількістю людей, за якими стежить користувач, то коефіцієнт підписників відносно малий. У той же час ймовірність того, що пов'язаний обліковий запис є спамом визначається високим.

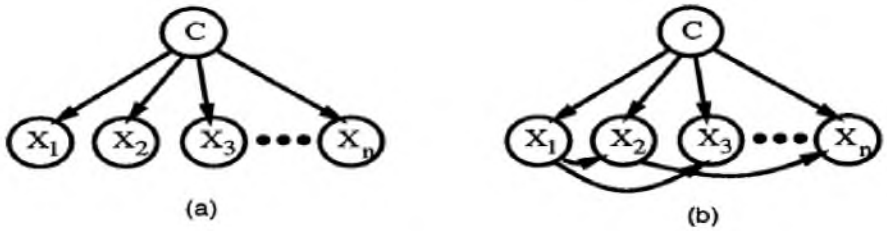

Рисунок 2.4 – Схема відношення характеристик у (а) Наївному Байесівському класифікаторі та (б) Байесівський класифікатор з відношеннями між особливостей

Наукова стаття [44] використовує класифікатор з двох шарів, щоб проаналізувати твіти на англійській та іспанській мовах. Схему класифікатора можна побачити на рисунку 2.5. Перший шар використовує метод опорних векторів (Support Vector Machines) з ядром радіальної базисної функції (RBF kernel) і екземпляром AdaBoost (Adaptive Boosting – алгоритм адаптивного прискорення), другий шар підсумовує прогнози попереднього шару за допомогою класифікатору м'якого голосування (Voting Classifier). Для іспаномовних твітів було використано один шар методу опорних векторів (SVM).

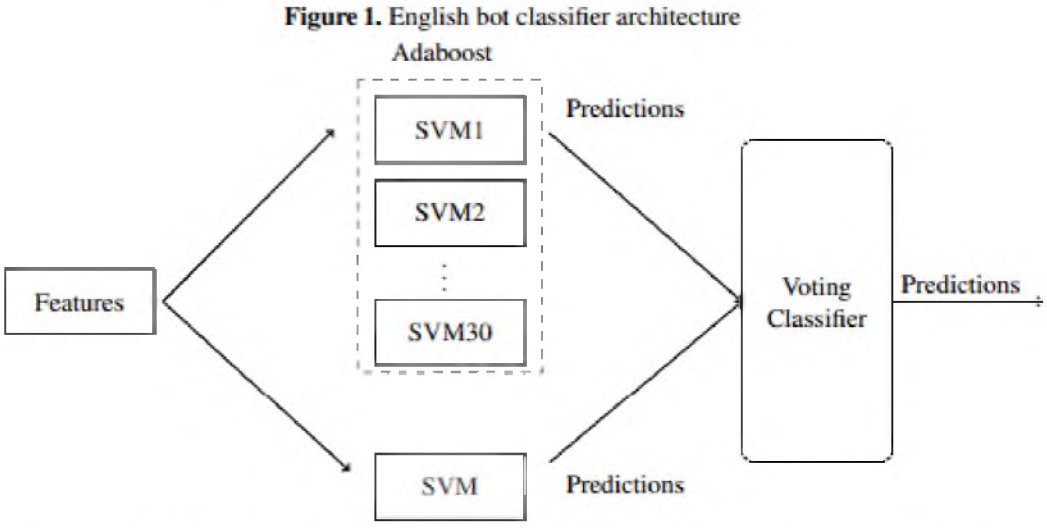

Рисунок 2.5 – Архітектура класифікатору англомовних ботів

Латентний семантичний аналіз (LSA) – це статистичний підхід до визначення зв'язків між словами за допомогою їхнього контексту використання в документах. LSA, починаючи з Tf-Idf (Term frequency – Inverse Document Frequency), застосовує декомпозицію сингулярного значення (Singular Value Decomposition) зі зменшеним рангом до матриці Tf-Idf, щоб зменшити кількість рядків, зберігаючи структуру подібності між стовпцями.

Під час аналізу ботів у цій статті були використані наступні особливості Твіттер акаунтів:

– Кількість смайликів,емодзі (emoji) у твіті.

– Кількість посилань у твіті.

– Кількість хештегів (hashtag) у твіті.

– Довжина твіту.

– Довжина ретвіту.

– Кількість крапок з комою.

– Оцінка косинусної подібності твітів. Ідея полягає в тому, що косинус-подібність ботів вища, ніж у людей.

– Аналіз настрою твіту. Проводиться аналіз настроїв для кожного твіту, використовується середня нейтральна оцінка настрою та середня складна оцінка.

– Спотворення тексту. Спотворення використовується, щоб підкреслити використання спеціальних символів і знаків пунктуації.

Як зазначає дослідження [45] Метод ЛСА пройшов випробування та підтвердив свою ефективність у таких напрямках обробки природної мови як моделювання концептуальних знань людини; інформаційний пошук, при реалізації якого ЛСА показує набагато кращі результати порівняно зі звичайними векторними методами.

2.1.4. Семантичний аналіз тексту

Семантичний аналіз тексту - аналіз одиниць лексичного рівня, тобто слів. До складу цього різновиду аналізу можуть входити різні словники, які містять лексичний репертуар та морфемну структуру лексичних одиниць, словозмінні парадигми тощо. Застосовують і так звані безсловникові методи, які передбачають здійснення розбиття слова на морфеми за заданими алгоритмами, забезпечення віднесення слів до словозмінних парадигматичних класів, виконання процедури лематизації (редукції текстових словоформ до початкових, вихідних форм тощо).

На цьому ж етапі частково здійснюється і граматичний аналіз, тобто ідентифікація лексико-граматичних класів та значень граматичних категорій текстових слів.

Наукова стаття [46] розглядає наступну проблему семантичного аналізу повідомлень шляхом використання ключових слів. Різноманітність тем не дозволяє провести повноцінний аналіз оцінок слів, так як терміни в одній ситуації можуть бути індикаторами спаму, в той час як для іншого контексту такі слова є необхідними.

У науковій статті [43] досліджувались функції на основі вмісту з трьох характеристик: кількість повторюваних твітів, кількість HTTP-посилань і кількість відповідей/згадок. «По-перше, обліковий запис може вважатися акаунтом що розсилає спам, якщо він публікує дубльований вміст в одному обліковому записі. Повторювані твіти виявляються шляхом вимірювання відстані Левенштейна (також відомої як відстань редагування) між двома різними твітами, опублікованими одним обліковим записом. По-друге, спам-боти намагаються публікувати шкідливі посилання у своїх твітах, щоб спонукати законних користувачів натиснути. Оскільки Twitter дозволяє публікувати повідомлення лише до 140 символів, деякі служби та програми для скорочення URL-адрес стають широко використовуваними і популярними. Скорочена URL-адреса приховує цільову адресу, і, як наслідок, сприяє спам-акаунтам у розіграшах, фішингу або приховуванні партнерів.»

Наукова стаття [44] використовує наступні семантичні особливості акаутнів:

– Оцінка косинусної подібності твітів. Під час розгляду косинусної подібності текстів твітів, кожне слово пов’язане з окремим виміром, а твіти характеризуються вектором, де значення кожного виміру відповідає кількості разівпояви слова у твіті. Тоді косинус подібності дає корисну оцінку того, наскільки подібні два твіти у словах теми. Ідея полягає в тому, що косинус-подібність ботів вища між ботами, ніж між людьми.

– Аналіз настрою твіту. Проводиться аналіз настроїв для кожного твіту, використовується середня нейтральна оцінка настрою та середня складна оцінка.

– Спотворення тексту. Спотворення використовується, щоб підкреслити використання спеціальних символів і знаків пунктуації.

2.2. Існуючі програмні рішення для визначення ботів у соціальних мережах

BOTORNOT (Botometer) [47] – це алгоритм машинного навчання, навчений обчислювати оцінку, де низькі оцінки вказують на ймовірні облікові записи людей, а високі – на облікові записи ботів. Щоб підрахувати оцінку, Botometer порівнює обліковий запис із десятками тисяч позначених прикладів. Під час перевірки облікового запису дані понад тисячі ознак передаються в API Botometer: профіль облікового запису, кількість друзів, структури соціальної мережі, час активності, мови, настроїв публікацій тощо. Зібрані характеристики профілю використовуються моделями машинного навчання для обчислення балів ботів.

Були розглянуті наукові статті [48, 49]. Botometer використовує навчання з вчителем для класифікатора. Можливе використання навчання без вчителя, однак але вони дозволяють лише виявити специфічну, заздалегідь визначену поведінку. Тому вони не підійшли для створення загального інструменту виявлення.

Botometer розглядає понад 1000 функцій, які можна класифікувати на шість класів: профіль користувача, друзі, мережа, час, контент і мова, а також настрій твіту. Таким чином обліковий запис може бути представлений вектором номерів функцій, так як особливості акаунту кодуються як числа, що дозволяє класифікаторам машинного навчання обробляти інформацію.

BotHunter [50] алгоритм захисту мережі, розроблений, щоб виявити запущене в системі зловмисне програмне забезпечення, орієнтоване на координацію. Він базується на алгоритмі кореляції мережевого діалогу (network dialog correlation), розробленому в рамках дослідницької програми Cyber-TA Лабораторією комп'ютерних наук SRI International. Це ПЗ відстежує двосторонні потоки зв'язку між хостами внутрішньої мережі та Інтернетом. Він класифікує обмін даними

мережі, як потенційний життєвий цикл поточного зараження шкідливим програмним забезпеченням. BotHunter використовує Snort [51] як механізм генерації діалогових подій, який було модифіковано та налаштовано для проведення процесу класифікації діалогів. Події діалогув мережі надходять безпосередньо в окремий механізм кореляції діалогів, де BotHunter зіставляє шаблони створення діалогів кожного хоста з абстрактною моделлю життєвого циклу зараження шкідливим програмним забезпеченням. Після отримання достатньої кількості доказів, щоб визнати хост інфікованим, BotHunter створює профіль зараження.

2.3. Порівняльна характеристика методів захисту

На основі досліджених методів та лгоритмів було проведено аналіз їх результативності, який можна побачити у таблиці 2.1.

Таблиця 2.1 – Порівняння запропонованих методів пошуку соціальних ботів

№	Наукова робота	Використані методи	Результа-тивність	Мови
1	2	3	4	5
1	«Bot and gender detection of Twitter accounts using distortion and LSA» [44]	Латентний семантичний аналіз, метод опорних векторів	0.94	Англійська
			0.90	Іспанська
2	«Naive-Bayesian classification for bot detection in Twitter» [39]	Байесівський класифікатор	0.81	Англійська
			0.88	Іспанська
3	«Fake news detection in social media» [52]	Наївний Байєсівськи класифікатор, метод опорних векторів	Не визначена	Англійська
4	«Bot and gender identification in Twitter using word and character N-grams» [53]	Метод головних компонент, використання N-грам	0.92	Англійська
			0.91	Іспанська

5	«Twitter bots and gender detection using tf-idf»[54]	Term frequency-Inverse document frequency	0.91	Англійська
6	«Identifying Twitter bots using a convolutional neural network» [55]	Згорткова нейронна мережа (Convolutional Neural Network)	0.9	Англійська
7	«Deep Neural Networks for Bot Detection» [22]	AdaBoost Classifier (Metadata-only + SMOTENN)	0.92	Англійскька
8		Global Vectors for Word Representation(тільки для Twitter) (200D GloVE), Contextual Long Short Term Memory модель	0.96	Англійскька
9		Random Forest Classifier (Metadata-only)	0.8	Англійскька

Під час вибору класифікаторів слід привернути увагу на Random Forest і AdaBoost, які постійно забезпечували найкращу продуктивність у всіх тестах виявлення ботів на рівні твітів.

Якщо порівнювати рекурентні семантичні нейронні мережі і ймовірнісні нейронні мережі, рекурентні семантичні нейронні мережі мають більшу продуктивність і потужність у задачах класифікації та кластеризації зразків текстів. Крім цього, практична реалізація рекурентних семантичних мереж не завжди можлива через використання значних обчислювальних ресурсів.

Для навчання классифікаторів для соціальної мережі Twitter можливе використання зразків з бази даних «TwiBot-20 dataset» [56].

2.4. Висновки

За результатами аналізу було виділено і рекомандовано використання наступних методів:

1) Класифікатор AdaBoost з використанням техніки передискретизації синтетичної меншості (SMOTE) і з покращенням даних через – редаговані найближчі сусіди (ENN) і Tomek Links.

2) Використання латентного семантичного аналізу і метод опорних векторів(SVM).

3) Рекурентна нейронна мережа з контекстною довготривалою короткочасною пам’яттю (RNN Contextual LSTM).

До недостатків слід визначити обмеженість мови і відсутність тестування точності алгоритмів на інших мовах.

Для створення запиту на вилучення проаналізованних акаунтів зловмисних соціальних ботів в мережі Twitter необхідно буде створити запит на форумі для обговорень питань безпеки і аутентифікації [57] з зазначенням виконаного аналізу і підстав у вигляді «порушення Політики маніпуляції платформою і спамом». Рекомендується використати автоматизовану форму для скарг [58]. Політика по відношенню маніпуляції платформою і спамом «забороняє штучно роздмухувати листування або заважати їх ходу шляхом використання кількох облікових записів або координації своїх дій з іншими людьми з метою порушення Правил Twitter» [59].

Якщо аналіз проводиться на інших платформах передачі інформації, необхідно створити запит на корпоративну пошту організації або запит у службу підтримки з метою вилучення зловмисних облікових записів і повідомлень.

Для визначення доцільності запровадження запропонованих рішень необхідно проаналізувати їх економічні показники.

РОЗДІЛ 3. ЕКОНОМІЧНИЙ РОЗДІЛ.

3.1 Розрахунок (фінансових) капітальних витрат

Капітальні витрати будуть включати в себе наступні витрати:

– вартість розробки проекту інформаційної безпеки;

– витрати на залучення зовнішніх консультантів;

– вартість створення основного і додаткового ПЗ;

– витрати на первісні закупівлі апаратного забезпечення;

– витрати на інтеграцію системи;

– витрати на навчання фахівців інформаційної безпеки і обслуговуючого персоналу.

3.1.1 Визначення трудомісткості розробки та опрацювання ПЗ

Трудомісткість визначається тривалістю кожної робочої операції згідно формули (3.1).

$$t = t_{\text{тз}} + t_{\text{в}} + t_{\text{а}} + t_{\text{пр}} + t_{\text{опр}} + t_{\text{д}}, \text{ годин}, \quad (3.1)$$

де $t_{\text{тз}}$ – тривалість складання технічного завдання на розробку ПЗ;

$t_{\text{тз}} = 25$ годин;

$t_{\text{в}}$ – тривалість визначення ТЗ;

$t_{\text{а}}$ – тривалість визначення блок-схеми алгоритму;

$t_{\text{пр}}$ – тривалість програмування за готовою блок-схемою;

$t_{\text{опр}}$ – тривалість опрацювання програми на ПК;

$t_{\text{д}}$ – тривалість підготовки технічної документації на ПЗ;

Умовна кількість операторів у програмі визначається за формулою (3.2).

$$Q = q \cdot c(1 + p)\text{ , штук,} \qquad (3.2)$$

де q – очікувана кількість операторів;

q = 202;

c – коефіцієнт складності програми;

c = 1.4;

p – коефіцієнт корекції програми в процесі її опрацювання;

p = 0.1.

$$Q = 202 \cdot 1.4(1 + 0{,}1) = \ 311.08 \text{ штук,}$$

Тривалість визначення технічного завдання оцінюється за формулою (3.3).

$$t_{\text{в}} = \frac{Q \cdot B}{(75 \ldots 85) \cdot k}\text{ , годин,} \qquad (3.3)$$

де B – коефіцієнт збільшення тривалості етапу внаслідок недостатнього опису завдання;

B = 1.3;

k – коефіцієнт, що враховує кваліфікацію програміста і визначається стажем роботи за фахом;

k = 1.

$$t_{\text{в}} = \frac{311.08 \cdot 1.3}{(80) \cdot 1} = 5.05 \text{ годин,}$$

Тривалість розробки блок-схеми алгоритму визначається за формулою (3.4).

$$t_{\text{а}} = \frac{Q}{(20 \ldots 25) \cdot k}\text{ , годин,} \qquad (3.4)$$

$$t_{\text{а}} = \frac{311.08}{(22) \cdot 1} = 14.14\text{ , годин,}$$

Тривалість складання програми за готовою блок-схемою визначається за формулою (3.5).

$$t_{\text{пр}} = \frac{Q}{(20...25)\cdot k}\text{, годин,} \qquad (3.5)$$

$$t_{\text{пр}} = \frac{311.08}{(22)\cdot 1} = 14.14\text{ , годин,}$$

Тривалість опрацювання програми на ПК визначається за формулою (3.6).

$$t_{\text{опр}} = \frac{1{,}5\cdot Q}{(4...5)\cdot k}\text{, годин,} \qquad (3.6)$$

$$t_{\text{опр}} = \frac{1.5\cdot 311.08}{(4.5)\cdot 1} = 103.69 \text{ годин,}$$

Тривалість підготовки технічної документації визначається за формулою (3.7).

$$t_{\text{д}} = \frac{Q}{(15...20)\cdot k} + \frac{Q}{(15...20)}\cdot 0.75\text{ , годин,} \qquad (3.7)$$

$$t_{\text{д}} = \frac{311.08}{(17)\cdot 1} + \frac{311.08}{(17)}\cdot 0.75 = 32.02 \text{ годин,}$$

$$t = 25 + 5.05 + 14.14 + 14.14 + 103.69 + 32.02 = 194.04 \text{ годин}$$

3.1.2 Розрахунок витрат на створення програмного продукту

Витрати на створення програмного продукту $К_{пз}$ визначаються за формулою (3.8):

$$К_{пз} = З_{зп} + З_{мч}\text{ , грн,} \qquad (3.8)$$

де $З_{зп}$ – заробітна плата виконавця програмного забезпечення, грн/год;

$З_{мч}$ – вартість машинного часу, що необхідний для розробки програмного забезпечення, год.

Заробітна плата виконавця враховує основну і додаткову ЗП, а також соціальні потреби і визначається за формулою (3.9).

$$З_{зп} = t \cdot З_{пр}\ , \text{грн}, \qquad (3.9)$$

де t – загальна тривалість створення ПЗ, годин;

t = 194.04 год;

$З_{пр}$ – середньогодинна заробітна плата програміста з нарахуванням, грн/годину.

Згідно [60] заробітна плата програміста 24316 грн. $З_{пр}$ = 24316/40 = 607.9 грн/годину.

$$З_{зп} = 194.04 \cdot 607.9 = 117957 \text{ грн.}$$

За формулою (3.10) визначимо вартість машинного часу для налагодження програми на ПК.

$$З_{мч} = t_{опр} \cdot С_{мч} + t_{д}\ , \text{грн}, \qquad (3.10)$$

де $t_{опр}$ – трудомісткість налагодження програми на ПК, годин;

$t_{опр}$ = 103.69 год;

$t_д$ – трудомісткість підготовки документації на ПК, годин;

$t_д$ = 32.02 год;

$С_{мч}$ – вартість 1 години машинного часу ПК, грн/година.

Вартість 1 години машинного часу визначається за формулою (3.11).

$$C_{мч} = P \cdot t_{нал} \cdot C_e + \frac{Ф_{зал} \cdot Н_а}{F_р} + \frac{К_{лпз} \cdot Н_{апз}}{F_р}\ , \text{грн}, \qquad (3.11)$$

де Р – встановлена потужність ПК, кВт;

Р = 0,4 кВт;

$t_{нал}$ – кількість задіяних робочих станцій;

$t_{нал} = 2$;

C_e – тариф на електричну енергію, грн/кВт*година;

$C_e = 1.68$ грн./(кВ*год);

$Ф_{зал}$ – залишкова вартість ПК на поточний рік, грн.;

$Н_а$ – річна норма амортизації на ПК, частки одиниці;

$$Н_а = 1/5 = 0.2$$

$Н_{апз}$ – річна норма амортизації на ліцензійне програмне забезпечення, частки одиниці;

$$Н_{апз} = 1/2 = 0.5$$

$К_{лпз}$ – вартість ліцензійного програмного забезпечення, грн.;

$F_р$ –річний фонд робочого часу (за 40-годинного робочого тижня $F_р = 1920$).

Вартість ПК = 25520 грн., мінімальний термін корисної служби = 60 місяців.

Накопичена амортизація = (25520 * 40)/60 = 17013 грн.

Залишкова вартість $Ф_{зал}$ = 25520 – 17013 = 8507 грн.

В табл. 3.2 визначена вартість закупівель ліцензійного програмного забезпечення.

$К_{лпз} = 0$ грн.

$$C_{мч} = 0.4 * 2 * 1.68 + \frac{8507 \cdot 0.2}{1920} + \frac{0 \cdot 0.5}{1920} = 2.23 \text{ грн./година}$$

$$З_{мч} = 194.04 \cdot 2.23 + 32.02 = 464.73 \text{ грн.}$$

$$К_{пз} = 117957 + 464.73 = 118422 \text{ грн.}$$

3.1.3 Капітальні (фіксовані) витрати на створення системи інформаційної безпеки

На проектування і впровадження проектних рішень кваліфікаційної роботи вирахуємо капітальні витрати, формула (3.12).

$$К = К_{рп} + К_{зпз} + К_{пз} + К_{аз} + К_{навч} + К_{н}\text{ , грн,} \qquad (3.12)$$

де $К_{рп}$ – вартість розробки проекту системи інформаційної безпеки та залучення для цього зовнішніх консультантів, грн.;

$К_{рп}$ = 18206 грн.;

$К_{зпз}$ – вартість закупівель ліцензійного основного та додаткового ПЗ, грн.;

$К_{зпз}$ = 0 грн.;

$К_{пз}$ – вартість розробки системи інформаційної безпеки, грн.;

$К_{пз}$ = 118422 грн.;

$К_{аз}$ – вартість закупівлі апаратного забезпечення та допоміжних матеріалів, грн.;

$К_{аз}$ = 25520 грн.;

$К_{навч}$ – витрати на навчання технічних фахівців і обслуговуючого персоналу, грн.;

$К_{навч}$ = 20000 грн.;

$К_{н}$ – витрати на встановлення обладнання та налагодження системи інформаційної безпеки, грн;

$К_{н}$ = 10000 грн.

$$К = 18206 + 0 + 118422 + 25520 + 20000 + 10000 = 192148 \text{ грн.}$$

3.2 Розрахунок експлуатаційних витрат

Для визначення витрат на функціонування запропонованої системи, формула (3.13), порахуємо витрати на оновлення і модернізацію, витрати на керування, формула (3.14), та витрати від активності користувачів.

$$С = С_{в} + С_{к} + С_{ак} + С_{пн}, \text{ грн.} \quad (3.13)$$

де $С_{в}$ – витрати на оновлення й модернізацію системи інформаційної безпеки;

Так як в вартість ліцензії програмного забезпечення входить постійне підтримання і оновлення до нових версій – витрати на оновлення і модернізацію не виникають.

$С_в$ = 20000 грн.;

$С_к$ – витрати на керування системи інформаційної безпеки;

$С_{ак}$– витрати викликані активністю користувачів системи;

Для використання системи інформаційної безпеки на ІТС необхідно 3 рази за рік провести тренінги обслуговуючого персоналу за 8000 грн.

$С_{ак}$ = 24000 грн;

$С_{пн}$ – витрати на перенавчання витрати машинного часу і на персонал.

Витрати на керування системи інформаційної безпеки визначаються за формулою (3.14).

$$С_к = С_н + С_а + С_з + С_{єв} + С_{ел} + С_о + С_{тос}, \text{ грн.}, \quad (3.14)$$

де $С_н$ – витрати на навчання адміністративного персоналу й кінцевих користувачів;

$С_н = 2 \cdot 12000 = 24000$ грн.;

$С_а$ – річний фонд амортизаційних відрахувань;

$С_а$ = 0/2 = 0 грн.;

$С_з$ – річний фонд заробітної плати інженерно-технічного персоналу на год на ставку;

$С_{єв}$ – витрати єдиного внеску на загальнообов’язкове соціальне страхування;

$С_{ел}$ – вартість електроенергії, що споживається апаратурою системи інформаційної безпеки протягом року;

$С_о$ – витрати на залучення сторонніх організацій для виконання деяких видів обслуговування, навчання та сертифікацію обслуговуючого персоналу;

$С_о$ = 0 грн;

$С_{тос}$ – витрати на технічне й організаційне адміністрування та сервіс системи інформаційної безпеки, грн.

$$\text{Стос} = 192148 \cdot 0.03 = 5764.44 \text{ грн.};$$

$$С_з = З_{осн} + З_{дод}, \text{ грн.}, \quad (3.15)$$

де $З_{осн}$ – основна заробітна плата, грн./рік;

$З_{дод}$ – додаткова заробітна плата, грн./рік.

Так як виконання робіт вимагає залучення спеціаліста з інформаційної можна брати спеціаліста на 0.5 ставки.

$$\text{Зосн} = 24316 \cdot 0.5 \cdot 12 = 145896 \text{ грн.};$$

$$\text{Здод} = 0{,}1 \cdot \text{Зосн} = 14589.6 \text{ грн.};$$

$$\text{Сз} = 145896 + 14589.6 = 160485.6 \text{ грн.}$$

$$\text{Сєв} = 160485.6 \cdot 0{,}22 = 35306.8 \text{ грн.}$$

Вартість споживання електроенергії апаратурою визначається за формулою (3.16).

$$С_{ел} = P \cdot F_р \cdot Ц_е, \text{ грн.}, \quad (3.16)$$

де Р – встановлена потужність апаратури системи інформаційної безпеки;

Р = 0,4 кВт.

$F_р$ – річний фонд робочого часу системи інформаційної безпеки;

$F_р$ = 1920 год.

$Ц_е$ – тариф на електроенергію;

$Ц_е$ = 1,68 грн./кВт за годину.

$$С_{ел} = 1.3 \cdot 1920 \cdot 1.68 = 4193.3 \text{ грн.}$$

$$С_к = 24000 + 20000 + 160485.6 + 35306.8 + 4193.3 + 0 + 2305.5 = 246291.2 \text{ грн.}$$

Так як використання нейронних мереж може потребувати перенавчання моделей, необхідно вирахувати додаткове споживання електроенергії. Додаткові витрати на персонал були включені в $С_в$.

Припустимо, що для перенавчання необхідно 30% від часу використання системи інформаційної безпеки.

$$С_{пн} = 1.3 \cdot 1920 \cdot 0.3 \cdot 1.68 = 1258 \text{ грн.}$$

$$C = 20000 + 246291.2 + 24000 + 1258 = 291550 \text{ грн.}$$

3.3 Оцінка величини збитку

Під час аналізу економічного збитку, яку зазнає Україна внаслідок атак від соціальних ботів, слід брати за увагу соціально-психологічний вплив і соціально-економічний збиток внаслідок втрати частин територій, які були підпорядковані Україні. Втрата підприємств зазнає значної соціальної шкоди, так як втрачаються робочі місця і рівень життя через це, економічні втрати – втрата прибутку який могло принести таке підприємство, демографічні – втрата можливостей населення перебувати на територіях без заробітної плати і як наслідок імміграція та еміграція, а також репутаційні, так як на міжнародному ринку компанії і інвестори не зацікавлені вести справи у зоні бойових дій.

Візьмемо як приклад «Авдіївський коксохімічний завод», на якому перебувало 3973 працівників.

Як зазначає проект «Економічна правда»: «Понад 95% його продукції йшли на внутрішній ринок. Зокрема, завод забезпечував коксом ММК ім. Ілліча та "Азовсталь". 13 березня підприємство зазнало масованого обстрілу. Снаряди потрапили в склад, коксові цехи, смолоперегінний та вуглепідготовчий цехи, зупинилася робота ТЕЦ.» За інформацією [61] близько 3тис. тон доменного коксу вироблялось підприємством щорічно.

Упущена вигода від простою атакованого вузла або сегмента корпоративної мережі визначається за формулою (3.17):

$$U = П_{п} + П_{в} + V, \text{ грн,} \qquad (3.17)$$

де $П_п$ – оплачувані втрати робочого часу та простої співробітників атакованого вузла або сегмента корпоративної мережі, грн.;

$П_в$ – вартість відновлення працездатності вузла або сегмента корпоративної мережі (переустановлення системи, зміна конфігурації та ін.), грн.;

V – втрати від зниження обсягу продажів за час простою атакованого вузла або сегмента корпоративної мережі, грн.

Втрати робочого часу і простою співробітників визначаються за формулою (3.18):

$$П_п = \frac{\sum З_c}{F} t_п, \text{ грн.}, \quad (3.18)$$

де $З_с$ – заробітна плата співробітників атакованого вузла або сегмента корпоративної мережі, грн. за місяць;

F – місячний фонд робочого часу (при 40-а годинному робочому тижні становить 176 годин), годин;

$t_п$ – час простою вузла або сегмента корпоративної мережі внаслідок атаки, години;

$t_п$ = 2080 години.

Середня заробітна плата співробітників згідно [61] – 115 тис. грн на рік.

$\sum З_c = 115000 \cdot 3973 = 456895$ тис. грн.;

$П_п = \frac{456895000}{176} \cdot 2080 = 5399668181$ грн. = 5400 млн.грн.

Знайдемо витрати від зниження обсягу продажів за час простою згідно формули (3.19):

$$V = \frac{О}{F_г} \cdot t_п, \text{ грн.}, \quad (3.19)$$

де $F_г$ – річний фонд часу роботи організації (52 робочих тижнів, 5-ти денний робочий тиждень, 8-ми годинний робочий день);

$F_г$ = 2080 годин;

О – обсяг продажів атакованого вузла або сегмента корпоративної мережі, грн. у рік;

О = 737.5 млн. дол. у рік = 27119.46 млн.грн. у рік. за нинішнім курсом НБУ(1 дол. = 36,77 грн.)

$$V = \frac{27119.5}{2080} \cdot (2080) = 27120 \text{ млн. грн.};$$

$$U = 27120 \text{ млн. грн.} + 5400 \text{ млн. грн.} = 32520 \text{ млн. грн.}$$

Величина загального збитку повинна бути скорегована від соціально-психологічного, демографічного і репутаційного впливу збитку.

3.4 Загальний ефект від впровадження системи інформаційної безпеки

З урахуванням ризиків порушення інформаційної безпеки можна визначити загальний ефект від впровадження системи інформаційної безпеки за формулою (3.20).

$$E = U \cdot R - C \tag{3.20}$$

де U – загальний збиток, грн;

R – очікувана ймовірність атаки на вузол або сегмент корпоративної мережі, частки одиниці;

R = 0.01%;

С – щорічні витрати на експлуатацію системи інформаційної безпеки, грн;

$$E = 32520000000 * 0.0001 - 291550 = 2{,}960{,}450 \text{ грн.}$$

3.5 Визначення та аналіз показників економічно ефективності системи

Коефіцієнт повернення інвестицій ROSI показує, скільки запобігає підприємство можливих втрат від атаки на сегмент корпоративної мережі від

впровадження запропонованого рішення. Визначимо коефіцієнт ROSI за формулою (3.21):

$$ROSI = \frac{E}{K}, \text{частки одиниць}, \quad (3.21)$$

де Е – загальний ефект від впровадження системи інформаційної безпеки, грн.;

К – капітальні інвестиції за варіантами, що забезпечили цей ефект, грн.

$ROSI_1 = \frac{2{,}960{,}450}{192148} = 15.4$;

Проект побудови системи інформаційної безпеки вважається доцільним за умови перевищення бажаного значення показника ефективності $E_н$, формула (3.22).

$$ROSI > E_н, \quad (3.22)$$

Для підприємства інформаційна безпека здійснюється за рахунок позикових коштів, тобто банківського кредиту. В такому разі показник ефективності визначається за формулою (3.23):

$$E_н = \frac{N_{кр} + N_{інф}}{100}, \quad (3.23)$$

де $N_{кр}$ – банківська кредитна ставка, %;

$N_{кр} = 0.24$ %;

$N_{інф}$ – річний рівень інфляції, %;

$N_{інф} = 100.7$ %.

$E_н = \frac{0.24+100.7}{100} = 1$;

Визначимо термін окупності капітальних інвестицій за рахунок впровадження системи інформаційної безпеки, формула (3.24).

$$T_0 = \frac{K}{E} = \frac{1}{ROSI}, \text{ років.} \quad (3.24)$$

$$T_0 = \frac{1}{15.4} = 0.065 \text{ роки},$$

Термін окупності ПЗ T_0 = 0.78 місяць, в той час як без його використання Україна зазнає значних збитків 32520 млн. грн.

3.6 Висновки до економічного розділу

В економічному розділі було проаналізовано основні економічні показники для впровадження системи інформаційної безпеки і визначено, що вона є економічно доцільною. Було визначені наступні показники:

- капітальні витрати становлять 192148 грн.;
- експлуатаційні витрати становлять 291550 грн.;
- загальний збиток від атаки складає 32520 млн. грн.грн.;
- ефект від впровадження системи інформаційної безпеки становить 2960450 грн;
- термін окупності капітальних інвестицій складає 0.78 місяця.

ВИСНОВКИ

У першому розділі кваліфікаційної роботи був проаналізований сучасний контекст України, за висновками якого була визначена необхідність у дослідженні і боротьбі з соціальними ботами.

В рамках другого розділу був проведений аналіз сучасних робіт з детекції і класифікації ботів, були зіставлені та серед представлених методів були запропоновані найбільш ефективні з зазначенням їх недоліків.

Під час виконання економічного розділу був розрахований варіант створення системи інформаційної безпеки з визначених методів, економічні показники якого складали: капітальні витрати – 192148 грн., експлуатаційні витрати – 291550 грн. В результаті аналізу потенційних загроз загальний збиток внаслідок виникнення інциденту інформаційної безпеки склав щонайменше 32520 млн. грн. для визначеного підприємства. Загальний ефект від впровадження системи безпеки становить 2960450 грн. з терміном окупності капітальних інвестицій в 0.78 місяця.

Для повної боротьби с соціальними ботами необхідно подальший аналіз сфери соціальних мереж з використанням актуальних на території України мов – української та російської.

ПЕРЕЛІК ПОСИЛАНЬ

1. Наукова стаття «Weaponized Health Communication: Twitter Bots and Russian Trolls Amplify the Vaccine Debate» [Електронний ресурс] – Режим доступу до ресурсу: https://ajph.aphapublications.org/doi/full/10.2105/AJPH.2018.304567

2. Підручник «НАРИС ТЕОРІЇ І ПРАКТИКИ ІНФОРМАЦІЙНО–ПСИХОЛОГІЧНИХ ОПЕРАЦІЙ» Військовий інститут Київського національного університету імені Тараса Шевченка [Електронний ресурс] – Режим доступу до ресурсу: https://shron1.chtyvo.org.ua/Balabin_Viktor/Narys_teorii_i_praktyky_informatsiino-psykholohichniikh_operatsii.pdf?PHPSESSID=s9ka1relfhilruftpuuretjuc0

3. Науковий журнал «What is Hybrid Warfare?» Author(s): Erik Reichborn-Kjennerud and Patrick Cullen [Електронний ресурс] – Режим доступу до ресурсу: https://www.jstor.org/stable/pdf/resrep07978.pdf?refreqid=excelsior%3Ad1143b6f018200a6fc577868e4b44596&ab_segments=&origin=&acceptTC=1

4. Науковий журнал «Making Sense of Hybrid Warfare» [Електронний ресурс] – Режим доступу до ресурсу: https://www.jstor.org/stable/26326441#metadata_info_tab_contents

5. Дослідження «Reuters Institute Digital News Report 2017» [Електронний ресурс] – Режим доступу до ресурсу: https://reutersinstitute.politics.ox.ac.uk/sites/default/files/Digital News Report 2017 web_0.pdf

6. Наукова стаття «Tweets in Time of Conflict: A Public Dataset Tracking the Twitter Discourse on the War Between Ukraine and Russia» [Електронний ресурс] – Режим доступу до ресурсу: https://arxiv.org/pdf/2203.07488.pdf

7. Щорічне опитування USAID-Internews «Ставлення населення до ЗМІ та споживання різних типів медіа у 2020 р.» [Електронний ресурс] – Режим доступу до ресурсу: https://internews.in.ua/wp-content/uploads/2020/10/2020-Media-Consumption-Survey-FULL-FIN-Ukr-1.pdf

8. Дослідження «ОПОРИ» від «Київського міжнародного інституту соціології» [Електронний ресурс] – Режим доступу до ресурсу:

https://oporaua.org/report/polit_ad/24068-mediaspozhivannia-ukrayintsiv-v-umovakh-povnomasshtabnoyi-viini-opituvannia-opori

9. Наукова стаття «Bots and online hate during the COVID-19 pandemic: case studies in the United States and the Philippines» [Електронний ресурс] – Режим доступу до ресурсу: https://link.springer.com/article/10.1007/s42001-020-00087-4

10. Наукова стаття «DISINFORMATION AND SOCIAL BOT OPERATIONS IN THE RUN UP TO THE 2017 FRENCH PRESIDENTIAL ELECTION» [Електронний ресурс] – Режим доступу до ресурсу: https://arxiv.org/ftp/arxiv/papers/1707/1707.00086.pdf

11. Наукова стаття «Rumors, False Flags, and Digital Vigilantes: Misinformation on Twitter after the 2013 Boston Marathon Bombing» [Електронний ресурс] – Режим доступу до ресурсу: https://faculty.washington.edu/kstarbi/Starbird_iConference2014-final.pdf

12. Дослідження «Alternative Narratives of Crisis Events: Communities and Social Botnets Engaged on Social Media» [Електронний ресурс] – Режим доступу до ресурсу: https://dl.acm.org/doi/10.1145/3022198.3026307

13. Наукова стаття «Limited individual attention and online virality of low-quality information» [Електронний ресурс] – Режим доступу до ресурсу: https://arxiv.org/abs/1701.02694

14. Наукова стаття «The spread of misinformation by social bots» [Електронний ресурс] – Режим доступу до ресурсу: http://cs.furman.edu/~tallen/csc271/source/viralBot.pdf

15. Наукова стаття «Examining Emergent Communities and Social Bots Within the Polarized Online Vaccination Debate in Twitter» [Електронний ресурс] – Режим доступу до ресурсу: https://journals.sagepub.com/doi/full/10.1177/2056305119865465

16. Чат-бот «Eliza» [Електронний ресурс] – Режим доступу до ресурсу: https://web.njit.edu/~ronkowit/eliza.html

17. Чат-бот «Amazon's Alexa» [Електронний ресурс] – Режим доступу до ресурсу: https://alexa.amazon.com/

18. Чат-бот «Apple's Siri» [Електронний ресурс] – Режим доступу до ресурсу: https://www.apple.com/siri/

19. Чат-бот «Google Assistant» [Електронний ресурс] – Режим доступу до ресурсу: https://assistant.google.com/intl/en_uk/

20. Наукова стаття «СФЕРИ ЗАСТОСУВАННЯ ЧАТ-БОТІВ» Трофименко О. Г. [Електронний ресурс] – Режим доступу до ресурсу: http://dspace.onua.edu.ua/handle/11300/18203?locale-attribute=uk

21. Дослідження «Fake News Detection on Social Media: A Data Mining Perspective» [Електронний ресурс] – Режим доступу до ресурсу: https://arxiv.org/pdf/1708.01967.pdf

22. Науковій стаття «Deep neural networks for bot detection» [Електронний ресурс] – Режим доступу до ресурсу: https://www.sciencedirect.com/science/article/abs/pii/S0020025518306248

23. Наукова стаття «Weaponized Health Communication: Twitter Bots and Russian Trolls Amplify the Vaccine Debate» [Електронний ресурс] – Режим доступу до ресурсу: https://ajph.aphapublications.org/doi/10.2105/AJPH.2018.304567

24. Про інформацію: Закон України від 02.10.1992 р. № 2657-XII. Дата оновлення від 20.11.2022 р. [Електронний ресурс] – Режим доступу до ресурсу: https://zakon.rada.gov.ua/laws/show/2657-12#Text

25. Закон України «Про основні засади забезпечення кібербезпеки України» від 05.10.2017 р. № 2163-VII. Дата оновлення від 17.08.2022 р. [Електронний ресурс] – Режим доступу до ресурсу: https://zakon.rada.gov.ua/laws/show/2163-19#Text

26. Наукова стаття «ГІБРИДНА ВІЙНА В УМОВАХ ІНФОРМАЦІЙНОГО СУСПІЛЬСТВА» [Електронний ресурс] – Режим доступу до ресурсу: https://www.researchgate.net/profile/Roman-Dodonov/publication/341822503_Postmodern_Reception_of_the_Postmodernism_Methodology_in_the_Mass_Consciousness/links/5ed62bb892851c9c5e7264ab/Postmodern-Reception-of-the-Postmodernism-Methodology-in-the-Mass-Consciousness.pdf#page=89

27. Наукова стаття «ДОСЛІДЖЕННЯ МЕТОДІВ АНТИУКРАЇНСЬКОЇ РОСІЙСЬКОЇ ПРОПАГАНДИ В ІНФОРМАЦІЙНИХ ВІЙНАХ ПРОТИ УКРАЇНИ» [Електронний ресурс] – Режим доступу до ресурсу: https://jrnl.nau.edu.ua/index.php/UV/article/view/16894

28. Дослідження соціальної мережі Facebook (p.19-20) «Threat Report The State of Influence Operations 2017-2020» [Електронний ресурс] – Режим доступу до ресурсу: https://about.fb.com/wp-content/uploads/2021/05/IO-Threat-Report-May-20-2021.pdf

29. Стаття «Could Russian trolls have helped elect Donald Trump?» від журналу Washingtone post [Електронний ресурс] – Режим доступу до ресурсу: https://www.washingtonpost.com/news/posteverything/wp/2017/11/10/could-russian-trolls-have-helped-elect-donald-trump/

30. Публікація від команди захисту Twitter [Електронний ресурс] – Режим доступу до ресурсу: https://blog.twitter.com/common-thread/en/topics/stories/2021/four-truths-about-bots

31. Наукова стаття «Trends in Detection and Characterization of Propaganda Bots» [Електронний ресурс] – Режим доступу до ресурсу: https://scholarspace.manoa.hawaii.edu/server/api/core/bitstreams/5e607673-f560-4f14-8df6-3abef74c6980/content#page=6&zoom=100,65,166

32. Наукова стаття «The Rise of Social Bots» [Електронний ресурс] – Режим доступу до ресурсу: https://arxiv.org/pdf/1407.5225.pdf

33. Дослідження «Online Human-Bot Interactions: Detection, Estimation, and Characterization» [Електронний ресурс] – Режим доступу до ресурсу: https://arxiv.org/pdf/1703.03107.pdf

34. Наукова стаття «Demystifying Social Bots: On the Intelligence of Automated Social Media Actors» [Електронний ресурс] – Режим доступу до ресурсу: https://journals.sagepub.com/doi/full/10.1177/2056305120939264

35. Наукова стаття «Are social bots on Twitter political actors? Empirical evidence from Ukrainian social botnet» [Електронний ресурс] – Режим доступу до

ресурсу: https://www.aaai.org/ocs/index.php/ICWSM/ICWSM16/paper/view/13015/12793

36. Дослідження «DeBot: Twitter Bot Detection via Warped Correlation» [Електронний ресурс] – Режим доступу до ресурсу: https://www.researchgate.net/profile/Abdullah-Mueen/publication/308021270_DeBot_Twitter_Bot_Detection_via_Warped_Correlation/links/59dc10f1a6fdcc1ec89fad11/DeBot-Twitter-Bot-Detection-via-Warped-Correlation.pdf

37. Наукова стаття «Heterogeneity-Aware Twitter Bot Detection with Relational Graph Transformers» [Електронний ресурс] – Режим доступу до ресурсу: https://arxiv.org/pdf/2109.02927.pdf

38. Наукова стаття «BotRGCN: Twitter Bot Detection with Relational Graph Convolutional Networks» [Електронний ресурс] – Режим доступу до ресурсу: https://arxiv.org/pdf/2106.13092.pdf

39. Дослідження «Naive-Bayesian Classification for Bot Detection in Twitter» [Електронний ресурс] – Режим доступу до ресурсу: https://pan.webis.de/downloads/publications/papers/gamallo_2019.pdf

40. Наукова стаття «A logical calculus of the ideas immanent in nervous activity» [Електронний ресурс] – Режим доступу до ресурсу: https://link.springer.com/article/10.1007/BF02478259

41. Наукова стаття «Contextual Long Short-Term Memory Recurrent Neural Network Architectures for Large Scale Acoustic Modeling» [Електронний ресурс] – Режим доступу до ресурсу: https://storage.googleapis.com/pub-tools-public-publication-data/pdf/43905.pdf

42. Наукова стаття «A Bayesian Approach to Filtering Junk E-Mail» [Електронний ресурс] – Режим доступу до ресурсу: https://www.aaai.org/Papers/Workshops/1998/WS-98-05/WS98-05-009.pdf

43. Наукова стаття «Detecting Spam Bots in Online Social Networking Sites: A Machine Learning Approach» [Електронний ресурс] – Режим доступу до ресурсу: https://link.springer.com/content/pdf/10.1007/978-3-642-13739-6_25.pdf

44. Наукова стаття «Bot and Gender Detection of Twitter Accounts Using Distortion and LSA (Latent Semantic Analysis)» [Електронний ресурс] – Режим доступу до ресурсу: https://ceur-ws.org/Vol-2380/paper_210.pdf

45. Дсолідження «Relevance of research of programs for semantic analysis of texts and review of methods of their realization» [Електронний ресурс] – Режим доступу до ресурсу: https://ceur-ws.org/Vol-2292/paper05.pdf

46. Наукова стаття «АЛГОРИТМІЧНА МОДЕЛЬ АСОЦІАТИВНО-СЕМАНТИЧНОГО КОНТЕКСТНОГО АНАЛІЗУ ТЕКСТІВ ПРИРОДНОЮ МОВОЮ» [Електронний ресурс] – Режим доступу до ресурсу: https://core.ac.uk/reader/38468700

47. Програмне забезпечення виявленні ботів «BOTORNOT» [Електронний ресурс] – Режим доступу до ресурсу: https://botometer.osome.iu.edu/

48. Наукова стаття «BotOrNot: A System to Evaluate Social Bots» [Електронний ресурс] – Режим доступу до ресурсу: https://arxiv.org/pdf/1602.00975.pdf

49. Наукова стаття «Botometer 101: social bot practicum for computational social scientists» [Електронний ресурс] – Режим доступу до ресурсу: https://link.springer.com/article/10.1007/s42001-022-00177-5

50. Алгоритм захисту мережі BotHunter. [Електронний ресурс] – Режим доступу до ресурсу: https://www.metaflows.com/features/bothunter/

51. Превентивна система вторгнень Snort. [Електронний ресурс] – Режим доступу до ресурсу: https://www.snort.org/

52. «Fake news detection in social media» https://www.csustan.edu/sites/default/files/groups/University%20Honors%20Program/Journals/02_stahl.pdf

53. Наукова стаття «Bot and gender identification in Twitter using word and character N-grams» [Електронний ресурс] – Режим доступу до ресурсу: http://ceur-ws.org/Vol-2380/paper_65.pdf DOI: 10.13140/RG.2.2.28481.71528.

54. Наукова стаття «Twitter bots and gender detection using tf-idf» [Електронний ресурс] – Режим доступу до ресурсу: http://ceur-ws.org/Vol-2380/paper_253.pdf

55. Наукова стаття «Identifying Twitter bots using a convolutional neural network» [Електронний ресурс] – Режим доступу до ресурсу: http://ceur-ws.org/Vol-2380/paper_227.pdf

56. Зразки навчання классифікаторів у Twitter [Електронний ресурс] – Режим доступу до ресурсу: https://github.com/bunsenfeng/twibot-20

57. Форум для обговорень питань безпеки соціальної мережі Twitter [Електронний ресурс] – Режим доступу до ресурсу: https://twittercommunity.com/c/oauth/12

58. Автоматизована форма для скарг на повідомлення соціальної мережі Twitter [Електронний ресурс] – Режим доступу до ресурсу: https://help.twitter.com/en/forms/authenticity/spam

59. Політика по відношенню маніпуляції платформою і спамом соціальної мережі Twitter» [Електронний ресурс] – Режим доступу до ресурсу: https://help.twitter.com/ru/rules-and-policies/platform-manipulation

60. Статистика заробітної плати спеціалістів сфери «Інформація і телекомунікація» в Україні [Електронний ресурс] – Режим доступу до ресурсу: https://ukrstat.gov.ua/operativ/operativ2005/gdn/Zarp_ek_p/Zp_ek_p_u/arh_zpp_u.htm

61. Центр експерти про промисловість Авдіївського КХЗ [Електронний ресурс] – Режим доступу до ресурсу: https://web.archive.org/web/20200219072904/https://gmk.center/ua/manufacturer/avdiivskij-khz/

ДОДАТОК А. Відомість матеріалів кваліфікаційної роботи

№	Формат	Найменування	Кількість листів	Примітки
1	2	3	4	5
Документація				
1	А4	Реферат	1	-
2	А4	Список умовних скорочень	1	-
3	А4	Зміст	2	-
4	А4	Вступ	1	-
5	А4	Розділ 1. Стан питання. Постановка задачі	12	-
6	А4	Розділ 2. Спеціальна частина	18	-
7	А4	Розділ 3. Економічна частина	12	-
8	А4	Висновки	1	-
9	А4	Перелік посилань	7	-
10	А3, А4	Додаток А. Відомість матеріалів кваліфікаційної роботи	1	-
11	А3	ДОДАТОК Б. Перелік документів на оптичному носії	1	-
12	А4	ДОДАТОК В. Відгук керівника економічної частини	1	-
13	А4	ДОДАТОК Г. Відгук керівника кваліфікаційної роботи	1	-

ДОДАТОК Б. Перелік документів на оптичному носії

1) Пояснювальна_записка_Павлов.docx
2) Пояснювальна_записка_Павлов.pdf
3) Презентація_Павлов.pptx

ДОДАТОК В. Відгук керівника економічної частини

Відгук керівника економічного розділу:

Керівник розділу

(підпис) (ініціали, прізвище)