УДК 004.3-185.4; 004.7-185.4, 330.4, 336;
336.01; 336.11; 336.741.28; 336.7
УКПП
№ Державної реєстрації 0121U109559
Інв. №

Міністерство освіти і науки України
Сумський державний університет (СумДУ)
40007, м. Суми, вул. Р.-Корсакова, 2,
тел. (0542) 66-51-10, факс (0542) 33-40-49

ЗАТВЕРДЖУЮ
Проректор з наукової роботи
д-р фіз.-мат. наук, професор

_____А.М. Чорноус

ЗВІТ
ПРО НАУКОВО-ДОСЛІДНУ РОБОТУ
Національна безпека через конвергенцію систем фінансового моніторингу та кібербезпеки: інтелектуальне моделювання механізмів регулювання фінансового ринку

МОДЕРНІЗАЦІЯ ІНСТРУМЕНТАРІЮ ПРОТИДІЇ ЛЕГАЛІЗАЦІЇ КРИМІНАЛЬНИХ ДОХОДІВ ТА КІБЕРШАХРАЙСТВАМ

(проміжний)

Керівниця НДР
доцентка кафедри економічної кібернетики
д-р. екон. наук, доцентка Г.М. Яровенко

2022

Рукопис закінчений 20 грудня 2022 р.

Результати цієї роботи розглянуті науковою радою СумДУ, протокол від __.12.2022 р. № __

СПИСОК АВТОРІВ

Доцентка кафедри економічної кібернетики, д-рка екон. наук, доцентка (керівниця)	20.12.2022	Г.М. Яровенко (вступ, розділи 1, 3, підрозділи 2.2, 2.3, 4.1, висновки)
Відповідальна виконавиця Зав. кафедри економічної кібернетики, д-р екон. наук, професорка	20.12.2022	О.В. Кузьменко (підрозділи 4.1, 4.2)
Професор кафедри економічної кібернетики, д-р екон. наук, професор	20.12.2022	С.В. Лєонов (підрозділ 4.2)
Асистентка кафедри економічної кібернетики	20.12.2022	О.В. Колотіліна (підрозділ 3.3)
Асистент кафедри економічної кібернетики	20.12.2022	С.В. Миненко (підрозділ 2.1)
Аспірантка кафедри економічної кібернетики	20.12.2022	М.С. Рожкова (підрозділ 2.2)
Магістрантка кафедри економічної кібернетики	20.12.2022	В.В. Кобзенко (розділ 1)
Магістрантка кафедри економічної кібернетики	20.12.2022	А.О. Рапута (підрозділ 2.3)

РЕФЕРАТ

Звіт про НДР: 231 с., 158 рис., 14 табл., 43 формул, 140 джерел, 6 додатків.

ІНТЕЛЕКТУАЛЬНЕ МОДЕЛЮВАННЯ, КІБЕРБЕЗПЕКА, КІБЕРШАХРАЙСТВА, КОНВЕРГЕНЦІЯ, ЛЕГАЛІЗАЦІЯ, НАЦІОНАЛЬНА БЕЗПЕКА.

Об’єкт дослідження – система економічних відносин, що виникають між суб’єктами господарювання та регуляторами фінансового ринку, що виникають в процесі комплексного застосування засобів фінансового моніторингу та боротьби із кіберзлочинністю. Мета роботи – розвиток методології та міждисциплінарного методичного інструментарію протидії легалізації кримінальних доходів та кібершахрайствам, що дозволить напрацювати принципово нові, засновані на концептах поведінкової економіки та відокремлені від людського фактору, інтелектуальні алгоритмізовані регуляторні механізми, які уможливлять комплексне забезпечення економічної, фінансової та інформаційної складових національної безпеки держави, а також захисту прав споживачів фінансових послуг.

Методи дослідження: фундаментальні положення економічної теорії, макро- і мікроекономіки, теорії стратегічного управління, державного регулювання економіки, економіко-математичного моделювання. Інформаційно-фактологічна база дослідження: наукові праці вітчизняних та зарубіжних фахівців, Інтернет сайти з офіційно доступною статистикою макро-показників, показників щодо трендів кібератак.

В роботі проведено базовий, статистичний, кореляційний, декомпозиційний аналіз часових рядів; побудовано об’єднані регресії, регресії з детермінованими індивідуальними і випадковими ефектами, рекурентну нейронну мережу Long Short-Term Memory для прогнозування трьох видів кібератак; виконано аналіз поняття «протидія легалізації доходів, отриманих незаконним шляхом» в умовах діджиталізації суспільства; здійснено оцінку ризику конвергенції системи протидії відмивання грошей та кібербезпеки;

побудовано нейромережеву модель потенційної конвергенції системи кібербезпеки та протидії легалізації кримінальним доходам; сформовано кіберпрофілі жертв на основі гендерного аналізу; розроблено кіберпрофілі сучасних фінансових кіберзлочинів на основі кластерного аналізу; розроблено алгоритми розпізнавання поведінки кіберзлочинців на основі методів інтелектуального аналізу; розроблено моделі прогнозування кібершахрайських атак на комп'ютерні системи, мережеву та хмарну інфраструктуру фінансової установи; ударно-хвильові моделі впливу кібершахрайських атак на рівень фінансової безпеки.

ЗМІСТ

ВСТУП

Сьогодні однією із пріоритетних задач для українського суспільства є формування механізму захисту від внутрішніх загроз, потреба в якому загострюється в умовах ведення війни із зовнішнім ворогом. Ця необхідність виникає завдяки тому, що, з одного боку, наявність військових конфліктів в країні є одним з факторів підвищення привабливості країни для легалізації кримінальних доходів та фінансування тероризму. З іншого боку, зростають ризики кібершахрайських атак на різні об’єкти державної та недержавної інфраструктури. Зростання кібератак почало відбуватися ще до початку військової агресії з боку Росії. Так, 14.02.2022 було зафіксовано масштабну атаку на більше ніж 70 урядових сайтів [1], 15.02.2022 було атаковано банківські установи України [2]. За аналітичними даними, наданими DNS-платформою Quad9, виявлено, що у березні відбувалося значне зростання кібератак проти українців. З 121 мільйона зловмисних подій, що відбувалися у світі станом на 9 березня 2022 року, 4,6 мільйона пов’язані із Україною та Польщею, куди були переміщено 1,4 млн. українських громадян на початок березня 2022 року [3].

Окрім інформаційних атак українського населення здійснюються також й фінансові кібершахрайства. Так, 14.04.2022 було зафіксовано розповсюдження банківського трояна «IcedID» з метою збору персональних банківських даних українців. У квітні 2022 року було виявлено інтернет-шахрайство, що здійснювалося за допомогою фіктивної сторінки у соціальних мережах для збору фінансової допомоги з країн ЄС, що вимагало сплату платежів із порушенням конфіденційності даних платіжних карток [4].

Наведені приклади свідчать про те, що проблема протидії фінансовим і кібершахрайствам є актуальною і повинна вирішуватися на різних рівнях державного управління. Це вимагає системного підходу, який передбачає необхідність конвергенції систем протидії фінансовим і кібершахрайств, що можливо завдяки їх інформаційної, технічної, програмної та організаційної інтеграції як для держави в цілому, так й для окремих суб’єктів господарювання.

На необхідності процесів конвергенції для сфер відмивання незаконних доходів та кібершахрайств наголошує Office of Law Enforcement Support Financial Crimes Enforcement Network [5], що повинно відбуватися на рівні відповідних підрозділів протидії легалізації кримінальних доходів, фінансування тероризму, кібербезпеки та безпосередньо самого бізнесу. Також дану проблематику висвітлюють у своїх звітах світові консалтингові компанії Deloitte та PwC [6-7].

Але на даному етапі виникає потреба не тільки у усвідомленні необхідності таких інтеграційних процесів, але також у формуванні та модернізації комплексу інструментів протидії легалізації кримінальних доходів та кібершахрайствам. Саме цей аспект окресленої проблеми є важливим для реалізації в сучасних умовах, оскільки є необхідність у розвитку науково-методичних підходів щодо формування конвергованої системи фінансового моніторингу та кіберзахисту, яка б надавала можливість безперервно реагувати на щоденні фінансові та кіберзагрози. Найбільш ефективними інструментами є статистичні методи, які дозволяють проводити оцінку на основі великих масивів даних, з урахуванням часових, просторових або інших характеристик та факторних ознак. Серед них слід виділити інструменти Data Mining або інтелектуального аналізу даних, застосування яких сприяє формуванню якісних прогнозів, визначенню різного роду оцінок, моделюванню в умовах швидкої зміни даних, тощо. Даний етап науково-дослідної роботи як раз і буде присвячений модернізації інструментарію протидії легалізації кримінальних доходів та кібершахрайствам для забезпечення стійких конвергенційних процесів систем фінансового моніторингу та кібербезпеки.

Окреслена проблема дозволила обрати об’єкт та предмет дослідження. Об’єкт дослідження – система економічних відносин, що виникають між суб’єктами господарювання та регуляторами фінансового ринку, що виникають в процесі комплексного застосування засобів фінансового моніторингу та боротьби із кіберзлочинністю.

Предмет дослідження – методологічні засади та методичний формування комплексних, попереджувальних інтелектуальних механізмів регулювання

фінансового ринку, що сприятимуть підвищенню національної безпеки в умовах цифровізації фінансового простору.

Відповідно до об’єкта та предмета дослідження було сформовано мету. Так, метою дослідження є розвиток методології та міждисциплінарного методичного інструментарію протидії легалізації кримінальних доходів та кібершахрайствам, що дозволить напрацювати принципово нові, засновані на концептах поведінкової економіки та відокремлені від людського фактору, інтелектуальні алгоритмізовані регуляторні механізми, які уможливлять комплексне забезпечення економічної, фінансової та інформаційної складових національної безпеки держави, а також захисту прав споживачів фінансових послуг.

Для реалізації поставленої мети необхідно було вирішити наступні завдання:

- провести базовий, статистичний, кореляційний, декомпозиційний аналіз часових рядів;
- побудувати об’єднані регресії, регресії з детермінованими індивідуальними і випадковими ефектами для змінних «Mail Anti Virus», «Kaspersky Anti-Spam», «Intrusion Detection Scan»;
- спрогнозувати тренди трьох видів кібератак на основі рекурентної нейронної мережі Long short-term memory та об’єднані регресії;
- провести аналіз поняття «протидія легалізації доходів, отриманих незаконним шляхом» в умовах діджиталізації суспільства;
- здійснити оцінку ризику конвергенції системи протидії відмивання грошей та кібербезпеки;
- побудувати нейромережеву модель потенційної конвергенції системи кібербезпеки та протидії легалізації кримінальним доходам;
- сформувати кіберпрофілі жертв на основі гендерного аналізу;
- розробити кіберпрофілі сучасних фінансових кіберзлочинів на основі кластерного аналізу;

- розробити алгоритми розпізнавання поведінки кіберзлочинців на основі методів інтелектуального аналізу;
- розроблено моделі прогнозування кібершахрайських атак на комп'ютерні системи, мережеву та хмарну інфраструктуру фінансової установи;
- розробити ударно-хвильову модель впливу кібершахрайських атак на рівень фінансової безпеки.

Методи дослідження – фундаментальні положення економічної теорії, макро- і мікроекономіки, теорії стратегічного управління, державного регулювання економіки, економіко-математичного моделювання (статистичний аналіз; декомпозиційний аналіз часових рядів; регресійний аналіз панельних даних; канонічний аналіз; гендерний аналіз; методи інтелектуального аналізу даних: кластерний аналіз, класифікаційне дерево рішень, нейронні мережі; методи експоненційного згладжування; модель Сєдова-Тейлора). Розрахунки в роботі проводилися із використанням мови програмування Python та аналітичних пакетів STATISTICA, Excel, Deductor Academic

Інформаційно-фактологічну базу дослідження сформували наукові праці вітчизняних та зарубіжних фахівців, Інтернет сайти з офіційно доступною статистикою макро-показників, показників щодо трендів кібератак.

Отримані у роботі результати впроваджені у навчальний процес, а саме при викладанні дисциплін «Введення до бізнес-аналітики», «Прогнозування соціально-економічних процесів», «Моделювання економіки».

За результатами НДР опубліковано: 9 статей у журналах, що індексується у базі даних Scopus та WoS; 8 фахових статей у виданнях України категорії Б; 2 – розділи у монографіях; опубліковано 6 тез міжнародних конференцій, з яких 3 індексуються в базі даних Scopus, 2 – Index Copernicus; отримано 4 свідоцтва про реєстрацію авторського права на твір; дипломи I та III ступеня у Всеукраїнському конкурсі студентських наукових робіт за напрямками «Економічна аналітика і статистика» та «Економічна кібернетика».

Звіт виконано на основі публікацій виконавців, перелік яких надано у списку літератури.

1 СТРУКТУРНИЙ БАГАТОШАРОВИЙ АНАЛІЗ ДЖЕРЕЛ КІБЕРАТАК

1.1 Базовий, статистичний, кореляційний, декомпозиційний аналіз часових рядів

Зростання рівня інформатизації та комп’ютеризації багатьох сфер життєдіяльності суспільства призвело до появи та розповсюдження такого явища як кіберзлочинність. Кіберзлочином вважається дія особи чи групи осіб, направлена на незаконне отримання персональних даних іншої фізичної особи, суб’єктів господарювання, державних органів, або порушення функціонування їх програмних та технічних засобів. Як правило, даний вид злочину здійснюються за допомогою комп’ютерних засобів та технологій.

Найбільш розповсюдженим видом кіберзлочину є кібератаки, які провадяться хакерами для досягнення економічних, політичних та соціальних цілей особи, груп осіб або держави. У 2020 році вони посідали п’яте місце у світі серед таких видів ризиків, як геополітичні, економічні, соціальні та навколишнього середовища, що робить їх досить серйозною проблемою для суспільства [8]. Серед кібератак виділяються такі види, як фішинг, DoS-атаки, розповсюдження шкідливого програмного забезпечення, Man-in-the-Middle, Zero-day exploit атаки, міжсайтовий скриптинг, логічні бомби, тощо. Їх головними характеристиками є непередбачуваність, стрімкість здійснення, масове охоплення об’єктів, висока ймовірність досягнення цілей, що робить їх швидкою та небезпечною зброєю в руках злочинців.

Результатом кібератак, як правило, є витік або втрата інформації. Так, у 2022 році найбільші втрати від кіберзлочинів відбулися у сфері охорони здоров’я (10,10 млн. дол. США), фінансовій індустрії (5,97 млн. дол. США), фармацевтичній галузі (5,01 млн. дол. США), технологічній сфері (4,97 млн. дол. США), енергетиці (4,72 млн. дол. США) та інших [9]. Також прогнозується, що у 2025 році кіберзлочинність буде коштувати компаніям приблизно 10,5 трлн. дол. США, що перевищуватиме втрати у 3,5 рази в порівнянні з 2015 роком [10].

Також слід зазначити, що кількість кібератак невпинно зростає. Наприклад, кількість їх випадків в результаті пандемії COVID-19 зросла на 600% [11].

Таким чином, проблема кіберзлочинів в цілому та кібератак зокрема є досить актуальною, потребує пошуку різних інструментів і методів її дослідження та протидії. Для цього ефективними є не тільки технічні та програмні засоби але й управлінські інструменти, такі як прогнозування трендів. Процес прогнозування є складним і включає різні етапи реалізації, одним з яких є підбір даних, здійснення їх попереднього аналізу та підготовки до розробки ефективних прогнозних моделей. Реалізації даного етапу й буде присвячене це дослідження.

Питання виявлення та протидії кібератак є актуальним перед усім для науковців, які займаються питаннями кібербезпеки. Але сьогодні дана проблема набуває міждисциплінарного значення, оскільки її наслідки спостерігаються в економіці, бізнесі, суспільстві, політиці, охороні здоров'я та інше. Тому вчені з різних наукових шкіл та напрямків намагаються вирішувати її з різних точок зору.

Стейсі П., Тейлор Р., Спанакі К. досліджували психологічні аспекти впливу кібератак, а саме емоційні реакції персоналу компаній [12]. Шендлер Р. та Гомес М. А. виявили, що кібератаки є джерелом суспільного ризику, що проявляється у зростанні рівня суспільної недовіри до уряду у випадку кіберзагроз [13]. Лонсдейл Д. Дж. перевіряв кібератаки на предмет їх благ для суспільства, що визначалося з точки зору поваги до людини, соціального благополуччя, безпеки та миру, а також солідарності [14]. Болпагні М. запропонував зведений індекс для вимірювання кіберризику та оцінив вплив соціо-економічних факторів на його зміни [15]. Сімонс Г., Даник Ю., Малярчук Т. намагалися вирішити дилему, породжену суперечністю правового регулювання із політичною та оперативною необхідністю управління ситуаціями, пов'язаними із кіберзагрозами [16].

Вівер Г. А., Феддерсен Б., Марла Л., Вей Д., Роуз А., Ван Моер М. вивчали економічні наслідки кібератак на прикладі морської транспортної системи та

застосували оптимізаційний підхід для оцінки взаємодії між кібератаками та відповідними інформаційними технологіями компанії [17]. Лерой І. запропонувала застосовувати інструменти управління репутацією компанії для відновлення вартості її акцій після здійснення кібератак [18]. Акото У. дослідив позитивний вплив кібератак на торгівельні операції країни, що проявляється у використанні секретів, які добуваються в результаті кібершпигунства на користь держави [19]. Лаллі Г. С., Шеперд Л. А., Медсестра Дж. Р. К., Ерола А., Епіфаніу Г., Мейпл К., Беллекенс Х. проаналізували період пандемії COVID-19 з точки зору кібератак та виявили тенденцію їх щоденного зростання [20].

Не дивлячись на те, що проблема кібератак є досить актуальною та практично значущою, існує потреба у розробці прогнозних моделей, які дозволять виявляти потенційні кіберзагрози для певних країн та застосовувати контрзаходи щодо їх попередження.

Базою для розробки будь-якої прогнозної моделі є побудова її концептуальної моделі. Вона представляє собою зображення процесу моделювання, як сукупності етапів, починаючи з виявлення та аналізу вхідних даних, які відображають проблеми дійсності, та завершуючи розрахунком прогнозів за обраною моделлю та перевіркою їх якості. Для прогнозування інформаційних трендів кібератак пропонуємо наступну концептуальну модель (Рисунок 1.1).

Представлена на рисунку 1.1 концептуальна модель прогнозування трендів кібератак передбачає виконання двох процесів – попереднього аналізу і підготовки даних та прогнозування. Перший процес є необхідним для досліджень подібного характеру, оскільки він дозволяє сформувати такий набір даних, від якості якого залежатимуть подальші дії щодо створення прогнозної моделі та отримання адекватних та точних прогнозів. Другий процес передбачає вибір математичних моделей, які відповідатимуть результатам, отриманим після попереднього аналізу та підготовки даних. Дане дослідження буде охоплювати результати здійснення першого процесу, передбаченого концептуальною моделлю. Другий процес висвітлюватиметься у наступному дослідженні.

Рисунок 1.1 – Концептуальна модель прогнозування трендів кібератак

На першому етапі було сформовано набір змінних для розробки прогнозної моделі трендів кібератак. Вхідною інформацією було обрано статистичні дані 40 країн світу (по 10 країн з Європи, Азії, Африки та по 5 країн з Північної та Південної Америки) за період з 14 серпня 2022 року до 13 вересня 2022 року, узятих з відкритого доступу Лабораторії Касперського. Вони представляють собою щоденну статистику про кількість кібератак, виявлених за допомогою спеціальних інструментів їх протидії, а саме:

- MAV (Mail Anti Virus) – поштовий антивірус, який показує потік даних шкідливих програм, виявлених серед нових об'єктів у поштових додатках. Він перевіряє вхідні повідомлення та запускає автоматичну перевірку при збереженні вкладених файлів на диск;
- KAS (Kaspersky Anti-Spam) – Касперський Анти-Спам, який показує підозрілий та небажаний поштовий трафік, виявлений за допомогою технологій репутаційної фільтрації «Лабораторії Касперського»;
- IDS (Intrusion Detection Scan) – система виявлення вторгнень, яка показує потік даних з виявлених мережевих атак.

Обрані дані є панельними, оскільки містять інформацію про одну і ту ж множину об’єктів за ряд послідовних періодів часу. Тобто маємо одні й ті самі дані щодо трьох видів кібератак для сорока країн за 30 днів. Відповідно, для кожного спостереження будуть вимірюватися декілька параметрів (регресійні змінних або ефектів) за кожен період часу. Оскільки в даному випадку всі показники відстежуються протягом однакової кількості періодів часу, то така панель є збалансованою.

На наступному кроці необхідно провести первинний аналіз початкових даних та здійснити відповідні маніпуляції для його підготовки до безпосередньої побудови прогнозної моделі.

Розрахунки для даного дослідження проводилися із використанням мови програмування Python. Для цього було використано ряд стандартних бібліотек для аналізу, візуалізації і моделювання даних, таких як: Pandas, Numpy, Scipy.stats, Statsmodels, Matplotlib, Seaborn, Linearmodels та інші.

Спочатку була проведена перевірка набору даних щодо наявності пропущених значень за допомогою функції isna(). Дана процедура необхідна для виявлення відсутніх даних, що робить вибірку неоднорідною. Результат її проведення показав, що набір не має пропущених даних і не потребує додаткових маніпуляцій по їх відновленню чи заміні.

Далі була проведена оцінка базових статистик, результати якої представлені на рисунку 1.2.

index	MAV	KAS	IDS
count	1240.0	1240.0	1240.0
mean	4647.270967741935	7617245.080645162	152111.00161290323
std	7700.3844928245235	20092268.916028455	251532.5601747635
min	1.0	3500.0	2.0
25%	286.0	140375.0	8927.0
50%	1630.5	764000.0	46237.0
75%	5174.0	4785625.0	213360.75
max	77612.0	181005000.0	2643943.0

Рисунок 1.2 – Результати розрахунку базових статистик для початкових даних

На рисунку 1.2 можна побачити, що набір даних складається з 1240 спостережень. Значення середньоквадратичного відхилення по всім трьом видам кібератак є дуже високим і значно перевищує середнє значення ряду, що говорить пре неоднорідність даних. Це пов’язано із тим, що деякі країни, які увійшли у вибірку, є більш атакованими, ніж інші. Також мінімальні та максимальні значення для всіх трьох видів кібератак мають суттєвий розкид – мінімальне є дуже маленьким числом, що свідчить про відсутність кібератак в даний момент часу для певної країни, а максимальне – дуже великим числом, що свідчить про значну активність кібератак в певному регіоні. У випадку MAV та IDS кібератак їх середні значення відповідають третьому квартилю, а у випадку KAS атаки – четвертому квартилю. Це свідчить про те, що кількість спостережень, відповідних найбільш активним фазам кібератак, складає приблизно 20-30% від загальної кількості. Тобто вони носять періодичний характер, який ймовірно залежить від часового періоду та від самої країни, на яку спрямована кібератака.

Оскільки панельні дані являються часовим рядом, необхідно дослідити їх декомпозицію та перевірити на відповідність нормальному розподілу. На рисунку 1.3 представлено декомпозицію інформаційних трендів кібератак, яка включає побудову графіків фактичних даних, трендової, сезонної та залишкової компонент.

Декомпозиції, представлені на рисунку 1.3, побудовано за адитивною моделлю, оскільки обрані тренди відповідають саме адитивному процесу. Це підтверджує випадковий розподіл їх залишків, які коливаються біля нуля. Візуальний аналіз трендової складової свідчить про її відсутність, але в даному випадку потрібні додаткові перевірки на стаціонарність, для чого використано тест Дики-Фулера. Графіки, які відповідають сезонним складовим вказують на можливість наявності даної компоненти в досліджуваних часових рядах.

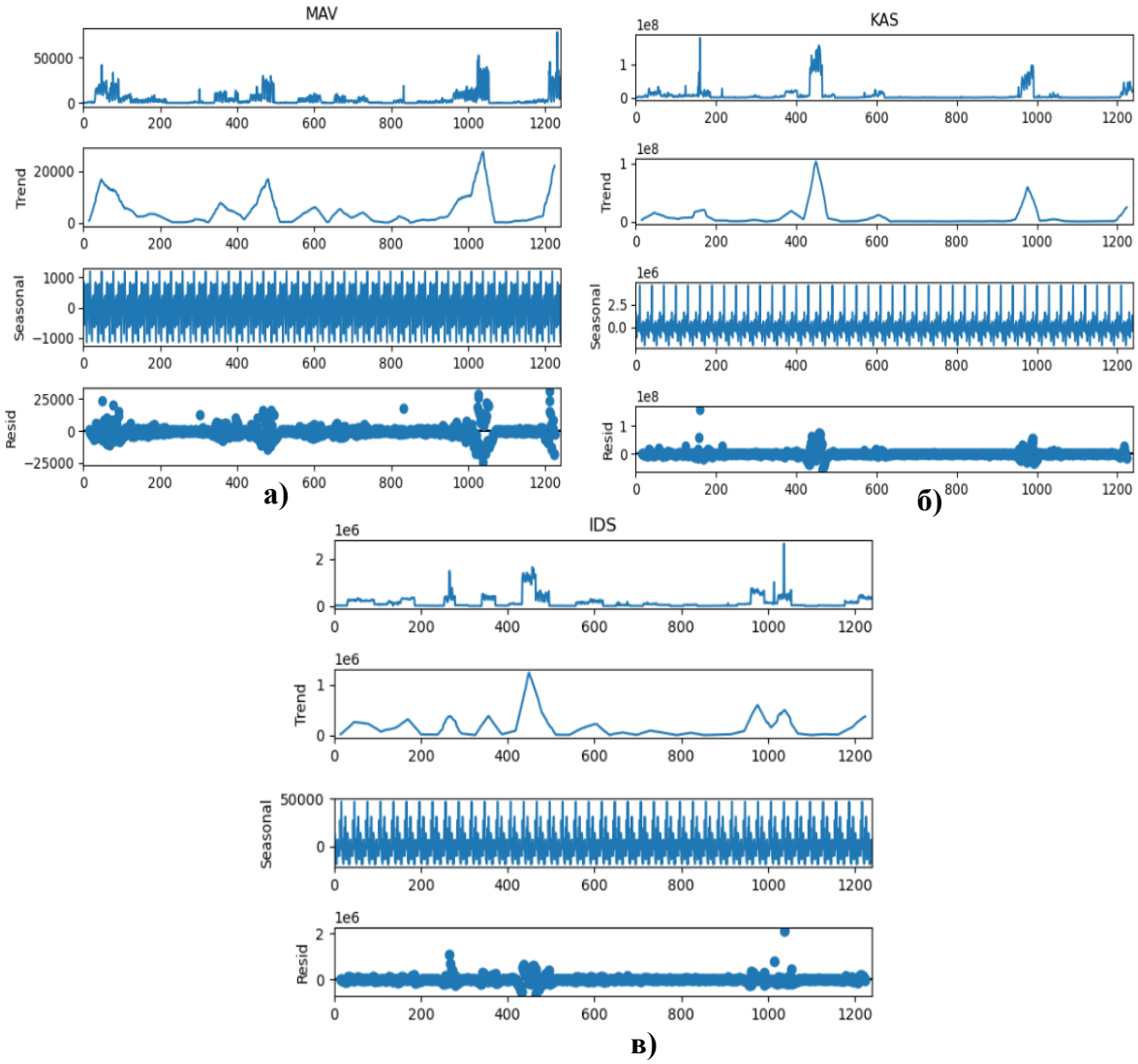


Рисунок 1.3 – Декомпозиція трендів для змінних: а) MAV; б) KAS; в) IDS

Результати перевірки досліджуваних рядів на стаціонарність представлені у таблиці 1.1.

Таблиця 1.1 – Результати тесту Дики-Фулера

Показники тесту	MAV	KAS	IDS
ADF	-7,1100	-36,6960	-36,0314
P-value	0,0000	0,0000	0,0000
Critical value 1%	-3,4357	-3,4356	-3,4356
Critical value 5%	-2,8639	-2,8639	-2,8639
Critical value 10%	2,5680	2,5680	2,5680
Висновок тесту	одиничних коренів немає, ряд є стаціонарним	одиничних коренів немає, ряд є стаціонарним	одиничних коренів немає, ряд є стаціонарним

Проведені тести перевірки рядів на стаціонарність показали, що вони є стаціонарними, тобто значення рядів не мають трендової складової. Для панельних даних у нашому випадку це означає, що у нас не буде виявлено ефекту хибної регресії, що дозволить будувати такі її різновиди, як об’єднана регресія, регресія з фіксованими та випадковими ефектами.

На наступному кроці проведемо перевірку часових рядів на нормальність, а саме їх відповідність нормальному розподілу. Для цього застосуємо два методи: метод побудови гістограм та обчислення тесту Харке-Бера. Результати проведеної процедури представлено на рисунку 1.4. Обидва методи показали, що вхідні дані не відповідають нормальному розподілу. Статистика тесту Харка-Бера завжди є позитивним числом, і якщо вона далека від нуля, а значення p-value менше 0,05, то це вказує на те, що вибіркові дані не відповідають нормальному розподілу. Також й візуальний аналіз графіків підтверджує даний висновок. Якщо гістограма має приблизно «дзвіноподібну форму», то дані вважаються нормально розподіленими. В нашому випадку змінні не мають такої форми, що свідчить про наявність асиметрії в даних і їх не відповідність нормальному розподілу.

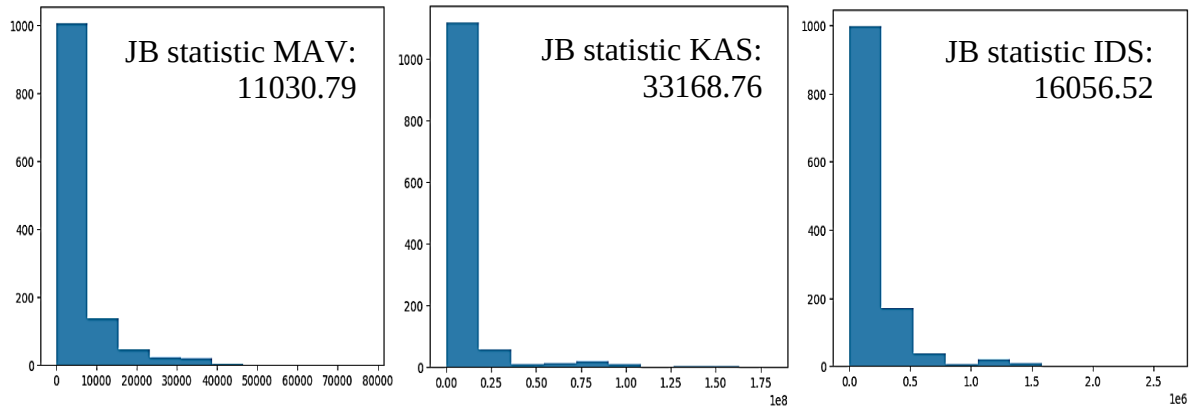


Рисунок 1.4 – Перевірка даних на відповідність закону нормального розподілу змінних

Для наближення даних до нормального розподілу можна виконати процедуру їх логарифмування, тобто здійснити перетворення незалежних

змінних “x” із використанням log(x). Отримані трансформовані дані візуалізовані на рисунку 1.5.

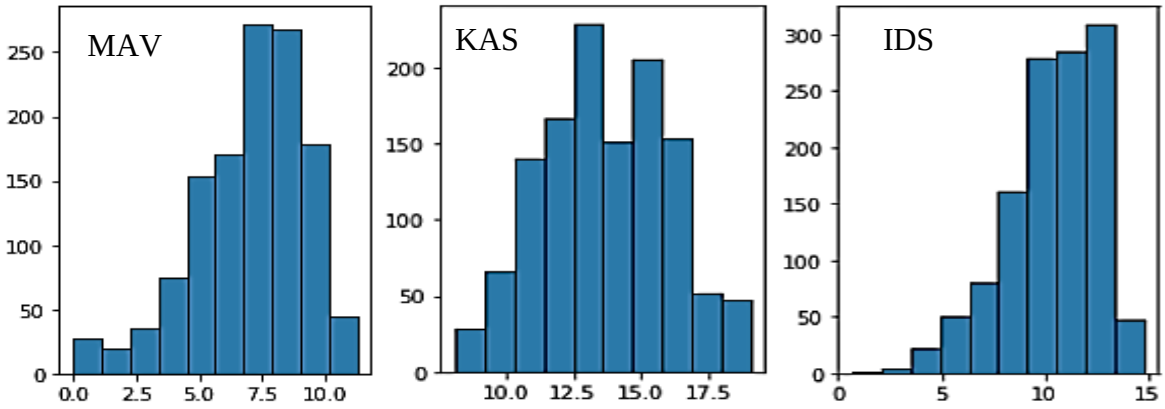


Рисунок 1.5. Результат трансформування незалежних змінних “x”

Хоча трансформація даних й не призвела до повній відповідності даних нормальному закону, але отримані розподіли, зображені на рисунку 1.5, є досить близькими, що, в принципі, дозволяє їх використання для побудови прогнозних моделей.

Дане дослідження присвячене проблематиці кібератак, кількість випадків яких зросла за останні роки. Їх наслідки є катастрофічними для бізнесу, фізичних осіб та держав в цілому. Саме тому виникає потреба у використанні інструментів щодо їх попередження та протидії, в якості яких можуть виступати моделі прогнозування. Для їх реалізації важливим етапом є аналіз та підготовка вхідних даних, що було проведено у даному досліджені. В якості бази емпіричних даних виступили три види часових трендів кібератак, які відслідковувалися за допомогою поштового антивірусу, Касперського Анти-Спаму та системи виявлення вторгнень.

У дослідженні було запропоновано концептуальну модель розробки прогнозних моделей кібератак, яка показує всі етапи процесу прогнозування. Розраховані базові статистики дозволили виявити неоднорідність даних. Встановлено, що це пов’язано із різним рівнем економічного розвитку країн, які було обрано для аналізу. Відповідно, деякі з них в більшій мірі ставали об’єктами кіберзагроз, інші – в меншій мірі. Проведена декомпозиція трендів дозволила

виявити, що дані не містять трендової складової, мають сезонність та зв'язок між змінними є адитивним. Перевірка на стаціонарність за допомогою розширеного тесту Дики-Фулера встановила, що аналізовані тренди є стаціонарними, тобто був підтверджений попередній висновок щодо відсутності трендової складової. Оскільки дані характеризуються нерівномірністю, то проведена перевірка на відповідність нормальному розподілу за допомогою тесту Харка-Бера підтвердила, що їх невідповідність. Для їх наближення до умов нормального розподілу було проведено трансформацію змінних “x” шляхом логарифмування.

Таким чином, проведені в статті процедури підготовки даних дозволяють побудувати прогнозні моделі, такі як об’єднану регресію, регресію з випадковим та фіксованим ефектами. Дані побудові буде реалізовано у подальших дослідженнях.

Пункт 1.1 було виконано із використанням матеріалів публікацій виконавців [22].

1.2 Побудова регресійних моделей для змінної «Mail Anti Virus»

1.2.1 Побудова об’єднаної регресійної моделі (Pooled OLS)

Регресійна модель Pooled OLS часто є хорошою відправною точкою та еталонною моделлю для кількох наборів панельних даних. Для побудови використовуємо OLS клас statsmodels для побудови та адаптації регресійної моделі OLS [23].

Для початку необхідно визначити залежну та незалежні змінні. Залежна змінна – це MAV, яка показує потік даних за шкідливими програмами, виявленими серед нових об'єктів у поштових додатках, незалежними змінними Data_num – показує порядок днів у місяці, так як Python не розуміє типу даних Дата, та перетворює їх в числа з 0 до кінцевого значення по порядку.

Також необхідно створити dummy змінні, які будуть виступати незалежними змінними, тобто кожна країна – це окрема булева змінна та приєднуємо їх до основної бази даних (рис. 1.6).

```
df_dummies = pd.get_dummies(df[unit_col_name])
df_panel_with_dummies = df.join(df_dummies)
df_panel_with_dummies
```

	Date	Country	Date_num	Country_Id	MAV	KAS	IDS	Afghanistan	Armenia	Azerbaijan	...	Togo	Tynisia	UK	Uganda	Ukraine	Uni Sta
0	2022-08-14	Ukraine	1	1	4.795791	13.978904	9.912150	0	0	0	...	0	0	0	0	1	
1	2022-08-15	Ukraine	2	1	5.379897	14.321123	9.913190	0	0	0	...	0	0	0	0	1	
2	2022-08-16	Ukraine	3	1	5.493061	14.300403	9.952897	0	0	0	...	0	0	0	0	1	
3	2022-08-17	Ukraine	4	1	6.345636	14.399956	9.977481	0	0	0	...	0	0	0	0	1	
4	2022-08-18	Ukraine	5	1	6.077642	14.313859	9.906084	0	0	0	...	0	0	0	0	1	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
1235	2022-09-09	Brazil	27	40	10.477851	17.168917	12.671064	0	0	0	...	0	0	0	0	0	
1236	2022-09-10	Brazil	28	40	8.645059	16.921710	12.578605	0	0	0	...	0	0	0	0	0	
1237	2022-09-11	Brazil	29	40	8.306966	16.747717	12.606053	0	0	0	...	0	0	0	0	0	
1238	2022-09-12	Brazil	30	40	10.256290	16.911541	12.725729	0	0	0	...	0	0	0	0	0	
1239	2022-09-13	Brazil	31	40	10.105653	16.701400	12.790911	0	0	0	...	0	0	0	0	0	

1240 rows × 47 columns

Рисунок 1.6 – Створення dummy змінних

Визначаємо залежну та незалежні змінні (рис. 1.7).

```
y = 'MAV'
x = df_panel_with_dummies.drop(['MAV', 'KAS', 'IDS', 'Date', 'Country'], axis=1)
```

Рисунок 1.7 – Вхідні змінні моделі

Будуємо об’єднану модель для змінної MAV (рис. 1.8).

Після побудови об’єднаної регресійної моделі отримуємо наступне рівняння регресії (1.1):

$$MAV = 6.77 + 0.0038 \cdot Data_{num} - 4.83 \cdot Afganistan + 5.11 \cdot Armenia + 5.22 \cdot Azerbaijan \ldots + Zandia \cdot 5.05 + 7.09 \cdot Zimbabwe + \epsilon \quad (1.1)$$

OLS Regression Results

Dep. Variable:	MAV	R-squared:	0.776
Model:	OLS	Adj. R-squared:	0.769
Method:	Least Squares	F-statistic:	104.0
Date:	Sun, 27 Nov 2022	Prob (F-statistic):	0.00
Time:	12:52:01	Log-Likelihood:	-1785.7
No. Observations:	1240	AIC:	3653.
Df Residuals:	1199	BIC:	3863.
Df Model:	40		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	6.7728	0.059	114.842	0.000	6.657	6.888
Date_num	0.0038	0.003	1.147	0.251	-0.003	0.010
Afghanistan	-1.9391	0.184	-10.527	0.000	-2.301	-1.578
Armenia	-1.6596	0.184	-9.010	0.000	-2.021	-1.298
Azerbaijan	-1.5556	0.184	-8.445	0.000	-1.917	-1.194
Brazil	2.7729	0.184	15.053	0.000	2.412	3.134
Canada	0.7021	0.184	3.812	0.000	0.341	1.064
Chile	0.8461	0.184	4.593	0.000	0.485	1.207
China	1.8793	0.184	10.202	0.000	1.518	2.241
Colombia	2.1974	0.184	11.929	0.000	1.836	2.559
Costa Rica	-0.7608	0.184	-4.130	0.000	-1.122	-0.399
Cuba	-2.4545	0.184	-13.325	0.000	-2.816	-2.093
Czech Republic	-0.2331	0.184	-1.266	0.206	-0.595	0.128
Egypt	1.2235	0.184	6.642	0.000	0.862	1.585
France	0.9883	0.184	5.365	0.000	0.627	1.350
Germany	2.8093	0.184	15.251	0.000	2.448	3.171
Hungary	0.7274	0.184	3.949	0.000	0.366	1.089
India	1.5464	0.184	8.395	0.000	1.185	1.908
Indonezia	1.8612	0.184	10.104	0.000	1.500	2.223
Iran	1.2874	0.184	6.989	0.000	0.926	1.649
Italy	2.3567	0.184	12.794	0.000	1.995	2.718
Japan	1.4038	0.184	7.620	0.000	1.042	1.765
Kenya	1.2794	0.184	6.945	0.000	0.918	1.641
Mexico	3.1120	0.184	16.894	0.000	2.751	3.473
Moldova	-2.0779	0.184	-11.280	0.000	-2.439	-1.716
Paraguay	-0.1887	0.184	-1.025	0.306	-0.550	0.173
Poland	0.6754	0.184	3.667	0.000	0.314	1.037
Slovakia	-2.0631	0.184	-11.200	0.000	-2.424	-1.702
Somalia	-5.6288	0.184	-30.557	0.000	-5.990	-5.267
South Korea	0.0173	0.184	0.094	0.925	-0.344	0.379
Sudan	-0.9073	0.184	-4.925	0.000	-1.269	-0.546
Tanzania	-0.1039	0.184	-0.564	0.573	-0.465	0.257
Togo	-3.0726	0.184	-16.680	0.000	-3.434	-2.711
Tynisia	0.2607	0.184	1.415	0.157	-0.101	0.622
UK	1.6226	0.184	8.808	0.000	1.261	1.984
Uganda	-0.3444	0.184	-1.870	0.062	-0.706	0.017
Ukraine	-0.5675	0.184	-3.081	0.002	-0.929	-0.206
United States	2.2485	0.184	12.206	0.000	1.887	2.610
Venezuala	-2.6326	0.184	-14.291	0.000	-2.994	-2.271
Vietnam	2.5526	0.184	13.857	0.000	2.191	2.914
Zamdia	-1.7266	0.184	-9.373	0.000	-2.088	-1.365
Zimbabwe	0.3189	0.184	1.731	0.084	-0.042	0.680

Omnibus:	245.994	Durbin-Watson:	1.499
Prob(Omnibus):	0.000	Jarque-Bera (JB):	441.868
Skew:	-1.216	Prob(JB):	1.12e-96
Kurtosis:	4.623	Cond. No.	1.72e+16

Рисунок 1.8 – Результати побудови об’єднаної моделі для змінної MAV

Щоб проаналізувати, чи є об’єднана модель OLS адекватною моделлю для нашої проблеми регресії, необхідно провести аналіз таких показників R-квадрат і F-тест, логарифм правдоподібності та балів AIC, а також опосередковано через аналіз залишків.

Скоригований R-квадрат, який вимірює частку загальної дисперсії в y, яка пояснюється X після врахування ступенів свободи, втрачених через включення змінних регресії, становить 0.769 або близько 76.9 %. Це, безумовно гарний результат.

F - тест для регресії, який вимірює спільну значущість параметрів моделі, дав тестову статистику 104.0 із значенням p = 0.00, що дозволяє зробити висновок, що оцінки коефіцієнтів моделі є спільно значущими при $p < 0.001$ [21].

Log-правдоподібність регресійної моделі - це спосіб вимірювання користі придатності для моделі. Чим вище значення лог-ймовірності, тим краще модель підходить для набору даних [25]. Log-правдоподібність моделі становить 1785.7, а показник AIC 3653. Ці значення придатності самі по собі не мають сенсу, якщо ми не порівняємо їх із показниками конкуруючої моделі [24].

Проаналізуємо залишкові похибки моделі для нормальності, гетероскедастичності та кореляції - трьох властивостей, які впливають на відповідність лінійної моделі.

Залишкова стандартна похибка - це міра, яка використовується для оцінки того, наскільки добре модель лінійної регресії відповідає даним [26]. Її результат представлений на рисунках 1.9-1.10.

```
print(pooled_olsr_model_results.resid)
```

```
0      -1.413282
1      -0.832959
2      -0.723579
3       0.125212
4      -0.146566
          ...
1235    0.829994
1236   -1.006582
1237   -1.348459
1238    0.597081
1239    0.442660
Length: 1240, dtype: float64
```

Рисунок 1.9 – Залишкові похибки моделі

```
print('Mean value of residual errors='+str(pooled_olsr_model_results.resid.mean()))
```

```
Mean value of residual errors=5.80431760003223e-15
```

Рисунок 1.10 – Середні значення похибок моделі

Це говорить нам про те, що регресійна модель прогнозує MAV із середньою похибкою близько 5.80 e-15.

1.2.2 Побудова моделі регресії фіксованих ефектів.

Для побудови моделі регресії фіксованих ефектів, необхідно створити фіктивні змінні (рис. 1.11).

```
unit_col_name= 'Country'
time_period_col_name='Date'
```

Рисунок 1.11 – Створення фіктивних змінних

Здійснюємо побудову регресії (рис. 1.12-1.13).

```
unit_names = ['Germany','Italy','UK' ,'Poland','France' ,'Hungary' ,'Moldova', 'Slovakia', 'Afghanistan','Indonezia','Japan','Ch
```

```
lsdv_expr = y_var_name + ' ~ '
i = 0
for X_var_name in X_var_names:
    if i > 0:
        lsdv_expr = lsdv_expr + ' + ' + X_var_name
    else:
        lsdv_expr = lsdv_expr + X_var_name
    i = i + 1
for dummy_name in unit_names[:-1]:
    lsdv_expr = lsdv_expr + ' + ' + dummy_name

print('Regression expression for OLS with dummies=' + lsdv_expr)
```

```
Regression expression for OLS with dummies=MAV ~ Date_num + Germany + Italy + UK + Poland + France + Hungary + Moldova + Slovak
ia + Afghanistan + Indonezia + Japan + China + Vietnam + Armenia + Azerbaijan + Iran + India + Zamdia + Kenya + Zimbabwe + Egyp
t + Sudan + Somalia + Tynisia + Togo + Uganda + Tanzania + Canada + Colombia + Mexico + Cuba + Paraguay + Chile + Brazil
```

Рисунок 1.12 – Побудова рівняння регресії

OLS Regression Results

Dep. Variable:	MAV	R-squared:	0.710
Model:	OLS	Adj. R-squared:	0.702
Method:	Least Squares	F-statistic:	84.29
Date:	Fri, 25 Nov 2022	Prob (F-statistic):	6.31e-295
Time:	13:51:22	Log-Likelihood:	-1946.1
No. Observations:	1240	AIC:	3964.
Df Residuals:	1204	BIC:	4149.
Df Model:	35		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	6.4514	0.105	61.309	0.000	6.245	6.658
Date_num	0.0038	0.004	1.010	0.313	-0.004	0.011
Germany	3.1307	0.229	13.680	0.000	2.682	3.580
Italy	2.6781	0.229	11.702	0.000	2.229	3.127
UK	1.9439	0.229	8.494	0.000	1.495	2.393
Poland	0.9968	0.229	4.356	0.000	0.548	1.446
France	1.3097	0.229	5.723	0.000	0.861	1.759
Hungary	1.0487	0.229	4.583	0.000	0.600	1.498
Moldova	-1.7565	0.229	-7.675	0.000	-2.206	-1.308
Slovakia	-1.7417	0.229	-7.611	0.000	-2.191	-1.293
Afghanistan	-1.6177	0.229	-7.069	0.000	-2.067	-1.169
Indonezia	2.1826	0.229	9.537	0.000	1.734	2.632
Japan	1.7251	0.229	7.538	0.000	1.276	2.174
China	2.2007	0.229	9.616	0.000	1.752	2.650
Vietnam	2.8740	0.229	12.558	0.000	2.425	3.323
Armenia	-1.3383	0.229	-5.848	0.000	-1.787	-0.889
Azerbaijan	-1.2342	0.229	-5.393	0.000	-1.683	-0.785
Iran	1.6088	0.229	7.030	0.000	1.160	2.058
India	1.8678	0.229	8.162	0.000	1.419	2.317
Zamdia	-1.4052	0.229	-6.140	0.000	-1.854	-0.956
Kenya	1.6007	0.229	6.995	0.000	1.152	2.050
Zimbabwe	0.6403	0.229	2.798	0.005	0.191	1.089
Egypt	1.5449	0.229	6.751	0.000	1.096	1.994
Sudan	-0.5859	0.229	-2.560	0.011	-1.035	-0.137
Somalia	-5.3074	0.229	-23.192	0.000	-5.756	-4.858
Tynisia	0.5821	0.229	2.544	0.011	0.133	1.031
Togo	-2.7512	0.229	-12.022	0.000	-3.200	-2.302
Uganda	-0.0231	0.229	-0.101	0.920	-0.472	0.426
Tanzania	0.2175	0.229	0.950	0.342	-0.232	0.666
Canada	1.0235	0.229	4.472	0.000	0.575	1.472
Colombia	2.5187	0.229	11.006	0.000	2.070	2.968
Mexico	3.4334	0.229	15.003	0.000	2.984	3.882
Cuba	-2.1331	0.229	-9.321	0.000	-2.582	-1.684
Paraguay	0.1327	0.229	0.580	0.562	-0.316	0.582
Chile	1.1674	0.229	5.101	0.000	0.718	1.616
Brazil	3.0943	0.229	13.521	0.000	2.645	3.543

Omnibus:	246.665	Durbin-Watson:	1.173
Prob(Omnibus):	0.000	Jarque-Bera (JB):	603.618
Skew:	-1.067	Prob(JB):	8.43e-132
Kurtosis:	5.670	Cond. No.	303.

Рисунок 1.13 – Регресія фіксованих змінних для змінної MAV

Рівняння регресії з фіксованими змінними виглядає наступним чином (1.2):

$$\begin{aligned} \mathrm{MAV} = 6.45 \ + 0.0038 \cdot \mathrm{Data}_{\mathrm{num}} - 3.13 \cdot \mathrm{Germany} + 2.68 \cdot \mathrm{Italy} + \cdots \\ + 2.17 \cdot \mathrm{Cuba} + 1.18 \cdot \mathrm{Chile} + 3.09 \cdot \mathrm{Brazil} + \ \epsilon \end{aligned} \quad (1.2)$$

Скоригований R-квадрат, який вимірює частку загальної дисперсії в y, яка пояснюється X після врахування ступенів свободи, втрачених через включення змінних регресії, становить 0.702 або близько 70.2 %. Це, безумовно гарний результат.

F - тест для регресії, який вимірює спільну значущість параметрів моделі, дав тестову статистику 84.29 із значенням p = 0.00, що дозволяє зробити висновок, що оцінки коефіцієнтів моделі є спільно значущими при $p < 0.001$.

Log-правдоподібність моделі становить 1946.1, а показник AIC 3964. Ці значення придатності самі по собі не мають сенсу, якщо ми не порівняємо їх із показниками конкуруючої моделі.

Проаналізуємо залишкові похибки підігнаної моделі для нормальності, гетероскедастичності та кореляції - трьох властивостей, які впливають на відповідність лінійної моделі [27] (рис. 1.14).

```
print(lsdv_model_results.resid)
print('Mean value of residual errors='+str(lsdv_model_results.resid.mean()))
```

```
0      -1.659387
1      -1.079064
2      -0.969684
3      -0.120892
4      -0.392670
          ...
1235    0.829994
1236   -1.006582
1237   -1.348459
1238    0.597081
1239    0.442660
Length: 1240, dtype: float64
Mean value of residual errors=4.410210548010421e-13
```

Рисунок 1.14 - Залишкові похибки моделі

1.2.3 Побудова моделі регресії випадкових ефектів.

Першим кроком для побудови моделі випадкових ефектів, необхідно розрахувати $\sigma^2\_ \epsilon$ і $\sigma^2\_ \mu$ - дисперсії компонентів похибки μ і ϵмоделі фіксованих ефектів та об'єднаної моделі та знайти різницю між ними (рис. 1.15).

```
sigma2_epsilon = lsdv_model_results.ssr/(n*T-(n+k+1))
print('sigma2_epsilon = ' + str(sigma2_epsilon))
```

```
sigma2_epsilon = -45.283594237287836
```

```
sigma2_pooled = pooled_olsr_model_results.ssr/(n*T-(k+1))
print('sigma2_pooled = ' + str(sigma2_pooled))
```

```
sigma2_pooled = -646.7554791838006
```

```
sigma2_u = sigma2_pooled - sigma2_epsilon
print('sigma2_u = ' + str(sigma2_u))
```

```
sigma2_u = -601.4718849465128
```

Рисунок 1.15 – Розрахунок значень дисперсії компонентів похибки моделей

Обчислюємо середні значення y та X для кожної групи (тобто кожної одиниці i) на панелі даних таким чином (рис. 1.16).

```
df_group_means = df_panel_with_dummies.groupby(unit_col_name).mean()

print(df_group_means)
```

Country	Date_num	Country_id	MAV	KAS	IDS
Afghanistan	16.0	11.0	4.894210	9.706546	7.178635
Armenia	16.0	17.0	5.173679	11.371538	8.767991
Azerbaijan	16.0	18.0	5.277701	11.792697	8.177036
Brazil	16.0	40.0	9.606234	16.871789	12.799997
Canada	16.0	31.0	7.535446	14.592936	11.297106
Chile	16.0	39.0	7.679380	13.193566	11.915792
China	16.0	15.0	8.712614	18.418690	14.017562
Colombia	16.0	33.0	9.030679	13.974321	11.785556
Costa Rica	16.0	37.0	6.072505	11.769029	9.744092
Cuba	16.0	36.0	4.378815	9.496146	6.163067
Czech Republic	16.0	10.0	6.600173	14.659354	10.278660
Egypt	16.0	24.0	8.056791	12.712312	11.350442
France	16.0	6.0	7.821634	16.351361	12.645950
Germany	16.0	2.0	9.642643	16.479932	12.442787
Hungary	16.0	7.0	7.560679	13.582649	9.446409
India	16.0	20.0	8.379729	16.105214	12.285693
Indonezia	16.0	12.0	8.694488	14.663417	12.816168

Рисунок 1.16 – Обчислення середніх значень

Наступним кроком є зменшення середніх значень всіх значень y та X для кожної одиниці, використовуючи масштабовану версію відповідного середнього значення для конкретної групи, обчислених на другому кроці (рис. 1.17).

```
theta = 1 - math.sqrt(c/(u))
print('theta = ' + str(theta))
```

```
theta = 0.9539504917492148
```

Рисунок 1.17 – Обчислення показника Тета

Тепер будуємо модель випадкових ефектів (рис. 1.18):

```
                          OLS Regression Results
==============================================================================
Dep. Variable:                    MAV   R-squared:                       0.640
Model:                            OLS   Adj. R-squared:                  0.630
Method:                 Least Squares   F-statistic:                     67.02
Date:                Sat, 26 Nov 2022   Prob (F-statistic):          1.25e-241
Time:                        11:07:05   Log-Likelihood:                -12222.
No. Observations:                1240   AIC:                         2.451e+04
Df Residuals:                    1207   BIC:                         2.468e+04
Df Model:                          32
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const       3.939e+04   6090.642      6.467      0.000    2.74e+04    5.13e+04
Date_num       19.4694     14.863      1.310      0.190      -9.691      48.629
Germany     3.204e+05   1.92e+04     16.649      0.000    2.83e+05    3.58e+05
Italy       2.373e+05   1.92e+04     12.329      0.000       2e+05    2.75e+05
UK          8.045e+04   1.92e+04      4.180      0.000    4.27e+04    1.18e+05
Poland      1.593e+04   1.92e+04      0.828      0.408   -2.18e+04    5.37e+04
France      3.191e+04   1.92e+04      1.658      0.098   -5849.990    6.97e+04
Hungary     1.161e+04   1.92e+04      0.603      0.547   -2.62e+04    4.94e+04
Moldova    -3.576e+04   1.92e+04     -1.858      0.063   -7.35e+04    1999.229
Slovakia   -3.566e+04   1.92e+04     -1.853      0.064   -7.34e+04    2097.694
Indonezia   1.271e+05   1.92e+04      6.602      0.000    8.93e+04    1.65e+05
Japan       5.866e+04   1.92e+04      3.048      0.002    2.09e+04    9.64e+04
China       1.136e+05   1.92e+04      5.903      0.000    7.58e+04    1.51e+05
Vietnam     3.244e+05   1.92e+04     16.855      0.000    2.87e+05    3.62e+05
Armenia    -3.277e+04   1.92e+04     -1.703      0.089   -7.05e+04    4983.837
Azerbaijan -3.298e+04   1.92e+04     -1.714      0.087   -7.07e+04    4777.133
Iran        4.386e+04   1.92e+04      2.279      0.023    6101.491    8.16e+04
India       8.245e+04   1.92e+04      4.284      0.000    4.47e+04    1.2e+05
Zamdia     -3.369e+04   1.92e+04     -1.751      0.080   -7.15e+04    4066.157
Kenya       7.124e+04   1.92e+04      3.702      0.000    3.35e+04    1.09e+05
Zimbabwe    1869.2433   1.92e+04      0.097      0.923   -3.59e+04    3.96e+04
Egypt       4.335e+04   1.92e+04      2.252      0.024    5587.241    8.11e+04
Sudan      -2.884e+04   1.92e+04     -1.498      0.134   -6.66e+04    8923.321
Somalia    -3.958e+04   1.92e+04     -2.057      0.040   -7.73e+04   -1825.540
Tynisia     1.175e+04   1.92e+04      0.611      0.542    -2.6e+04    4.95e+04
Togo       -3.803e+04   1.92e+04     -1.976      0.048   -7.58e+04    -268.025
Canada      1.125e+04   1.92e+04      0.584      0.559   -2.65e+04    4.9e+04
Colombia    1.893e+05   1.92e+04      9.836      0.000    1.52e+05    2.27e+05
Mexico      5.526e+05   1.92e+04     28.711      0.000    5.15e+05    5.9e+05
Cuba       -3.736e+04   1.92e+04     -1.941      0.052   -7.51e+04    397.263
Chile       1.626e+04   1.92e+04      0.845      0.398   -2.15e+04    5.4e+04
Brazil      4.359e+05   1.92e+04     22.647      0.000    3.98e+05    4.74e+05
Ukraine    -2.337e+04   1.92e+04     -1.214      0.225   -6.11e+04    1.44e+04
==============================================================================
Omnibus:                      482.475   Durbin-Watson:                   1.422
Prob(Omnibus):                  0.000   Jarque-Bera (JB):            29755.691
Skew:                           0.959   Prob(JB):                         0.00
Kurtosis:                      26.921   Cond. No.                     2.62e+03
==============================================================================
```

Рисунок 1.18 – Побудова моделі регресії випадкових ефектів

Отримана модель матиме вигляд формули (1.3):

$$\begin{aligned} \mathrm{MAV} = 39390 + 19.47 \cdot \mathrm{Data}_{\mathrm{num}} - 3.204\mathrm{e}+04 \cdot \mathrm{Germany} + 2.373\mathrm{t} \\ + 05 \cdot \mathrm{Italy} + \cdots + 4.359\mathrm{e}+05 \cdot \mathrm{Brazil} - 2.337\mathrm{e}+04 \\ \cdot \mathrm{Ukraine} + \epsilon \end{aligned} \tag{1.3}$$

Обчислена дисперсія σ²_u було оцінено як -601.47, а σ²_є було оцінено як -45.28. Таким чином, частка загальної дисперсії, яка може бути віднесена до випадкового ефекту окремої одиниці, дорівнює (1.4):

$$\frac{-601.47}{-601.47+(-45.28)} = 0{,}93 \qquad (1.4)$$

Це означає, що на 93% присутній випадковий ефект у моделі, але дивлячись на показник скоригованого R, який дорівнює 0.63, можна зробити висновок, що ця модель не є кращою за модель фіксованого ефекту та об'єднану модель.

Ми можемо використовувати тест Breusch-Pagan LM для перевірки значущості випадкового ефекту [28].

Нульова гіпотеза тесту Бреуша-Пагана LM полягає в тому, що одинична дисперсія σ²_u дорівнює нулю (рис. 1.19).

```
df_pooled_olsr_resid_with_unitnames = pd.concat([df_data[unit_col_name],pooled_olsr_model_results.resid], axis=1)
df_pooled_olsr_resid_group_means = df_pooled_olsr_resid_with_unitnames.groupby(unit_col_name).mean()
ssr_grouped_means=(df_pooled_olsr_resid_group_means[0]**2).sum()
ssr_pooled_olsr=pooled_olsr_model_results.ssr
LM_statistic = (n*T)/(2*(T-1))*math.pow(((T*T*ssr_grouped_means)/ssr_pooled_olsr - 1),2)
print('BP LM Statistic='+str(LM_statistic))
alpha=0.05
chi2_critical_value=st.chi2.ppf((1.0-alpha), 1)
print('chi2_critical_value='+str(chi2_critical_value))
```

```
BP LM Statistic=18.008298755186722
chi2_critical_value=3.841458820694124
```

Рисунок 1.19 -Перевірка значущості моделі випадкового ефекту

Тестова статистика тесту LM (18,0083) більша, ніж критичне значення Chi-squared = 3,84146 при α=0,05, що означає, що випадковий ефект є значущим при альфа 0,05.

1.3 Побудова регресійних моделей для змінної «Kaspersky Anti-Spam»

1.3.1 Побудова об’єднаної регресійної моделі (Pooled OLS)

Необхідно визначити залежну та незалежні змінні. Залежна змінна – це KAS, яка показує потік даних з виявлених мережевих атак. Визначаємо вхідні дані моделі (рис. 1.20):

```
y = 'KAS'
x = df_panel_with_dummies.drop(['MAV', 'KAS', 'IDS', 'Date', 'Country'], axis=1)
```

Рисунок 1.20 – Вхідні змінні моделі

Статистично незначущі фактори було усунуто з моделі та модель побудували знову (рис. 1.21).

Після побудови об’єднаної регресійної моделі, отримуємо наступне рівняння регресії (1.5):

$$\begin{aligned} \mathrm{KAS} = {} & 13.02 + 0.0203 \cdot \mathrm{Data}_{\mathrm{num}} - 3.64 \cdot \mathrm{Afganistan} - 1.97 \cdot \mathrm{Armenia} \\ & - 1.55 \cdot \mathrm{Azerbaijan} + \cdots - 2.37 \cdot \mathrm{Zandia} - 2.2 \cdot \mathrm{Zimbabwe} \\ & + \epsilon \end{aligned} \quad (1.5)$$

OLS Regression Results

Dep. Variable:	KAS	R-squared:	0.923
Model:	OLS	Adj. R-squared:	0.920
Method:	Least Squares	F-statistic:	357.2
Date:	Sat, 26 Nov 2022	Prob (F-statistic):	0.00
Time:	13:38:14	Log-Likelihood:	-1211.8
No. Observations:	1240	AIC:	2506.
Df Residuals:	1199	BIC:	2716.
Df Model:	40		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	13.0191	0.037	350.683	0.000	12.946	13.092
Date_num	0.0203	0.002	9.787	0.000	0.016	0.024
Afghanistan	-3.6376	0.116	-31.369	0.000	-3.865	-3.410
Armenia	-1.9726	0.116	-17.011	0.000	-2.200	-1.745
Azerbaijan	-1.5515	0.116	-13.379	0.000	-1.779	-1.324
Brazil	3.5276	0.116	30.421	0.000	3.300	3.755
Canada	1.2488	0.116	10.769	0.000	1.021	1.476
Chile	-0.1506	0.116	-1.299	0.194	-0.378	0.077
China	5.0745	0.116	43.760	0.000	4.847	5.302
Colombia	0.6301	0.116	5.434	0.000	0.403	0.858
Costa Rica	-1.5751	0.116	-13.583	0.000	-1.803	-1.348
Cuba	-3.8480	0.116	-33.184	0.000	-4.076	-3.621
Czech Republic	1.3152	0.116	11.342	0.000	1.088	1.543
Egypt	-0.6319	0.116	-5.449	0.000	-0.859	-0.404
France	3.0072	0.116	25.933	0.000	2.780	3.235
Germany	3.1358	0.116	27.041	0.000	2.908	3.363
Hungary	0.2385	0.116	2.056	0.040	0.011	0.466
India	2.7610	0.116	23.810	0.000	2.534	2.989
Indonezia	1.3192	0.116	11.377	0.000	1.092	1.547
Iran	1.1763	0.116	10.144	0.000	0.949	1.404
Italy	2.2894	0.116	19.743	0.000	2.062	2.517
Japan	3.3639	0.116	29.009	0.000	3.136	3.591
Kenya	-0.3822	0.116	-3.296	0.001	-0.610	-0.155
Mexico	1.2818	0.116	11.054	0.000	1.054	1.509
Moldova	-0.3385	0.116	-2.919	0.004	-0.566	-0.111
Paraguay	-1.0199	0.116	-8.795	0.000	-1.247	-0.792
Poland	2.3224	0.116	20.027	0.000	2.095	2.550
Slovakia	-0.5354	0.116	-4.617	0.000	-0.763	-0.308
Somalia	-2.5877	0.116	-22.315	0.000	-2.815	-2.360
South Korea	1.6567	0.116	14.287	0.000	1.429	1.884
Sudan	-2.3620	0.116	-20.369	0.000	-2.590	-2.135
Tanzania	-2.2058	0.116	-19.022	0.000	-2.433	-1.978
Togo	-2.4489	0.116	-21.118	0.000	-2.676	-2.221
Tynisia	-0.5313	0.116	-4.582	0.000	-0.759	-0.304
UK	1.8676	0.116	16.106	0.000	1.640	2.095
Uganda	-0.0887	0.116	-0.765	0.445	-0.316	0.139
Ukraine	1.2904	0.116	11.128	0.000	1.063	1.518
United States	4.4757	0.116	38.597	0.000	4.248	4.703
Venezuala	-0.5113	0.116	-4.409	0.000	-0.739	-0.284
Vietnam	1.9832	0.116	17.102	0.000	1.756	2.211
Zandia	-2.3716	0.116	-20.452	0.000	-2.599	-2.144
Zimbabwe	-2.1955	0.116	-18.933	0.000	-2.423	-1.968

Omnibus:	87.758	Durbin-Watson:	1.297
Prob(Omnibus):	0.000	Jarque-Bera (JB):	383.362
Skew:	0.139	Prob(JB):	5.68e-84
Kurtosis:	5.710	Cond. No.	1.72e+16

Рисунок 1.21 - Об’єднана регресійна модель для змінної KAS

Скоригований R-квадрат, який вимірює частку загальної дисперсії в y, яка пояснюється X після врахування ступенів свободи, втрачених через включення змінних регресії, становить 0,920 або близько 92,0 %. Це, безумовно гарний результат.

F - тест для регресії, який вимірює спільну значущість параметрів моделі, дав тестову статистику 357,2 із значенням p = 0,00, що дозволяє зробити висновок, що оцінки коефіцієнтів моделі є спільно значущими при $p < 0,001$.

Log-правдоподібність моделі становить – 1211,8, а показник AIC становить 2506.

Проаналізуємо залишкові похибки (рис. 1.22):

```
print(pooled_olsr_model_results.resid)
print('Mean value of residual errors='+str(pooled_olsr_model_results.resid.mean()))
```

```
0       -0.350885
1       -0.028983
2       -0.070020
3        0.009218
4       -0.097199
           ...
1235     0.073640
1236    -0.193884
1237    -0.388195
1238    -0.244688
1239    -0.475146
Length: 1240, dtype: float64
Mean value of residual errors=-6.661767911502407e-14
```

Рисунок 1.22 – Залишкові похибки моделі

Це говорить нам про те, що регресійна модель прогнозує KAS із середньою похибкою близько -6,66e-14.

1.3.2 Побудова моделі регресії фіксованих ефектів

Визначаємо залежну та незалежні змінні (рис. 1.23):

```
y_var_name = 'KAS'
X_var_names = df.drop(['MAV', 'KAS', 'IDS', 'Date', 'Country'], axis=1)
```

Рисунок 1.23 - Визначення вхідних даних моделі

Визначаємо всі країни, які будуть незалежними змінними, які впливають на залежну та будуємо рівняння регресії (1.24-1.25):

```
lsdv_expr = y_var_name + ' ~ '
i = 0
for X_var_name in X_var_names:
    if i > 0:
        lsdv_expr = lsdv_expr + ' + ' + X_var_name
    else:
        lsdv_expr = lsdv_expr + X_var_name
    i = i + 1
for dummy_name in unit_names[:-1]:
    lsdv_expr = lsdv_expr + ' + ' + dummy_name

print('Regression expression for OLS with dummies=' + lsdv_expr)
```

```
Regression expression for OLS with dummies=KAS ~ Date_num + Germany + Italy + UK + Poland + France + Hungary + Moldova + Slovak
ia + Afghanistan + Indonezia + Japan + China + Vietnam + Armenia + Azerbaijan + Iran + India + Zamdia + Kenya + Zimbabwe + Egyp
t + Sudan + Somalia + Tynisia + Togo + Uganda + Tanzania + Canada + Colombia + Mexico + Cuba + Paraguay + Chile + Brazil
```

Рисунок 1.24 – Побудова рівняння регресії

Рівняння регресії з фіксованими змінними виглядає наступним чином (1.6):

$$\mathrm{KAS} = 14.13 + 0.0203 \cdot \mathrm{Data}_{\mathrm{num}} - 2.03 \cdot \mathrm{Germany} - 1.18 \cdot \mathrm{Italy} - \\ -1.25 \cdot \mathrm{Chile} + \cdots + 2.42 \cdot \mathrm{Brazil} + \epsilon \quad (1.6)$$

Далі розглянемо коефіцієнти для цікавих фіктивних змінних, що представляють вплив на конкретну країну. Ми спостерігаємо, що відрізок регресії, який представляє специфічний для країни ефект для України (пропущена змінна), становить 0,822 і є статистично значущим (це означає, що його значення для населення оцінюється як відмінне від нуля), при p-значенні 0,000 .

Скориговане значення R-квадрат дорівнює 0,817, або 81,7% - значення показує дуже хорошу відповідність між незалежними змінними та залежною.

OLS Regression Results

Dep. Variable:	KAS	R-squared:	0.822
Model:	OLS	Adj. R-squared:	0.817
Method:	Least Squares	F-statistic:	158.6
Date:	Wed, 23 Nov 2022	Prob (F-statistic):	0.00
Time:	14:20:08	Log-Likelihood:	-1728.9
No. Observations:	1240	AIC:	3530.
Df Residuals:	1204	BIC:	3714.
Df Model:	35		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	14.1277	0.088	159.958	0.000	13.954	14.301
Date_num	0.0203	0.003	6.463	0.000	0.014	0.026
Germany	2.0272	0.192	10.554	0.000	1.650	2.404
Italy	1.1808	0.192	6.147	0.000	0.804	1.558
UK	0.7590	0.192	3.952	0.000	0.382	1.136
Poland	1.2138	0.192	6.319	0.000	0.837	1.591
France	1.8986	0.192	9.884	0.000	1.522	2.275
Hungary	-0.8701	0.192	-4.530	0.000	-1.247	-0.493
Moldova	-1.4471	0.192	-7.534	0.000	-1.824	-1.070
Slovakia	-1.6440	0.192	-8.559	0.000	-2.021	-1.267
Afghanistan	-4.7462	0.192	-24.709	0.000	-5.123	-4.369
Indonezia	0.2106	0.192	1.097	0.273	-0.166	0.587
Japan	2.2553	0.192	11.741	0.000	1.878	2.632
China	3.9659	0.192	20.647	0.000	3.589	4.343
Vietnam	0.8746	0.192	4.553	0.000	0.498	1.251
Armenia	-3.0812	0.192	-16.041	0.000	-3.458	-2.704
Azerbaijan	-2.6601	0.192	-13.849	0.000	-3.037	-2.283
Iran	0.0677	0.192	0.352	0.725	-0.309	0.445
India	1.6524	0.192	8.603	0.000	1.276	2.029
Zamdia	-3.4802	0.192	-18.118	0.000	-3.857	-3.103
Kenya	-1.4908	0.192	-7.761	0.000	-1.868	-1.114
Zimbabwe	-3.3041	0.192	-17.202	0.000	-3.681	-2.927
Egypt	-1.7405	0.192	-9.061	0.000	-2.117	-1.364
Sudan	-3.4706	0.192	-18.068	0.000	-3.847	-3.094
Somalia	-3.6962	0.192	-19.243	0.000	-4.073	-3.319
Tynisia	-1.6399	0.192	-8.537	0.000	-2.017	-1.263
Togo	-3.5575	0.192	-18.521	0.000	-3.934	-3.181
Uganda	-1.1973	0.192	-6.233	0.000	-1.574	-0.820
Tanzania	-3.3143	0.192	-17.255	0.000	-3.691	-2.937
Canada	0.1402	0.192	0.730	0.466	-0.237	0.517
Colombia	-0.4784	0.192	-2.491	0.013	-0.855	-0.102
Mexico	0.1732	0.192	0.902	0.367	-0.204	0.550
Cuba	-4.9566	0.192	-25.805	0.000	-5.333	-4.580
Paraguay	-2.1285	0.192	-11.081	0.000	-2.505	-1.752
Chile	-1.2592	0.192	-6.556	0.000	-1.636	-0.882
Brazil	2.4190	0.192	12.594	0.000	2.042	2.796

Omnibus:	150.602	Durbin-Watson:	0.605
Prob(Omnibus):	0.000	Jarque-Bera (JB):	1164.189
Skew:	0.256	Prob(JB):	1.58e-253
Kurtosis:	7.719	Cond. No.	303.

Рисунок 1.25 – Регресія фіксованих змінних для змінної IDS

1.3.3 Побудова моделі регресії випадкових ефектів.

Першим кроком для побудови моделі випадкових ефектів, необхідно розрахувати σ^2_ϵ і σ^2_μ - дисперсії компонентів похибки μ і ємоделі фіксованих ефектів та об'єднаної моделі та знайти різницю між ними (рис. 1.26).

```
sigma2_epsilon = lsdv_model_results.ssr/(n*T-(n+k+1))
print('sigma2_epsilon = ' + str(sigma2_epsilon))
```

```
sigma2_epsilon = -31.901267130742646
```

```
sigma2_pooled = pooled_olsr_model_results.ssr/(n*T-(k+1))
print('sigma2_pooled = ' + str(sigma2_pooled))
```

```
sigma2_pooled = -256.29655163768166
```

```
sigma2_u = sigma2_pooled - sigma2_epsilon
print('sigma2_u = ' + str(sigma2_u))
```

```
sigma2_u = -224.39528450693902
```

Рисунок 1.26 – Розрахунок значень дисперсії компонентів похибки моделей

Обчислюємо середні значення y та X для кожної групи (тобто кожної одиниці i) на панелі даних таким чином та розраховуємо показник Тета (рис. 1.27):

```
theta = 1 - math.sqrt(c/(u))
print('theta = ' + str(theta))
```

```
theta = 0.9367805559430118
```

Рисунок 1.27 – Обчислення показника Тета

Тепер будуємо модель випадкових ефектів(рис. 1.28)

OLS Regression Results

Dep. Variable:	KAS	R-squared:	0.662
Model:	OLS	Adj. R-squared:	0.653
Method:	Least Squares	F-statistic:	73.98
Date:	Sat, 26 Nov 2022	Prob (F-statistic):	2.80e-258
Time:	13:41:40	Log-Likelihood:	-21938.
No. Observations:	1240	AIC:	4.394e+04
Df Residuals:	1207	BIC:	4.411e+04
Df Model:	32		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	1.168e+08	1.12e+07	10.414	0.000	9.48e+07	1.39e+08
Date_num	8.201e+04	3.76e+04	2.183	0.029	8321.468	1.56e+05
UK	-4.902e+07	3.54e+07	-1.384	0.167	-1.19e+08	2.05e+07
Germany	1.259e+08	3.54e+07	3.555	0.000	5.64e+07	1.95e+08
Italy	-9.15e+06	3.54e+07	-0.258	0.796	-7.87e+07	6.04e+07
Poland	-5.396e+06	3.54e+07	-0.152	0.879	-7.49e+07	6.41e+07
France	2.023e+08	3.54e+07	5.711	0.000	1.33e+08	2.72e+08
Hungary	-9.153e+07	3.54e+07	-2.584	0.010	-1.61e+08	-2.2e+07
Moldova	-1.096e+08	3.54e+07	-3.094	0.002	-1.79e+08	-4.01e+07
Slovakia	-1.115e+08	3.54e+07	-3.148	0.002	-1.81e+08	-4.2e+07
Indonezia	-7.33e+07	3.54e+07	-2.069	0.039	-1.43e+08	-3.8e+06
Japan	1.699e+08	3.54e+07	4.795	0.000	1e+08	2.39e+08
China	1.511e+09	3.54e+07	42.647	0.000	1.44e+09	1.58e+09
Vietnam	-3.76e+07	3.54e+07	-1.061	0.289	-1.07e+08	3.19e+07
Armenia	-1.163e+08	3.54e+07	-3.284	0.001	-1.86e+08	-4.68e+07
Azerbaijan	-1.159e+08	3.54e+07	-3.270	0.001	-1.85e+08	-4.64e+07
Iran	-7.528e+07	3.54e+07	-2.125	0.034	-1.45e+08	-5.77e+06
India	5.721e+07	3.54e+07	1.615	0.107	-1.23e+07	1.27e+08
Zamdia	-1.17e+08	3.54e+07	-3.303	0.001	-1.86e+08	-4.75e+07
Kenya	-1.094e+08	3.54e+07	-3.089	0.002	-1.79e+08	-3.99e+07
Zimbabwe	-1.168e+08	3.54e+07	-3.298	0.001	-1.86e+08	-4.73e+07
Egypt	-1.107e+08	3.54e+07	-3.126	0.002	-1.8e+08	-4.12e+07
Sudan	-1.169e+08	3.54e+07	-3.299	0.001	-1.86e+08	-4.74e+07
Somalia	-1.171e+08	3.54e+07	-3.307	0.001	-1.87e+08	-4.76e+07
Tynisia	-1.103e+08	3.54e+07	-3.114	0.002	-1.8e+08	-4.08e+07
Togo	-1.17e+08	3.54e+07	-3.303	0.001	-1.87e+08	-4.75e+07
Canada	-6.731e+07	3.54e+07	-1.900	0.058	-1.37e+08	2.19e+06
Colombia	-9.538e+07	3.54e+07	-2.693	0.007	-1.65e+08	-2.59e+07
Mexico	-5.944e+07	3.54e+07	-1.678	0.094	-1.29e+08	1.01e+07
Cuba	-1.179e+08	3.54e+07	-3.327	0.001	-1.87e+08	-4.84e+07
Chile	-1.084e+08	3.54e+07	-3.061	0.002	-1.78e+08	-3.89e+07
Brazil	2.629e+08	3.54e+07	7.421	0.000	1.93e+08	3.32e+08
Ukraine	-7.687e+07	3.54e+07	-2.170	0.030	-1.46e+08	-7.37e+06

Omnibus:	1386.189	Durbin-Watson:	0.725
Prob(Omnibus):	0.000	Jarque-Bera (JB):	118570.476
Skew:	5.516	Prob(JB):	0.00
Kurtosis:	49.618	Cond. No.	1.92e+03

Рисунок 1.28 – Побудова моделі регресії випадкових ефектів

Рівняння регресії з випадковим ефектом виглядає наступним чином (1.7):

$$KAS = 1.168e + 08 + 8.201e + 04 \cdot Data_{num} - 4.902e + 07 \cdot UK1. + 259e + 08 \cdot Germany + \cdots - 7.687 \cdot Ukraine + \epsilon \quad (1.7)$$

Обчислена дисперсія σ²_u було оцінено як - -224,4, а σ²_є було оцінено як – 31,9. Таким чином, частка загальної дисперсії, яка може бути віднесена до випадкового ефекту окремої одиниці, дорівнює (1.8):

$$\frac{-224.4}{-224.7+(-31.9)} = 0.87 \quad (1.8)$$

Це означає, що на 87% присутній випадковий ефект у моделі, але дивлячись на показник скоригованого R, який дорівнює 0,653, можна зробити висновок, що ця модель не є кращою за об’єднану модель та модель фіксованого ефекту.

Тестова статистика тесту LM (18,0083) більша, ніж критичне значення Chi-squared = 3,84146 при α=0,05, що означає, що випадковий ефект є значущим при альфа 0,05.

1.4 Побудова регресійних моделей для змінної «Intrusion Detection Scan»

1.4.1 Побудова об’єднаної регресійної моделі (Pooled OLS).

Необхідно визначити залежну та незалежні змінні.

Залежна змінна – це IDS, яка показує потік даних з виявлених мережевих атак. Визначаємо вхідні дані моделі (рис. 1.29):

```
y = 'LN_IDS'
x = df_data.drop(['MAV', 'KAS', 'IDS', 'Date', 'Country','LN_IDS'], axis=1)
```

Рисунок 1.29 – Вхідні змінні моделі

Результати моделі представлені на рисунку 1.30.

OLS Regression Results

Dep. Variable:	LN_IDS	R-squared:	0.959
Model:	OLS	Adj. R-squared:	0.958
Method:	Least Squares	F-statistic:	703.9
Date:	Sat, 26 Nov 2022	Prob (F-statistic):	0.00
Time:	12:22:28	Log-Likelihood:	-787.96
No. Observations:	1240	AIC:	1658.
Df Residuals:	1199	BIC:	1868.
Df Model:	40		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Date_num	0.0022	0.001	1.479	0.139	-0.001	0.005
Afghanistan	7.1437	0.087	82.387	0.000	6.974	7.314
Armenia	8.7331	0.087	100.716	0.000	8.563	8.903
Azerbaijan	8.1421	0.087	93.901	0.000	7.972	8.312
Brazil	12.7651	0.087	147.216	0.000	12.595	12.935
Canada	11.2622	0.087	129.884	0.000	11.092	11.432
Chile	11.8809	0.087	137.019	0.000	11.711	12.051
China	13.9827	0.087	161.258	0.000	13.813	14.153
Colombia	11.7506	0.087	135.517	0.000	11.581	11.921
Costa Rica	9.7092	0.087	111.973	0.000	9.539	9.879
Cuba	6.1282	0.087	70.674	0.000	5.958	6.298
Czech Republic	10.2438	0.087	118.138	0.000	10.074	10.414
Egypt	11.3155	0.087	130.499	0.000	11.145	11.486
France	12.6110	0.087	145.440	0.000	12.441	12.781
Germany	12.4079	0.087	143.097	0.000	12.238	12.578
Hungary	9.4115	0.087	108.540	0.000	9.241	9.582
India	12.2508	0.087	141.285	0.000	12.081	12.421
Indonezia	12.7813	0.087	147.403	0.000	12.611	12.951
Iran	11.7913	0.087	135.985	0.000	11.621	11.961
Italy	12.2955	0.087	141.800	0.000	12.125	12.466
Japan	9.6814	0.087	111.653	0.000	9.511	9.851
Kenya	10.5971	0.087	122.213	0.000	10.427	10.767
Mexico	12.9486	0.087	149.332	0.000	12.778	13.119
Moldova	8.9293	0.087	102.980	0.000	8.759	9.099
Paraguay	9.1457	0.087	105.475	0.000	8.976	9.316
Poland	11.6987	0.087	134.918	0.000	11.529	11.869
Slovakia	12.0494	0.087	138.963	0.000	11.879	12.220
Somalia	4.3781	0.087	50.491	0.000	4.208	4.548
South Korea	11.2754	0.087	130.036	0.000	11.105	11.446
Sudan	10.6387	0.087	122.693	0.000	10.469	10.809
Tanzania	9.3409	0.087	107.726	0.000	9.171	9.511
Togo	5.7743	0.087	66.593	0.000	5.604	5.944
Tynisia	10.6471	0.087	122.791	0.000	10.477	10.817
UK	11.0886	0.087	127.882	0.000	10.919	11.259
Uganda	8.6322	0.087	99.553	0.000	8.462	8.802
Ukraine	9.7300	0.087	112.214	0.000	9.560	9.900
United States	13.2521	0.087	152.833	0.000	13.082	13.422
Venezuala	10.2241	0.087	117.912	0.000	10.054	10.394
Vietnam	12.8515	0.087	148.213	0.000	12.681	13.022
Zandia	6.9104	0.087	79.696	0.000	6.740	7.081
Zimbabwe	8.1591	0.087	94.097	0.000	7.989	8.329

Omnibus:	599.721	Durbin-Watson:	1.077
Prob(Omnibus):	0.000	Jarque-Bera (JB):	14763.177
Skew:	-1.692	Prob(JB):	0.00
Kurtosis:	19.562	Cond. No.	238.

Рисунок 1.30 – Об’єднана регресійна модель для змінної IDS

Після побудови об’єднаної регресійної моделі, отримуємо наступне рівняння регресії (1.9):

$$\begin{aligned} IDS = 00.22 + 7.14 \cdot Afganistan + 8.73 \cdot Armenia + 6.91 \cdot Zandia + \\ +8.16 \cdot Zimbabwe\ 6 + \epsilon \end{aligned} \quad (1.9)$$

Скоригований R-квадрат становить 0,958 або 95,8 %. Це, безумовно гарний результат.

F - тест для регресії дав тестову статистику 703,9 із значенням p 0,00, що дозволяє зробити висновок, що оцінки коефіцієнтів моделі є значущими при p < 0.001. оцінки коефіцієнтів моделі є спільно значущими при p < 0,001.

Log-правдоподібність моделі становить -787,96, а показник AIC 1658.

Проаналізуємо залишкові похибки моделі для нормальності, гетероскедастичності та кореляції - трьох властивостей, які впливають на відповідність лінійної моделі (рис. 1.31).

```
print(pooled_olsr_model_results.resid)
print('Mean value of residual errors='+str(pooled_olsr_model_results.resid.mean()))
```

```
0       0.179944
1       0.178803
2       0.216327
3       0.238730
4       0.165151
          ...
1235   -0.152933
1236   -0.247574
1237   -0.222308
1238   -0.104814
1239   -0.041814
Length: 1240, dtype: float64
Mean value of residual errors=6.778448770025915e-15
```

Рисунок 1.31 – Залишкові похибки моделі

1.4.2 Побудова моделі регресії фіксованих ефектів.

Визначаємо залежну та незалежні змінні (рис. 1.32):

```
y_var_name = 'LN_IDS'
X_var_names = df.drop(['MAV', 'KAS', 'IDS', 'Date', 'Country', 'Country_id'], axis=1)
```

Рисунок 1.32 - Визначення вхідних даних моделі

Визначаємо всі країни, які будуть незалежними змінними, які впливають на залежну та будуємо рівняння регресії (рис. 1.33):

```
lsdv_expr = y_var_name + ' ~ '
i = 0
for X_var_name in X_var_names:
    if i > 0:
        lsdv_expr = lsdv_expr + ' + ' + X_var_name
    else:
        lsdv_expr = lsdv_expr + X_var_name
    i = i + 1
for dummy_name in unit_names[:-1]:
    lsdv_expr = lsdv_expr + ' + ' + dummy_name

print('Regression expression for OLS with dummies=' + lsdv_expr)
```

```
Regression expression for OLS with dummies=LN_IDS ~ Date_num + Germany + Italy + UK + Poland + France + Hungary + Moldova + Slo
vakia + Afghanistan + Indonezia + Japan + China + Vietnam + Armenia + Azerbaijan + Iran + India + Zamdia + Kenya + Zimbabwe + E
gypt + Sudan + Somalia + Tynisia + Togo + Uganda + Tanzania + Canada + Colombia + Mexico + Cuba + Paraguay + Chile + Brazil
```

Рисунок 1.33 – Побудова рівняння регресії

OLS Regression Results

Dep. Variable:	LN_IDS	R-squared:	0.914
Model:	OLS	Adj. R-squared:	0.912
Method:	Least Squares	F-statistic:	366.4
Date:	Sat, 26 Nov 2022	Prob (F-statistic):	0.00
Time:	12:38:52	Log-Likelihood:	-1248.3
No. Observations:	1240	AIC:	2569.
Df Residuals:	1204	BIC:	2753.
Df Model:	35		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	10.7391	0.060	179.151	0.000	10.621	10.857
Date_num	0.0022	0.002	1.023	0.307	-0.002	0.006
Germany	1.6688	0.130	12.801	0.000	1.413	1.925
Italy	1.5564	0.130	11.938	0.000	1.301	1.812
UK	0.3496	0.130	2.681	0.007	0.094	0.605
Poland	0.9596	0.130	7.361	0.000	0.704	1.215
France	1.8720	0.130	14.359	0.000	1.616	2.128
Hungary	-1.3276	0.130	-10.183	0.000	-1.583	-1.072
Moldova	-1.8097	0.130	-13.882	0.000	-2.066	-1.554
Slovakia	1.3103	0.130	10.051	0.000	1.055	1.566
Afghanistan	-3.5954	0.130	-27.579	0.000	-3.851	-3.340
Indonezia	2.0422	0.130	15.665	0.000	1.786	2.298
Japan	-1.0577	0.130	-8.113	0.000	-1.313	-0.802
China	3.2436	0.130	24.880	0.000	2.988	3.499
Vietnam	2.1125	0.130	16.204	0.000	1.857	2.368
Armenia	-2.0060	0.130	-15.387	0.000	-2.262	-1.750
Azerbaijan	-2.5970	0.130	-19.920	0.000	-2.853	-2.341
Iran	1.0522	0.130	8.071	0.000	0.796	1.308
India	1.5117	0.130	11.596	0.000	1.256	1.767
Zamdia	-3.8287	0.130	-29.368	0.000	-4.084	-3.573
Kenya	-0.1420	0.130	-1.089	0.276	-0.398	0.114
Zimbabwe	-2.5799	0.130	-19.790	0.000	-2.836	-2.324
Egypt	0.5764	0.130	4.422	0.000	0.321	0.832
Sudan	-0.1004	0.130	-0.770	0.441	-0.356	0.155
Somalia	-6.3610	0.130	-48.793	0.000	-6.617	-6.105
Tynisia	-0.0919	0.130	-0.705	0.481	-0.348	0.164
Togo	-4.9648	0.130	-38.083	0.000	-5.221	-4.709
Uganda	-2.1069	0.130	-16.161	0.000	-2.363	-1.851
Tanzania	-1.3982	0.130	-10.725	0.000	-1.654	-1.142
Canada	0.5231	0.130	4.013	0.000	0.267	0.779
Colombia	1.0116	0.130	7.759	0.000	0.756	1.267
Mexico	2.2095	0.130	16.948	0.000	1.954	2.465
Cuba	-4.6109	0.130	-35.369	0.000	-4.867	-4.355
Paraguay	-1.5934	0.130	-12.222	0.000	-1.849	-1.338
Chile	1.1418	0.130	8.758	0.000	0.886	1.398
Brazil	2.0260	0.130	15.541	0.000	1.770	2.282

Omnibus:	251.203	Durbin-Watson:	0.548
Prob(Omnibus):	0.000	Jarque-Bera (JB):	2683.788
Skew:	0.608	Prob(JB):	0.00
Kurtosis:	10.104	Cond. No.	303.

Рисунок 1.34 – Регресія фіксованих змінних для змінної IDS

Рівняння регресії з фіксованими змінними виглядає наступним чином (1.10):

$$\text{IDS} = 10.74 + 00.22 \cdot \text{Germany} + 1.53 \cdot \text{Italy} + \ldots + 1.14 \cdot \text{Chile} + 2.03 \cdot \text{Brazil} + \epsilon \quad (1.10)$$

Скориговане значення R-квадрат дорівнює 0,912, або 91,2% - значення показує дуже хорошу відповідність між незалежними змінними та залежною.

F - тест для регресійного аналізу перевіряє, чи всі коефіцієнти моделі є спільно значущими, і, отже, чи відповідність моделі FE краща, ніж у попередній моделі. Статистика F-тесту 366,4 є значною при p < 0,00, що означає, що відповідність моделі справді краща, ніж в об'єднаній регресійній моделі.

1.4.3 Побудова моделі регресії випадкових ефектів.

Першим кроком для побудови моделі випадкових ефектів, необхідно розрахувати σ^2_ϵ і σ^2_μ - дисперсії компонентів похибки μ і є моделі фіксованих ефектів та об'єднаної моделі та знайти різницю між ними (рис. 1.35).

```
sigma2_epsilon = lsdv_model_results.ssr/(n*T-(n+k+1))
print('sigma2_epsilon = ' + str(sigma2_epsilon))
```

```
sigma2_epsilon = -14.695039254289002
```

```
sigma2_pooled = pooled_olsr_model_results.ssr/(n*T-(k+1))
print('sigma2_pooled = ' + str(sigma2_pooled))
```

```
sigma2_pooled = -129.37891349913338
```

```
sigma2_u = sigma2_pooled - sigma2_epsilon
print('sigma2_u = ' + str(sigma2_u))
```

```
sigma2_u = -114.68387424484438
```

Рисунок 1.35 – Розрахунок значень дисперсії компонентів похибки моделей

Обчислюємо середні значення y та X для кожної групи (тобто кожної одиниці і) на панелі даних таким чином та розраховуємо показник Тета (рис. 1.36):

```
theta = 1 - math.sqrt(c/(u))
print('theta = ' + str(theta))
```

```
theta = 0.939969296454986
```

Рисунок 1.36 – Обчислення показника Тета

Тепер будуємо модель випадкових ефектів (рис. 1.37):

OLS Regression Results

Dep. Variable:	IDS	R-squared:	0.741
Model:	OLS	Adj. R-squared:	0.734
Method:	Least Squares	F-statistic:	107.7
Date:	Sat, 26 Nov 2022	Prob (F-statistic):	0.00
Time:	13:37:43	Log-Likelihood:	-16342.
No. Observations:	1240	AIC:	3.275e+04
Df Residuals:	1207	BIC:	3.292e+04
Df Model:	32		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	1.451e+06	1.3e+05	11.194	0.000	1.2e+06	1.71e+06
Date_num	249.9978	412.073	0.607	0.544	-558.462	1058.458
Germany	2.798e+06	4.09e+05	6.835	0.000	1.99e+06	3.6e+06
Italy	2.332e+06	4.09e+05	5.698	0.000	1.53e+06	3.14e+06
UK	-3.164e+05	4.09e+05	-0.773	0.440	-1.12e+06	4.87e+05
Poland	7.927e+05	4.09e+05	1.937	0.053	-1.04e+04	1.6e+06
France	3.732e+06	4.09e+05	9.117	0.000	2.93e+06	4.53e+06
Hungary	-1.241e+06	4.09e+05	-3.032	0.002	-2.04e+06	-4.38e+05
Moldova	-1.318e+06	4.09e+05	-3.219	0.001	-2.12e+06	-5.14e+05
Slovakia	4.456e+06	4.09e+05	10.886	0.000	3.65e+06	5.26e+06
Indonezia	4.802e+06	4.09e+05	11.733	0.000	4e+06	5.61e+06
Japan	-1.174e+06	4.09e+05	-2.868	0.004	-1.98e+06	-3.71e+05
China	1.913e+07	4.09e+05	46.746	0.000	1.83e+07	1.99e+07
Vietnam	5.699e+06	4.09e+05	13.924	0.000	4.9e+06	6.5e+06
Armenia	-1.341e+06	4.09e+05	-3.275	0.001	-2.14e+06	-5.38e+05
Azerbaijan	-1.383e+06	4.09e+05	-3.379	0.001	-2.19e+06	-5.8e+05
Iran	8.704e+05	4.09e+05	2.126	0.034	6.73e+04	1.67e+06
India	2.23e+06	4.09e+05	5.449	0.000	1.43e+06	3.03e+06
Zamdia	-1.436e+06	4.09e+05	-3.508	0.000	-2.24e+06	-6.33e+05
Kenya	-5.993e+05	4.09e+05	-1.464	0.143	-1.4e+06	2.04e+05
Zimbabwe	-1.392e+06	4.09e+05	-3.402	0.001	-2.2e+06	-5.89e+05
Egypt	-6192.2912	4.09e+05	-0.015	0.988	-8.09e+05	7.97e+05
Sudan	-7.24e+05	4.09e+05	-1.769	0.077	-1.53e+06	7.9e+04
Somalia	-1.452e+06	4.09e+05	-3.548	0.000	-2.26e+06	-6.49e+05
Tynisia	-7.084e+05	4.09e+05	-1.731	0.084	-1.51e+06	9.47e+04
Togo	-1.448e+06	4.09e+05	-3.537	0.000	-2.25e+06	-6.45e+05
Canada	-1.019e+05	4.09e+05	-0.249	0.803	-9.05e+05	7.01e+05
Colombia	1.092e+06	4.09e+05	2.668	0.008	2.89e+05	1.89e+06
Mexico	6.769e+06	4.09e+05	16.538	0.000	5.97e+06	7.57e+06
Cuba	-1.446e+06	4.09e+05	-3.532	0.000	-2.25e+06	-6.43e+05
Chile	1.065e+06	4.09e+05	2.602	0.009	2.62e+05	1.87e+06
Brazil	4.64e+06	4.09e+05	11.336	0.000	3.84e+06	5.44e+06
Ukraine	-1.161e+06	4.09e+05	-2.836	0.005	-1.96e+06	-3.58e+05

Omnibus:	1540.323	Durbin-Watson:	0.882
Prob(Omnibus):	0.000	Jarque-Bera (JB):	307346.453
Skew:	6.282	Prob(JB):	0.00
Kurtosis:	79.097	Cond. No.	2.02e+03

Рисунок 1.37 – Побудова моделі регресії випадкових ефектів

Рівняння регресії з випадковим ефектом виглядає наступним чином (1.11):

$$\begin{aligned} \text{IDS} = {} & 1.451\text{e} + 06 + 249.998 \cdot \text{Data}_{\text{num}} + 2.798\text{t} + 06 \cdot \text{Germany} \\ & + 3.322\text{t} + 06 \cdot \text{Italy} + \ldots + 4.64\text{t} + 06 \cdot \text{Chile} - 1.16\text{t} + 06 \\ & \cdot \text{Ukraine} + \epsilon \end{aligned} \quad (1.11)$$

Обчислена дисперсія σ²_u було оцінено як - -114,68, а σ²_є було оцінено як – 14,70. Таким чином, частка загальної дисперсії, яка може бути віднесена до випадкового ефекту окремої одиниці, дорівнює (1.12):

$$\frac{-114.68}{-114.68+(-14.70)} = 0{,}89 \quad (1.12)$$

Це означає, що на 89% присутній випадковий ефект у моделі, але дивлячись на показник скоригованого R, який дорівнює 0,734, можна зробити висновок, що ця модель не є кращою за об’єднану модель та модель фіксованого ефекту.

1.5 Оцінка отриманих результатів

Проведемо оцінку отриманих результатів побудованих регресій для кожного виду кібератак. В таблиці 1.2 представлені результати для MAV.

Таблиця 1.2 – Порівняння результатів побудованих моделей для MAV

Показник	Pooled OLS Regression Model	The Fixed Effects Regression Model	The Random Effects Regression Model
Скоригований R^2	0,769	0,702	0,640
Log-Likelihood	-1785,7	-1946,1	-12222
AIC	3653	3964	24510
MRE	5,804e-15	4,410e-13	1,017e-10

Скоригований R-квадрат об’єднаної моделі (Pooled OLS) дорівнює 76.9% значно кращий порівняно з моделями фіксованого та випадкового ефекту. Pooled OLS також забезпечує невелике збільшення логарифмічної ймовірності до -

1785.7 порівняно з іншими моделями, я також показник AIC теж показав кращий результат. Середня похибка залишків теж є найменшою для об'єднаної моделі. Можна зробити висновок, що об'єднана модель є найкращою для опису залежної змінної MAV.

В таблиці 1.3 представлені результати для KAS.

Таблиця 1.3 – Порівняння результатів побудованих моделей для KAS

Показник	Pooled OLS Regression Model	The Fixed Effects Regression Model	The Random Effects Regression Model
Скоригований R^2	0,920	0,817	0,653
Log-Likelihood	-1277,8	-1728,9	-21938
AIC	2506	3530	43940
MRE	-6,66e-14	8,52e-13	3,92e-08

Скоригований R-квадрат об'єднаної моделі (Pooled OLS) дорівнює 92,0% значно кращий порівняно з моделями фіксованого та випадкового ефекту. Pooled OLS також забезпечує невелике збільшення логарифмічної ймовірності до -1277.8 порівняно з іншими моделями, я також показник AIC теж показав кращий результат. Середня похибка залишків теж є найменшою для об'єднаної моделі. Можна зробити висновок, що об'єднана модель є найкращою для опису залежної змінної KAS.

В таблиці 1.4 представлені результати для IDS.

Таблиця 1.4 – Порівняння результатів побудованих моделей для IDS

Показник	Pooled OLS Regression Model	The Fixed Effects Regression Model	The Random Effects Regression Model
Скоригований R^2	0,958	0,912	0,734
Log-Likelihood	-787,96	-1248,3	-16342
AIC	1658	2569	32750
MRE	6,77e-15	6,51e-13	1,41e-08

Скоригований R-квадрат об'єднаної моделі (Pooled OLS) дорівнює 95,8% значно кращий порівняно з моделями фіксованого та випадкового ефекту. Pooled OLS також забезпечує невелике збільшення логарифмічної ймовірності до -787.96 порівняно з іншими моделями, я також показник AIC теж показав кращий

результат. Середня похибка залишків теж є найменшою для об’єднаної моделі. Можна зробити висновок, що об’єднана модель є найкращою для опису залежної змінної IDS.

1.6 Прогнозування трендів кібератак

1.6.1 Прогнозування трендів кібератаки виду MAV

Визначивши найкращу модель для моделювання виду кібератаки MAV, спрогнозуємо тренд за допомогою об’єднаної моделі (Pooled model) та LSTM моделі.

Перш ніж побудувати прогноз необхідно поділити базу даних на дві частини - тестову та тренувальну. Головна проблема та складність цього процесу полягає у тому, як правильно поділити панельні дані на дві частини. Розподіл відбувається для кожної країни, а потім об’єднуємо тестову та тренувальну частини для кожної країни (рис. 1.38).

```
def train_test_split(data):
    size=int(len(data)*0.8)
    # for train data will be collected from each country's data which index is from 0-size (80%)
    x_train =data.drop(['MAV'], axis=1).iloc[0:size]
    # for test data will be collected from each country's  data which index is from size to the end (20%)
    x_test = data.drop(['MAV'], axis=1).iloc[size:]
    y_train=data['MAV'].iloc[0:size]
    y_test=data['MAV'].iloc[size:]
    return x_train, x_test,y_train,y_test
```

```
country=list(set(dt.Countries))
# loop each station and collect train and test data
X_train=[]
X_test=[]
Y_train=[]
Y_test=[]
for i in range(0,len(country)):
    data=dt[dt['Countries']==country[i]]
    x_train, x_test,y_train,y_test=train_test_split(data)
    X_train.append(x_train)
    X_test.append(x_test)
    Y_train.append(y_train)
    Y_test.append(y_test)
```

Рисунок 1.38 – Розподіл бази даних та тренувальну та тестову частини

Після розподілу бази даних на тестову та тренувальну частину, ідентифікуємо тренувальні та тестові залежні та незалежні змінні (рис. 1.39) [29]:

```
from sklearn.preprocessing import LabelEncoder
encoder = LabelEncoder()
#combine x train and y train as train data
train_data=pd.DataFrame()
train_data[X_train.columns]=X_train
train_data[Y_train.columns]=Y_train
train_data['Countries']= encoder.fit_transform(train_data['Countries'])
#combine x test and y test as test data
test_data=pd.DataFrame()
test_data[X_test.columns]=X_test
test_data[Y_test.columns]=Y_test
test_data['Countries']= encoder.fit_transform(test_data['Countries'])
# using the function to obtian reshaped x_train,x_test,y_train,y_test
x_train,x_test,y_train,y_test=reshape_data(train_data,test_data)
```

Рисунок 1.39 – Ідентифікація тренувальних та тестових змінних

Всі дані підготовлені до побудови прогнозної моделі на основі об’єднаної моделі та LSTM моделі [30].

Спочатку побудуємо об’єднану прогнозу модель з використанням тестового та тренувального розподілу бази даних (рис. 1.40).

```
y_pred = model.predict(X_test)
df_results = pd.DataFrame({'Actual': y_test, 'Predicted': y_pred})
df_results
```

	Actual	Predicted
780	1.609438	1.174482
817	8.043021	7.272133
363	9.110631	8.806825
308	6.848005	6.625916
1205	7.132498	7.771581
...	...	...
609	8.336390	8.479036
332	5.587249	4.932872
1088	4.828314	4.211598
1137	3.367296	6.138822
149	7.920810	7.571089

Рисунок 1.40 – Результати побудови прогнозної моделі на основі Pooled regression

Одним із способів оцінити, наскільки добре регресійна модель відповідає набору даних, є обчислення середньої квадратичної похибки кореня (RMSE), яка є метрикою, яка повідомляє нам середню відстань між прогнозованими значеннями від моделі та фактичними значеннями в наборі даних [31].

Чим нижче RMSE, тим краще дана модель здатна «підігнати» набір даних (1.13):

$$\mathrm{RMSE} = \frac{\sqrt{\sum(\mathrm{P}-\mathrm{A})^2}}{\mathrm{n}} \qquad (1.13)$$

де: P – прогнозне значення;

A – значення спостереження;

n – розмір вибірки.

Тому розраховуємо середню квадратичну похибку кореня та коефіцієнт детермінації для прогнозної моделі (рис. 1.41):

```
from sklearn.metrics import r2_score, mean_squared_error
RMSE = np.sqrt(mean_squared_error(y_test, y_pred))
r2 = r2_score(y_test, y_pred)
print('RMSE:', RMSE, 'R2:',r2)
```

```
RMSE: 1.1298507283393997 R2: 0.7140165385759949
```

Рисунок 1.41 – Розрахунок показників для прогнозної моделі для змінної MAV

Середню квадратичну похибку кореня становить 1,12985 та коефіцієнт детермінації дорівнює 0,71, що являється досить гарним результатом.

Тепер необхідно побудувати нову прогнозну модель - LSTM модель для порівняння та обрати кращу модель для прогнозування явища кібератак.

Всі дані підготовлені до побудови LSTM моделі. LSTM модель буде будуватися за допомогою Sequential() задачі (рис. 1.42) [32].

Sequential (Класифікація послідовностей) - це задача прогностичного моделювання, де є певна послідовність входів у просторі або часі, і завдання полягає в тому, щоб передбачити категорію для послідовності.

```
model = Sequential()
model.add(LSTM(60, activation='sigmoid',input_shape=(x_train.shape[1], x_train.shape[2])))
model.add(Dense(1))
model.compile(loss='mae', optimizer='adam')
# fit network
history = model.fit(x_train, y_train, epochs=1000, batch_size=64, verbose=0, shuffle=False)
```

```
# make a prediction
y_test_pre=model.predict(x_test)
```

```
9/9 [==============================] - 0s 875us/step
```

```
# make a prediction
y_test_pre=model.predict(x_test)
y_test_pre.shape,y_test.shape
```

```
9/9 [==============================] - 0s 750us/step
((279, 1), (279,))
```

Рисунок 1.42 – Побудова прогнозної LSTM моделі

Після побудови прогнозної моделі отримуємо наступні результати (рис. 1.43).

```
pa=pd.DataFrame()
pa['Data']=X_test.reset_index().Data.iloc[1:-1]
pa['Prediction']=[i[0] for i in y_test_pre][1:]
pa['Actual Values']=y_test[:-1]
pa.head()
```

	Data	Prediction	Actual Values
1	2022-09-08	5.344671	4.553877
2	2022-09-09	5.325020	4.418841
3	2022-09-10	4.153235	2.772589
4	2022-09-11	5.082784	3.912023
5	2022-09-12	5.840456	4.867534

Рисунок 1.43 – Прогнозні дані для змінної MAV

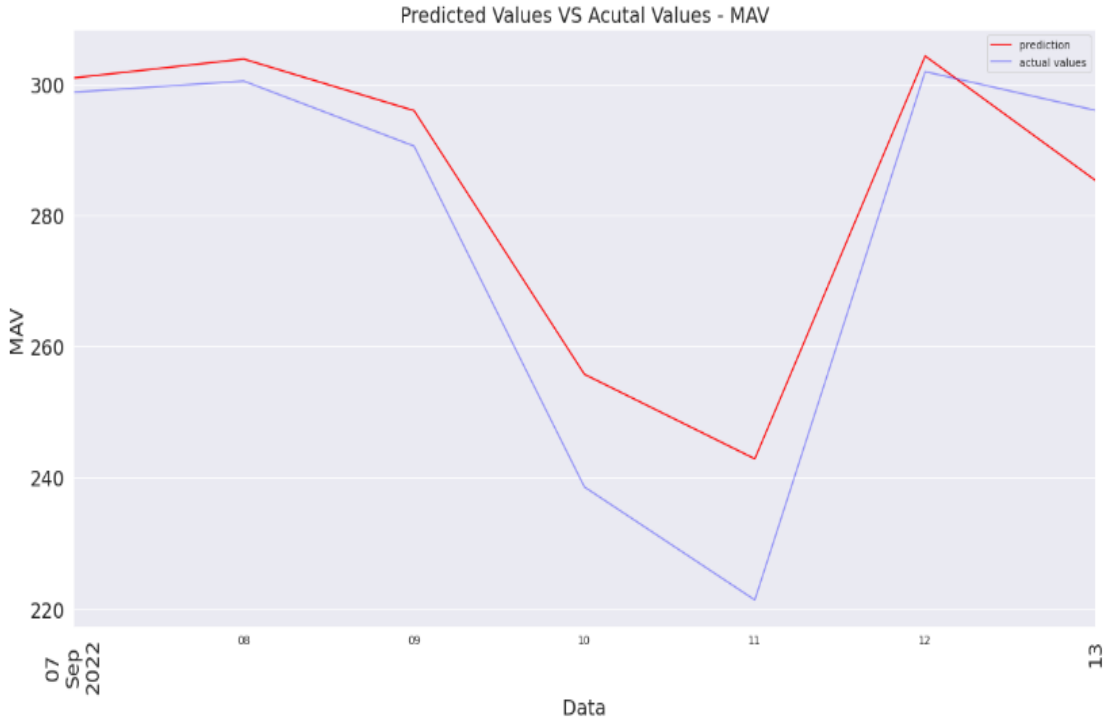


Рисунок 1.44 – Результати побудови прогнозної LSTM моделі

```
print(RMSE(y_test[:-1],[i[0] for i in y_test_pre][1:]))
```

```
0.5821494070053101
```

```
from sklearn.metrics import mean_squared_error, r2_score, mean_absolute_error
print('R2 Score: ', r2_score(y_test, y_test_pre))
```

```
R2 Score:  0.5151603123958282
```

Рисунок 1.45 – Розрахунок показників для прогнозної моделі для змінної MAV

Середню квадратичну похибку кореня становить 0,58 та коефіцієнт детермінації дорівнює 0,51, що являється теж досить гарним результатом. Але спираючись на той факт, що коефіцієнт детермінації не є повністю надійним показником для порівняння моделей, використаємо середню квадратичну похибку, яка повідомляє нам середню відстань між прогнозованими значеннями від моделі та фактичними значеннями в наборі даних.

Тому кращою прогнозною моделлю для змінної MAV є LSTM модель. Спробуємо побудувати прогнозну LSTM модель для кожної країни (рис. 1.46).

```
def normalization_train_test_split(country):
    scaler = PowerTransformer(method='yeo-johnson', standardize=True)
    scaled = scaler.fit_transform(country.drop(columns=['Country','Date']))
    # create dateframe fir scaled data
    scaled_df=pd.DataFrame(data=scaled,columns=country.drop(columns=['Country','Date']).columns)
    scaled_df['MAV']=list(country.MAV)
    X_train, X_test,Y_train,Y_test=train_test_split(scaled_df)
    #combine x train and y train as train data
    train_data=pd.DataFrame()
    train_data[X_train.columns]=X_train
    train_data['MAV']=Y_train
    #combine x test and y test as test data
    test_data=pd.DataFrame()
    test_data[X_test.columns]=X_test
    test_data['MAV']=Y_test

    # using the function to obtian reshaped x_train,x_test,y_train,y_test
    x_train,x_test,y_train,y_test=reshape_data(train_data,test_data)
    return x_train, x_test,y_train,y_test
```

```
# Loop through top 10 countries' data
#
for i in range(len(top_10_country_names)):
     # obtian one country's data
    country=df_data[df_data.Country==top_10_country_names[i]]
    # train test split, nomalization and reshape the data
    x_train, x_test,y_train,y_test=normalization_train_test_split(country)
    # model
    model = Sequential()
    model.add(LSTM(60, activation='sigmoid',input_shape=(x_train.shape[1], x_train.shape[2])))
    model.add(Dense(1))
    model.compile(loss='mae', optimizer='adamax')
    # fit network
    history = model.fit(x_train, y_train, epochs=2000, batch_size=128,  verbose=0, shuffle=False)
    # make a prediction
    y_test_pre=model.predict(x_test)
    #RMSE
    rmse=RMSE(y_test[:-1],[i[0] for i in y_test_pre][1:])
    print('{} - RMSE: {}'.format(top_10_country_names[i],rmse))
    #create new dataframe for plot
    pa=pd.DataFrame()
    pa['Date']=list(country.Date.iloc[int(len(country)*0.8):])[1:-1]
    pa['Prediction']=[i[0] for i in y_test_pre][1:]
    pa['Actual Values']=list(y_test[:-1])

    plt.figure(figsize=(20,10))
    pa.groupby('Date')['Prediction'].sum().plot(kind='line',label='prediction',color='red',alpha=1)
    pa.groupby("Date")['Actual Values'].sum().plot(kind='line',label='actual values',color='blue',alpha=0.4)
    plt.xticks(rotation=90,size=20)
    plt.yticks(size=20)

    plt.ylabel('MAV',fontsize=20)
    plt.xlabel('Date',fontsize=20)
    plt.title('Predicted Values VS Acutal Values - MAV in {}'.format(top_10_country_names[i]),fontsize=20)
    plt.legend()
```

Рисунок 1.46 – Побудова прогнозної моделі для кожної країни

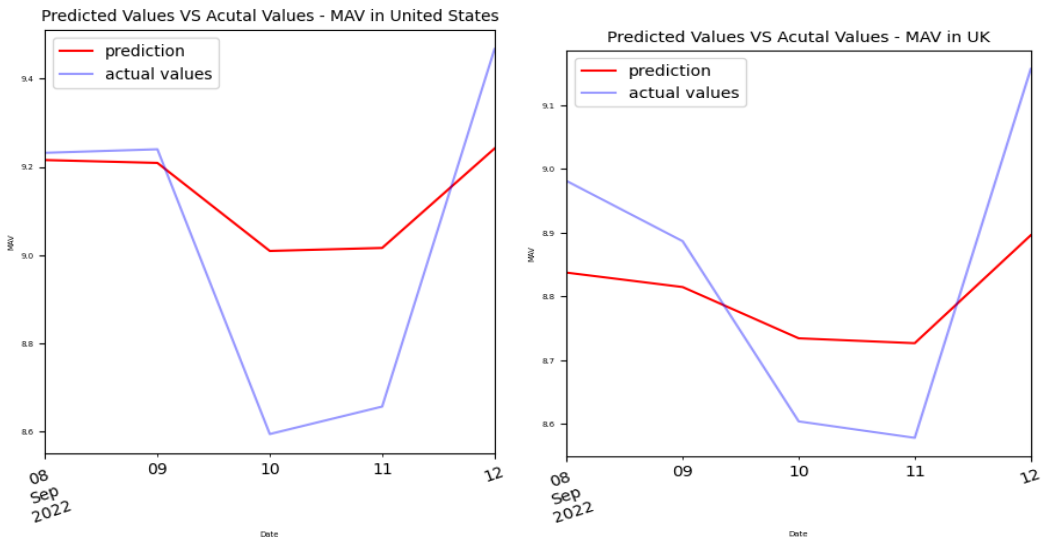


Рисунок 1.47 – Прогноз значення MAV для США та Великобританії

Отримані гарні результати для країн. Прикладом прогнозу для країн виступають США та Великобританія за значенням середньої квадратичної похибки 0,266 та 0,163 відповідно.

1.6.2 Прогнозування трендів кібератаки виду KAS

Визначивши найкращу модель для моделювання виду кібератаки KAS, спрогнозуємо тренд за допомогою об'єднаної моделі (Pooled model) та LSTM моделі.

Спочатку побудуємо об'єднану прогнозу модель з використанням тестового та тренувального розподілу бази даних (рис. 1.48):

```
y_pred = model.predict(X_test)
df_results = pd.DataFrame({'Actual': y_test, 'Predicted': y_pred})
df_results
```

	Actual	Predicted
999	14.939141	13.788484
651	9.952278	12.723177
1023	13.138237	14.310919
250	12.398757	12.584448
860	11.362103	11.019525
...	...	...
631	11.532728	10.876397
473	15.882500	15.127464
916	11.497812	11.197719
760	10.373491	11.069807
173	16.437204	16.313346

Рисунок 1.48 – Результати побудови прогнозної моделі на основі Pooled regression

Одним із способів оцінити, наскільки добре регресійна модель відповідає набору даних, є обчислення середньої квадратичної похибки кореня (RMSE). Тому розраховуємо середню квадратичну похибку кореня та коефіцієнт детермінації для прогнозної моделі (рис. 1.49):

```
from sklearn.metrics import r2_score, mean_squared_error
RMSE = np.sqrt(mean_squared_error(y_test, y_pred))
r2 = r2_score(y_test, y_pred)
print(RMSE, r2)
```

```
0.6147591605646721 0.9297993080835679
```

Рисунок 1.49 – Розрахунок показників для прогнозної моделі для змінної KAS

Середню квадратичну похибку кореня становить 0,6148 та коефіцієнт детермінації дорівнює 0,9298, що являється досить гарним результатом.

Тепер необхідно побудувати нову прогнозну модель - LSTM модель для порівняння та обрати кращу модель для прогнозування явища кібератак. Після побудови прогнозної моделі отримуємо наступні результати (рис. 1.50).

	Date	Prediction	Actual Values
1	2022-09-08	16.024281	15.841156
2	2022-09-09	16.052122	15.875068
3	2022-09-10	15.683013	15.471450
4	2022-09-11	15.695789	15.488205
5	2022-09-12	15.915629	15.735956

Рисунок 1.50 – Прогнозні дані для змінної KAS

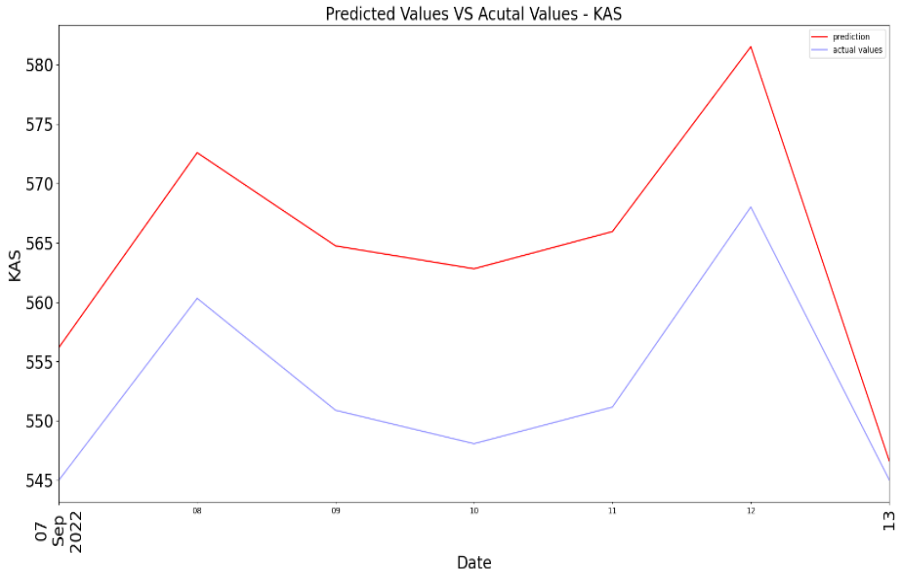


Рисунок 1.51 – Результати побудови прогнозної LSTM моделі,

```
print(RMSE(y_test[:-1],[i[0] for i in y_test_pre][1:]))
```

```
0.505295611130907
```

```
from sklearn.metrics import mean_squared_error, r2_score, mean_absolute_error
print('R2 Score: ', r2_score(y_test, y_test_pre))
```

```
R2 Score:  0.7116407077885306
```

Рисунок 1.52 – Розрахунок показників для прогнозної моделі для змінної KAS

Середню квадратичну похибку кореня становить 0,51 та коефіцієнт детермінації дорівнює 0,71, що являється теж досить гарним результатом. Але спираючись на той факт, що коефіцієнт детермінації не є повністю надійним показником для порівняння моделей, використаємо середню квадратичну похибку, яка повідомляє нам середню відстань між прогнозованими значеннями від моделі та фактичними значеннями в наборі даних.

Тому кращою прогнозною моделлю для змінної KAS є LSTM модель. Спробуємо побудувати прогнозну LSTM модель для кожної країни (рис.1.53).

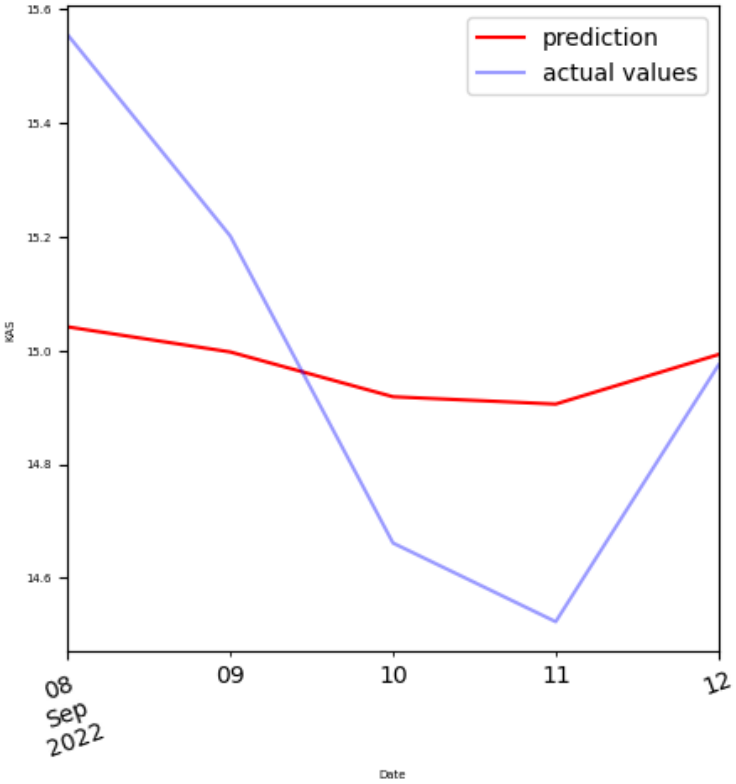


Рисунок 1.53 – Прогноз значення KAS для України

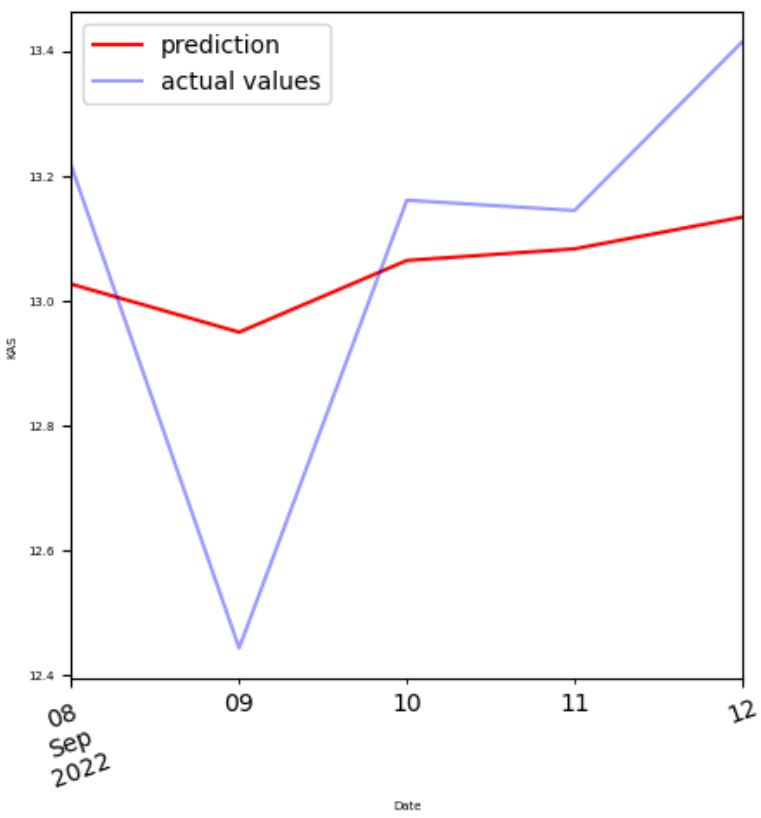


Рисунок 1.54 – Прогноз значення KAS для Тунісу

Отримані гарні результати для країн. Прикладом прогнозу для країн виступають Україна та Туніс за значенням середньої квадратичної похибки 0,329 та 0,277 відповідно.

1.6.3 Прогнозування трендів кібератаки виду IDS

Визначивши найкращу модель для моделювання виду кібератаки IDS, спрогнозуємо тренд за допомогою об’єднаної моделі (Pooled model) та LSTM моделі.

Спочатку побудуємо об’єднану прогнозу модель з використанням тестового та тренувального розподілу бази даних (рис. 1.55).

```
y_pred = model.predict(X_test)
df_results = pd.DataFrame({'Actual': y_test, 'Predicted': y_pred})
df_results
```

	Actual	Predicted
729	11.349959	11.356542
925	9.655667	9.319099
1070	10.403202	10.244445
249	9.120963	12.228848
961	13.244149	13.253300
...	...	...
880	9.081597	8.644902
99	11.076465	11.119290
332	7.352441	7.202662
400	9.927790	9.709871
418	11.359040	11.312384

Рисунок 1.55 – Результати побудови прогнозної моделі на основі Pooled regression

Одним із способів оцінити, наскільки добре регресійна модель відповідає набору даних, є обчислення середньої квадратичної похибки кореня (RMSE).

Тому розраховуємо середню квадратичну похибку кореня та коефіцієнт детермінації для прогнозної моделі (рис. 1.56):

```
from sklearn.metrics import r2_score, mean_squared_error
RMSE = np.sqrt(mean_squared_error(y_test, y_pred))
r2 = r2_score(y_test, y_pred)
print(RMSE, r2)
```

```
0.5360467594061045 0.9429578957759858
```

Рисунок 1.56 – Розрахунок показників для прогнозної моделі для змінної IDS

Середню квадратичну похибку кореня становить 0,536 та коефіцієнт детермінації дорівнює 0,943, що являється досить гарним результатом.

Тепер необхідно побудувати нову прогнозну модель - LSTM модель для порівняння та обрати кращу модель для прогнозування явища кібератак. Після побудови прогнозної моделі отримуємо наступні результати (рис. 1.57).

	Date	Prediction	Actual Values
1	2022-09-08	9.170294	9.013839
2	2022-09-09	8.823923	8.629629
3	2022-09-10	8.770832	8.567126
4	2022-09-11	9.114446	8.941807
5	2022-09-12	9.222954	9.060099

Рисунок 1.57 – Прогнозні дані для змінної IDS

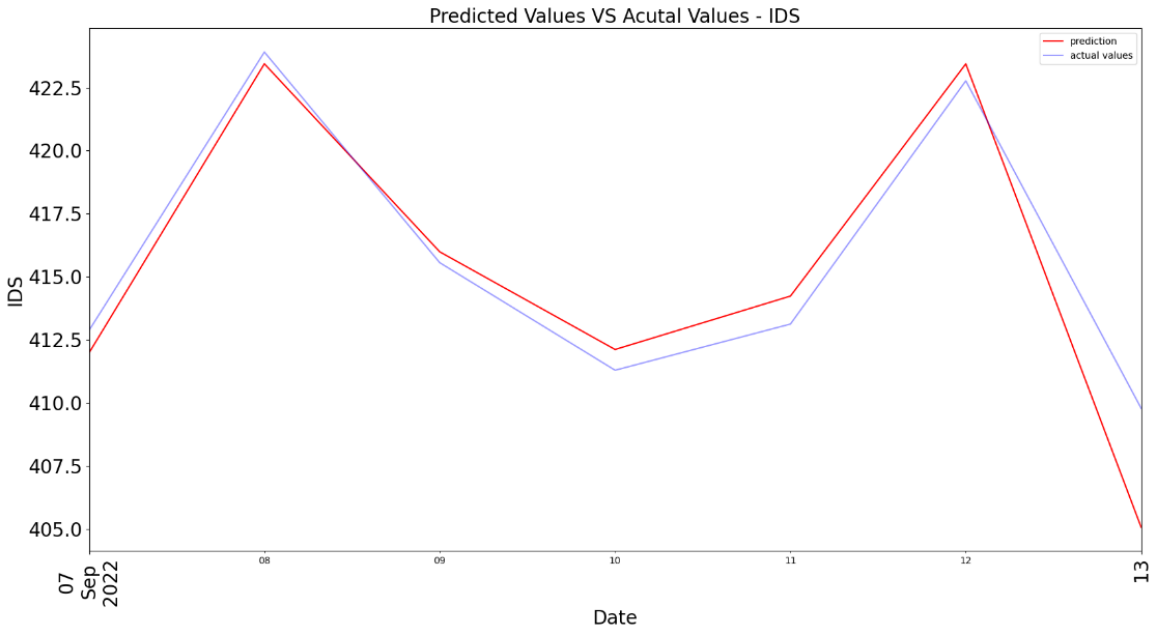


Рисунок 1.58 – Результати побудови прогнозної LSTM моделі

```
print(RMSE(y_test[:-1],[i[0] for i in y_test_pre][1:]))
```

```
0.4595955566642357
```

```
from sklearn.metrics import mean_squared_error, r2_score, mean_absolute_error
print('R2 Score: ', r2_score(y_test, y_test_pre))
```

```
R2 Score:  0.7257339291988435
```

Рисунок 1.59 – Розрахунок показників для прогнозної моделі для змінної IDS

Середню квадратичну похибку кореня становить 0,459 та коефіцієнт детермінації дорівнює 0,726, що являється теж досить гарним результатом.

Але спираючись на той факт, що коефіцієнт детермінації не є повністю надійним показником для порівняння моделей, використаємо середню квадратичну похибку, яка повідомляє нам середню відстань між прогнозованими значеннями від моделі та фактичними значеннями в наборі даних. Тому кращою прогнозною моделлю для змінної KAS є LSTM модель.

Спробуємо побудувати прогнозну LSTM модель для кожної країни (рис. 1.60-1.61).

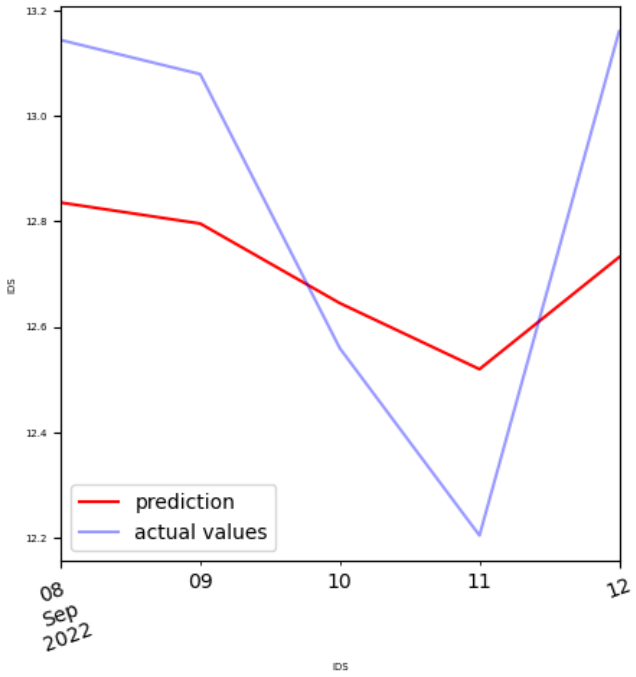


Рисунок 1.60 – Прогноз значення KAS для В’єтнаму

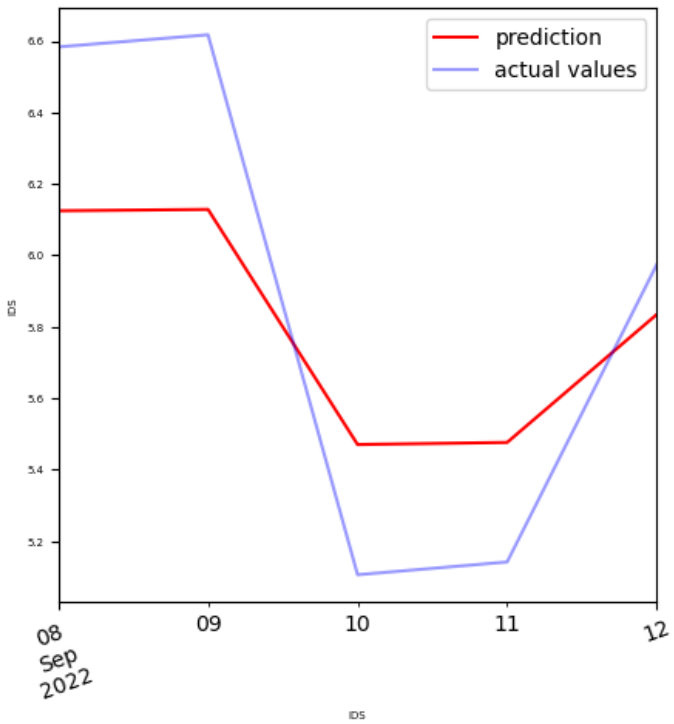


Рисунок 1.61 – Прогноз значення KAS для Того

Отримані гарні результати для країн. Прикладом прогнозу для країн виступають В’єтнам та Того за значенням середньої квадратичної похибки 0,304 та 0,377 відповідно.

У даному дослідженні було здійснено аналіз, моделювання та прогнозування трендів кібератак за допомогою побудови математичних моделей та регресії для панельних даних, а саме об’єднаної моделі, моделі фіксованих ефектів, випадкових ефектів та LSTM моделі. Відповідні розрахунки було проведено із використанням сучасної мови програмування Python. Вважаємо, що побудована об’єднана модель буде одним із найкращих методів, що дозволяє змоделювати тренди кібератак та продемонструвала гарні результати скоригованого коефіцієнту детермінації, залишкових похибок моделі та параметру Акайка (AIC). Також було побудовано прогнозну модель на основі об’єднаної та LSTM модель. На останньому етапі було проведено порівняння побудованих прогнозних моделей для кожної незалежної змінної, в результаті чого найкращі результати продемонструвала LSTM, не зважаючи на гірший результат скоригованого коефіцієнту детермінації, середня квадратна похибка кореня показала кращий результат, що означає кращу здатність «підігнати» дані набору даних для прогнозування. Щоб мати постійне уявлення про ймовірні кібератаки, результати повинні регулярно доповнюватися, оновлюватися для використання їх у реальних умовах, що дозволить вчасно реагувати на злочинні дії та попереджати їх виникнення.

Пункти 1.2-1.6 було виконано із використанням матеріалів публікацій виконавців [33].

2 МУЛЬТИСЕРВІСНА МОДЕЛЬ КОМПЛЕКСНОЇ ОЦІНКИ ТА ПРІОРИТЕЗАЦІЇ РИЗИКІВ ЛЕГАЛІЗАЦІЇ КРИМІНАЛЬНИХ ДОХОДІВ ТА КІБЕРРИЗИКІВ

2.1 Теоретичні основи до розуміння сутності поняття «протидія легалізації доходів, отриманих незаконним шляхом» в умовах діджиталізації суспільства

Незважаючи на динамічний розвиток суспільства та активні трансформаційні зміни глобальної економіки, реальний та фінансовий сектори досі включають офіційну та тіньову складову. В офіційній економіці всі господарські операції, що здійснюються економічними агентами знаходяться під наглядом та контролем держави. Кожен рух коштів регулюється нормативними актами, чітко відбувається в межах фінансової системи та підлягає оподаткуванню. Функціонування в офіційному секторі економіки дозволяє учасникам користуватись певними привілеями, а саме відсутністю адміністративних та кримінальних покарань, формування ділової репутації, а також реалізація тактичних і стратегічних цілей організації.

Поруч з офіційним сектором економіки існує тіньова економіка, тобто та, яка знаходиться поза полем зору держави. Відповідно до даних Міністерства економіки України, у 2021 році розміри тіньового сектору економіки України були на рівні 32% від обсягу офіційного ВВП [34].

Тіньовий сектор економіки привабливий для недобросовісних суб’єктів господарювання, які хочуть зменшити свої витрати та для злочинців, грошові потоки яких прямо пов’язані з незаконною діяльністю (грабежі, крадіжки, обіг наркотиків, шахрайство, фінансування тероризму тощо).

Через тісну сплетеність зазначених секторів, у суб’єктів тіньової економіки постійно виникає необхідність надати фінансовим ресурсам, накопиченим у тіньовій економіці вигляду «чистих», офіційних. Тільки в офіційному секторі економіки є можливість отримувати товари і послуги найширшого спектру,

оскільки в тіньовому секторі за нелегальні кошти можна отримати тільки нелегальний товар.

Значний обсяг тіньової економіки носить деструктивний характер для економічної безпеки національної економіки: від деформації податкової та бюджетної сфер, до неможливості побудувати адекватну макроекономічну політику, інвестиційний клімат, стабільну грошово-кредитну систему [35].

Легалізація доходів, отриманих незаконним шляхом спричиняє підвищення інфляції, зниження рівня валютної безпеки держави, підрив довіри до банківського сектору, зростання обсягу тіньової економіки [36] та зниження рівня довіри до країни на міжнародному ринку (через ризик потрапляння до чорного списку ФАТФ [37]). Через це держава прикладає значних зусиль для боротьби з легалізацією незаконних доходів.

Процес легалізації доходів, отриманих незаконним шляхом складається з трьох етапів: розміщення, розшарування та інтеграція.

Розміщенням прийнято вважати введення коштів, отриманих незаконним шляхом у офіційну систему фінансових операцій. Розміщення може бути реалізоване через придбання акцій чи інших цінних паперів, оформлення депозиту, здійснення банківського переказу тощо. Основна мета розміщення – вивести кошти від прямого зв’язку з безпосереднім злочином, внаслідок якого вони були отримані.

Розшарування коштів, раніше розміщених у офіційну систему фінансових операцій направлене на подальше приховування зв’язку з коштів та початковим злочином. Реалізується через перепродаж цінних паперів, придбання чи продаж нерухомості чи інших товарів, переведення коштів за кордон чи на інші рахунки. Під час розшарування незаконні кошти можуть переплітатись із законними.

Коли незаконно отримані кошти настільки зливаються з офіційно отриманими, що вся сума виглядає законно отриманою говорять про настання етапу інтеграції незаконних доходів. Протягом цього етапу злочинці отримують на перший погляд законні підстави володіння активами. В такому випадку тільки ретельне розслідування дозволяє встановити зв’язок між інтегрованими

коштами після розміщення та розшарування і незаконно отриманими внаслідок злочину.

Цифровізація фінансово-економічних відносин відкрила нові можливості для злочинців у сфері легалізації доходів, отриманих незаконним шляхом. З одного боку, цифровий слід грошей стало простіше відслідковувати, саме тому одним із напрямків протидії легалізації доходів є контроль та орієнтування на поступову відмову від готівки. З іншого боку, зловмисники можуть створювати десятки акаунтів у платіжних системах та робити сотні транзакцій не виходячи з дому. Генерація великих масивів даних в такому випадку логічно веде до застосування сучасних методів аналізу даних, заснованих на глибокому та машинному навчанні.

Методи легалізації незаконних доходів можуть бути різні: застосування недосконалості платіжних систем, створення та використання фіктивних підприємств, контрабанда, використання банківських переказів, укладання договорів псевдострахування, використання ломбардів та кредитних спілок, використання криптовалюти.

Проте, з огляду нашого дослідження, їх доцільно поділити на дві групи: ті які не зазнали значних змін від цифровізації відносин, та ті, які виникли внаслідок цифровізації або ж зазнали значної модернізації внаслідок неї.

До першої групи можна віднести:

1. Використання контрабандного способу легалізації доходів, отриманих незаконним шляхом полягає у махінаціях при декларуванні готівки, дорогоцінних банківських металів. Вчиняючи злочинні дії, порушуючи митні правила, шляхом декларування у іншій країні готівки в іноземній валюті як особистих заощаджень та ухилення від декларування їх при проходженні митниці в Україні реалізується розшарування доходів. Заплутаність митних правил в різних країнах, помилки, халатність чи злочинний умисел під час проходження особою митного контролю, використання «обхідних шляхів» на державному кордоні підвищує ризик легалізації доходів, отриманих незаконним шляхом [38]. Безумовно митні органи значно збільшили власні інструменти

контролю з розвитком цифровізації, так зараз сформовані значні бази даних перевірки митної вартості та інші системи контролю за товарами, проте в межах схеми легалізації ці зміни не здійснюють суттєвий вплив.

2. Окремим прикладом легалізації доходів, отриманих незаконним шляхом є використання страхових компаній. Зловмисниками здійснюється підробка страхових випадків, оформлення договорів псевдострахування, ухилення від оподаткування [39]. З точки зору діджиталізаційних процесів, які хоч і впливають на страхову сферу, проте сутнісно схеми легалізації не зазнали змін.

3. Використання ломбардів для обміну предметів розкоші та інших цінних активів на готівку та навпаки. Відповідно суттю схем є готівковий обіг, якого майже не стосується цифровізація.

До другої групи варто віднести:

1. Платіжні системи. За допомогою платіжних систем зловмисники здійснюють перерахунок коштів з рахунків на інші рахунки, в тому числі закордон, реалізуючи розшарування незаконних доходів. Використання мережі підставних осіб дозволяє зменшити розміри транзакції, уникаючи їх підозрілості, що ускладнює контроль та можливості виявлення фактів легалізації. Розвиток цифровізації спричинив зростання кількості платіжних систем та поширення доступу до них з будь-якої країни. Реєстрація гаманців у платіжних системах здійснюється за простішою ніж у банках процедурою.

2. Конвертаційні центри. Створення та використання конвертаційних центрів дозволяє зловмисникам перетворювати безготівкові кошти в готівкові, що спричиняє втрату їх цифрового сліду. Афілійовані з підприємствами особи, які мають право прийняття рішень, в разі їх залучення до процесів легалізації доходів, отриманих незаконним шляхом допомагають заплутувати походження коштів шляхом переуступки права вимоги по боргам, надання чи отримання благодійної допомоги. Цим самим замінюючи реальне походження коштів. Іншим способом використання конвертаційного центру – фіктивна господарська діяльність. Шляхом оформлення фіктивних платіжних доручень, отримання

кредитів на діяльність, отримання коштів за державними цільовими програмами підтримки бізнесу здійснюється як розміщення, так і розшарування і інтеграція незаконних доходів [40].

Наразі, зареєструвати підприємницьку діяльність в Україні можна в режимі онлайн. Додатково до цього, цифровізація зумовила виникнення цілого ряду видів економічної діяльності, пов'язаної з інформаційними технологіями. Наприклад, розробка програмного забезпечення є по суті видом інтелектуальної діяльності, для підтримки якої не потрібно значного статутного капіталу чи основних фондів. Підприємства з орієнтацією на розробку програмного забезпечення підходять для легалізації незаконних доходів, оскільки легко імітувати факт надання послуг чи виконання робіт.

3. Банки. Банківські перекази мають бути під пильним контролем з точки зору протидії легалізації. Через банківські перекази найчастіше відбувається заплутування джерел походження коштів. Значна кількість транзакцій між фізичними особами з подальшим їх переказом чи виведення у готівку, перерахунок за роботи чи послуги фіктивного підприємства, придбання цінних паперів через банки є шляхами легалізації незаконних доходів через банківські установи. Розвиток мобільного банкінгу та фінтех прискорили процес розшарювання доходів, отриманих незаконним шляхом.

4. Віртуальні активи. Останнім часом серед злочинців часто стала використовуватись криптовалюта. Електронні кошти слугують засобом розшарування, оскільки можливість створення багатьох криптогаманців за короткий період часу з мінімальною, на відміну від банку, індентифікацією особи. Здійснення тисяч транзакцій за допомогою сотень криптогаманців розмиває походження коштів і стає складно пов'язати кошти з первинним злочином [41].

Окремим підвидом можливих маніпуляцій за допомогою блокчейн-середовища є NFT. Унікальний в мережі, невзаємозамінний блокчейн токен (NFT) дозволяє надати цінності та закріпити авторське право за будь-яким цифровим об'єктом. Відтак, це дозволяє завищувати ціну на цифровий актив за

допомогою аукціону та надавати злочинним коштам ознак легальності. Обсяг продажів NFT у 2020 році складав 250 млн дол. США, а у 2021 році уже понад 2 млрд дол. США [42].

Розуміючи це, криптобіржі вдаються до заходів протидії легалізації незаконних доходів, які базуються на встановленні ризиковості транзакції на основі аналізу співпраці особи з даркнетом чи сумнівними біржами.

Ускладнює роботу з протидії легалізації доходів, отриманих незаконним шляхом через цифрові активи недосконаліть регулюючого законодавства у цій сфері, що дає змогу вільно користуватись різними способами обігу криптовалюти.

У користувачів криптовалюти є декілька шляхів конвертувати цифрові гроші у гривні на банківський рахунок чи за допомогою готівки.

Обмінники криптовалют надають сервіс для швидкого обміну криптовалюти на фіатні кошти чи на інші криптовалюти. Діяльність таких обмінників не заборонена, проте ЗУ «Про віртуальні активи» ще не набрав чинності.

Крипто біржі дозволяють купувати та продавати різні види криптоактивів. В Україні діє декілька криптобірж, найбільші з них: KUNA, WhiteBIT, BTC TRADE UA, QMALL. Проте, через середовище інтернет та недосконаліть регуляції на міжнародному ринку крипто бірж поняття кордонів для віртуальних активів розмивається. Відтак, є можливість переводити грошові суми кому-завгодно за кордоном. Криптобіржі дозволяють P2P обмін, за курсом, обговореним у онлайн-чаті, що дає ширші можливості для легалізації незаконних доходів.

Криптовалюти можна обмінювати через електронні платіжні системи, такі як PayPal, Perfect Money, Payeer, Portmone, Rupay. З березня 2022 року PayPal почав в Україні повноцінну роботу [43]. З одного боку це дозволило швидко переказувати кошти фізичним особам з-за кордону, оминаючи тривалий час очікування банківського переказу. А з іншого боку – надало можливість

зловмисникам заплутувати шляхи переказу незаконних доходів: криптовалюта – електронна платіжна система – банківський рахунок.

Використання криптоматів дозволяє переводити готівку одразу в криптовалюту [44]. Таким способом готівка, яка найчастіше використовується злочинцями обмінюється в криптовалюту, з чого починається процес легалізації незаконних доходів.

Питання впливу цифровізації на легалізацію незаконних доходів та її протидію все частіше стає предметом наукових досліджень.

Таким чином, справедливо зробити висновок, що тіньовий сектор економіки активно розвивається паралельно із розвитком суспільства, а цифровізація значно вплинула на трансформаційні процеси легалізації кримінальних доходів збільшивши їх різноманітність та механізми реалізації. Виходячи з цього, державні органи влади повинні випереджаючими темпами розвивати власну систему протидії легалізації доходів одержаних незаконним шляхом, а також вивчати випереджаючи міжнародні практики та проводити їх імплементацію у національне законодавство.

Переходячи до дослідження теоретичної сутності поняття «протидія легалізації доходів отриманих незаконним шляхом» зауважимо, що даний процес запропоновано реалізувати з використанням програмного продукту VOSviewer. Отже, провівши бібліометричний аналіз наукових публікацій, що індексуються наукометричною базою даних Scopus за допомогою програмного продукту VOSviewer розглянемо зв’язки між категоріями, які досліджують провідні науковці світу. На рисунку А.1 додатку А представлена мапа зі взаємозв’язками поняття «протидія легалізації доходів, отриманих незаконним шляхом (anti-money laundering) з іншими категоріями, відповідно до публікацій за 2012-2022 роки. Дослідження дозволило виділити 6 основних кластерів, які на рисунку А.1 відповідають зеленому, червоному, фіолетовому, блакитному, рожевому та коричневому кольорам, та 2 другорядних кластери: сірий та помаранчевий. Варто зауважити, що чим більше публікацій, які містять ключові

слова з терміном «anti-money laundering», тим більший прямокутник за площею на рисунку А.1 вони займають.

Відповідно до додатку А рисунку А.1, можна простежити зв’язок поняття «протидія легалізації доходів, отриманих незаконним шляхом» з такими категоріями як легалізація грошей, легалізація, тероризм та фінансування тероризму, ризик менеджмент, банківський нагляд, злочин, комплаєнс. Значна частина зв’язків присвячена економіко-математичним методам та машинному навчанню: нейронні мережі, кластеризація, ряди Маркова, бази даних, машинне навчання, великі дані.

Окремо, по фіолетовому кластеру можемо прослідкувати направленість досліджень, пов’язаних із сучасними інформаційними технологіями: блокчейн, розумний контракт, розподілений реєстр, криптовалюта та інформаційні сервіси.

Продовжуючи дослідження у часовому контексті, що відображений на рисунку А.2 додатку А, робимо висновок, що використання економетричних методів та моделей досліджувалось науковцями поряд з протидією легалізації доходів, отриманих незаконним шляхом, починаючи з 2012 року. Цьому свідчать фіолетовий колір блоків «clustering», «data sets», «neural networks», «database systems». А дослідження останніх років саме присвячені зв’язку криптовалюти та протидії легалізації незаконних доходів «cryptocurrency anti-money laundering», «block-chain», «distributed ledger», «smart contract», «art market». Зелено-жовтий відтінок свідчить, що особливу увагу цій тематиці почали приділяти у 2020-2022 роках. Окремо варто виділити наявність зв’язку в публікаціях присвячених протидії легалізації незаконних доходів та методами збереження конфіденційності «privacy-preserving techniques». Оскільки розвиток цифровізації з одного боку генерує велику кількість даних, які стають публічні в інтернеті, що порушує принципи конфіденційності, а з іншого боку, дотримання конфіденційності допомагає злочинцям реалізовувати схеми легалізації незаконних доходів. Тому значна частина науковців приділяють увагу питанням конфіденційності в мережі та протидії легалізації незаконних доходів.

Аналіз візуалізаційних карт, зображених на рисунках А.1 та А.2 підтверджує необхідність продовжувати дослідження у сфері взаємозв'язків між цифровими технологіями та протидією легалізації доходів, отриманих незаконним шляхом.

Зосереджуючись на безпосередньому взаємозв'язку цифровізації та протидії легалізації доходів, отриманих незаконним шляхом, можна виділити наступну залежність (рисунок А.3 додатку А): «anti-money laundering» – «digitalization» – «money laundering» – «cryptocurrency». Цифровізація найчастіше в наукових працях зустрічається у взаємозв'язку з протидією легалізації доходів, отриманих незаконним шляхом. Тоді як у контексті легалізації доходів, отриманих незаконним шляхом частіше зустрічається термін криптовалюта.

Переходячи до аналізу наукового доробку в сфері протидії легалізації доходів, отриманих незаконним шляхом варто зазначити, що дана тема є популярною як серед науковців-економістів, так і серед науковців у галузі юриспруденції.

Вплив легалізації доходів, отриманих незаконним шляхом на сталий розвиток досліджували Z. Dobrowolski та L. Sulkowski [45]. Визначення ризику легалізації доходів, отриманих незаконним шляхом для бізнес-сектору стало основою для дослідження J. Ferwerda та E.R. Kleemans [46].

Застосування методів data mining для протидії легалізації доходів, отриманих незаконним шляхом досліджували A. Salehi, M. Ghazanfari та M. Fathian [47]. Натомість A.I. Canhoto [48] досліджувала можливості використання методів машинного навчання для протидії легалізації доходів, отриманих незаконним шляхом.

Переходячи до дослідження наукових робіт присвячених впливу цифровізації на роботу суб'єктів системи протидії легалізації кримінальним доходам, зазначимо, що безумовно діджиталізація підвищує якість фінансового моніторингу в банках [49]. та сприяє ефективнішому впровадженню рекомендацій FATF у сфері anti-money laundering. До схожого висновку дійшли

і K. Said та D.Karimi [50], зазначаючи що автоматизація банківських процесів підвищує фінансову безпеку банку та покращує фінансовий моніторинг. Особливо вагомий вплив діджиталізації на легалізацію незаконних доходів прослідковується через зменшення кількості готівки в обігу [51]. Останні дослідження свідчать що фінтех широко впроваджується в банках в країнах що розвиваються [52], що дозволяє набагато швидше приєднатись фінансовим установам даних країн до глобальної системи протидії легалізації кримінальним доходам. Значна група науковців – K. Djalilov та J. Hölscher [53]; K. Djalilov та C. Hartwell [54] дійшла висновку про те, що впровадження цифрових технологій в комплексі покращує можливості банківського середовища та безумовно впливає на можливість оперативного виявлення незаконних операцій. Окрім зазначеного, не можна нехтувати тим, що цифровізація пришвидшує роботу співробітника відповідного органу системи протидії легілації: інтерактивний пошук замінює довге перелистування папірців, тобто економляться людино-години роботи [55, 56].

A. Addo та PK. Senyo [57] у своїх дослідженнях зупиняються на вивченні впливу цифровізації на правоохоронну діяльність, а саме боротьбу з корупцією. Так, вони визначають, що цифровізація є потужним антикорупційним інструментом. Приклади застосування цифрових технологій у прокурорській діяльності досліджував Denis de Castro Halis [58], автор зазначає, що засоби цифровізації одночасно з підвищенням швидкості обробки документів формують й засади до їх безпосереднього контролю, забезпечуючи незалежність слідства та відповідність його законам. Окремо варто розглянути застосування сучасних технологій у судовій діяльності. Автор Yu. Mulyana [59] зазначає, що цифровізація судів починається з електронного документообігу, але не обмежується ним: всі процеси пов’язані зі справами можуть обслуговуватись в електронному суді, або наприклад суди можуть проводитись у онлайн-форматі, що пришвидшує процес.

Легалізацію доходів, отриманих незаконним шляхом за допомогою криптовалюти досліджував C. Wronka, зосереджуючись не тільки на аналізі

сутності легалізації, а й можливих превентивних заходів їй протидії [60]. D. Dupuis та K. Gleason займались питанням необхідності регулювання обігу криптовалюти, в контексті протидії легалізації незаконних доходів [61].

Отже, справедливо зробити висновок, що в сучасному науковому середовищі протидії легалізації доходів отриманих незаконним шляхом в умовах діджиталізації суспільства відводиться значна роль та розглядають в своїй більшості як комплекс заходів з попередження, виявлення та подальшого покарання злочинних дій, спрямованих на приховання чи маскування незаконного походження коштів або іншого майна.

Пункт 2.1 було виконано із використанням матеріалів публікацій виконавців [6271].

2.2 Оцінка ризику конвергенції системи протидії відмивання грошей та кібербезпеки

З кожним роком роль цифрових технологій у різноманітних сферах життя зростає швидкими темпами. Їх запровадження та вдосконалення відбувається не лише на рівні конкретного підприємства, фізичної особи чи організації, але й на рівні держави. В цифровому просторі розміщено велику кількість даних, такі як особисті документи громадян, їх персональна інформація, що знаходиться, як у відкритих джерелах (соціальні мережі), так і закритих (дані міграційних служб, реєстрів відділів соціального захисту, податкових служб, банківських установ і т.д.), фінансова звітність підприємств, їх установчі документи, тощо. Наявність такого простору вимагає досить високого рівня захисту даних. І в ідеальному середовищі цей захист має відбуватися не лише окремо в обмеженій установі, але й охоплювати більш широкий спектр – рівень країн та міжнародних спільнот. Кібербезпека є не лише однією із найважливіших складових національної безпеки країни, але й закладає основи міжнародних відносин між державами по всьому світу.

Оцінка ризиків, пов’язаних з кіберзагрозами, на сьогодні є одним із пріоритетних напрямків діяльності теоретиків та практиків. Проведення

аналітики щодо їх виявлення та усунення є надзвичайно складним завданням, оскільки дана група ризиків є динамічною. Ризики такого типу здатні модифікуватись та адаптуватися до змін системи ще задовго до виникнення можливості їх передбачення та усунення. А методи, що діють для одного кіберризику, можуть бути зовсім недієвими для іншого, тому навіть найменший ризик може спричинити значні втрати і призвести до краху цілої системи. За останні роки рівень та масштаби кіберзлочинів невпинно зростають, а збитки від них значно переважають збитки від торгівлі наркотиками та зброєю, чого не спостерігалося ще п’ять років тому. Банківські та фінансові установи останнім часом найбільше потерпають саме від кіберзлочинців, ніж, наприклад, від зміни на фондових біржах чи боргових криз. Це все зумовлює потребу в удосконаленні системи боротьби проти різного виду кібернетичних шахрайств. Тому методи боротьби із кіберзлочинністю розвиваються на законодавчому рівні багатьох країн, що дозволяє підтримувати загальний рівень безпеки країни.

В сучасному світі темі вивчення систем протидії кібернетичним злочинам та фінансовим махінаціям присвячено багато праць, як вітчизняних, так і закордонних вчених. Так, однією із найбільш ґрунтовних робіт вважається робота М. Еллінга, який досліджував кіберризики та можливості їх протидії [63]. Важливу роль у дослідженні кібербезпеки та її взаємодії із фінансовою системою вивчали Ю. Кожедуб, К. Семенова та інші [64]. Значний внесок у дослідження кіберризиків мають приватні консалтингові компанії, які мають на меті практичне застосування напрацювань. Серед них варто виділити Deloitte, AON, IBM, тощо [65]. Саме ці компанії в останні роки сформували велику базу знань, що допомагає не лише науковцям, але й підприємствам та урядам у власній діяльності.

Національний індекс кібербезпеки (NCSI)– один із збірних показників, який є одним із основних характеристик стану інформаційної безпеки [66]. Для розрахування даного індексу враховується значна кількість показників. Найважливішим серед них є стан законодавства, яке пов’язане з охороною даних та кібербезпекою, оскільки саме юридичне забезпечення дозволяє підготувати

основу для реалізації стабільних заходів щодо протидії злочинам. Враховується також частота виникнення кіберінцидентів, рівень освіти громадян в сфері кібербезпеки, види та ефективність заходів щодо захисту персональних даних громадян, щодо реагування на кібератаки та можливості зниження кіберризиків. Даний індекс включає в себе результативність та кількісне виявлення загального рівня боротьби з кіберзлочинністю. Саме цей індекс враховується при оцінці інвестиційної привабливості країни в цілому та є відображенням стабільності та надійності її інформаційної системи.

Поряд із забезпеченням сталого розвитку кібербезпеки та підтримки її на належному рівні значну роль в утворенні та формуванні національної безпеки є фінансова. Фінансова безпека є підґрунтям для побудови міцної економічної безпеки та конкурентоспроможної економіки в цілому. Основним завданням фінансової безпеки є побудова сприятливого середовища, як правового, так і економічного, а також підтримка інституційної інфраструктури, яка дозволить стимулювати розвиток потенційно життєздатних підприємств та запустить інвестиційні процеси, що допоможуть підтримувати загальний фінансовий рівень держави.

Фінансова система є однією із найбільш схильною до негативних впливів внутрішнього та зовнішнього середовищ, а також є вразливою до більшості видів ризиків. Великим дестабілізуючим фактором, що знижує рівень фінансової безпеки є дестабілізація бюджету України, особливо в нинішніх умовах, коли він розподіляється нерівномірно і має значний дефіцит. Великою проблемою є збільшення грошового обігу поза банківською системою країни, особливо в операціях з валютою. Це свідчить про недовіру до внутрішнього ринку країни та сприяє збільшенню рівня тінізації економіки та відмиванню коштів. Відмивання коштів та подальша легалізація таких доходів є великою проблемою не лише для України, але й для багатьох країн, що розвиваються. Розвиток технологій та вдосконалення кібератак дозволило злочинцям знаходити нові методи у відмиванні коштів та отримувати прибуток, використовуючи не лише традиційні гроші, але й криптовалюту та інші види електронних грошей. Це значно

підвищило потенційні ризики та можливі збитки від їх настання. Тому оцінка фінансової стабільності та методів протидії відмиванню коштів є важливою складовою національної безпеки Так, основним індексом для оцінки є індекс протидії відмиванню коштів (AML–індекс) [67]. Вперше даний показник був розрахований та застосований на базі Базельського інституту управління у 2012 році. Даний інститут є незалежним асоційованим інститутом Базельського університету та діє під керівництвом Програми Організації Об'єднаних Націй в сфері попередження злочинності та кримінального правосуддя [68]. Він застосовується для виміру та оцінки ризиків, що пов'язані з відмиванням коштів та спонсоруванням тероризму. Результати всіх досліджень є незалежними та використовуються різноманітними державними, міжнародними, комерційними, фінансовими установами для оцінки можливостей настання ризиків та способів їх уникнення. Даний індекс є корисним, оскільки він враховує не рівень корупції та кримінальної діяльності, що пов'язана з відмиванням коштів та фінансуванням тероризму, а більше ризики їх виникнення та розвитку. Даний індекс включає в себе різні показники, що мають різну спрямованість та значимість в оцінці. Тому показники, що застосовуються при визначенні індексу, розподіляються по основним категоріям та мають свою вагу. Це дозволяє отримати всебічну картину та впроваджувати подальші заходи для мінімізації чи усунення зазначених ризиків.

Якість системи протидії відмиванню коштів та фінансування тероризму складає 65 % від загального індексу. Тут враховуються політичні, правові, соціально-економічні методи, спрямовані на забезпечення та протидію даному виду злочинності. Ризик корупції у державі складає 10 %, фінансова прозорість також 10%, а загальна прозорість та підзвітність має 5 % . Ще 10 % становлять правові та політичні ризики, оскільки саме від них залежить стан фінансової системи щодо тероризму та усі можливі наслідки такої діяльності. Зазначений індекс охоплює усю економічну, політичну та фінансову систему, що дозволяє отримати чіткі та показові результати, необхідні для подальшого розвитку країни. Отже, можна зазначити, що для кожної країни Національний індекс

кібербезпеки та AML – індекс мають власне значення і базуються на власних показниках [69]. Але в загальному підсумку деякі країни мають схожі тенденції до розвитку і мають загальні риси в підтримці сукупної національної безпеки. Необхідно проаналізувати декілька країн і визначити, чи дійсно існують великі розбіжності між схожими країнами і на чому саме ґрунтується можлива подібність.

Для аналізу в даному випадку використовуються методи кластеризації, оскільки саме вони дозволять отримати групи країн, які матимуть подібні дані та схожі риси.

Silhouette analysis використовується для визначення відстані поділу між отриманими кластерами. Він відображає, наскільки близько розташована кожна точка в одному кластері відповідно до інших кластерів. Показники даного аналізу мають діапазон від -1 до 1. Коефіцієнти біля 1 вказують на те, що вибірка знаходиться далеко від сусідніх кластерів. Тобто розподіл є найбільш чітким і придатним для аналізу. Значення 0 відображає, що вибірка знаходиться на межі прийняття рішення між двома сусідніми кластерами або дуже близько до неї, а від’ємні значення вказують на те, що ці вибірки могли бути призначені неправильному кластеру [70].

Після проведення оцінки Silhouette для здійснення кластеризації індексом AML (рис. 2.1), можна зробити висновок, що найбільш оптимальною кількістю кластерів є вісім, оскільки оцінка в даному випадку є найвищою. Тому для проведення кластеризації методом K-means необхідно використати саме таку кількість кластерів.

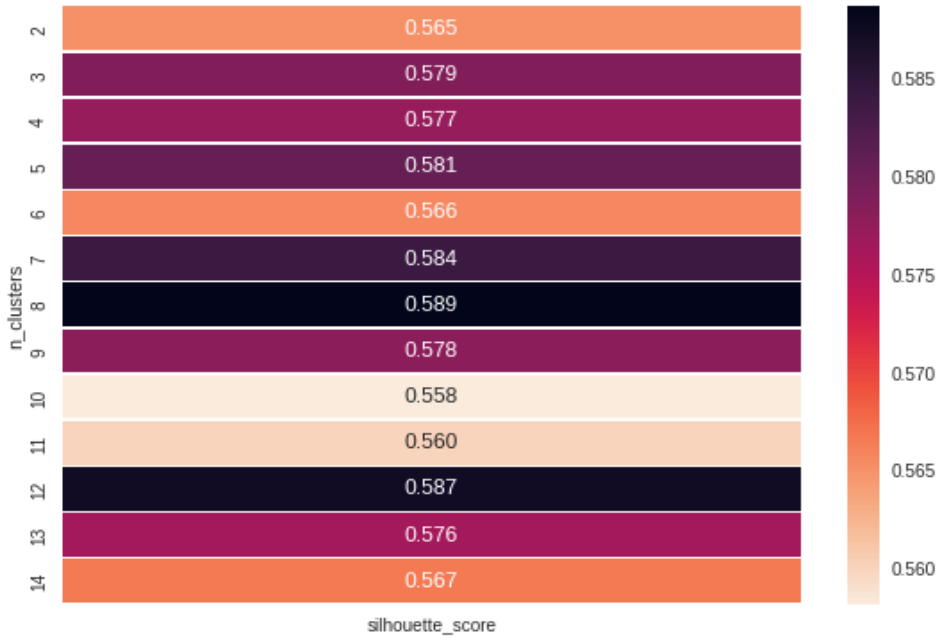


Рисунок 2.1 – Оцінка «Silhouette» щодо оптимального вибору кластерів країн за індексом AML

Рисунок 2.2 показує розподіл даних по кластерам в результаті Silhouette-оцінювання. Результати розподілені за кластерами із відсутністю вийнятків, то можна сказати, що всі дані кластерів є однорідними.

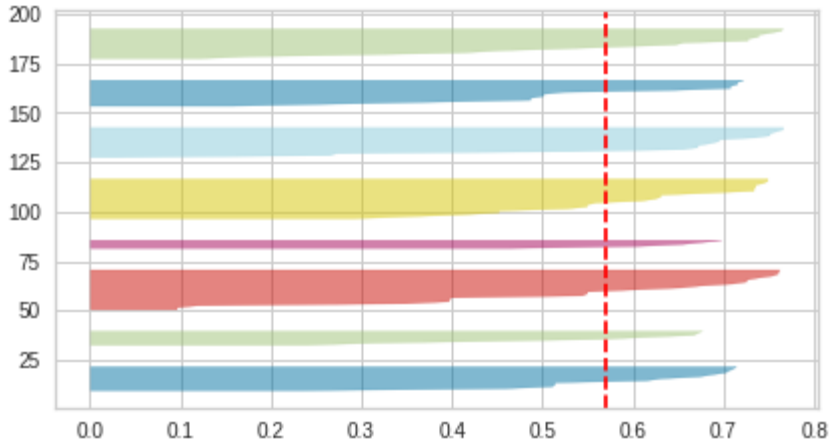


Рисунок 2.2 – Результати Silhouette-analysis

На рисунку 2.3 представлений результат проведеної кластеризації країн за індексом AML.

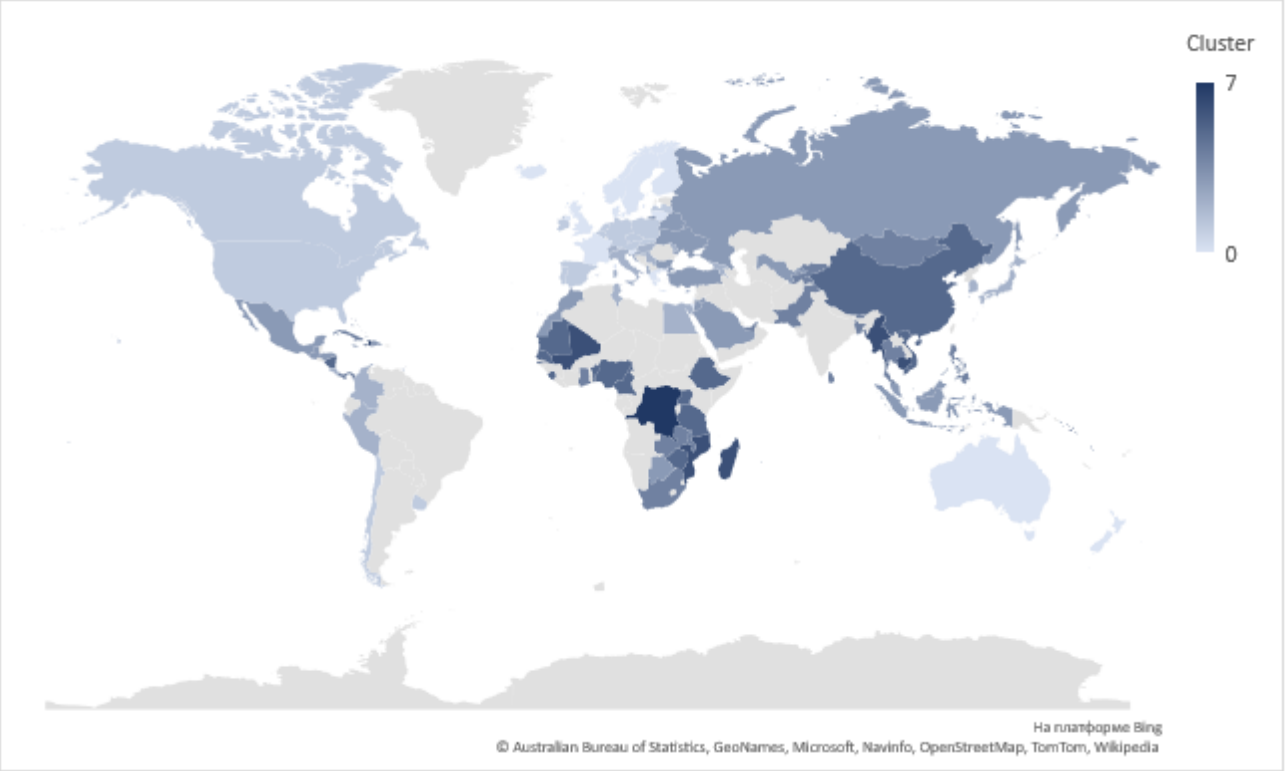


Рисунок 2.3 – Карта кластерів країн за індексом AML

Розподіл на кластери дозволив отримати досить показові групи, які надають змогу зробити певні оцінки. Як можна побачити, країни, які є достатньо розвиненими та мають високі показники в забезпеченні фінансової безпеки, віднесені до однієї групи. Так, найкращі показники мають такі країни, як Австралія, Велика Британія, Данія, Норвегія, Фінляндія, Нова Зеландія Греція, Ізраїль, Франція, Литва, Словенія, Ісландія та Швеція. Вони мають високий рівень протидії відмиванню коштів і фінансування тероризму та впроваджують ефективні рішення у боротьбі з легалізацією кримінальних доходів. Вони не є привабливими для злочинців, оскільки мають потужні системи захисту у фінансових установах.

Україна входить до третьої групи за показником ризиковості щодо відмивання кримінальних доходів. Це обумовлено тим, що наша країна має досить високий рівень тінізації та корумпізації економіки, що в сукупності з військовими діями створює сприятливе підґрунтя для зменшення ризиків для легалізації нелегальних коштів. Ці процеси гальмують розвиток в економіці та

соціальній сфері. До даної групи увійшли ще 21 країна, серед яких слід зазначити Гондурас, Туреччину, Сейшельські острови, Барбадос, Ямайку та інші.

Країни, які є сприятливими для легалізації кримінальних доходів є країни 6 та 7 кластерів. Сюди відносяться Малі, Камбоджі, Мадагаскар, Мозамбік, М'янма, Гаїті та Демократична республіка Конго. Перелічені країни відносяться до найменш розвинених. Реформи, що проводяться в них, спрямовані на внутрішні сфери і не приносять відповідних позитивних результатів для їх розвитку. Режими цих країн є досить обмежувальними, а рівень розвитку економіки нестабільним.

Розглянемо результати кластеризації щодо «Національного показника кібербезпеки» (рис. 2.4). Найвище значення оцінки відповідає кількості кластерів, яка дорівнює двом. Але дана кількість кластерів не дозволяє виявити більш детальні групи країн, тому для проведення Silhouette-analysis оберемо кластери з найвищими оцінками – 4, 5, 6, 7.

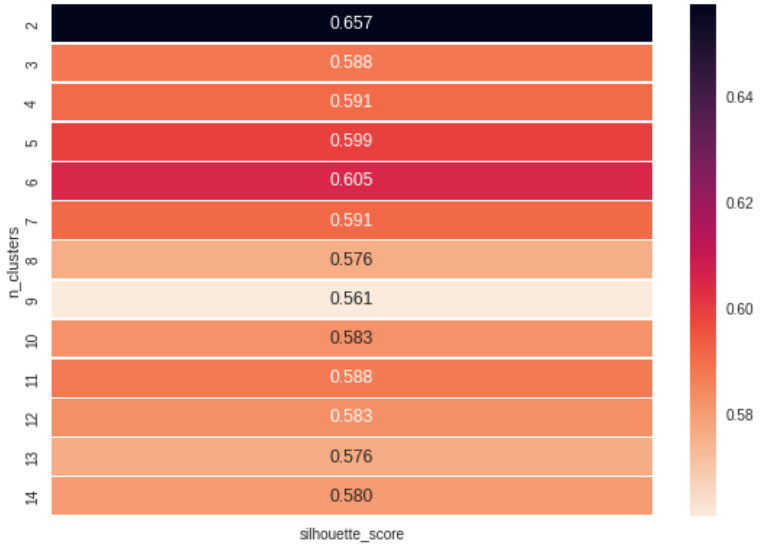


Рисунок 2.4 – Оцінка «Silhouette» щодо оптимального вибору кластерів країн за індексом NSCI

Результати Silhouette-analysis (рис. 2.5) показують, що кластеризація країн із використанням 4, 5, 6 або 7 кластерів дозволить отримати групи з однорідними

даними. Тому, оберемо кількість кластерів 6, оскільки цьому значенню відповідає найвища оцінка Silhouette.

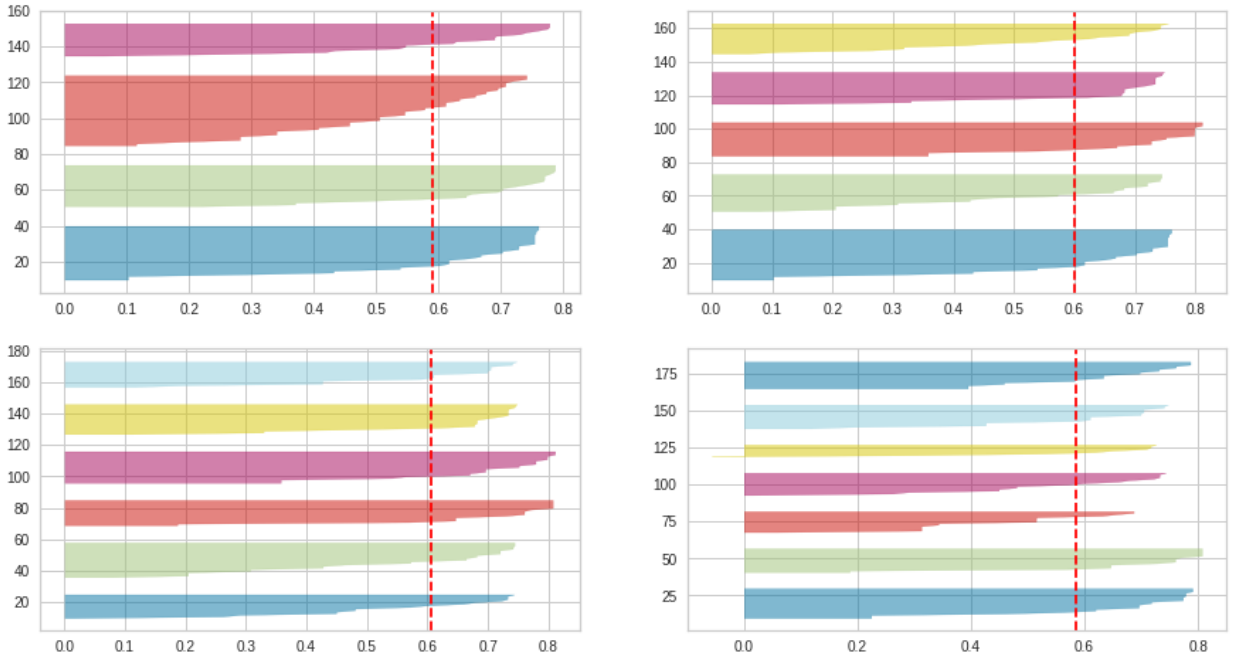


Рисунок 2.5 – Результати Silhouette-analysis

На рисунку 2.6 представлений результат проведеної кластеризації країн за індексом NSCI після проведеного Silhouette-analysis.

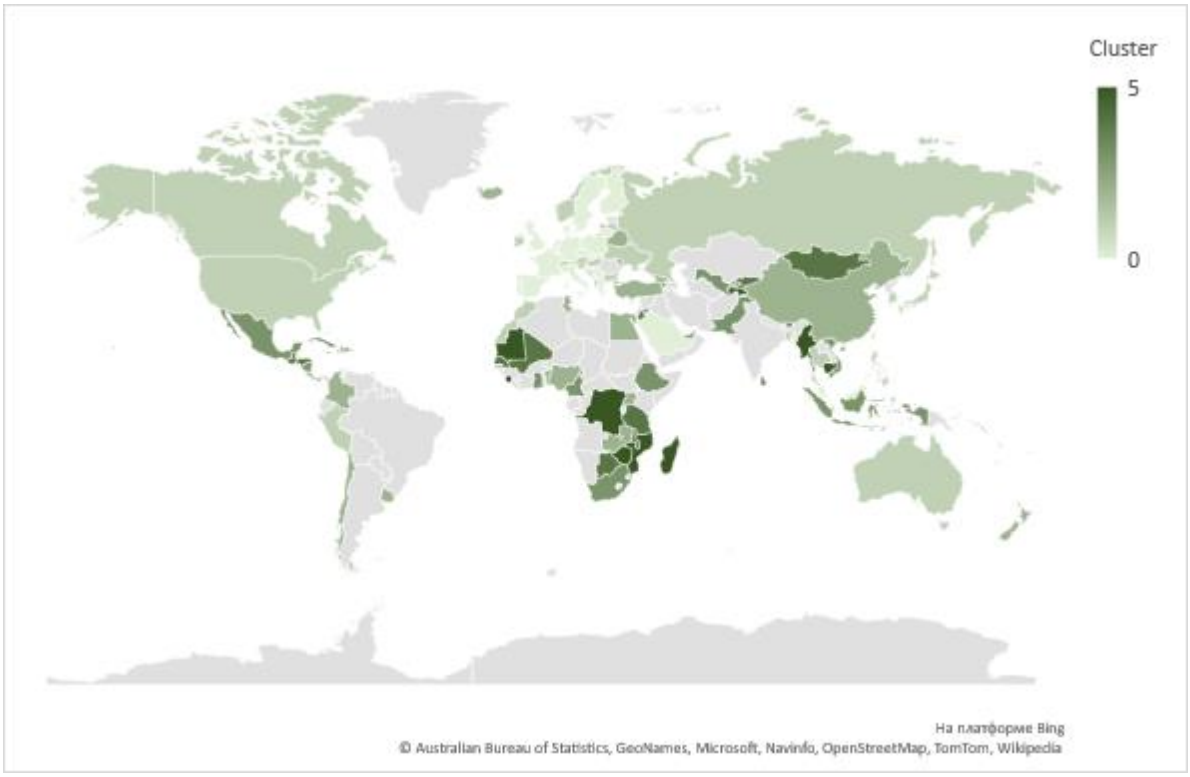


Рисунок 2.6 – Карта кластерів країн за індексом NSCI

Як можна побачити, до групи з найбільш високо розвиненим рівнем кібербезпеки відносяться країни нульового кластеру: Греція, Бельгія, Нідерланди, Німеччина, Іспанія, Малайзія, Саудівська Аравія, Сербія, Хорватія, Італія, Польща, Словаччина, Португалія, Чехія, Велика Британія, Данія, Франція, Литва, Швеція та Фінляндія. Такі результати свідчать, що ці країни мають високий рівень національної кібербезпеки, який передбачає організацію потужного комплексу інформаційного, програмного, технічного та організаційного забезпечення з питань кіберзахисту. Так само, можна побачити, що найнижчі показники мають Конго, Мадагаскар, Мозамбік, Гаїті, Камбоджі, Зімбабве, Таджикистан, Вануату, тощо. Тобто ці країни є менш розвиненими і потребують вкладень і модифікації не лише окремої системи кіберзахисту, але прогресивних державних заходів в цілому щодо розвитку її політичної та соціально-економічної сфер.

Для визначення ризиків конвергенції системи протидії фінансовим та кібершахрайствам доцільно впровадити комплексну оцінку, яка б в собі поєднувала можливості країн щодо кіберзахисту від різного роду загроз та їх потенціал щодо зниження ризиків відмивання кримінальних доходів та фінансування тероризму. Пропонуємо визначити інтегральний індекс конвергенції наступним чином:

– здійснюється нормалізація AML – індекса за критерієм Севіджа, як дестимулятора (формула 2.1):

$$x_{ij}^{*} = \frac{x_j^{max} - x_{ij}}{x_j^{max} - x_j^{min}} \tag{2.1}$$

– здійснюється нормалізація NSCI за природньою нормалізацією як стимулятора (формула 2.2):

$$x_{ij}^{*} = \frac{x_{ij} - x_j^{min}}{x_j^{max} - x_j^{min}} \tag{2.2}$$

– інтегральний індекс конвергенції утворюється як середньогеометричне (формула 2.3):

$$G_m = \left(\prod_{i=1}^{n} \widetilde{x_{ik}} \right)^{1/n} \tag{2.3}$$

Розглянемо результати кластеризації щодо «Інтегрального показника конвергенції» (рис. 2.7). Найвище значення оцінки відповідає кількості кластерів, яка дорівнює двом. Але така кількість кластерів не дозволить отримати уявлення про ризик конвергенції. Тому для проведення Silhouette-analysis оберемо кластер з наступною найвищою оцінкою – 9.

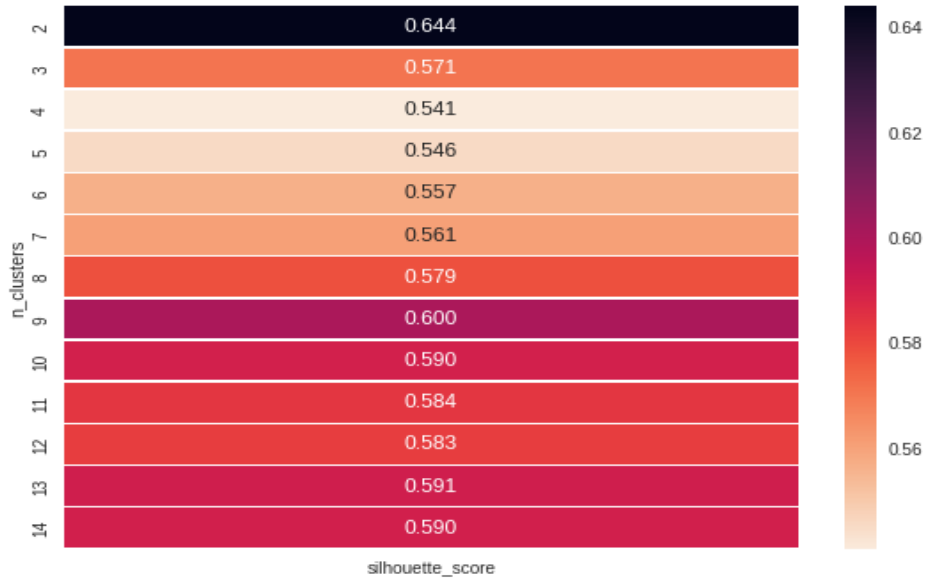


Рисунок 2.7 – Оцінка «Silhouette» щодо оптимального вибору кластерів країн за умови конвергенції системи протидії відмивання грошей та кібербезпеки

Результати Silhouette-analysis (рис. 2.8) показують, що кластеризація країн із використанням 9 кластерів дозволить отримати групи з однорідними даними, що дозволяє провести кластерний аналіз за методом k-means.

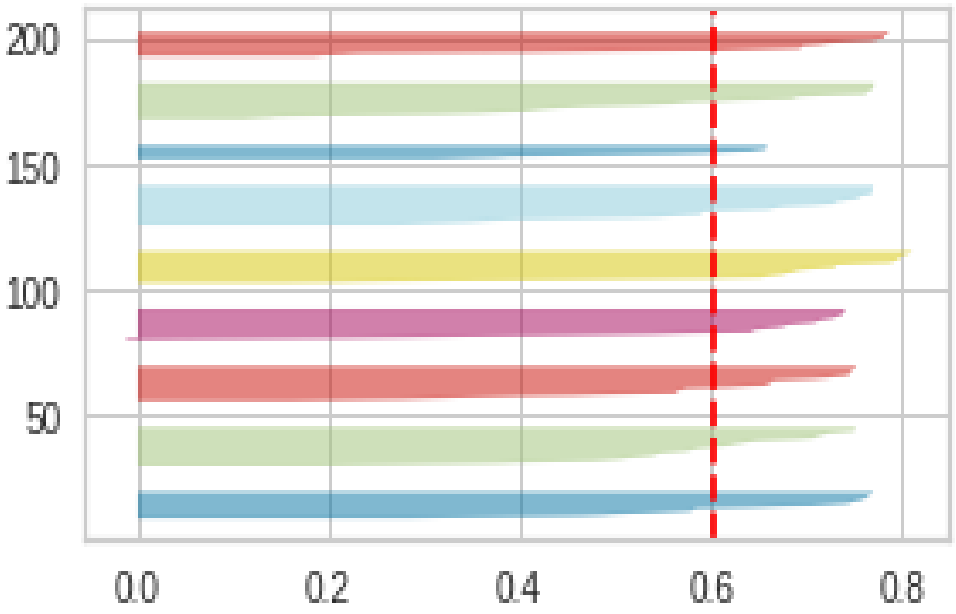


Рисунок 2.8 – Результати Silhouette-analysis

На рисунку 2.9 представлений результат проведеної кластеризації країн за «Інтегральним індексом конвергенції» після проведеного Silhouette-analysis.

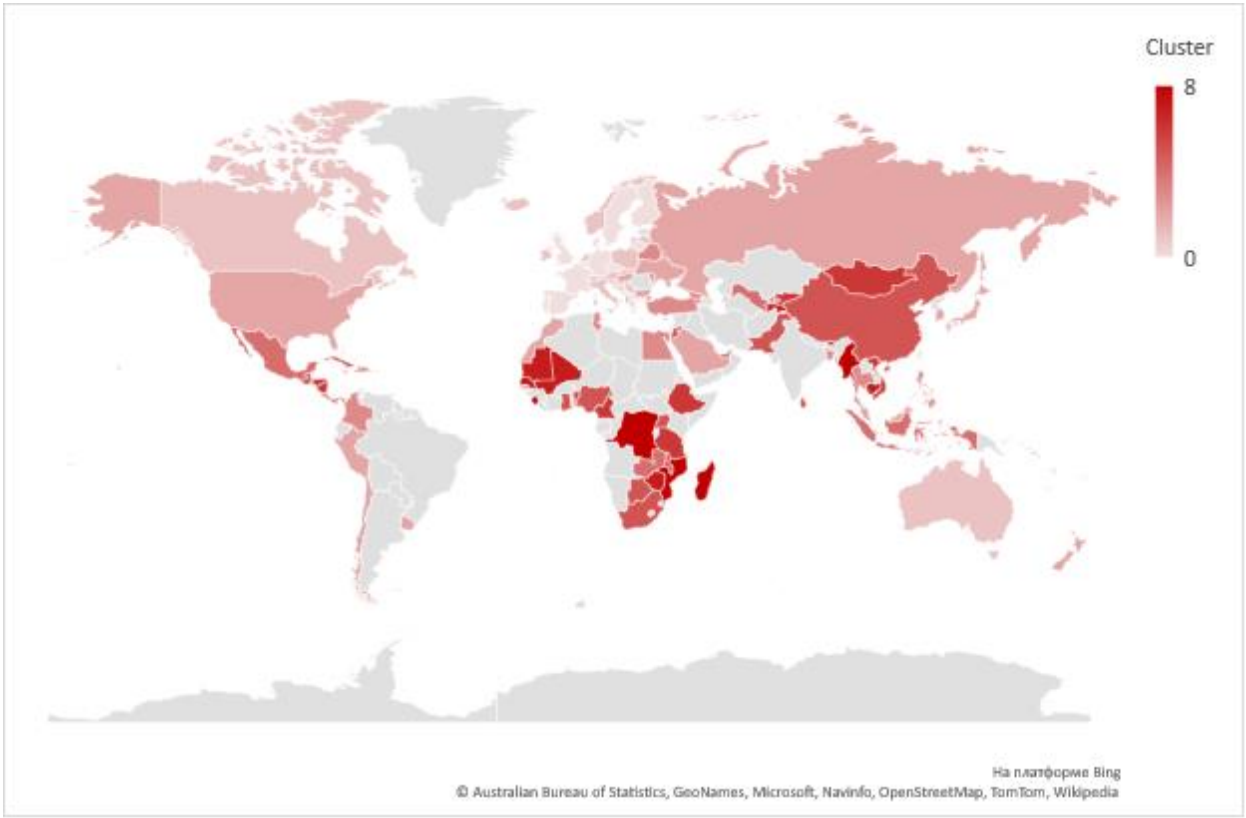


Рисунок 2.9 – Карта кластерів країн за рівнем конвергенції систем протидії відмиванню кримінальних доходів та кібербезпеки

Проведення кластеризації на основі інтегрального індексу конвергенції систем показує розподіл країн за можливостями протидіяти кіберзагрозам та фінансовим злочинам. Чим ближче його значення до 1, тим нижчий рівень ризику конвергенції, тобто в країнах створені сприятливі умови, які дозволяють інтегрувати систему кіберзахисту та систему фінансового моніторингу. Якщо значення даного показника наближається до 0, то це свідчить про неготовність країни до конвергенції двох систем, що може бути викликано сформованими сприятливими умовами для розвитку фінансової та кіберзлочинності.

Виходячи з отриманих даних сформуємо критерії ризику в залежності від отриманих кластерів для індексу конвергенції. Результати представлені в таблиці 2.1. Таблиця містить усереднені значення AML та NSCI для відповідного кластеру. Країни, що відносяться до 0-го кластеру генерують найнижчий ризик конвергенції. Відповідно, країни 8-го кластеру генерують найвищий ризик. Застосування даних карти 2.9 та таблиці 2.1 дозволить сформувати висновки щодо потенційних ризиків конвергенції систем протидії фінансовим та кіберзлочинам.

Таблиця 2.1 – Ідентифікація ризику в залежності від кластерів індексу конвергенції

Низький ризик		Помірний ризик		Високий ризик	
Кластери	AML / NSCI	Кластери	AML / NSCI	Кластери	AML / NSCI
0	3,65 / 88,42	3	5,09 / 55,95	6	6,07 / 22,94
1	4,09 / 72,96	4	5,30 / 41,43	7	6,30 / 12,51
2	4,72 / 67,61	5	5,83 / 37,29	8	7,75 / 9,31

Карта кластерів 2.9 показує, що найбільш сприятливі умови для конвергенції сформовані в 12 країнах, таких як Німеччина, Бельгія, Португалія, Іспанія, Чехія, Греція, Велика Британія, Данія, Франція, Литва, Швеція, Фінляндія. Ці країни генерують найнижчий ризик. Що стосується України, то її було віднесено до 2-го кластеру. У даному випадку вона має високий ризик легалізації кримінальних доходів, що компенсується розвиненим рівнем кібербезпеки. Це дозволяє зменшувати ризики відмивання коштів за рахунок

можливостей системи кіберзахисту. Тобто Україна має значний потенціал для створення взаємодії фінансової та інформаційної систем та підтримки їх безпеки. Рівень її системи відмивання коштів та легалізації доходів значно може скорочуватися за рахунок безпекового потенціалу.

В країнах, що віднесені до 6-8 кластерів, сформовані найбільш несприятливі умови, оскільки вони знаходяться на нижчій стадії економічного і соціального розвитку. Сюди відносять найменш розвинені країни Африки, Азії та острівні держави.

На основі отриманих даних побудуємо прогнозну модель ризику конвергенції, яка дозволить визначити відповідний його рівень за рахунок зміни умов конвергенції системи кібербезпеки та фінансового моніторингу. Для побудови класифікаційного дерева рішень було використано мову програмування Python. Модель було визначено на основі коефіцієнту Джині.

Результат класифікаційної моделі представлений на рисунку 2.10.

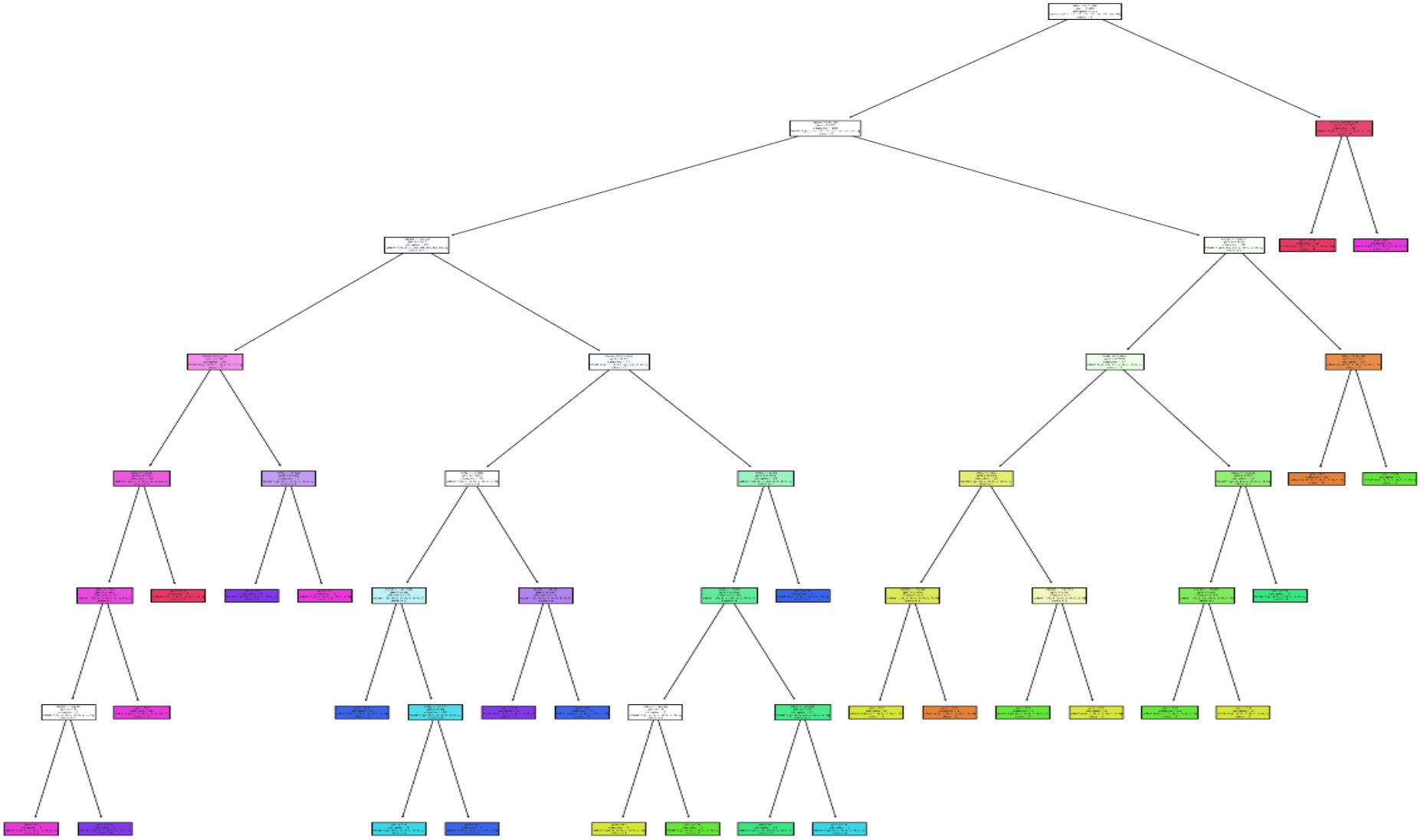


Рисунок 2.10 – Класифікаційна модель прогнозування ризиків конвергенції системи протидії фінансовим та кіберзлочинам

Оцінка якості побудованого дерева представлена на рисунку 2.11.

```
Confusion Matrix:
[[4 0 0 0 0 0 0 0 0]
 [2 3 0 0 0 0 0 0 0]
 [0 2 0 0 0 0 0 0 0]
 [0 0 0 2 0 0 0 0 0]
 [0 0 0 0 7 0 0 0 0]
 [0 0 0 0 0 3 0 0 0]
 [0 0 0 0 0 1 3 0 0]
 [0 0 0 0 0 0 0 2 1]
 [0 0 0 0 0 0 0 0 1]]
Classification Report:
              precision    recall  f1-score   support

           0       0.67      1.00      0.80         4
           1       0.60      0.60      0.60         5
           2       0.00      0.00      0.00         2
           3       1.00      1.00      1.00         2
           4       1.00      1.00      1.00         7
           5       0.75      1.00      0.86         3
           6       1.00      0.75      0.86         4
           7       1.00      0.67      0.80         3
           8       0.50      1.00      0.67         1

    accuracy                           0.81        31
   macro avg       0.72      0.78      0.73        31
weighted avg       0.79      0.81      0.78        31

Accuracy: 0.8064516129032258
```

Рисунок 2.11 – Оцінка якості класифікаційної моделі прогнозування ризиків конвергенції системи протидії фінансовим та кіберзлочинам

В цілому загальна точність моделі є високою і відповідає приблизно 81%. Хоча даний показник не є гарним для моделей такого рівня, але це пов’язано із тим, що вона передбачає класифікацію значної кількості груп. Наприклад, для другого кластеру модель не зможе зробити жодного передбачення, хоча інші рівні ризику вона передбачатиме на рівні вище середнього.

Отже, як можна побачити стабільність систем захисту як проти кіберризиків, так і проти відмивання фінансових коштів, залежить від рівня розвитку країни та комплексу заходів щодо мінімізації та попередження можливих загроз. Неможливо створити єдину модель, яка б задовольняла потреби усіх країн та організацій, оскільки не дивлячись на можливі загальні тенденції, кожна система має власну основу і власні вразливі місця, до яких можуть адаптуватися різні види атак. Тому для підвищення рівня захищеності

даних та зниження рівня відмивання коштів необхідно застосовувати комплексні заходи. В першу чергу необхідно будувати міцну законодавчу базу, яка дозволить правоохоронним органам та організаціям вільно ділитися інформацією та швидко протидіяти можливим злочинам. Необхідно застосовувати всі можливі заходи безпеки в інформаційному просторі, такі як аутентифікація користувачів, використання цифрового підпису тощо. Саме комплексні дії дозволять пристосуватися до мінливого середовища та забезпечити надійний захист і підвищення рівня національної безпеки.

Пункт 2.2 було виконано із використанням матеріалів публікацій виконавців [71].

2.3 Побудова нейромережевої моделі потенційної конвергенції системи кібербезпеки та протидії легалізації кримінальним доходам

2.3.1 Статистичний аналіз потенційної конвергенції системи кібербезпеки та протидії фінансовим злочинам

Аналіз базових статистик було проведено з допомогою статистичних методів обробки даних, їх систематизації, наочного представлення як таблиць і графіків, а також кількісний опис даних з допомогою системи статистичних показників. Статистичний аналіз проводився з допомогою мови програмування Python. Python - високорівнева мова програмування загального призначення з динамічною строгою типізацією та автоматичним управлінням пам'яттю, орієнтована на підвищення продуктивності розробника, читання коду та його якості, а також на забезпечення переносимості написаних на ньому програм [72]. У даній роботі мова програмування Python була використана для розрахунку базових статистик і візуалізації даних конвергенції системи кібербезпеки та протидії фінансовим злочинам [73].

Першим кроком був імпорт необхідних бібліотек, таких як: Pandas, Numpy, Preprocessing, Matplotlib.Pyplot і деякі інші [74, 75, 76].

Для відображення даних було використано функцію .head() (рис. 2.12).

```
df.head()
```

	Country	GCI	ICTDI	NRI	NCSI	DDL	PSI	GEI	EDB	CI	CPI	GTI	FCI
0	Australia	89	82	79	59.74	80.49	1.00	1.60	80.14	42.55	77	2.827	244.358302
1	Austria	83	80	77	68.83	78.67	0.91	1.45	78.54	20.41	76	1.852	310.412705
2	Bahrain	59	76	73	25.97	74.43	-0.84	0.18	68.03	36.96	36	3.883	490.706709
3	Barbados	17	73	0	15.58	73.10	0.92	0.43	56.78	51.31	68	0.000	230.952985
4	Belgium	81	78	77	85.71	77.62	0.41	1.17	71.71	42.17	75	4.060	212.965184

Рисунок 2.12 – Відображення вхідних даних

На рисунку 2.12 представлено дані, які характеризують потенційний процес конвергенції системи кібербезпеки та протидії фінансовим злочинам. Представлено 14 стовпців, перший стовпець вказує на порядковий номер, стовпець під назвою «Country» містить перелік країн, з 3 по 14 стовпці представлено такі статистичні дані, як: GCI – Глобальний індекс кібербезпеки; ICTDI – Індекс розвитку інформаційно-комунікаційних технологій; NRI – Індекс мережевої готовності; NCSI – Національний індекс кібербезпеки; DDL – Рівень цифрової трансформації; PSI – Індекс політичної стабільності; GEI – Індекс ефективності уряду; EDB – Індекс легкості ведення бізнесу; CI – Індекс злочинності; GTI – Глобальний індекс тероризму; CPI – Індекс споживчих цін; FCI – Індекс фінансової таємниці.

Наступним етапом був розрахунок базових статистик для кожного з вказаних показників. До основних статистик відноситься: загальна кількість спостережень, середнє значення, стандартне відхилення, мінімальне значення, і максимальне значення (рис. 2.13).

```
df.describe()
```

	GCI	ICTDI	NRI	NCSI	DDL	PSI	GEI	EDB	CI	CPI	GTI	FCI
count	76.000000	76.000000	76.000000	76.000000	76.000000	76.000000	76.000000	76.000000	76.000000	76.000000	76.000000	76.000000
mean	66.078947	65.078947	61.894737	54.255000	65.576184	0.322763	0.633684	70.199342	42.055000	55.342105	2.143276	284.696287
std	24.289786	18.071534	19.904826	23.195725	13.973113	0.779067	0.848850	10.234283	14.360367	18.745704	2.317043	279.032311
min	2.000000	0.000000	0.000000	3.900000	28.100000	-1.860000	-1.580000	30.850000	13.100000	18.000000	0.000000	27.860721
25%	56.000000	56.000000	57.000000	35.060000	57.725000	-0.227500	0.040000	64.265000	33.897500	40.500000	0.052250	127.230995
50%	75.000000	69.500000	63.500000	57.140000	66.815000	0.465000	0.495000	71.825000	40.170000	55.000000	1.011500	208.255223
75%	85.000000	79.000000	77.000000	71.755000	78.145000	0.950000	1.272500	78.095000	49.292500	72.250000	3.958000	355.705963
max	93.000000	90.000000	86.000000	96.100000	85.130000	1.540000	2.230000	86.590000	83.600000	88.000000	7.568000	1589.573888

Рисунок 2.13 – Базова статистика

Середнє значення вибірки характеризує розташування значень випадкової величини та вказує на центр розсіювання даних. Стандартне відхилення – це найпоширеніший показник розсіювання значень випадкової величини щодо її математичного очікування. Мінімальне значення - вказує на найменше значення вибірки на вказаному інтервалі даних. Максимальне значення, відповідно, вказує на найбільше значення вибірки.

З малюнку 2.13 видно, що загальна кількість спостережень для кожного показника складає 76. Середнє значення, стандартне відхилення, максимальне і мінімальне значення різняться.

Візуалізація даних - це представлення даних у вигляді, який забезпечує найефективнішу роботу людини, яка їх вивчає. Візуалізація даних знаходить широке застосування у багатьох сферах, таких як: наукових та статистичних дослідженнях, у педагогічному дизайні для навчання та тестування, у новинних зведеннях та аналітичних оглядах [77]. Візуалізація даних допомагає досягати результату, оцінювати значення інформації чи даних. Під візуалізацією даних мається на увазі представлення інформації у графічній формі, наприклад, у вигляді кругової діаграми, графіка або візуального представлення іншого типу [77]. Графіки зручно використовувати, якщо потрібно зобразити характер чи загальну тенденцію розвитку явища чи явищ. Лінії зручні і за зображенні кількох

динамічних рядів їхнього порівняння, коли потрібно порівняння темпи зростання [78]. Гістограми є одним із найважливіших інструментів аналізу даних. Подання результатів спостережень з допомогою дозволяє оцінити ряд статистичних показників, зробити висновки про функції розподілу і визначити можливі відхилення, і навіть порівняти набори даних [78].

Для візуалізації даних Pandas було використано бібліотеку Matplotlib. З її допомогою можна з легкістю будувати діаграми [79]. За допомогою вже імпортованого модуля Matplotlib.Pyplot та метода Plot() були побудовані графіки 2.14-2.15.

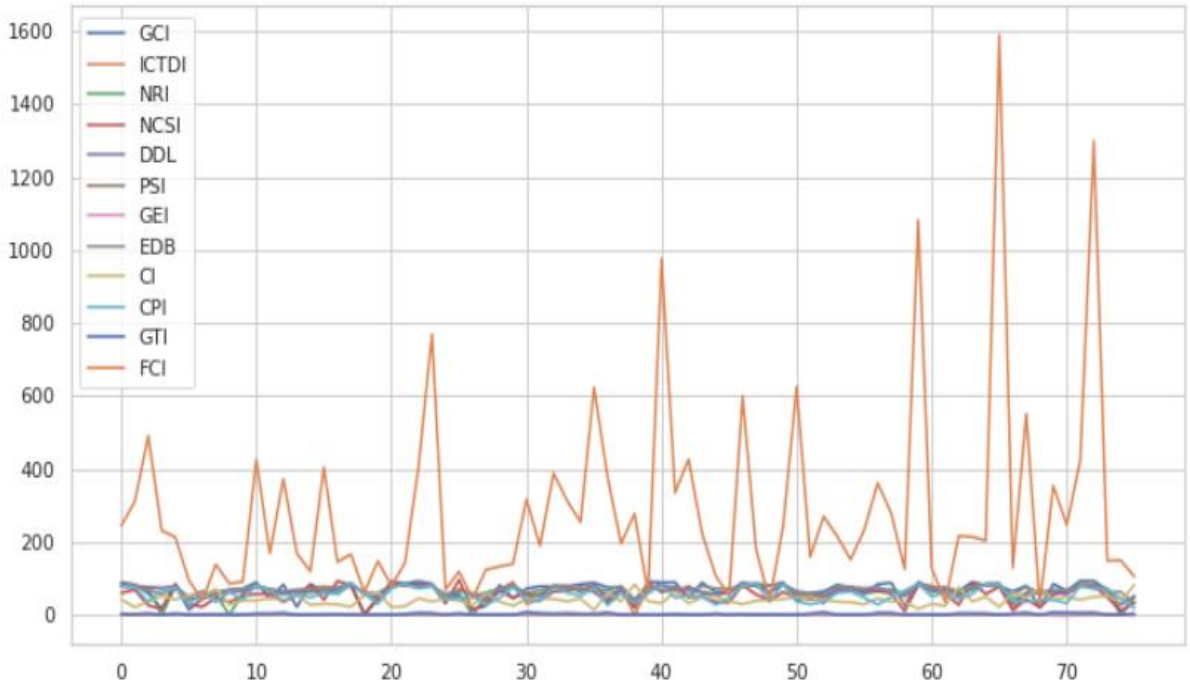


Рисунок 2.14 – Графік розподілу даних

На рисунку 2.14 зображено розподіл всіх даних. Як видно, показники, окрім FCI, знаходяться приблизно в одних межах. Показник FCI відрізняється, значення цієї характеристики значно більше, ніж значення інших. Максимальне і мінімальне значення цього показника дорівнює 1589,573888, 27,860721, відповідно. З метою кращого розуміння розподілу інших показників був зроблений ще один графік, він описує всі вхідні дані, окрім FCI (рис. 2.15).

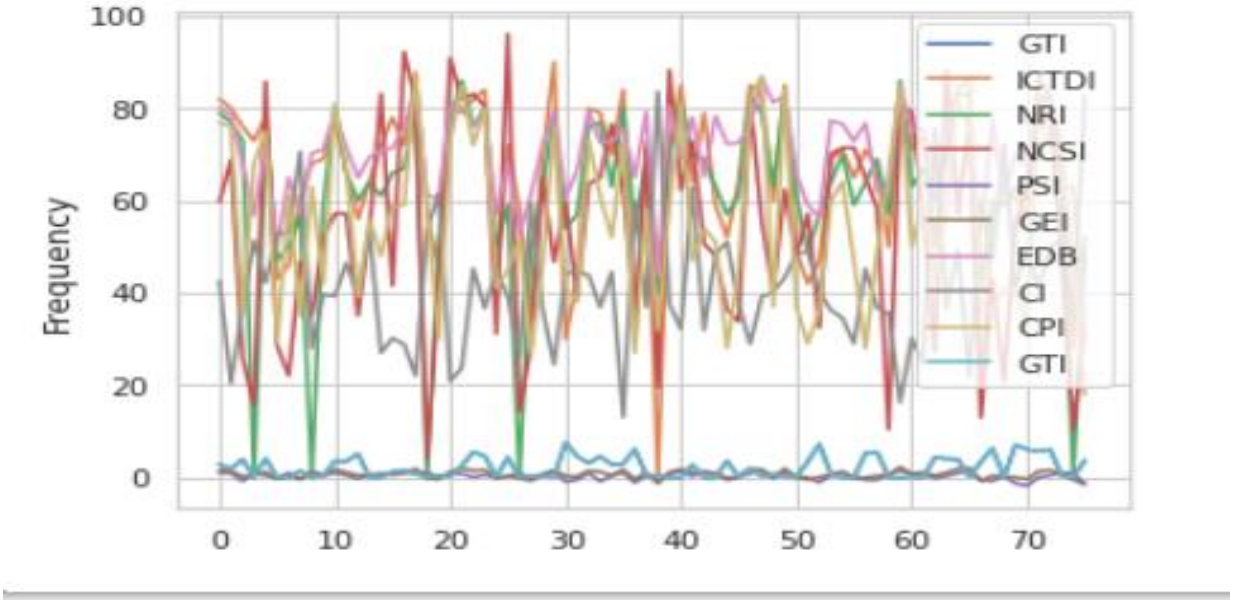


Рисунок 2.15 – Графік розподілу даних

Рисунок 2.15 описує розподіл даних. Загальна кількість спостережень описаних на рисунку – 76, значення показників варіюються від 0 до 100.

Наступник кроком є побудова гістограми для кожного показника окремо (рис. 2.16).

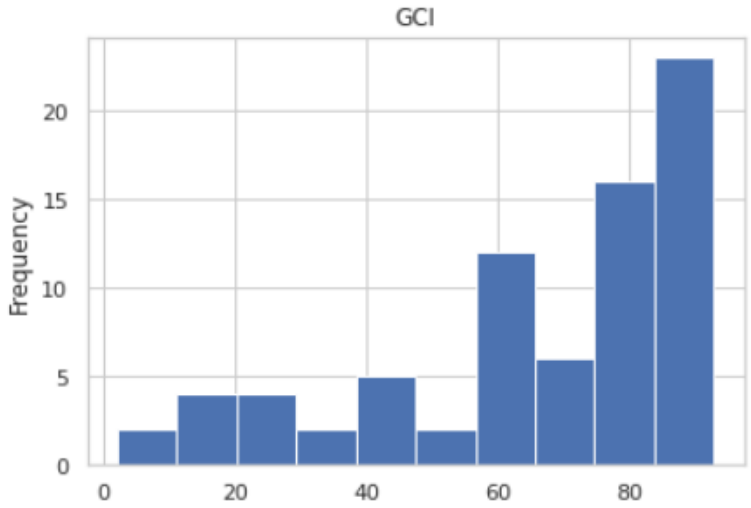


Рисунок 2.16 – Гістограма розподілу показника «GCI»

Як видно з рисунку 2.16, більшість значень знаходяться на проміжку 60-93. Кількість спостережень від 0 до 60 повторюється набагато менше.

На рисунку 2.17 показано, що найбільша кількість спостережень зосереджена на проміжку від 70 до 80. Тобто, найчастіше у вибірці повторюється саме ці значення.

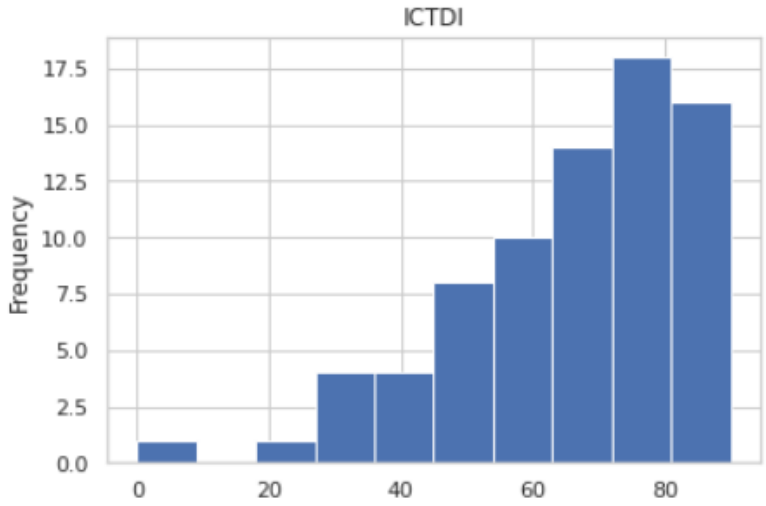


Рисунок 2.17 – Гістограма розподілу показника «ICTDI»

Рисунок 2.18 описує характеристики «NRI». Найбільші значення зосереджені на проміжку від 50 до 90, і повторюються більшу кількість разів, ніж значення від 0 до 50.

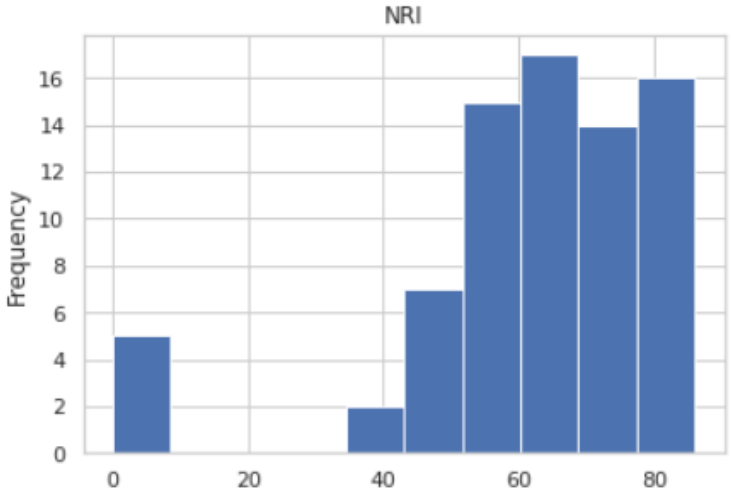


Рисунок 2.18 – Гістограма розподілу показника «NRI»

На рисунку 2.19, порівнюючи з попередніми, дані розподіляються більш рівномірно, найчастіше повторюються значення починаючи від 40 і до 90. Інші значення зустрічаються не так часто.

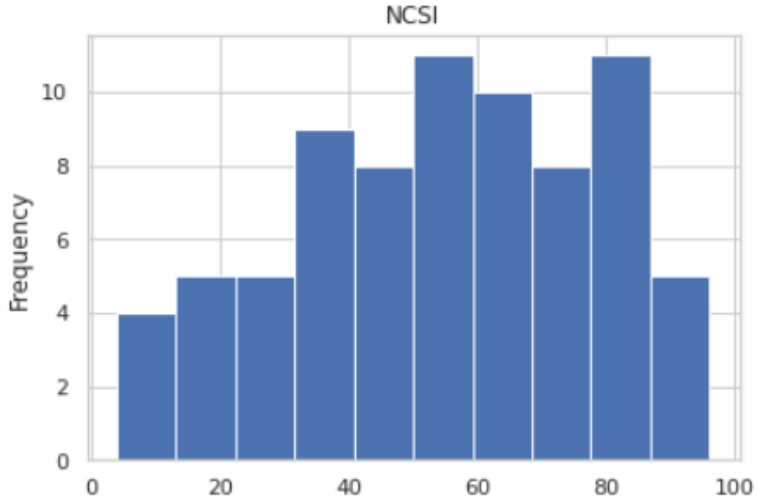


Рисунок 2.19 – Гістограма розподілу показника «NCSI»

На рисунках додатку Б зображено особливості розподілу факторів «PSI», «EDB», «CI» , «CPI», «GTI», відповідно, які також демонструють нерівномірність розподілу змінних.

Канонічний аналіз - це багатовимірний метод аналізу, що стосується визначення взаємозв'язків між групами змінних у наборі даних. Його основна мета – пошук максимальних кореляційних зв'язків між групами вихідних змінних [80]. Канонічний аналіз даних був проведений за допомогою програми STATISTICA (рис. 2.20).

	Canonical Analysis Summary (Convergensy.sta) Canonical R: .96635 Chi²(35)=297.75 p=0.0000	
N=76	Left Set	**Right Set**
No. of variables	5	7
Variance extracted	100.000%	85.9457%
Total redundancy	58.7705%	58.7748%
Variables: 1	GCI	PSI
2	ICTDI	GEI
3	NRI	EDB
4	NCSI	CI
5	DDL	CPI
6		GTI
7		FCI

Рисунок 2.20 – Результати канонічного аналізу по всім змінним

Отримане значення канонічного R достатньо велике і дорівнює 0.96635. Це свідчить про те, що наявний сильний кореляційний зв'язок між факторами, які

характеризують системи кібербезпеки та запобігання фінансовим злочинам. Критерій Пірсона, який у даному випадку дорівнює 297,75, підтверджує статистичну значимість коефіцієнта кореляції, і рівень значущості даного коефіцієнта не перевищує 0,05 (p=0,0000). Значення лівої множини, яка була сформована з індексів системи кібербезпеки, дорівнює 58,7705%. Дане значення говорить про те, що фактори, описані у правій множині, які характеризують рівень протидії фінансовим злочинам, пояснюють на 58,7705% мінливість факторів системи кібербезпеки. Система протидії фінансовим злочинам певною мірою залежить від системи кібербезпеки в країні. Фактори системи кібербезпеки на 58,7748% пояснюють мінливість факторів, які характеризують рівень протидії фінансовим шахрайствам. З проведеного аналізу показників видно, що фактори системи кібербезпеки мають вплив на процес протидії фінансовим злочинам.

Наступним етапом був аналіз впливу кожного фактору системи кібербезпеки на протидію фінансовим шахрайствам. На рисунку 2.21 показані результати канонічного аналізу фактору системи кібербезпеки GCI та факторів протидії фінансовим шахрайствам.

	Canonical Analysis Summary (Convergensy.sta) Canonical R: .62413 Chi²(7)=34.794 p=.00001	
N=76	**Left Set**	Right Set
No. of variables	1	7
Variance extracted	100.000%	18.2672%
Total redundancy	38.9537%	7.11577%
Variables: 1	GCI	PSI
2		GEI
3		EDB
4		CI
5		CPI
6		GTI
7		FCI

Рисунок 2.21 – Результати канонічного аналізу.

Як видно, отримане значення канонічного R не є достатньо великим(R = 0,62413). Це говорить про низьким кореляційний зв'язок між глобальним індексом кібербезпеки та запобігання фінансовим злочинам. Критерій Пірсона,

який дорівнює 34,794, також підтверджує статистичну незначущість коефіцієнта кореляції. Значення лівої множини дорівнює 38,9537% і говорить про те, що фактори, які описують протидію фінансовим злочинам, на 38,9537% пояснюють мінливість глобального індексу кібербезпеки.

Як видно з рисунка 2,21, фактори протидії фінансовим злочинам у дуже незначному відсотку залежать від обраного фактору системи кібербезпеки. Фактор глобального індексу кібербезпеки лише на 7,11577% пояснює мінливість факторів протидії фінансовим злочинам. Отримане значення говорить по те, що вплив глобального індексу кібербезпеки на протидію фінансовим шахрайствам низький і не має значущого впливу на систему протидії фінансовим шахрайствам.

На рисунку В.1 додатку В наведені результати аналізу впливу індексу розвитку інформаційно-комунікаційних технологій на протидію фінансовим шахрайствам. Отримане значення R є низьким, це говорить про низький кореляційний зв'язок між обраними факторами. Коефіцієнт Пірсона дорівнює 60,519, що підтверджує статистичну незначущість коефіцієнта кореляції. Значення факторів лівої множини дорівнює 57,6170%, тобто фактори протидії фінансовим шахрайствам на 57,6170% описують мінливість фактору системи кібербезпеки. Значення факторів правої множини дорівнює 12,6134%, тобто індекс розвитку інформаційно-комунікаційних технологій на 12,6134% описує мінливість факторів протидії фінансовим злочинам. Отримані значення вказують на те, що вплив індексу розвитку інформаційно-комунікаційних технологій на протидію фінансовим злочинам є, але він не настільки великий.

Рисунок В.2 додатку В описує вплив індексу мережевої готовності на фактори протидії фінансовим злочинам. Значення отриманого R є низьким, і це говорить про низький кореляційний зв'язок між факторами. Коефіцієнт Пірсона підтверджує припущення статистичної незначущість коефіцієнта кореляції. Значення факторів лівої множини складає 50,5428%. Це говорить про те, що фактори протидії фінансовим шахрайствам на 50,5428% описують мінливість індексу мережевої готовності. Значення факторів правої множини дорівнює

7,99112%. Це говорить про те, що індекс мережевої готовності на 7,99112% описує мінливість факторів протидії фінансовим злочинам. Отримані результати вказують на невисокий вплив фактору мережевої готовності на фактори протидії фінансовим злочинам.

Кореляційний зв'язок між факторами протидії фінансовим злочинам та національним індексом кібербезпеки є слабким, про це каже коефіцієнт кореляції, який дорівнює 0,74559 (рис. В.3). Критерій Пірсона дорівнює 57,225 і підтверджує, що коефіцієнт кореляції не є статистичну значущим. Значення надмірності для лівої множини, яка складається з фактору системи кібербезпеки, а саме фактору «Національний індекс кібербезпеки» дорівнює 55,5897%. Це означає, що фактори правої множини, які складаються з індексів протидії фінансовим злочинам, на 55,5897% пояснюють мінливість системи кібербезпеки. Протидія фінансовим шахрайствам частково залежить від національного індексу кібербезпеки, оскільки фактор системи кібербезпеки на 23,4440% описує мінливість системи протидії фінансовим злочинам. Хоча отримані значення є помірними, але цього є достатньо для доказу невеликого впливу показника «Національний індекс кібербезпеки» на протидію фінансовим шахрайствам в країнах.

На рисунку В.4 зображено результати аналізу впливу фактору «Рівень цифрової трансформації» на протидію фінансовим злочинам в країнах. Як видно, отримане значення канонічного R=0,95472. Це говорить про те, що існує сильний кореляційний зв'язок між факторами, які характеризують рівень цифрової трансформації та протидію фінансовим злочинам. Критерій Пірсона, який дорівнює 170,94, і рівень значимості якого не перевищує 0,05 (p=0,0000), підтверджує статистичну значущість коефіцієнта кореляції. Значення надмірності для лівої множини, яка складається з фактору системи кібербезпеки, а саме «Рівень цифрової трансформації», дорівнює 91,1491%. Цей свідчить про те, що фактори правої множини, які описують протидію фінансовим злочинам, на 91,1491% пояснюють мінливість індексу рівня цифрової трансформації, що свідчить про високе значення впливу. Процес протидії фінансовим шахрайствам

в країнах залежить від кіберзахисту фінансових систем, так як індекс рівня цифрової трансформації на 39,7615% описує мінливість факторів, які характеризують протидію фінансовим шахрайствам у країнах. Отримане значення є високим і це говорить про те, що система кібербезпеки (індекс рівня цифрової трансформації) має сильний вплив на протидію фінансовим махінаціям. Отже, в процесі аналізу впливу факторів кібербезпеки було виявлено один індекс, який має сильний вплив на фактори, які характеризують протидії фінансовим злочинам, цим фактор є «Рівень цифрової трансформації».

2.3.2 Усунення мультиколінеарності факторів із використанням методу головних компонентів

Метод головних компонентів — це техніка для зменшення розмірності наборів даних, підвищення інтерпретації, але водночас мінімізації втрати інформації. Це робиться шляхом створення нових некорельованих змінних, які послідовно максимізують дисперсію [81].

Метод головних компонентів виконаний на мові програмування Python. Його необхідність викликана високим рівнем кореляції між змінними, які було відібрано в результаті канонічного аналізу (рис. 2.22).

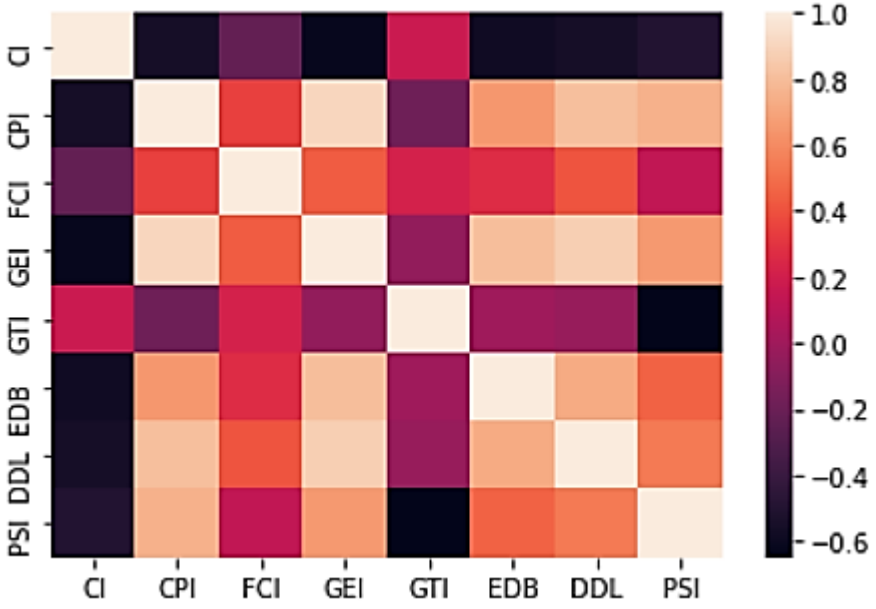


Рисунок 2.22 – Кореляційна матриця

Показники, які були вибрані, мають високий рівень мультиколлінеарності. Оскільки показники мультиколінеарні, то подальша робота з цими даними неможлива, тому й необхідно використати метод головних компонентів. Після побудови його графіка та виконання розрахунків видно, що отримано всього 4 статистично значущі компоненти. Як видно з рисунків 2.23-2.24, 4 компоненти накопичують варіацію 93% і рівень значущості кожного з них не повинен бути меншим за 0,05.

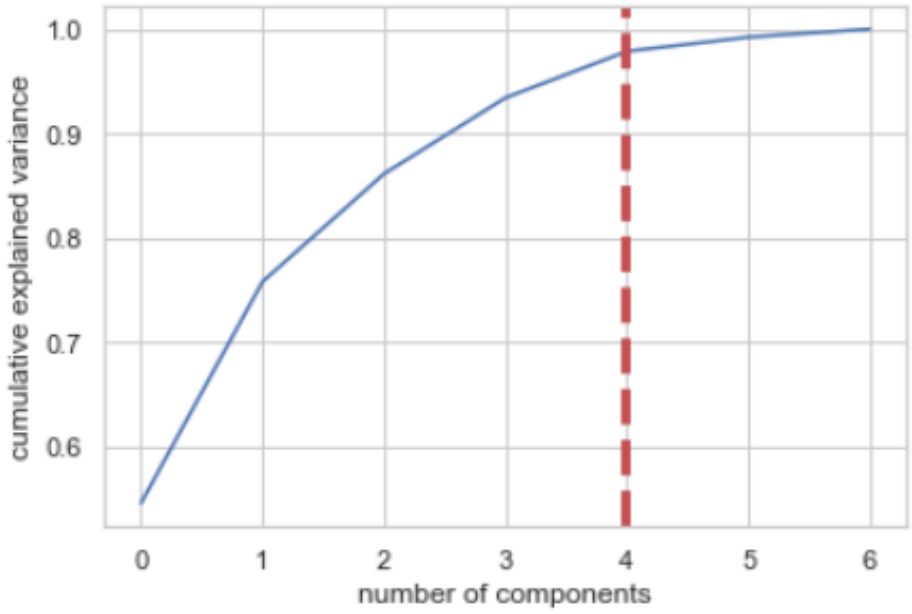


Рисунок 2.23 – Графік нових компонентів

	Cumulative Variance Ratio	Explained Variance Ratio
0	0.545733	0.545733
1	0.757906	0.212173
2	0.861269	0.103363
3	0.934204	0.072935

Рисунок 2.24 – Результати розрахунку

Результат отриманих значень головних компонент представлений на рисунку 2.24, які будуть використані для побудови нейронної мережі.

2.3.3 Побудова нейромережевої моделі

Нейронна мережа — це серія алгоритмів, які створені для розпізнавання основних зв'язків в наборі даних. Нейронні мережі можуть адаптуватися до зміни вхідних даних; таким чином мережа генерує найкращий можливий результат без необхідності перепроектувати критерії виводу. Результати розрахованих коефіцієнтів нейронної мережі представлені на рисунку 2.25.

```
array([[ 8.75068112e-01,  1.14594642e+00,  9.77763578e-01,
         3.46989096e-02,  1.16302930e+00,  2.97045324e-01,
        -1.45524418e-01],
       [ 7.58777731e-01,  9.68062577e-01,  8.20387503e-01,
        -1.51728868e+00,  1.10932929e+00, -1.26545664e-01,
         9.27752513e-02],
       [-1.50242411e+00, -5.38020594e-01, -2.13376592e-01,
        -3.57153423e-01, -1.03867137e+00,  7.55826948e-01,
         7.43208829e-01],
       [ 7.71698884e-01, -2.41547529e-01, -1.31992712e+00,
         6.48764461e-01,  6.79729155e-01, -9.31151315e-01,
        -1.93885806e-01],
       [ 1.12720062e-01,  6.36012744e-01,  1.48588382e-01,
         8.06136284e-03,  1.05562927e+00,  8.32725004e-01,
        -2.58779088e-01],
       [-7.78839522e-01, -1.13096672e+00, -1.97008703e+00,
         7.59520577e-01, -1.41457149e+00, -9.31151315e-01,
        -6.84996207e-01],
```

Рисунок 2.25 – Фрагмент розрахованих значень компонент

Для побудови нейромережевої сітки була використана активаційна функція ReLU. ReLU - це нелінійна функція активації. Ця функція є найчастіше використовуваною функцією. Її використовують для згорткових нейронних мереж та глибокого навчання для всіх шарів, крім вихідного.

На рисунку 2.26 зображено результати нейромережевої сітки, яка показує фактичні значення та прогнозовані значення, а також оцінку. Виходить, що критерій детермінації дорівнює за результатами тесту 0,799 та оцінка тренувального коефіцієнта детермінації дорівнює 0,821 (рис. 2.27). Модель містить три шари, в кожному шарі міститься по 45 вузлів. Її характеристики наведені у додатку Г.

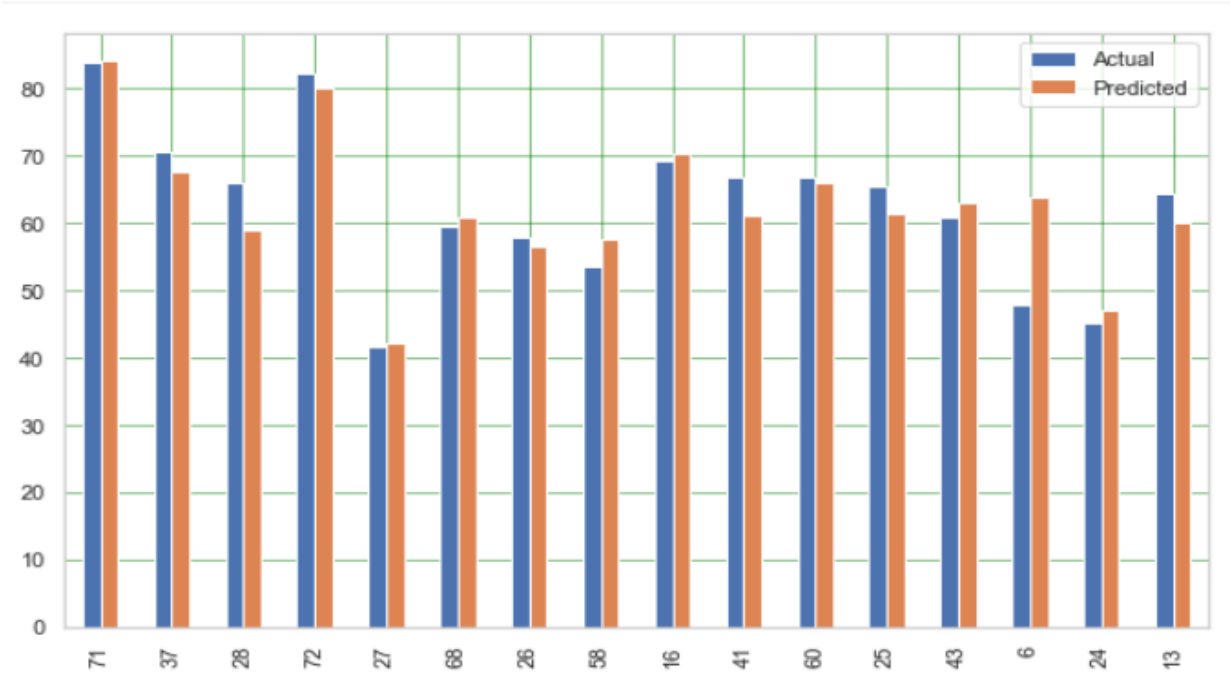


Рисунок 2.26 – Нейромережева сітка

```
Mean Absolute Error: 3.47495173501873
Mean Squared Error: 26.07682730734608
Root Mean Squared Error: 5.106547493889201
Test R^2 Score : 0.799
Training R^2 Score : 0.821
```

Рисунок 2.27 – Результати розрахунку

Такі оцінки як, середня абсолютна помилка (яка дорівнює 3,47495173501873) і середня квадратична помилка (яка дорівнює 26,07682730734608) описують порівняння якість прогнозованих даних і фактичних. Як видно з рисунка 2.27, помилки є незначними і це говорить про високу якість прогнозу.

Наступний етап – побудова кривої втрат. Одним з найбільш часто використовуваних графіків для налагодження нейронної мережі є крива втрат під час навчання (рис. 2.28).

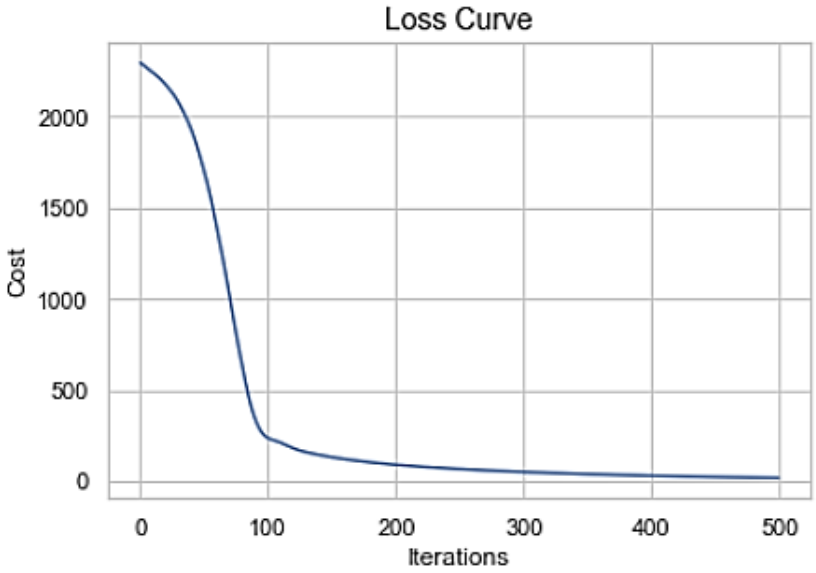


Рисунок 2.28 – Графік втрат

Дана крива вимірює помилку моделі і показує «наскільки погано працює модель». Усі ці показники фіксують продуктивність моделі, тому чим вони вищі, тим кращою стає модель. В даному випадку кількість втрат знижується, крива з часом зменшується і вирівнюється до певного рівня.

Коефіцієнти, які були отримані для нейромережевої сітки зображені на рисунку 2.29-2.30.

```
[(4, 45), (45, 45), (45, 45), (45, 1)]
[array([[-0.26004826, -0.12029681,  0.20590884, -0.06945696, -0.3113571 ,
         0.21721447, -0.30771875, -0.10858792, -0.2631856 , -0.21179765,
         0.03922925,  0.34470119,  0.00092399, -0.03584113, -0.21201222,
         0.39837075,  0.21987484, -0.09763449, -0.34011139, -0.13194794,
        -0.32285526,  0.35387933,  0.0824541 , -0.32409041, -0.31574998,
         0.01327073, -0.374209  , -0.2635563 ,  0.23914328, -0.16130783,
        -0.08460061, -0.0878057 , -0.27154981, -0.22982154,  0.18132776,
         0.0032163 , -0.04726616, -0.38168097, -0.05618342, -0.0240545 ,
        -0.01425243,  0.09723332,  0.25665729, -0.27056029,  0.04194362],
       [-0.37137422,  0.29138754,  0.4493272 , -0.09557417,  0.32288134,
        -0.40760883,  0.12893772, -0.281831  , -0.24050574,  0.45293821,
        -0.12016031,  0.40809908,  0.20684486, -0.14231219, -0.36514533,
        -0.4131254 ,  0.36199315, -0.07692699, -0.33685647,  0.35394341,
        -0.43567026,  0.46121382,  0.26172665,  0.32925829,  0.08731418,
        -0.21769305,  0.37403948, -0.31338637,  0.16617805,  0.32513892,
         0.27063393, -0.14837558, -0.05335017, -0.26485246,  0.47923464,
        -0.23381631,  0.43610618, -0.272976  , -0.34756817, -0.39744876,
         0.23763593, -0.327351  , -0.09516424, -0.16217758,  0.4214515 ],
       [-0.34914899,  0.03479934, -0.38569757, -0.05618054, -0.32536316,
         0.3083773 , -0.50888427,  0.03729243,  0.12824579, -0.29934257,
         0.18971017,  0.03014801, -0.32204607, -0.06726018,  0.35458359,
         0.03199527, -0.16551574,  0.27497358, -0.08863406, -0.07647509,
        -0.34039611,  0.16389043,  0.15351833,  0.11692195, -0.07506765,
        -0.0310238 ,  0.02180154,  0.3428187 , -0.37137128, -0.03728706,
        -0.46024447,  0.12702506, -0.55350416,  0.06861506,  0.0110601 ,
         0.08716832, -0.06640432, -0.26295906, -0.23210827, -0.11435545,
        -0.39570294, -0.34034274,  0.08491891, -0.00243018,  0.32409415],
       [ 0.28956792,  0.6455377 ,  0.28963336, -0.00493871, -0.44261102,
        -0.53542088, -0.18523631,  0.47829689,  0.0411324 ,  0.1874174 ,
         0.42292073, -0.36328156, -0.1570596 , -0.23801606, -0.36803248,
        -0.15363543, -0.09741371,  0.60318768,  0.2947905 , -0.05998706,
        -0.11174835,  0.19493713,  0.42810767,  0.1572119 , -0.05909069,
        -0.21803694,  0.13179591, -0.24389973, -0.12475218,  0.03903671,
        -0.3136397 , -0.33920583,  0.0564947 , -0.28340583, -0.40138855,
         0.41286151, -0.02299765, -0.15389607, -0.32725482, -0.36204329,
        -0.24209945, -0.63771505,  0.04305627,  0.13866181,  0.3861812 ]]),
```

```
array([[-2.09553329e-03, -2.83383686e-01, -1.89348098e-01, ...,
         4.10836329e-09,  1.78167763e-01, -2.12532525e-02],
       [-1.56179435e-07,  1.68390463e-01, -4.12112578e-02, ...,
        -7.84874525e-03,  3.62979784e-01,  1.52828725e-01],
       [-5.30306052e-13, -3.59982154e-01, -4.34660216e-01, ...,
        -8.13580897e-03,  2.33873914e-01,  1.51226849e-01],
       ...,
       [-1.00980314e-02,  2.62556370e-01,  1.98985453e-02, ...,
        -4.36060544e-09,  3.31775193e-01, -1.40769837e-01],
       [-7.81491532e-03, -1.47737230e-01, -4.39706741e-01, ...,
        -1.85035924e-03, -1.51867423e-01,  2.72673594e-01],
       [-1.27720020e-03,  1.45879762e-01,  1.30810459e-01, ...,
        -5.32661953e-05,  3.22773706e-01,  2.96754380e-01]]),
array([[ 6.67087614e-03,  2.57824102e-04, -2.10369387e-04, ...,
         1.30100327e-05,  3.10330738e-10, -3.63647439e-03],
       [-2.22895452e-02,  3.96579367e-02, -2.10459127e-01, ...,
         2.29254822e-12,  1.34435617e-01,  1.47489465e-01],
       [-7.98080545e-02,  6.39805649e-02, -2.07827714e-01, ...,
         6.69372477e-02, -2.34963646e-01,  4.64588216e-01],
       ...,
       [-5.49662702e-04,  9.97029747e-04,  2.77062508e-04, ...,
         3.03950320e-13, -3.10835098e-10, -6.57735393e-05],
       [-1.54747232e-01, -6.46594534e-02,  7.36793926e-02, ...,
        -1.13950265e-01, -9.87774397e-02,  3.27221960e-01],
       [ 1.27878535e-02, -8.27450917e-02, -7.72087166e-02, ...,
        -2.36712608e-01,  2.10399132e-01,  5.84909485e-02]]),
```

Рисунок 2.29 – Коефіцієнти нейромережевої сітки

```
array([[-2.25687092e-01],          [-1.49110418e-01],
       [-4.25735261e-02],          [-3.89642835e-02],
       [-3.20274899e-02],          [ 4.35854289e-01],
       [-2.52926024e-01],          [ 1.95112094e-01],
       [ 4.52188309e-01],          [ 4.82352862e-01],
       [ 3.49553936e-01],          [-9.72168594e-02],
       [-8.57636211e-03],          [ 3.83305037e-01],
       [-3.17339902e-01],          [-1.74542005e-02],
       [-1.04082447e-01],          [-2.59554935e-01],
       [-2.69530392e-02],          [-5.32480436e-02],
       [-2.28642096e-01],          [-3.13696354e-01],
       [-2.27235498e-01],          [-3.34995432e-01],
       [-2.59030798e-01],          [-3.39534231e-01],
       [-2.94106680e-01],          [ 4.27805082e-01],
       [ 4.95360488e-01],          [ 4.69683121e-01],
       [ 2.99785542e-01],          [ 6.01660508e-05],
       [-3.01470028e-01],          [ 4.53446623e-02],
       [ 4.66513231e-01],          [-1.43142808e-01],
       [-2.97192536e-01],          [-1.58161829e-01],
       [-1.10176387e-01],          [ 1.62127858e-01],
       [ 2.18239491e-01],          [-2.67300605e-01],
       [ 3.80657163e-01],          [-1.48890230e-01],
                                   [ 4.75176427e-01]])]
```

Рисунок 2.30 – Коефіцієнти нейромережевої сітки

На малюнку 2.26 зображено графік нейронної мережі, параметри моделі були підібрані вручну, як видно, мережа вийшла досить гарною, і передбачувані значення дуже близькі до реальних значень. Існує так само і метод пошуку параметрів моделі по решітці. Цей метод передбачає, що підбір гіперпараметрів задаються вручну, потім виконується повний перебір. Популярною реалізацією цього методу є Grid Search із sklearn. У цій роботі був виконаний метод підбору параметрів, три варіації нейронної мережі були випробувані і був обраний найкращий варіант мережі (рис. 2.31). Результати розрахунку критерію детермінації, середньої абсолютної помилки, середньої квадратичної помилки представлені на рисунку 2.32.

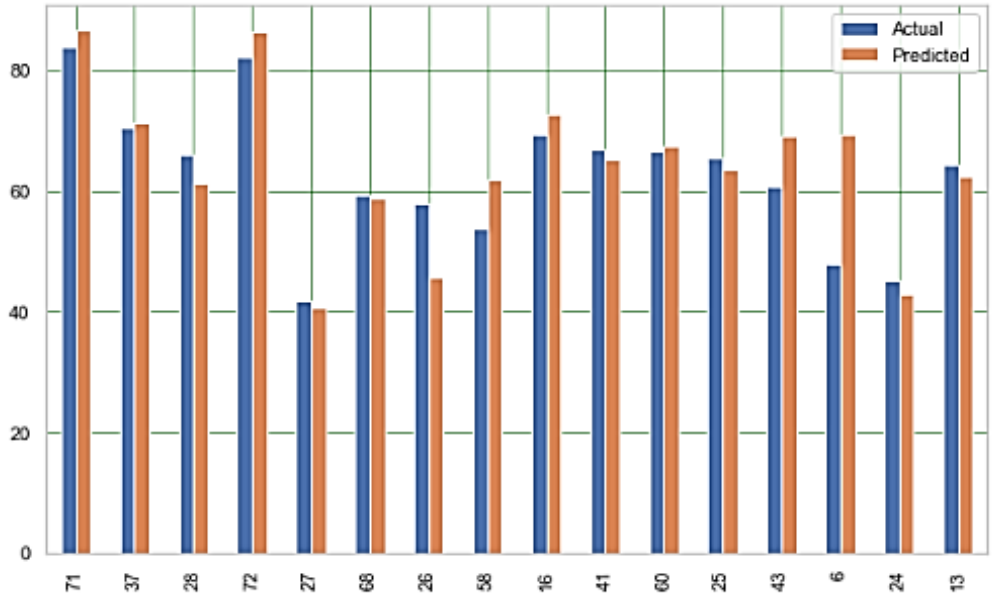


Рисунок 2.31 – Нейромережева сітка

```
Mean Absolute Error: 4.766978809140717
Mean Squared Error: 51.81089511376816
Root Mean Squared Error: 7.197978543575145
Test R^2 Score : 0.601
Training R^2 Score : 0.889
```

Рисунок 2.32 – Результати розрахунку

Як видно з рисунка 2.31, передбачувані значення мають доволі високі відхилення від реальних значень. Отриманий коефіцієнт детермінації невисокий, середня абсолютна помилка (4,766978809140717) і середня квадратична помилка (51,81089511376816). Отже, перша нейромережева модель є кращою для подальшого використовування і прогнозування.

2.3.4 Регресійний аналіз

Регресійний аналіз — це набір статистичних процесів для оцінки зв'язків між залежною змінною та однією або кількома незалежними змінними [25]. Регресійний аналіз в основному використовується для двох концептуально різних цілей. По-перше, регресійний аналіз широко використовується для передбачення та прогнозування, де його використання суттєво перекривається сферою машинного навчання [24]. По-друге, у деяких ситуаціях регресійний аналіз можна використовувати для висновку про причинно-наслідкові зв’язки між незалежними та залежними змінними [23].

Для аналізу та прогнозування впливу факторів системи кібербезпеки на протидію фінансовим шахрайствам було проведено регресійний аналіз (рис. 2.33). Проведений регресійний аналіз показав, що більшість параметрів має значення P, яке перевищую 0,05. Відповідно, ці параметри не мають впливу на протидію фінансовим шахрайствам.

Наступний етап – проведення відбору параметрів, значення P яких нижче 0,05. Відібраними параметрами для регресії є : «EDB», «CPI», «FCI». FCI-цей параметр був залишений тому що індекс фінансової таємності все ж таки важливий для системи протидії фінансовим злочинам. Відхилення значення 0,06 від 0,05 незначне, ґрунтуючись на цьому, було прийнято рішення залишити цей

показник і провести регресійний аналіз, ґрунтуючись на трьох факторах (рис. 2.33).

OLS Regression Results

Dep. Variable:	DDL	R-squared (uncentered):	0.989
Model:	OLS	Adj. R-squared (uncentered):	0.988
Method:	Least Squares	F-statistic:	917.1
Date:	Thu, 20 Jan 2022	Prob (F-statistic):	1.94e-65
Time:	17:50:40	Log-Likelihood:	-254.77
No. Observations:	76	AIC:	523.5
Df Residuals:	69	BIC:	539.9
Df Model:	7		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
PSI	-4.2121	2.741	-1.537	0.129	-9.680	1.255
GEI	3.3332	2.472	1.349	0.182	-1.598	8.264
EDB	0.5366	0.069	7.748	0.000	0.398	0.675
CI	0.0723	0.071	1.016	0.313	-0.070	0.214
CPI	0.4296	0.103	4.177	0.000	0.224	0.635
GTI	-0.5404	0.650	-0.831	0.409	-1.838	0.757
FCI	0.0047	0.003	1.376	0.173	-0.002	0.012

Omnibus:	8.057	Durbin-Watson:	2.162
Prob(Omnibus):	0.018	Jarque-Bera (JB):	13.461
Skew:	-0.286	Prob(JB):	0.00119
Kurtosis:	4.981	Cond. No.	1.48e+03

Рисунок 2.32 – Результати регресійного аналізу

OLS Regression Results

Dep. Variable:	DDL	R-squared (uncentered):	0.989
Model:	OLS	Adj. R-squared (uncentered):	0.988
Method:	Least Squares	F-statistic:	2126.
Date:	Thu, 20 Jan 2022	Prob (F-statistic):	6.24e-71
Time:	17:53:51	Log-Likelihood:	-257.13
No. Observations:	76	AIC:	520.3
Df Residuals:	73	BIC:	527.3
Df Model:	3		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
EDB	0.6010	0.047	12.902	0.000	0.508	0.694
CPI	0.3881	0.059	6.634	0.000	0.272	0.505
FCI	0.0061	0.003	1.908	0.060	-0.000	0.013

Omnibus:	16.234	Durbin-Watson:	2.211
Prob(Omnibus):	0.000	Jarque-Bera (JB):	49.537
Skew:	-0.490	Prob(JB):	1.75e-11
Kurtosis:	6.832	Cond. No.	35.5

Рисунок 2.33 – Результати регресійного аналізу

У результаті проведення регресійного аналізу маємо прогнозовані значення для тестового та тренувального набору даних (рис. 2.34-2.37).

```
38    40.289868
13    64.451372
61    68.930772
14    62.065520
45    61.376066
9     59.607550
11    69.969857
25    59.177610
49    83.789158
62    56.873089
34    65.434820
32    78.437056
73    65.492145
46    81.183942
29    78.009303
24    50.685918
dtype: float64
```

Рисунок 2.34 – Прогнозовані значення для тестового набору даних

```
Mean Absolute Error: 2.7086619105116534
Mean Squared Error: 11.779761215591133
Root Mean Squared Error: 3.4321656742632825
```

Рисунок 2.35 – Результати розрахунку

Середня абсолютна помилка для тестового набору даних складає 2,7086619105116534, що є допустимим. Середня квадратична помилка прогнозу складає 11,779761215591133, середньоквадратична помилка складає 3,432165674263282.

```
                1    78.597861
71  83.116014  25    59.177610
7   48.714471  28    61.899997
5   41.986723  3     61.929266
24  50.685918  57    57.911556
38  40.289868  47    86.897676
9   59.607550  46    81.183942
14  62.065520  58    63.895372
37  71.341739  39    71.660891
53  71.067008  0     79.543090
64  83.077241  43    68.212335
16  69.491275  74    56.676619
62  56.873089  59    90.679383
67  63.852648  6     62.943810
55  63.432625  63    70.461803
68  52.700596  34    65.434820
30  54.270308  18    59.179764
```

Рисунок 2.36 – Прогнозовані значення для тренувального набору даних

```
Mean Absolute Error: 4.424836546812158
Mean Squared Error: 43.873275551329534
Root Mean Squared Error: 6.623690478225076
```

Рисунок 2.37 – Результати розрахунку

Середня абсолютна помилка для тренувального набору даних складає 4,424836546812158. Середня квадратична помилка прогнозу складає 43,873275551329534, середньоквадратична помилка складає 6,623690478225076.

Проведений регресійний аналіз допоміг виявити фактори, які в поєднанні з індексом рівня цифрової трансформації є важливими в процесі протидії фінансовим шахрайствам у країнах. Цими факторами є індекс легкості ведення бізнесу, індекс споживчих цін, індекс фінансової таємниці.

На сьогодні проблема впливу розвитку технологій та цифровізації економіки на зростання кількість кібершахрайств у сфері фінансів у всьому світі постає перед людством. Темпи розвитку технологій і якості кіберсистем підвищується, і, звичайно, зростає кількість шахрайств у різних сферах, а особливо у сфері фінансів. На даний момент системи фінансів не мають достатнього кіберзахисту, і є вразливими у час інформаційних технологій. Інформація, гроші у різних валютах, цінні папери можуть бути викраденими хакерами з різних куточків світу. З метою покращення якості функціонування фінансової сфери, а також для зменшення кількості кіберзлочинів і протидії фінансовим махінаціям був проведений аналіз потенційної конвергенції системи кібербезпеки та протидії фінансовим шахрайствам. У ході виконання роботи статистичний та візуальний аналіз описали фактори, які можуть впливати на рівень захищеності фінансової сфери. За допомогою канонічного аналізу був виявлений фактор «DDL», який характеризує рівень цифрової трансформації у сфері кібербезпеки. Процес протидії фінансовим шахрайствам в країнах залежить від кіберзахисту фінансових систем, а саме від індексу рівня цифрової трансформації. Цей висновок був зроблений на основі розрахунків, зроблених за допомогою канонічного аналізу. У роботі було використано два методу аналізу – нейромережева сітка і регресійний аналіз. Це було зроблено з метою порівняння двох видів аналізу та виявити, який з них краще описує потенційний процес конвергенції системи кібербезпеки та протидію фінансовим шахрайствам. Основуючись на цих видах аналізу, було виявлено, що важливими показниками в сфері кібербезпеки та протидії фінансовим шахрайствам є

фактори рівня цифрової трансформації і три показника, які характеризують індекс легкості ведення бізнесу, індекс споживчих цін, індекс фінансової таємниці, відповідно.

Пункт 2.3 було виконано із використанням матеріалів публікацій виконавців [82].

3 АЛГОРИТМИ РОЗПІЗНАВАННЯ ПОВЕДІНКИ КІБЕРШАХРАЇВ

3.1 Формування кіберпрофілю жертви: гендерний аналіз

За останні двадцять років у світі спостерігається бурхливий розвиток науково-технічного прогресу завдяки Четвертій промисловій революції. Його результати призвели до повної інтеграції та впровадження різноманітних технологій у всі сфери життя суспільства. На державному рівні набувають поширення інструменти, починаючи з електронного уряду та демократії і закінчуючи технологіями розумного міста. Це сприяє сталому економічному та соціальному розвитку цілих регіонів і веде до підвищення рівня життя населення. Прогрес торкнувся і бізнес-сфери. Неможливо уявити діяльність будь-якого суб'єкта господарювання без використання інформаційних систем. Наприклад, компанії використовують Системи планування ресурсів підприємства для повної автоматизації бізнес-процесів, Customer Relationship Management Systems для інтеграції відносин з клієнтами в діяльність компанії, Business Intelligent Systems для аналізу діяльності та прийняття рішень тощо. Крім того, комп’ютери та цифрові технології вплинули на життя багатьох людей у світі. Завдяки їм з’явилися можливості дистанційної роботи, заробітку в Інтернеті, здійснення платежів через мобільні додатки, отримання інформації з різних куточків світу, не виходячи з дому, навчання в престижних світових університетах, не відвідуючи їх фізично тощо.

Усі ці приклади свідчать про позитивний вплив техніки на існування та розвиток суспільства та держави. Але водночас набуло поширення таке явище, як кіберзлочинність, яка передбачає здійснення протиправних дій щодо фізичної чи юридичної особи, і навіть держави, із застосуванням комп’ютерних технологій. У результаті відбувається незаконне привласнення або порушення персональних даних, що призводить до втрати фінансових ресурсів особами, компаніями та державними установами. Статистичний аналіз даних щодо вартості витоку даних через кіберзлочинність показує її поступове зростання, тобто у 2022 році вона склала 4,35 млн доларів, що на 24,29% більше, ніж у 2014

році [83]. При цьому найбільш значні обсяги збитків характерні для таких галузей, як охорона здоров'я (10,1 млн. дол. США), фінансова (5,97 млн. дол. США), фармацевтична (5,01 млн. дол. США), технологічна (4,97 млн. дол. США) та ін. енергетики (4,72 млн дол. США) [84]. Ця інформація відображає не тільки фінансові втрати компаній, але й ідентифікує їх клієнтів, оскільки втрата персональних даних призводить до збільшення кількості жертв кібершахраїв після таких витоків даних. За даними дослідження Comparitech, 71,1 мільйона людей щороку страждають від кіберзлочинів; тобто близько 1% населення планети коли-небудь стикалося з цим видом злочинності [85]. Ці статистичні дані свідчать про те, що кіберзлочинність є глобальною проблемою, яка потребує відповідних заходів боротьби з нею та протидії. Тому в даному контексті постає необхідність визначити аспекти, критично важливі для формування портрета потенційної жертви кібершахрая. Поведінка жінок і чоловіків може відрізнятися в ситуаціях, коли вони стикаються з подібними злочинами. Крім того, ймовірний вплив може сформувати ставлення чоловіків і жінок до використання засобів особистого кіберзахисту. Такі аспекти потребують відповідного аналізу та дослідження, проведеного в даній науковій роботі.

Кіберзлочини набули актуальності в науковому середовищі з кінця 90-х років 20 століття, коли персональні комп'ютери стали доступні багатьом користувачам світу, а явище кібершахрайства почало набувати масового характеру. На сьогоднішній день сформувалося багато наукових напрямків, які досліджують різні аспекти цієї проблеми. Найбільша кількість досліджень відноситься до галузі інформатики. Варто згадати публікації, які пропонують методи виявлення кіберзлочинів, такі як машинне навчання [86], нечітка логіка та аналіз даних [87], нейронні мережі [88], блокчейни [89], тощо.

Також виділено напрямок дослідження психологічних та соціальних аспектів кіберзлочинності. Дюпон і Холт обґрунтували критичну роль людського фактору в кіберзлочинах [90]. Лазар та ін. досліджували різницю у сприйнятті кібершахрайства жінками та чоловіками. Вони прийшли до висновку, що психосоціальні кіберзлочини є більш гендерними, тоді як соціально-

економічні шахрайства жодним чином не залежать від статі [91]. Юрина Коннолі та Борріон висунули гіпотезу та перевірили свідомий вибір жертви заплатити викуп шахраю [92]. Witsenboer та ін. представила результати анкетування школярів, які дали змогу зробити висновок про необхідність набуття відповідних знань щодо використання засобів індивідуального захисту, починаючи зі шкільного віку [93]. Бретт Друрі та ін. доведено, що популярність соціальних мереж призвела до їх використання для кібершахрайства, оскільки вони можуть охопити велику аудиторію, створюючи широкі можливості для злочинців [94]. Лі та ін. досліджував ризики жертв фішингу та запропонував теоретичну перспективу розуміння впливу звичайних дій в Інтернеті на можливості кіберзлочинців [95].

Незважаючи на значний внесок науковців у дослідження проблем кіберзлочинності, питання їх гендерного аналізу та визначення критичних характеристик для формування портрета жертв кібершахрайства потребує додаткового дослідження.

Для цього дослідження автор використав результати опитування «Спеціальний Євробарометр 499: ставлення європейців до кібербезпеки (кіберзлочинності)», проведеного Європейською комісією у 2019 році [96]. Дані охоплюють респондентів з 28 європейських країн, таких як Бельгія, Данія, Греція, Іспанія, Фінляндія, Франція, Ірландія, Італія, Люксембург, Нідерланди, Австрія, Португалія, Швеція, Німеччина, Велика Британія, Болгарія, Кіпр, Чехія, Естонія, Угорщина, Латвія, Литва, Мальта, Польща, Румунія, Словаччина, Словенія та Хорватія. Загалом було опитано 27607 осіб, з них 48,44% чоловіків та 51,56% жінок. Найбільша кількість респондентів у Німеччині (16,00%), Великій Британії (13,00%), Франції (13,00%), Італії (12,00%), Іспанії (9,00%) та Польщі (8,00%).

Вибрані дані стосуються питань про типи пристроїв, якими користуються респонденти, типи їхньої онлайн-діяльності, обізнаність про ризики кіберзлочинності, особистий досвід щодо можливостей і результатів впливу на різні види кібершахрайства та інструменти для боротьби з потенційним

шахрайством. Ця інформація була обрана для проведення гендерного аналізу та формування портрету потенційної жертви кіберзлочину, щоб виділити можливі категорії осіб з урахуванням їх гендерного розподілу за типом пристроїв, обізнаності про можливі ризики, впливу методів захисту тощо.

Основним методом цього дослідження є гендерний аналіз. Організація з безпеки та співробітництва в Європі визначає цю концепцію як «збір та аналіз даних, дезагрегованих за статтю, щоб виявити будь-який різний вплив дії на жінок і чоловіків, а також наслідки гендерних ролей і обов'язків. Це також передбачає якісний аналіз які допомагають з'ясувати, як і чому виникли ці різні ролі, відповідальність і вплив» [97]. Його реалізація також передбачає порівняння та оцінку поточних і майбутніх подій з урахуванням гендерних особливостей [98].

Застосування цього аналізу вимагає наступних етапів. На першому етапі формується та розподіляється за статтю база даних емпіричних даних. Далі вибираються найбільш значущі характеристики, за якими буде проходити порівняння. На наступному етапі дані візуалізуються за допомогою діаграм і графіків. Ці результати аналізуються та формуються відповідні висновки щодо гендерних відмінностей у досліджуваних явищах.

Результати опитування показали, що у повсякденному житті найбільша кількість респондентів переважно використовують смартфони та персональні комп'ютери для виходу в Інтернет (рис. 3.1). І чоловіки, і жінки більше віддають перевагу смартфонам, ніж комп'ютерам через їх мобільність для виконання багатьох повсякденних завдань, хоча відсоток жінок у цьому випадку вищий. Чоловіки активніше за жінок користуються комп'ютерами, телевізорами та ігровими приставками.

Для виконання яких завдань респонденти використовують смартфони, комп'ютери та інші пристрої? Аналіз активності в Інтернеті показує, що найпопулярнішими є листування електронною поштою, читання блогів і форумів, онлайн-банкінг, соціальні мережі, онлайн-магазини та миттєві повідомлення (рис. 3.2). Більше 50% чоловіків і жінок вибирають ці види.

Більшість жінок віддають перевагу соціальним мережам як засобу спілкування. Оскільки вони в основному користуються смартфонами (рис. 3.1), ця інформація підтверджує, що їх більше цікавить швидкий і мобільний зв'язок, враховуючи їхню зайнятість на роботі, вдома та з дітьми. Чоловіки є активними користувачами в усіх інших онлайн-завданнях. Оскільки вони в основному використовують різні пристрої (рис. 3.1), разом із широким спектром діяльності, можна зробити висновок, що чоловіки проводять більше вільного часу в Інтернеті і не обмежуються лише комунікативними формами спілкування, як жінки.

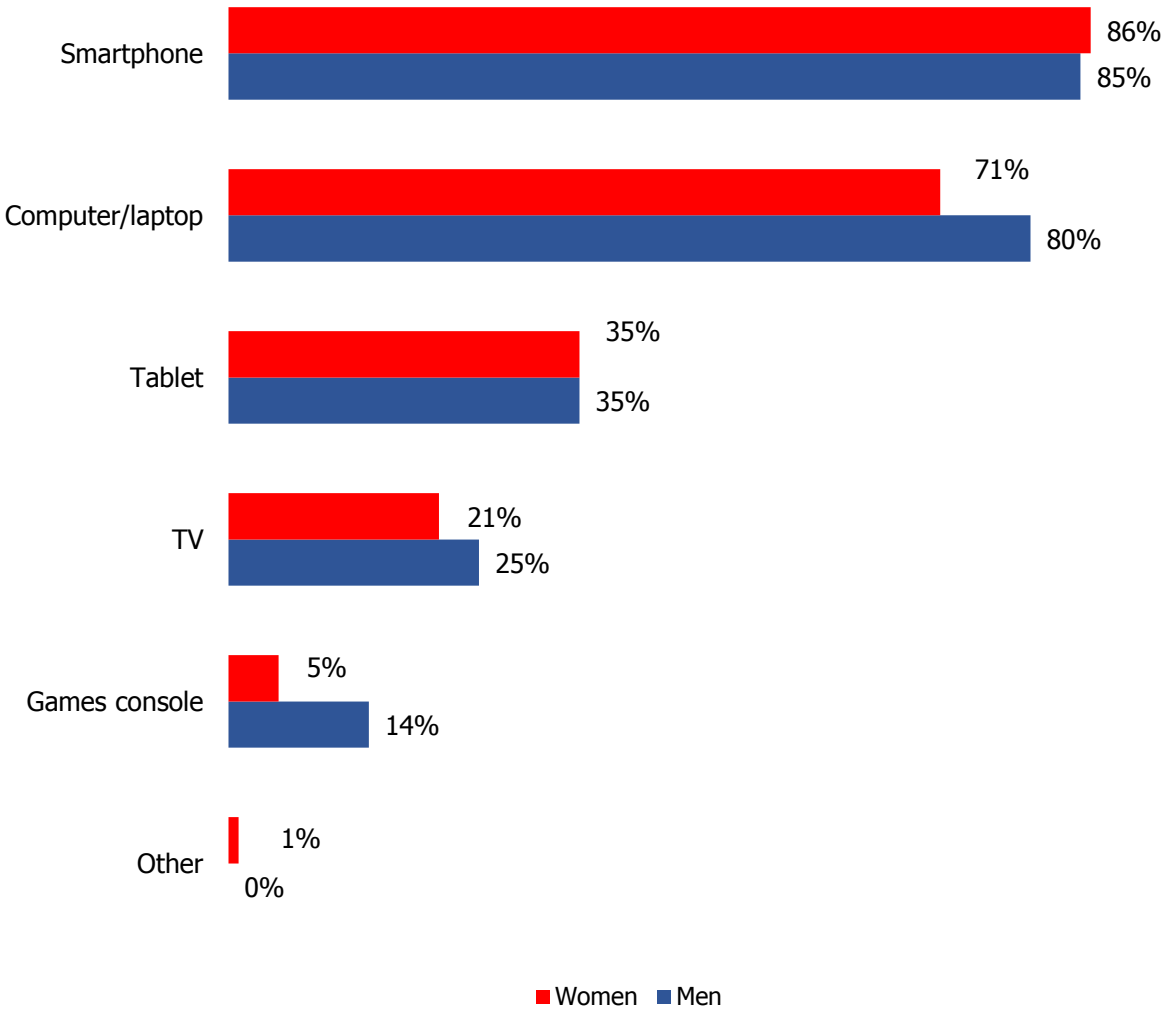


Рисунок 3.1 – Гендерний розподіл респондентів щодо використання пристроїв для доступу до Інтернету (створено автором за результатами опитування [96])

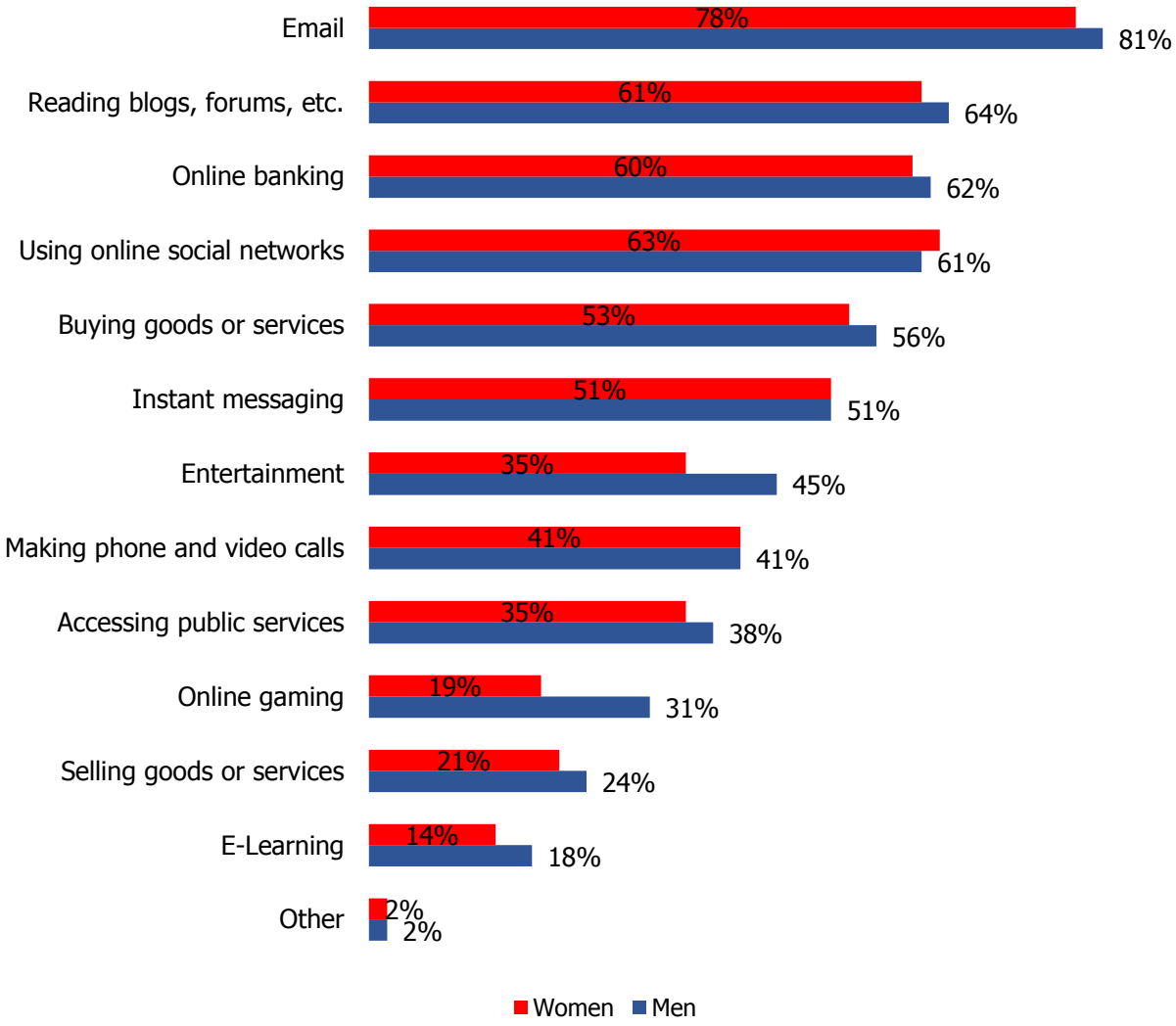


Рисунок 3.2 – Гендерний розподіл респондентів щодо їх активності в Інтернеті (створено автором за результатами опитування [96])

Високий відсоток активності електронної пошти можна віднести до найпопулярнішого кібершахрайства – фішингу, що відбувається шляхом надсилання електронних листів для отримання приватних даних користувача. Оскільки і чоловіки, і жінки переважно займаються цим видом діяльності, вони можуть швидко стати мішенню для кібершахраїв, хоча чоловіки будуть більш вразливими, ніж жінки.

Об’єктом кіберзагроз може бути кожен вид онлайн-діяльності. Тому кожен користувач повинен мати інформацію про ймовірність ризику стати жертвою кіберзлочинців. 55% чоловіків вважають себе більш обізнаними про ризики

кібершахрайства. З іншого боку, 55% жінок вважають себе менш обізнаними. Тобто чоловіки вважають себе більш впевненими у питаннях, пов'язаних із кіберзлочинністю, що може вплинути на зниження уваги до цієї проблеми. Оскільки жінки менш обізнані в цьому питанні, вони більше зосередяться на пошуку додаткових джерел інформації щодо особистого захисту.

Аналіз ситуацій, у яких респонденти постраждали від кіберзлочинності, показує, що чоловіки та жінки ставали жертвами у більшості випадків, коли отримували шахрайські повідомлення чи дзвінки (38% чоловіків та 33% жінок) або знаходили шкідливе програмне забезпечення на своїх комп'ютерах (31% чоловіків і 24% жінок) (рис. 3.3). Так, більше третини європейців хоча б раз стикалися з фішинговими, вішинговими та вірусними кібератаками.

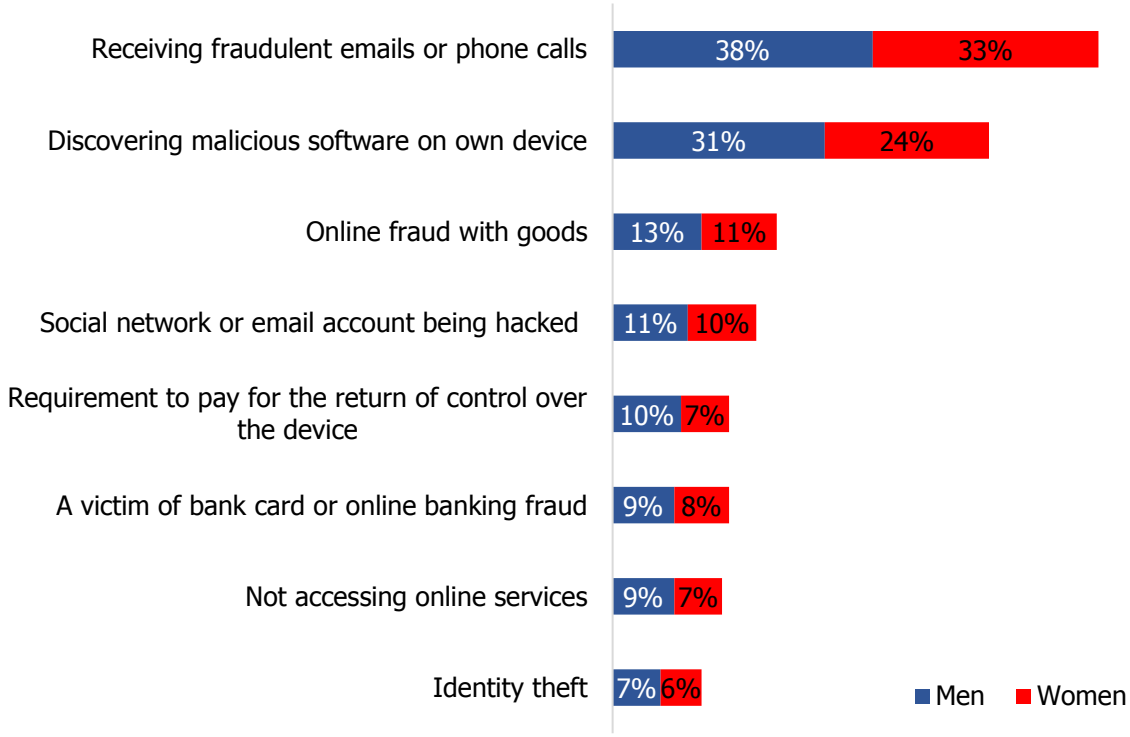


Рисунок 3.3 – Гендерний розподіл респондентів щодо ситуацій кібершахрайства (створено автором за результатами опитування [96])

Інші випадки стати жертвами відзначили 6%-13% користувачів, що значно менше, ніж для описаних вище ситуацій. Це пояснюється високими стандартами захисту інформації програмних додатків, за допомогою яких респонденти

здійснюють онлайн-діяльність. Наприклад, користувач здійснює оплату транзакції через мобільний додаток банку або купує товари в електронних магазинах. У більшості випадків системи банків або інтернет-магазинів реалізують додаткові методи захисту, які вимагають іншої аутентифікації або багатоетапної перевірки, які шахраю важко обійти. Оскільки злочинці, окрім комп'ютерних навичок, також використовують психологічні чинники, такі як неуважність, довірливість та низька поінформованість користувача, такий підхід впливає на поширеність «Отримання шахрайських електронних листів або телефонних дзвінків»

Незалежно від типу кібершахрайства, чоловіки частіше, ніж жінки, ставали жертвами. Найбільш істотна різниця відчувається в ситуаціях з фішингом, вішингом і вірусними атаками. Це можна пояснити багатьма факторами: низька концентрація на особистому захисті, обізнаність у питаннях кібербезпеки, рівень довіри, освіта користувачів, вік, соціальний статус, рівень доходу, кількість часу, проведеного в Інтернеті, тощо. Однак ця статистика показує, що чоловіки більше вразливіші до кіберзлочинців, ніж жінки.

Рисунок 3.4 демонструє рівень занепокоєння респондентів щодо ситуацій кіберзлочинності. У всіх випадках жінки більше, ніж чоловіки, стурбовані можливим кібершахрайством. Їхні найбільші побоювання викликають шахрайство з онлайн-банкінгом, крадіжка особистих даних і зараження шкідливим програмним забезпеченням. Незалежно від ситуації жінки можуть швидше відреагувати на потенційний злочин і вжити більш особистих заходів для протидії шахрайству.

На малюнку 3.5 показано, як респонденти захищаються від кібершахрайства. Жінки вважають за краще не відкривати пошту від незнайомців, не вводити особисту інформацію на сайтах, відвідувати лише надійні сайти, меншою мірою користуватися онлайн-сервісами та банкінгом. Чоловіки віддають перевагу антивірусним програмам, використовують різні і складні паролі, регулярно змінюють їх і налаштування безпеки, біометричні дані, менеджер паролів. Оскільки жінки рідше відвідують незнайомі сайти та не

відкривають незнайомі листи, це більшою мірою допомагає запобігти фішингу та отриманню особистих даних через фейкові сайти. Крім того, характер суб'єктивних оцінок, вжитих жінками, вказує на те, що вони не довіряють незнайомій або підозрілій інформації більше, ніж чоловіки, і віддають перевагу більш традиційним заняттям, ніж онлайн. Що стосується чоловіків, то, оскільки вони проводять багато часу в Інтернеті, вони більше, ніж жінки, намагаються використовувати більш складні та просунуті засоби особистого захисту, такі як біометрія, складні паролі та антивірусні програми.

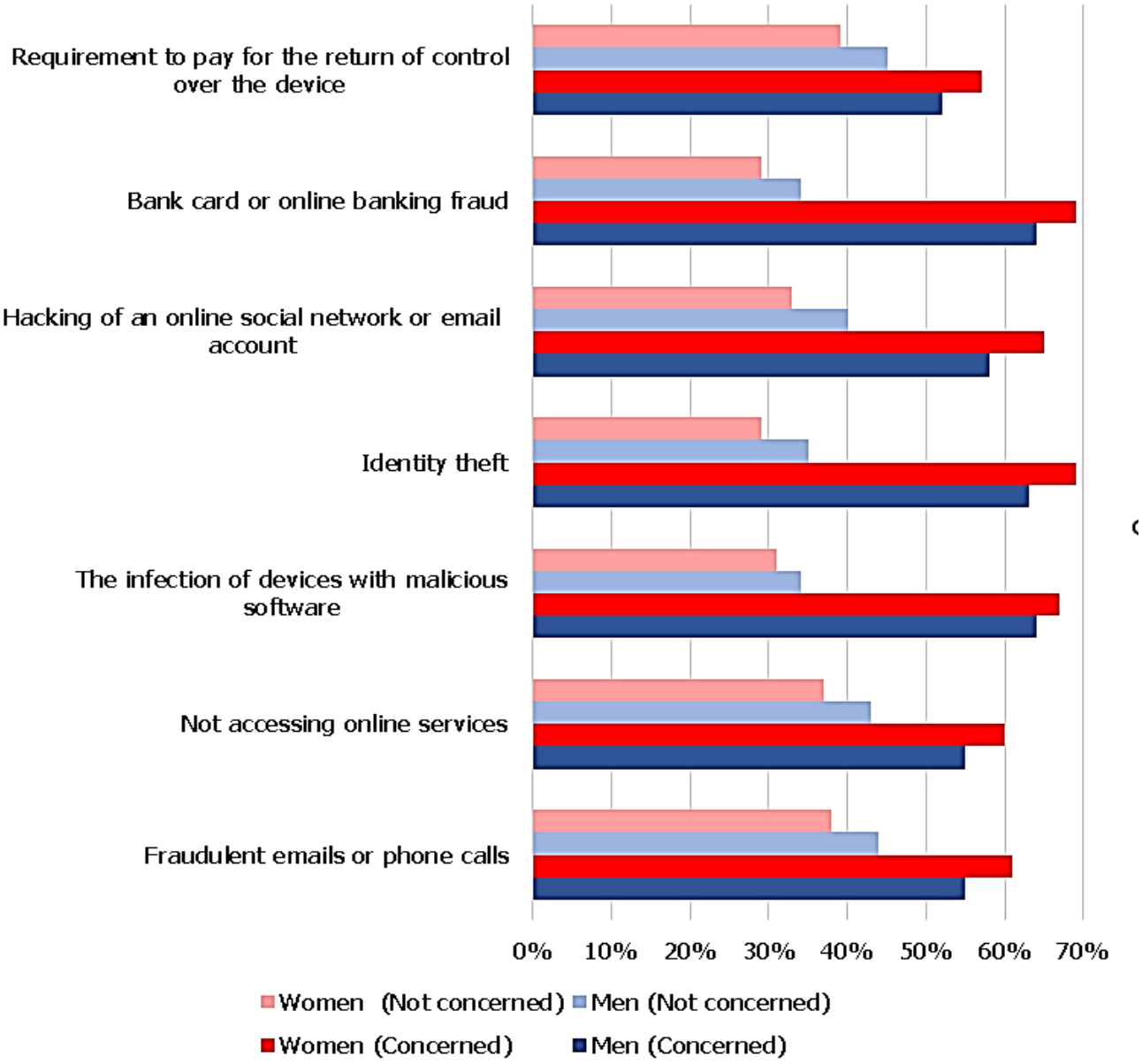


Рисунок 3.4 – Гендерний розподіл респондентів щодо впевненості в ситуаціях кібершахрайства (створено автором за результатами опитування [96])

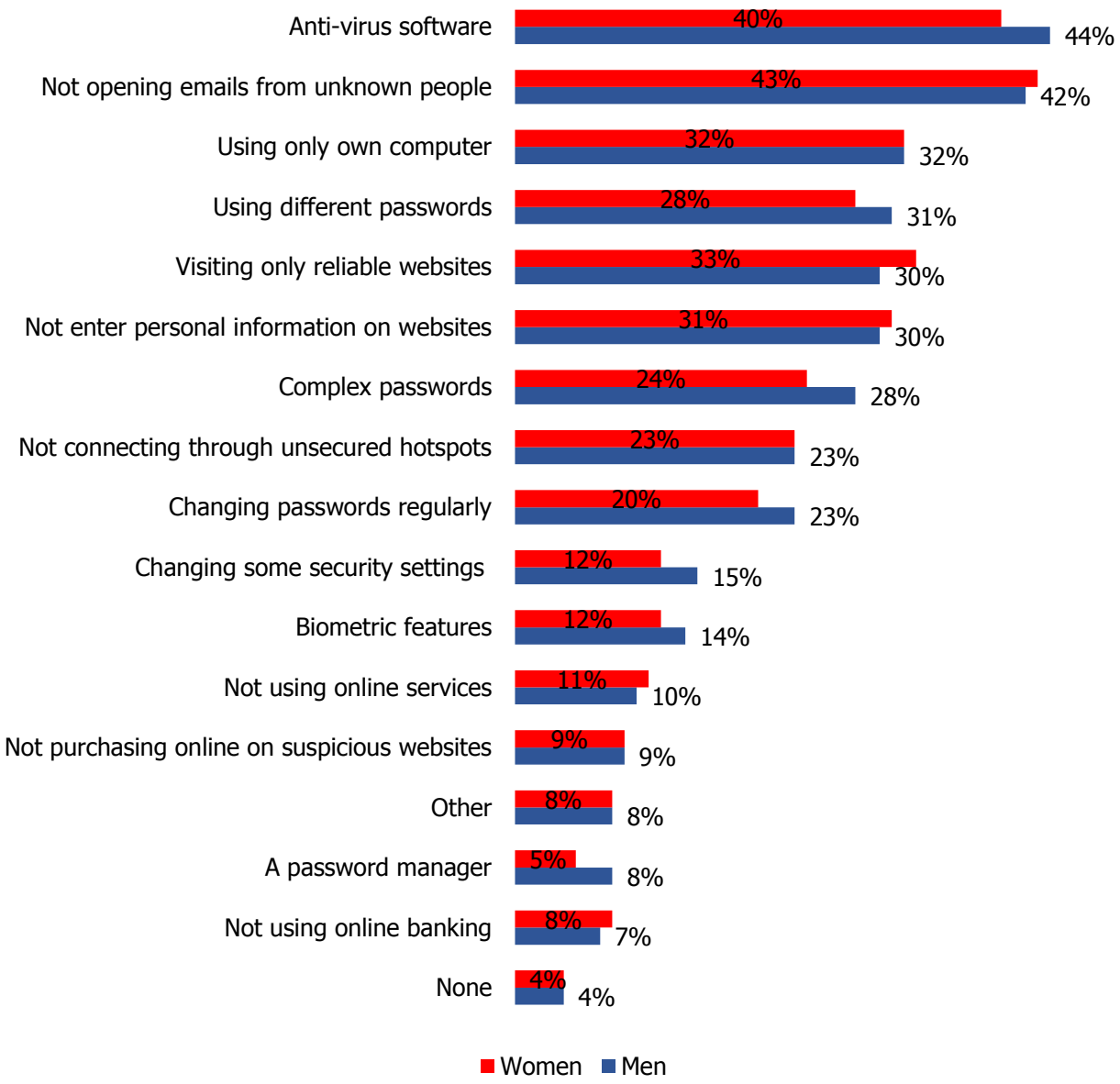


Рисунок 3.5 – Гендерний розподіл респондентів щодо способів захисту від кібершахрайства (створено автором за результатами опитування [96])

За результатами гендерного аналізу буде сформовано портрет ймовірної жертви кіберзлочинців. Це може бути як чоловік, так і жінка, але жертвами в більшості ситуацій є чоловіки, ніж жінки. Жертви чоловічої статі є користувачами комп'ютерів і займаються різноманітною онлайн-діяльністю, переважно електронною поштою, що збільшує ймовірність зіткнутися з фішингом. Жінки-жертви — користувачі смартфонів, які проводять більше часу в соціальних мережах, але вони також можуть стати об'єктами фішингу та

вішингу. Чоловіки-жертви вважають себе більш впевненими та усвідомлюють ризики кібершахрайства, ніж жінки-жертви, що може призвести до меншої уваги до підозрілої інформації. Менше жінок вважають себе добре поінформованими, що може вплинути на їхній підвищений інтерес до питань особистого захисту. Чоловіки-жертви менш стурбовані випадками шахрайства, які можуть змусити їх нехтувати різними заходами безпеки. Це впливає на їхній інтерес до впровадження методів захисту програмного забезпечення та ігнорування традиційних методів. Тобто ймовірною жертвою кібершахрая стане особа чоловічої статі, яка проводить значну кількість часу в Інтернеті, зі зниженою концентрацією уваги, підвищеною довірою до сторонньої інформації та впевненістю в технічних і програмних заходах безпеки.

Вирішення питань протидії кібершахрайству є актуальним як для наукової сфери, так і для практичної діяльності через швидку автоматизацію та цифровізацію багатьох процесів у державі, бізнесі, соціальному та особистому житті населення. Найефективнішим засобом боротьби з кібершахрайством є сучасні математичні алгоритми, методи та інформаційні технології. Кіберзлочинці випереджають системи захисту та постійно вдосконалюють свої засоби боротьби з кіберзлочинністю. Однак вони також враховують різні аспекти потенційних жертв, щоб збільшити ефект своїх злочинів. Тому, коли компанії будують стратегії безпеки, вкрай важливо оцінювати різні характеристики жертв кібершахрайства, наприклад, стать.

У цьому дослідженні було проведено гендерний аналіз потенційних жертв кіберзлочинів, для якого використано результати опитування Європейської комісії. Дослідження показало, що чоловіки використовують більше різноманітних електронних пристроїв у повсякденному житті, ніж жінки. У той же час вони зацікавлені в електронному листуванні, читанні блогів і форумів, онлайн-банкінгу, покупках, іграх, розвагах і освіті. Жінки віддають перевагу смартфонам і спілкуванню в соціальних мережах. Тобто чоловіки проводять більше часу в Інтернеті, ніж жінки, тому вони частіше стають жертвами кібершахраїв, ніж жінки. Жінки вважають, що їм не вистачає знань про

запобігання злочинам, і побоюються можливого шахрайства. Чоловіки вважають себе більш обізнаними про кіберризики, які можуть вплинути на їх довіру до інформації третіх сторін. Аналіз різних ситуацій кіберзлочинності показав, що чоловіки частіше стають жертвами, ніж жінки. Встановлено, що жінки дотримуються традиційних методів особистого захисту та не довіряють стороннім ресурсам, листам, сервісам і веб-сайтам. Сучасними методами особистої безпеки в основному користуються чоловіки. Тому, можливо, більше довіряють сторонній інформації в надії на системи захисту програмного забезпечення.

Підводячи підсумок, можна сказати, що потенційна жертва кібершахрайства – це людина, яка проводить значну кількість часу в Інтернеті, впевнена в заходах особистої безпеки та обізнаності про кіберризики, довіряє стороннім сервісам, сайтам і листам. Тобто може стати об’єктом вішингових, фішингових та вірусних кібератак. Результати гендерного аналізу можуть бути корисними при розробці систем моніторингу та кібербезпеки в компаніях, які практикують онлайн-платіжні транзакції, наприклад, у банках, компаніях електронної комерції, індустрії розваг тощо. Ця інформація також може бути використана для створення портретів потенційних жертв кіберзлочинів для різних ділових та державних секторів.

3.2 Розробка кіберпрофілів сучасних фінансових кіберзлочинів

У сучасному світі масштаби кіберзлочинності стрімко зростають. Це наслідок масової автоматизації різних сфер суспільства. Тому актуальною є проблема своєчасного виявлення та попередження кіберзлочинів, що дозволить суб’єктам господарювання та органам державної влади оперативно реагувати на масові загрози. На практиці використовується багато методів і засобів боротьби з кіберзлочинністю. Сьогодні є популярними програмні та кіберфізичні комплекси на основі сучасних математичних методів. Найбільш ефективними виявилися штучний інтелект, нейронні мережі, генетичні алгоритми, блокчейни,

роботи тощо. Але, на нашу думку, поряд із такими потужними інструментами варто використовувати й профілювання кіберзлочинців.

Профілювання — це процес розробки профілів злочинців, які визначають потенційні загрози на основі заздалегідь визначених характеристик. Як правило, вони формуються за вже відомими випадками кіберзлочинів. Профілі використовують ретроспективний підхід до відображення подій, що дозволяє постійно оновлювати їх у разі появи нових даних про нові кіберзлочини.

Профілюванням повинні займатися спеціальні підрозділи кібербезпеки суб'єктів господарювання або державних установ. Оскільки саме вони збирають статистичну інформацію щодо кіберзлочинів, загроз, порушень, шахрайств, тощо. Але на практиці цьому виду інструменту кіберзахисту приділяють не досить багато уваги. Це пов'язано з трудомісткістю збору та обробки інформації, що потребує постійної реалізації даних процесів. Оскільки кіберзлочини удосконалюються з часом завдяки застосуванню новітніх технологій, які впроваджують злочинці, то й процес профілювання повинен базуватися також на врахуванні великої кількості ознак кібершахраїв та злочинів. База даних ознак повинна поповнюватися новою інформацією, як результат здійснення різних видів кіберзагроз. При розробці профілів також треба враховувати, що вони корелюють як для злочину, так й для людини, яка його вчинила.

Основними характеристиками, які можна використовувати при побудові профілів, є поведінкові характеристики злочинців в Інтернеті. Ця необхідність викликана тим, що більшість фінансових та кіберзлочинів все ж таки реалізуються шляхом застосування онлайн-технологій. Це може відбуватися через соціальні мережі, Інтернет-магазини, блоги, онлайн-повідомлення, тощо. Відповідно, варто враховувати цю ознаку в процесі профілювання. Але тут є важливим в більшій мірі – частота відвідування конкретних інтернет-ресурсів, особливості використання тих чи інших інструментів і додатків, їх нестандартна та непередбачувана поведінка, проведення тіньових розрахунків та незаконних операцій з готівкою.

Можливе також використання такої характеристики, як відношення злочинця до державного законодавства країни. Усі злочинці нехтують дотриманням законів. Це пов’язано із тим, що більшість країн не мають чітко визначених заходів відповідальності за кіберзлочини. В основному за це передбачена адміністративна відповідальність, хоча в ряді країн, таких як США, Великобританія, за кіберзлочини впроваджена й кримінальна відповідальність. При визначенні профілю кіберзлочинця можна додавати й цю характеристику. Наприклад, на основі створеної бази даних кримінальних злочинців можна визначити тих, які мають спеціальну комп’ютерну освіту, або здійснювали кіберзлочини. Для таких зловмисників створюється «чорний лист», до якого можна постійно звертатися в процесі ідентифікації кіберзлочинців.

Географічне профілювання також є важливим, що дозволяє ідентифікувати злочинців за їхнім географічним розташуванням. Це можна зробити, відстежуючи джерела кібератак та IP-адреси, з яких надсилалися або надсилаються віруси, спам, кібератаки тощо. Існують спеціальні ресурси, наприклад, як Лабораторія Касперського, які в режимі онлайн демонструють різні види кібератак, які направлені на конкретні країни та які відбуваються з інших країн світу. Використовуючи схожий принцип та можливості подібних ресурсів, можна визначити потенційні країни-атакери, тобто ті, з яких найбільше всього відбувається кібератак у світі. А також можна вивити країни-жертви, тобто ті, які піддаються кібератакам найбільше. Використання цієї інформації допоможе ідентифікувати клієнтів або транзакції, які відбуваються з країн-атакерів.

При розробці кіберпрофілю важливо враховувати такий аспект, як мотивація зловмисників. Тут можна виділити два її види. Перший пов'язаний з усвідомленим наміром кіберзлочинця вчинити злочин. Сюди можна віднести прагнення швидко отримати матеріальну вигоду, позбутися почуття залежності чи обмеженості, цікавість і допитливість, почуття задоволення від здобуття влади, потреба у визнанні чи самоствердженні, а також бажання висловити чи стверджувати політичні чи соціальні позиції, інтереси. Як правило, такі злочинці

є хитрими та обізнаними в усіх аспектах кіберзлочинів і вони підуть до кінця, щоб отримати бажане. Це вид зловмисників дуже небезпечний для індивідів, бізнесу та держави. Їх виявлення потребуватиме більший обсяг зусиль та засобів, оскільки їх дії, як правило, є більш продуманими та організованими.

Другим мотиваційним напрямком виступають фінансові проблеми кіберзлочинця, пов'язані із необхідністю виплати кредитів, втратою роботи, низьким рівнем доходу, наявністю великої кількості дітей, хворих членів сім'ї, вирішенням проблем, пов'язаних із виплатою боргів в результаті азартних ігор, тощо. Як правило, такі зловмисники стають ними, виходячи з певних обставин в їхньому житті. Вони не можуть планувати такі злочини ідеально. Можливо, вони також не мають достатніх технічних знань для здійснення кіберзлочинів. Інструменти, якими вони користуються, є примітивними. Такі злочинці, як правило, вдаються до найбільш простих видів кіберзлочинів, наприклад, соціальна інженерія або розсилка програм-вимагачів. Знаходження таких індивідів є менш складним процесом, ніж у першому випадку, оскільки їх дії є типовими та вони слідують заздалегідь придбаної інструкції або відомій їм техніці злочину, наприклад, підглянутої в Інтернеті.

Звичайно, що процес профілювання кіберзлочинців для різних видів кіберзлочинів буде відрізнятися в певних деталях, але підходи його формування будуть однаковими. Він використовує три види заходів, орієнтовані на особи, зловмисне програмне забезпечення або злочини.

Підхід, орієнтований на людину, передбачає здійснення аналізу дій кіберзлочинців та їх особистих характеристик. З цією метою використовуються різні матеріали, опубліковані в джерелах масової інформації та в Інтернеті, а також інформація, узята з професійних баз даних кіберзлочинців. В більшій мірі ці дані будуть відображати види та інструменти злочинів, характеристики жертв, особливості здійснення протиправних дій, тощо. Дана інформація дозволить сформувати реальне уявлення щодо особистості, яка скоїла злочин. Звичайно, що цей портрет може бути індивідуальним в кожному певному випадку, але більшість кіберзлочинців все одно матимуть багато спільного.

Підхід, який орієнтується на зловмисне програмне забезпечення, передбачає використання знань щодо схожих програм, які застосовують або застосовували кіберзлочинці. Збирається картотека відповідних програм на основі даних попереднього досвіду щодо кіберзагроз. Також сюди можуть включати і те забезпечення, яке потрібне для виявлення і протидії кіберзлочинів. Їх ядро направлене на аналіз патернів, які ідентифікують ситуації, що можуть бути ймовірною загрозою. Саме ці алгоритми можуть застосовуватися для виявлення інших ситуацій за умов коректування відповідних шаблонів перевірки.

Третій захід направлений конкретно на певний випадок кіберзлочину. Він застосовується з урахуванням тих самих методів, які реалізуються в попередніх двох підходах. Тобто є потреба у створенні відповідної бази даних кримінальних випадків із залученням комп'ютерних технологій, де відображаються їх ключові характеристики. Поєднання трьох підходів дає більший ефект, оскільки одночасно спрямовується на суб'єкта, хто здійснює кіберзлочин, інструмент, за допомогою якого він реалізується, та об'єкт, на який спрямований злочин.

Для формування кіберпрофілів суттєве значення має використання криміналістичних методів ідентифікації традиційних злочинців, серед яких виділяють кримінально-розшукові, клінічні та статистичні методи. Підхід кримінального розслідування базується на проведенні різних видів експертизи для виявлення подібних випадків у минулому. Але такий підхід не буде ефективним на сто відсотків для формування профілю кіберзлочинців, оскільки цей вид злочинності є специфічним через швидкий розвиток технологій і методів, які використовують злочинці, що ускладнює їх ідентифікацію. Також він потребує постійного оновлення характеристик кіберзлочинців, кіберзлочинів та їх інструментів, але за часту це зробити складно за рахунок постійної модифікації технік та інструментів, які використовують злочинці. І завдяки цьому потрапити на місце злочину складно за рахунок віддаленості зловмисника або його маскування. Але в певних випадках проведення експертизи дозволятиме сформувати реальну картину подій.

Клінічний підхід забезпечує створення повної історії злочину, що дозволяє оцінити його основні характеристики. Його можна частково застосувати для виявлення кіберзлочинів, оскільки також є необхідність постійно оновлювати історію кіберзлочинів. Також може виникнути проблема у формуванні клінічної картини за рахунок відсутності відповідних фахівців, які б могли оцінити ознаки кіберзагрози. Деякі випадки навіть потребують сторонніх спеціалістів, які володіють знаннями та навиками використання відповідної мови програмування.

Найбільш ефективним є статистичний підхід, оскільки він передбачає використання спеціальних програмних засобів і статистичних методів, що дозволяє не тільки збирати відповідні характеристики, а й проводити відповідні розрахунки, які сприятимуть ідентифікації більш значної кількості злочинів. Коли інформації недостатньо, використовується непараметрична статистика. Коли доступна більш глибока інформація, регресії та байєсовські мережі можуть бути корисними в контексті профілювання. Методи класифікації та кластеризації використовуються для формування профілів із типовими ознаками та стереотипним уявленням про злочинців, підозрюваних, свідків і жертв.

Статистичні методи кіберпрофілювання дають більший ефект, ніж інші, оскільки дозволяють вираховувати різні статистичні характеристики, моделювати ситуації потенційних кіберзагроз, прогнозувати ймовірність виникнення кіберзлочину, тощо. Звичайно, що використання тільки подібних методів не дозволить виявляти ситуації загроз на сто відсотків, але їх використання у сукупності з іншими підходами та методами профілювання сприятиме комплексній оцінці ситуацій в цілому або формуванню дій попередження, які дозволять підвищити увагу до конкретних транзакцій, або запитів, або користувачів системи.

В даній роботі було проведено формування кіберпрофілю потенційних злочинців, які здійснюють незаконні дії щодо кредитних операцій. У банківському секторі це досить серйозна проблема на сьогодні, оскільки пов’язана із спрощенням умов кредитування та отримання коштів онлайн різним категоріям населення за допомогою різних мобільних додатків. Існують також

способи незаконно отримати кредити, використовуючи чужі дані, або цілеспрямовано отримати їх і не повернути, тому що у злочинця на меті здійснення кібершахрайства. Для реалізації даного завдання доцільно використати кластерний аналіз, який дозволить чітко сформувати профілі тих клієнтів, які будуть мати ознаки потенційної злочинної діяльності для банку.

Було застосовано такий різновид кластерного аналізу, як «Очікування-максимізація». Суть цього алгоритму полягає у виявленні кластерів (груп) клієнтів банку, які потенційно можуть бути пов’язані з кібершахрайством. Для його реалізації використовувалася клієнтська база даних одного з банків, яка містить понад 300 тис. спостережень. Кожен запис має 122 атрибути, які включають тип нерухомості клієнта, наявність автомобіля у клієнта, стать клієнта, кількість дітей, середній дохід і тип доходу клієнта, освіту, суму кредиту, суму щомісячний платіж тощо. В якості цільового атрибута використовується характеристика труднощів клієнта при виплаті кредиту. Якщо його значення дорівнює «1», то, відповідно, у клієнта виникають ускладнення, які можуть сигналізувати про можливе кібершахрайство. Якщо значення «0» відповідає клієнту, то його профіль не викликає жодних підозр з боку кібербезпеки банку. Фрагмент вхідних даних представлено на рисунку 3.6.

Набір даних містить багато спостережень, які мають пропущену інформацію або мають викиди та екстремальні значення. В результаті, було відібрано 51 зміну з 122 та проведено їх якість. Для реалізації цього підходу використано програмно-аналітичний пакет «Deductor Studio Academic». Результат оцінки на рисунку 3.7. З відібраних даних система виявила 4 атрибути є непридатними. Інші містять викиди та екстремальні значення. До них було застосовано процедури заповнення пропусків та усунення викидів, екстремальних значень. Результат даної процедури представлений на рисунку 3.8.

Column1	Row	Description
1	SK_ID_CURR	ID of loan in our sample
2	TARGET	Target variable (1 - client with payment difficulties: he/she had late payment more than X days on at least one of the first Y installments of the loan in our sample, 0 - all other cases)
5	NAME_CONTRACT_TYPE	Identification if loan is cash or revolving
6	CODE_GENDER	Gender of the client
7	FLAG_OWN_CAR	Flag if the client owns a car
8	FLAG_OWN_REALTY	Flag if client owns a house or flat
9	CNT_CHILDREN	Number of children the client has
10	AMT_INCOME_TOTAL	Income of the client
11	AMT_CREDIT	Credit amount of the loan
12	AMT_ANNUITY	Loan annuity
13	AMT_GOODS_PRICE	For consumer loans it is the price of the goods for which the loan is given
14	NAME_TYPE_SUITE	Who was accompanying client when he was applying for the loan
15	NAME_INCOME_TYPE	Clients income type (businessman, working, maternity leave,...)
16	NAME_EDUCATION_TYPE	Level of highest education the client achieved
17	NAME_FAMILY_STATUS	Family status of the client
18	NAME_HOUSING_TYPE	What is the housing situation of the client (renting, living with parents, ...)
19	REGION_POPULATION_RELATIVE	Normalized population of region where client lives (higher number means the client lives in more populated region)
20	DAYS_BIRTH	Client's age in days at the time of application
21	DAYS_EMPLOYED	How many days before the application the person started current employment
22	DAYS_REGISTRATION	How many days before the application did client change his registration
23	DAYS_ID_PUBLISH	How many days before the application did client change the identity document with which he applied for the loan
24	OWN_CAR_AGE	Age of client's car
25	FLAG_MOBIL	Did client provide mobile phone (1=YES, 0=NO)
26	FLAG_EMP_PHONE	Did client provide work phone (1=YES, 0=NO)
27	FLAG_WORK_PHONE	Did client provide home phone (1=YES, 0=NO)
28	FLAG_CONT_MOBILE	Was mobile phone reachable (1=YES, 0=NO)
29	FLAG_PHONE	Did client provide home phone (1=YES, 0=NO)
30	FLAG_EMAIL	Did client provide email (1=YES, 0=NO)
31	OCCUPATION_TYPE	What kind of occupation does the client have
32	CNT_FAM_MEMBERS	How many family members does client have
33	REGION_RATING_CLIENT	Our rating of the region where client lives (1,2,3)
34	REGION_RATING_CLIENT_W_CITY	Our rating of the region where client lives with taking city into account (1,2,3)
35	WEEKDAY_APPR_PROCESS_START	On which day of the week did the client apply for the loan

Рисунок 3.6 – Фрагмент початкових даних для формування кіберпрофілю

№	Столбец	Тип данных	Вид данных	Пропуски		Выбросы		Экстремальные		Кол-во уникальных	Качество данных	Резюме
				Кол-во	Действие	Кол-во	Действие	Кол-во	Действие			
6	FLAG_OWN_REALTY	ab Строковый	⋯ Дискретный							2	0,8999	Пригоден
7	CNT_CHILDREN	9.0 Вещественный	⋯ Дискретный			62	Заменять медианой	9	Заменять медианой	9	0,4115	Предобработка
8	AMT_INCOME_TOTAL	ab Строковый	⋯ Дискретный			441	Заменять наиболее ...	939	Заменять наиболее ...	416	0,5693	Предобработка
9	AMT_CREDIT	ab Строковый	⋯ Дискретный			1 906	Заменять наиболее ...	2 423	Заменять наиболее ...	2504	0,7854	Предобработка
10	AMT_ANNUITY	ab Строковый	⋯ Дискретный			2 673	Заменять наиболее ...			6119	0,8925	Предобработка
11	AMT_GOODS_PRICE	ab Строковый	⋯ Дискретный			1 832	Заменять наиболее ...	3 764	Заменять наиболее ...	426	0,6918	Предобработка
12	NAME_TYPE_SUITE	ab Строковый	⋯ Дискретный					1 479	Заменять наиболее ...	8	0,3171	Предобработка
13	NAME_INCOME_TYPE	ab Строковый	⋯ Дискретный			1 249	Заменять наиболее ...	10	Заменять наиболее ...	6	0,5800	Предобработка
14	NAME_EDUCATION_TYPE	ab Строковый	⋯ Дискретный					1 292	Заменять наиболее ...	5	0,4167	Предобработка
15	NAME_FAMILY_STATUS	ab Строковый	⋯ Дискретный			937	Заменять наиболее ...			5	0,7277	Предобработка
16	NAME_HOUSING_TYPE	ab Строковый	⋯ Дискретный			1 736	Заменять наиболее ...	1 817	Заменять наиболее ...	6	0,3284	Предобработка
17	REGION_POPULATION_RELATIVE	ab Строковый	⋯ Дискретный					3	Заменять наиболее ...	80	0,9255	Предобработка
18	DAYS_BIRTH	9.0 Вещественный	⋯ Непрерывный								0,9596	Пригоден
19	DAYS_EMPLOYED	9.0 Вещественный	⋯ Непрерывный								0,1327	Пригоден
20	DAYS_REGISTRATION	ab Строковый	⋯ Дискретный							9930	0,9792	Пригоден
21	DAYS_ID_PUBLISH	9.0 Вещественный	⋯ Непрерывный								0,9422	Пригоден
22	FLAG_MOBIL	0/1 Логический	⋯ Дискретный							1	0,0000	Непригоден
23	FLAG_EMP_PHONE	0/1 Логический	⋯ Дискретный							2	0,5307	Пригоден
24	FLAG_WORK_PHONE	0/1 Логический	⋯ Дискретный							2	0,7914	Пригоден
25	FLAG_CONT_MOBILE	0/1 Логический	⋯ Дискретный					45	Заменять наиболее ...	2	0,0191	Предобработка
26	FLAG_PHONE	0/1 Логический	⋯ Дискретный							2	0,8032	Пригоден
27	FLAG_EMAIL	0/1 Логический	⋯ Дискретный					1 374	Заменять наиболее ...	2	0,3087	Предобработка
28	OCCUPATION_TYPE	ab Строковый	⋯ Дискретный			327	Заменять наиболее ...	221	Заменять наиболее ...	19	0,7510	Предобработка
29	CNT_FAM_MEMBERS	Дата/Время	⋯ Дискретный	1	Заменять мед...	55	Ограничивать	14	Ограничивать	10	0,5593	Предобработка
30	REGION_RATING_CLIENT	9.0 Вещественный	⋯ Дискретный							3	0,6736	Пригоден
31	REGION_RATING_CLIENT_W_CITY	9.0 Вещественный	⋯ Дискретный							3	0,6660	Пригоден
32	WEEKDAY_APPR_PROCESS_START	ab Строковый	⋯ Дискретный							7	0,9718	Пригоден
33	HOUR_APPR_PROCESS_START	9.0 Вещественный	⋯ Непрерывный			33	Ограничивать				0,7937	Предобработка
34	REG_REGION_NOT_LIVE_REGION	0/1 Логический	⋯ Дискретный					433	Заменять наиболее ...	2	0,1268	Предобработка
35	REG_REGION_NOT_WORK_REGION	0/1 Логический	⋯ Дискретный					1 388	Заменять наиболее ...	2	0,3110	Предобработка
36	LIVE_REGION_NOT_WORK_REGION	0/1 Логический	⋯ Дискретный					1 056	Заменять наиболее ...	2	0,2538	Предобработка
37	REG_CITY_NOT_LIVE_CITY	0/1 Логический	⋯ Дискретный							2	0,5247	Пригоден
38	REG_CITY_NOT_WORK_CITY	0/1 Логический	⋯ Дискретный							2	0,8848	Пригоден
39	LIVE_CITY_NOT_WORK_CITY	0/1 Логический	⋯ Дискретный							2	0,7632	Пригоден
40	ORGANIZATION_TYPE	ab Строковый	⋯ Дискретный			978	Заменять наиболее ...	1 706	Заменять наиболее ...	58	0,6917	Предобработка
41	EXT_SOURCE_2	ab Строковый	⋯ Дискретный							21988	0,9900	Пригоден
42	FONDKAPREMONT_MODE	ab Строковый	⋯ Дискретный					1 548	Заменять наиболее ...	5	0,4928	Предобработка
43	HOUSETYPE_MODE	ab Строковый	⋯ Дискретный					255	Заменять наиболее ...	4	0,5332	Предобработка
44	TOTALAREA_MODE	ab Строковый	⋯ Дискретный					11 119	Заменять наиболее ...	2644	0,5123	Предобработка
45	WALLSMATERIAL_MODE	ab Строковый	⋯ Дискретный					1 562	Заменять наиболее ...	8	0,5746	Предобработка
46	EMERGENCYSTATE_MODE	ab Строковый	⋯ Дискретный					223	Заменять наиболее ...	3	0,6677	Предобработка
47	OBS_30_CNT_SOCIAL_CIRCLE	Дата/Время	⋯ Дискретный	13 097	Оставить без и...	263	Оставить без измен...			12	0,0000	Непригоден
48	DEF_30_CNT_SOCIAL_CIRCLE	Дата/Время	⋯ Дискретный	21 134	Оставить без и...	35	Оставить без измен...	9	Оставить без измен...	6	0,0000	Непригоден
49	OBS_60_CNT_SOCIAL_CIRCLE	Дата/Время	⋯ Дискретный	13 164	Оставить без и...	254	Оставить без измен...			12	0,0000	Непригоден
50	DEF_60_CNT_SOCIAL_CIRCLE	Дата/Время	⋯ Дискретный	22 030	Оставить без и...	95	Оставить без измен...	18	Оставить без измен...	5	0,0000	Непригоден

Рисунок 3.7 – Результат оцінки якості початкових даних

№	Столбец	Тип данных	Вид данных	Пропуски		Выбросы		Экстремальные		Кол-во уникальных	Качество данных	Резюме
				Кол-во	Действие	Кол-во	Действие	Кол-во	Действие			
1	SK_ID_CURR	9.0 Вещественный	— Непрерывный								0,9999	Пригоден
2	TARGET	9.0 Вещественный	⋯ Дискретный							1	0,0000	Непригоден
3	NAME_CONTRACT_...	ab Строковый	⋯ Дискретный							1	0,0000	Непригоден
4	CODE_GENDER	ab Строковый	⋯ Дискретный							2	0,9855	Пригоден
5	FLAG_OWN_CAR	ab Строковый	⋯ Дискретный							2	0,8875	Пригоден
6	FLAG_OWN_REALTY	ab Строковый	⋯ Дискретный							2	0,8999	Пригоден
7	CNT_CHILDREN	9.0 Вещественный	⋯ Дискретный							3	0,7405	Пригоден
8	AMT_INCOME_TOTAL	ab Строковый	⋯ Дискретный			873	Заменять наиболее ...	60	Заменять наиболее ...	46	0,7961	Предобработка
9	AMT_CREDIT	ab Строковый	⋯ Дискретный			1 445	Заменять наиболее ...	10 076	Заменять наиболее ...	571	0,7492	Предобработка
10	AMT_ANNUITY	ab Строковый	⋯ Дискретный			2 019	Заменять наиболее ...	16 129	Заменять наиболее ...	3446	0,8442	Предобработка
11	AMT_GOODS_PRICE	ab Строковый	⋯ Дискретный			643	Заменять наиболее ...	5 239	Заменять наиболее ...	49	0,7051	Предобработка
12	NAME_TYPE_SUITE	ab Строковый	⋯ Дискретный							2	0,5328	Пригоден
13	NAME_INCOME_TYPE	ab Строковый	⋯ Дискретный							3	0,7805	Пригоден
14	NAME_EDUCATION_...	ab Строковый	⋯ Дискретный							2	0,6379	Пригоден
15	NAME_FAMILY_STA...	ab Строковый	⋯ Дискретный							4	0,7415	Пригоден
16	NAME_HOUSING_TY...	ab Строковый	⋯ Дискретный							1	0,0000	Непригоден
17	REGION_POPULATI...	ab Строковый	⋯ Дискретный							79	0,9280	Пригоден
18	DAYS_BIRTH	9.0 Вещественный	— Непрерывный								0,9596	Пригоден
19	DAYS_EMPLOYED	9.0 Вещественный	— Непрерывный								0,1327	Пригоден
20	DAYS_REGISTRATION	ab Строковый	⋯ Дискретный							9930	0,9792	Пригоден
21	DAYS_ID_PUBLISH	9.0 Вещественный	— Непрерывный								0,9422	Пригоден
22	FLAG_MOBIL	0/1 Логический	⋯ Дискретный							1	0,0000	Непригоден
23	FLAG_EMP_PHONE	0/1 Логический	⋯ Дискретный							2	0,5307	Пригоден
24	FLAG_WORK_PHONE	0/1 Логический	⋯ Дискретный							2	0,7914	Пригоден
25	FLAG_CONT_MOBILE	0/1 Логический	⋯ Дискретный					45	Заменять наиболее ...	2	0,0191	Предобработка
26	FLAG_PHONE	0/1 Логический	⋯ Дискретный							2	0,8032	Пригоден
27	FLAG_EMAIL	0/1 Логический	⋯ Дискретный					1 374	Заменять наиболее ...	2	0,3087	Предобработка
28	OCCUPATION_TYPE	ab Строковый	⋯ Дискретный							13	0,8183	Пригоден
29	CNT_FAM_MEMBERS	Дата/Время	— Непрерывный								0,4611	Пригоден
30	REGION_RATING_C...	9.0 Вещественный	⋯ Дискретный							3	0,6736	Пригоден
31	REGION_RATING_C...	9.0 Вещественный	⋯ Дискретный							3	0,6660	Пригоден
32	WEEKDAY_APPR_P...	ab Строковый	⋯ Дискретный							7	0,9718	Пригоден
33	HOUR_APPR_PROC...	9.0 Вещественный	— Непрерывный								0,8521	Пригоден
34	REG_REGION_NOT_...	0/1 Логический	⋯ Дискретный					433	Заменять наиболее ...	2	0,1268	Предобработка
35	REG_REGION_NOT_...	0/1 Логический	⋯ Дискретный					1 388	Заменять наиболее ...	2	0,3110	Предобработка
36	LIVE_REGION_NOT_...	0/1 Логический	⋯ Дискретный					1 056	Заменять наиболее ...	2	0,2538	Предобработка
37	REG_CITY_NOT_LIV...	0/1 Логический	⋯ Дискретный							2	0,5247	Пригоден
38	REG_CITY_NOT_WO...	0/1 Логический	⋯ Дискретный							2	0,8848	Пригоден
39	LIVE_CITY_NOT_WO...	0/1 Логический	⋯ Дискретный							2	0,7632	Пригоден
✔ 40	ORGANIZATION_TYPE	ab Строковый	⋯ Дискретный			1 033	Заменять наиболее ...	726	Заменять наиболее ...	19	0,7594	Предобработка
41	EXT_SOURCE_2	ab Строковый	⋯ Дискретный							21988	0,9900	Пригоден
42	FONDKAPREMONT_...	ab Строковый	⋯ Дискретный							2	0,7367	Пригоден
43	HOUSETYPE_MODE	ab Строковый	⋯ Дискретный							2	0,9819	Пригоден
44	TOTALAREA_MODE	ab Строковый	⋯ Дискретный							1	0,0000	Непригоден
45	WALLSMATERIAL_M...	ab Строковый	⋯ Дискретный							3	0,8237	Пригоден
46	EMERGENCYSTATE...	ab Строковый	⋯ Дискретный							2	0,9920	Пригоден

Рисунок 3.8 – Кінцевий результат оцінки якості початкових даних

Хоча проведені процедури не дозволили повністю усунути всі недоліки, але якість масиву даних значно покращилася. Також було виявлено ще непридатні для аналізу дані, які будуть не враховані в подальшому моделюванні.

Далі було проведено кластерний аналіз «Очікування-максимізація», який використовувався як алгоритм для формування профілю кібершахраїв із кредитними операціями. Це ітераційний метод, який формує групи за максимальною правдоподібністю або максимальними апостеріорними оцінками параметрів. Перевагою цього алгоритму є пошук прихованих змінних, які можуть сильно впливати на формування цільової змінної, що може бути корисним, оскільки аналітик або фахівець з кібербезпеки можуть не виявити всіх можливих ознак. У результаті було сформовано 10 кластерів користувачів, які можна проаналізувати на предмет можливого кібершахрайства з кредитними операціями. Результати зв’язків між кластерами представлені на рисунку 3.9. На рисунку 3.9 можна побачити міцність кластерів, максимальну похибку розпізнання та середню похибку розпізнання.

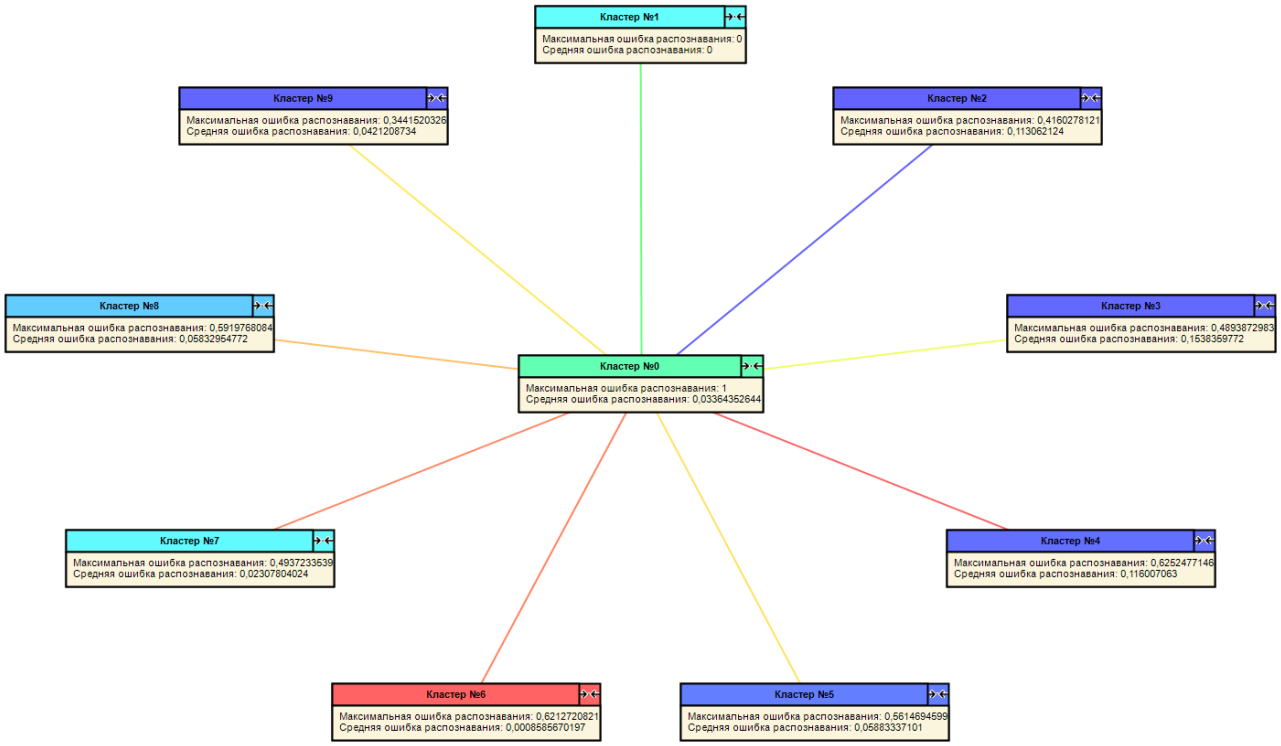


Рисунок 3.9 – Результати зв’язків між кластерами

Найбільш потужними є кластери № 4, 3, 1, 7, 9. Можна було б зменшити їх кількість до 5, але залишимо дану конфігурацію, оскільки це дозволить нам сформувати більш детальні профілі злочинців, виходячи із наповнюваності кластерів. За умови отримання нової інформації, такий обсяг профілів сприятиме їх уточненню та більш детальній класифікації кіберзлочинців.

Фрагмент отриманих профілів, виходячи з ознаки кластерів, представлено на рисунку 3.10. Всі інші частини профілів наведено у додатку Д.

Найбільш наповненими є перші п’ять кластерів. Тобто при ідентифікації потенційних кіберзлочинців, ймовірність того, що вони належатимуть даним кластерам є більшою, ніж у інших випадках. Вони формуватимуть основний набір характеристик, які відповідатимуть потенційним загрозам. Рисунок 3.10 показує, що значення типу будинку є невизначеним у більшості випадків (атрибут “HOUSETYPE_MODE”). Візуалізація наповненості кластерів клієнтів за даною характеристикою представлена на рисунку 3.11. Це пов’язано із відсутністю вхідних даних і зазначенням їх як невизначений тип.

Рисунок 3.10 – Фрагмент профілів кіберзлочинців

Але клієнти, які відносяться до 3, 1 та 7 кластеру і які були ідентифіковані, як шахраї, в переважній більшості мали «Block of Flats». Це можна пояснити тим, що більшість клієнтів банку є власниками цього типу нерухомості. Саме такі клієнти звертаються до банків за кредитами на купівлю будинку. Відповідно, вони можуть належати до групи кібершахраїв.

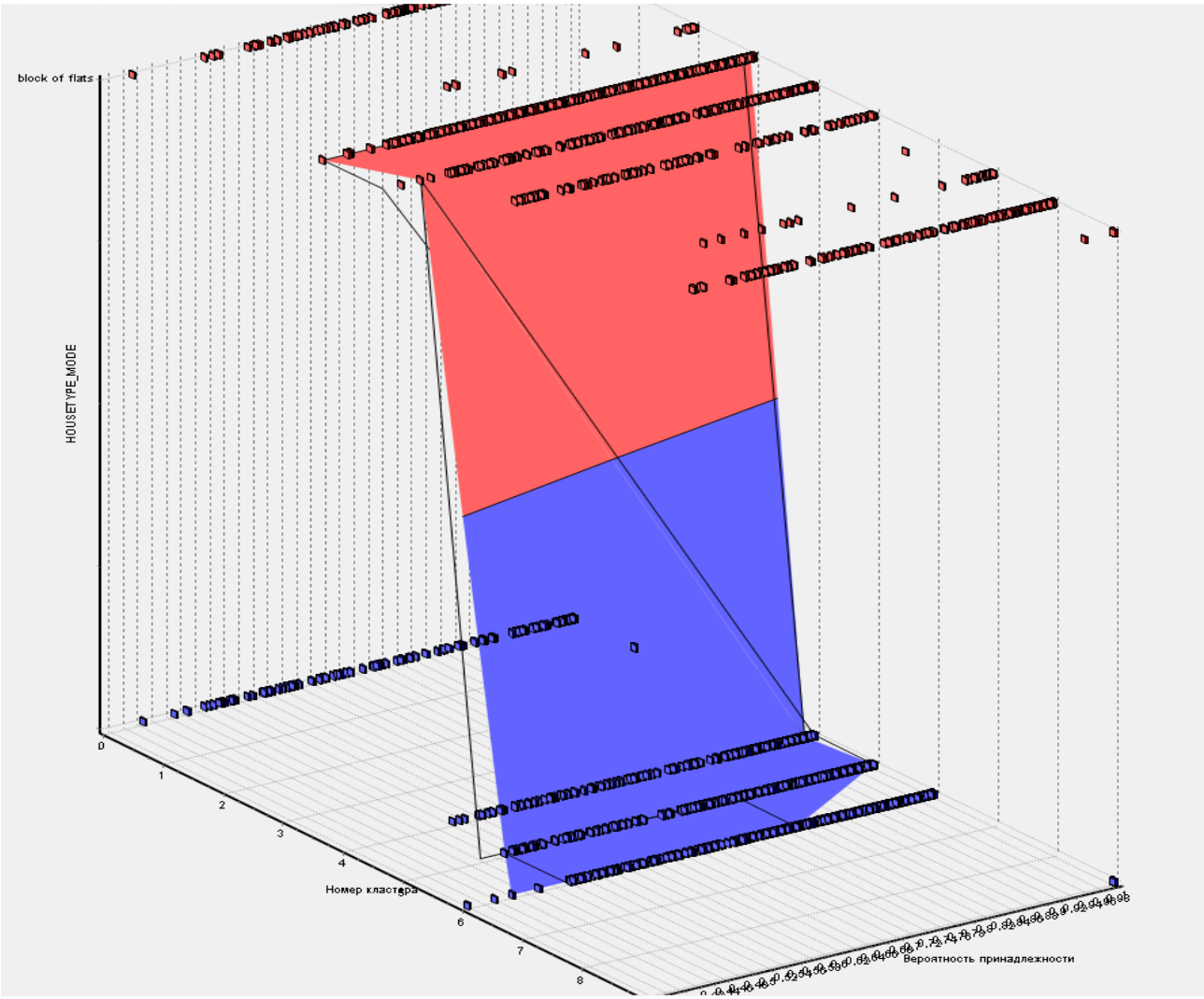


Рисунок 3.11 – Візуалізація характеристики “HOUSETYPE_MODE”

Якщо аналізувати змінну “FONDKAPREMONT_MODE”, то також видно, що умови якості житла, яким володіє індивід, є значними для 3, 1 та 7 кластерів. Візуалізація наповненості кластерів клієнтів за даною характеристикою представлена на рисунку 3.12. Для інших вони є невизначеними, томи при аналізі характеристик можуть бути не враховані. Але потенційні зловмисники, які відповідають кластерам 3, 1 та 7, можуть мати мотивацією до здійснення злочину

квартирне питання – або взяття коштів на поліпшення житлових умов, або зміни типу житла. Звісно, що якщо людина намагається узяти кредит за цими цілями, то це не означає, що їй банк повинен відмовити. Отримання інформації, що клієнт може бути потенційним шахраєм, оскільки попадає в даний кластер за цими ознаками, означатиме проведення додаткових перевірок та аналізу за іншими критеріями.

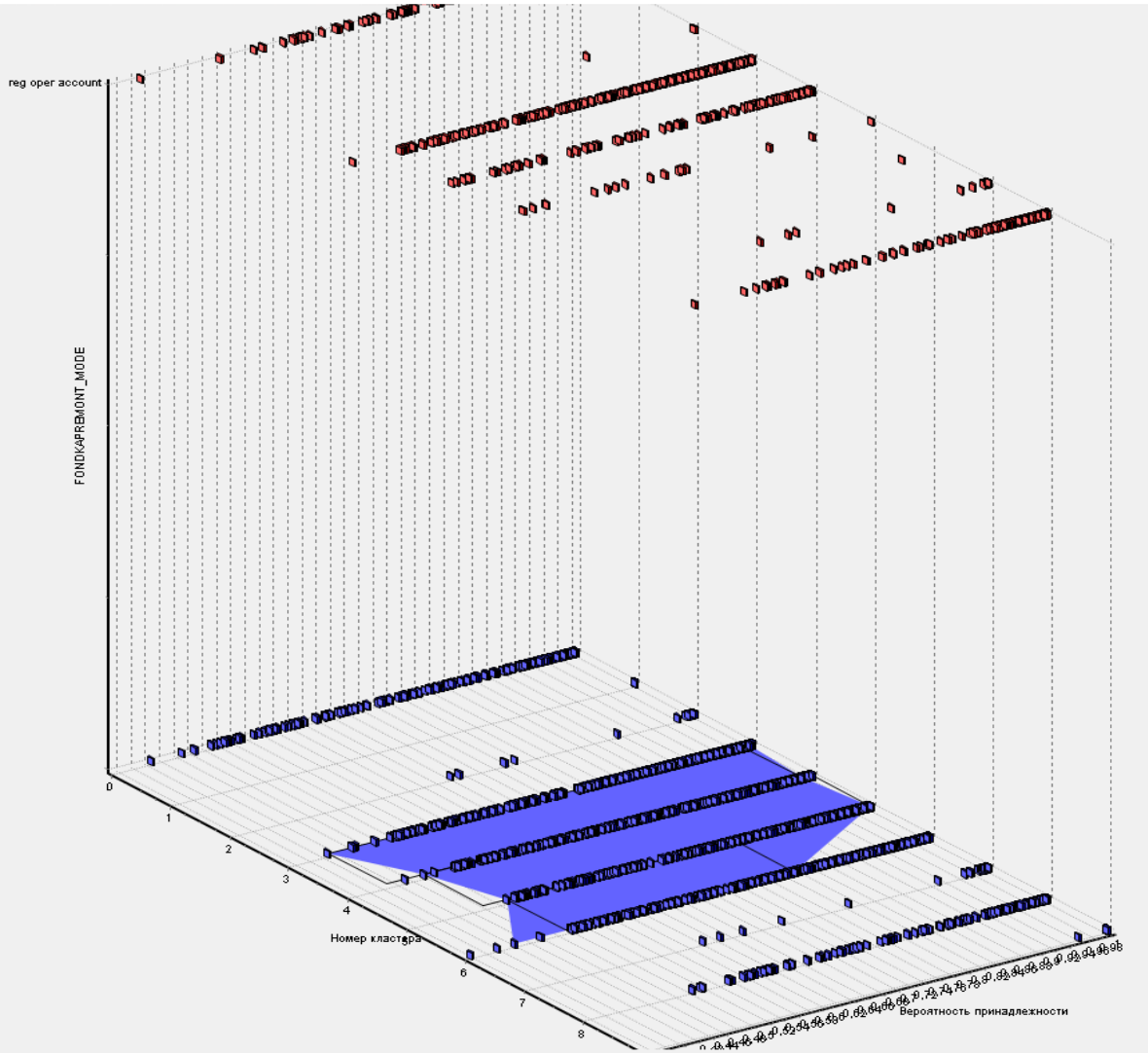


Рисунок 3.12 – Візуалізація характеристики “FONDKAPREMONT_MODE”

Що стосується наступної ознаки, такої як “ORGANIZATION_TYPE”, то вона свідчитиме про вид діяльності клієнта. Візуалізація наповненості кластерів клієнтів за даною характеристикою представлена на рисунку 3.13. Можна побачити, що 1-й кластер сформували переважно ті, чий вид діяльності ідентифікується як “XNA”. Оскільки використана для дослідження база даних не

містить розшифровок, то важко сказати, яка діяльність має дану абревіатуру. До 4, 3 та 7 кластерів увійшли ті, хто класифікується як “Business Entity Type 3”. Також 20,2% 4-го кластеру сформовані клієнтами, які є самозайнятими. Ймовірно, що ці зловмисники мають проблеми з їх видом діяльності, які потребують додаткових інвестицій, або фінансових засобів на розширення, або погашення боргів, тому вони й вдаються до злочинних схем із незаконним отриманням або неповерненням кредитних коштів.

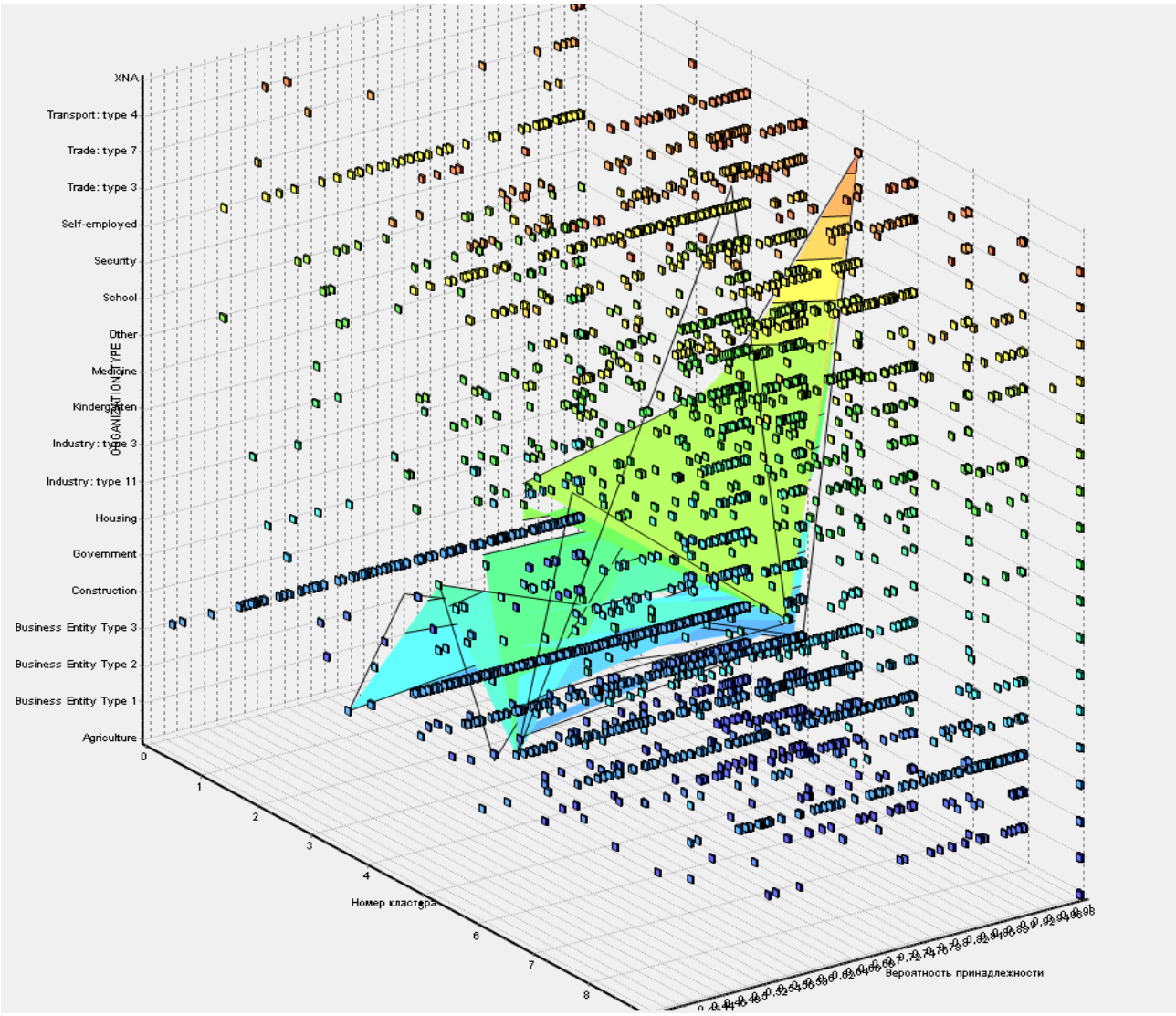


Рисунок 3.13 – Візуалізація характеристики “ORGANIZATION_TYPE”

При формуванні профіля важливою ознакою є сімейний статус “NAME_FAMILY_STATUS”. Візуалізація наповненості кластерів клієнтів за даною характеристикою представлена на рисунку 3.14. Значення кластерів показують, що переважна більшість клієнтів банку є одруженими / заміжніми.

На нашу думку, до аналізованого виду кібершахрайств можуть вдаватися саме сімейні люди, які звертаються до злочину завдяки неможливості утримувати сім'ю або наявності хворого члена сім'ї, або задля реалізації бажань, що потребують значних фінансових ресурсів.

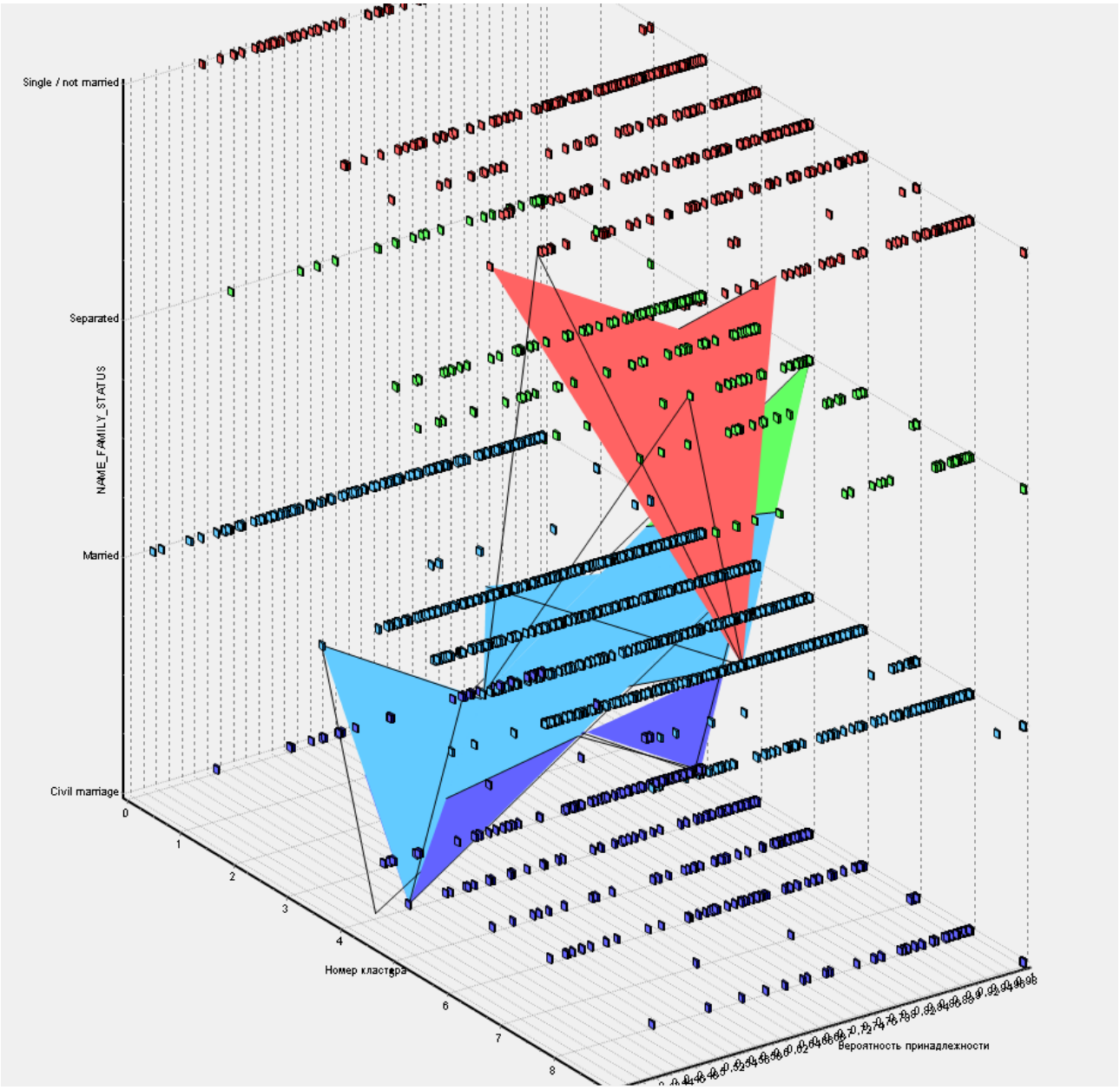


Рисунок 3.14 – Візуалізація характеристики "NAME_FAMILY_STATUS"

Цікавим виявилася така характеристика, як тип освіти. Візуалізація наповненості кластерів клієнтів за даною характеристикою представлена на рисунку 3.15. Виявилося, що шахрайством можуть займатися ті, хто завершили середню школу або мають вищу освіту. Переважно зловмисники відносяться до

першої категорії (завершили середню школу). Ці особистості можливо не мають певних досягнень у житті, тому для швидкої самореалізації вони можуть вдатися до найпростішого виду кібершахрайств, такого як шахрайства з кредитними операціями. Відповідно, такі злочинці слідують типовим схемам і можуть вираховуватися швидше. Але приблизно 10-20% клієнтів мають вищу освіту, при цьому кожен кластер містить таких потенційних злочинців. Ця категорія є достатньо вмотивованою, оскільки, можливо, намагаються здійснити злочин заради самоствердження або підтримки гострих почуттів. Виявляти таких кандидатів досить складно, оскільки вони мають не тільки освіту, але й роботу, що гарантує формуванню гарної кредитної історії.

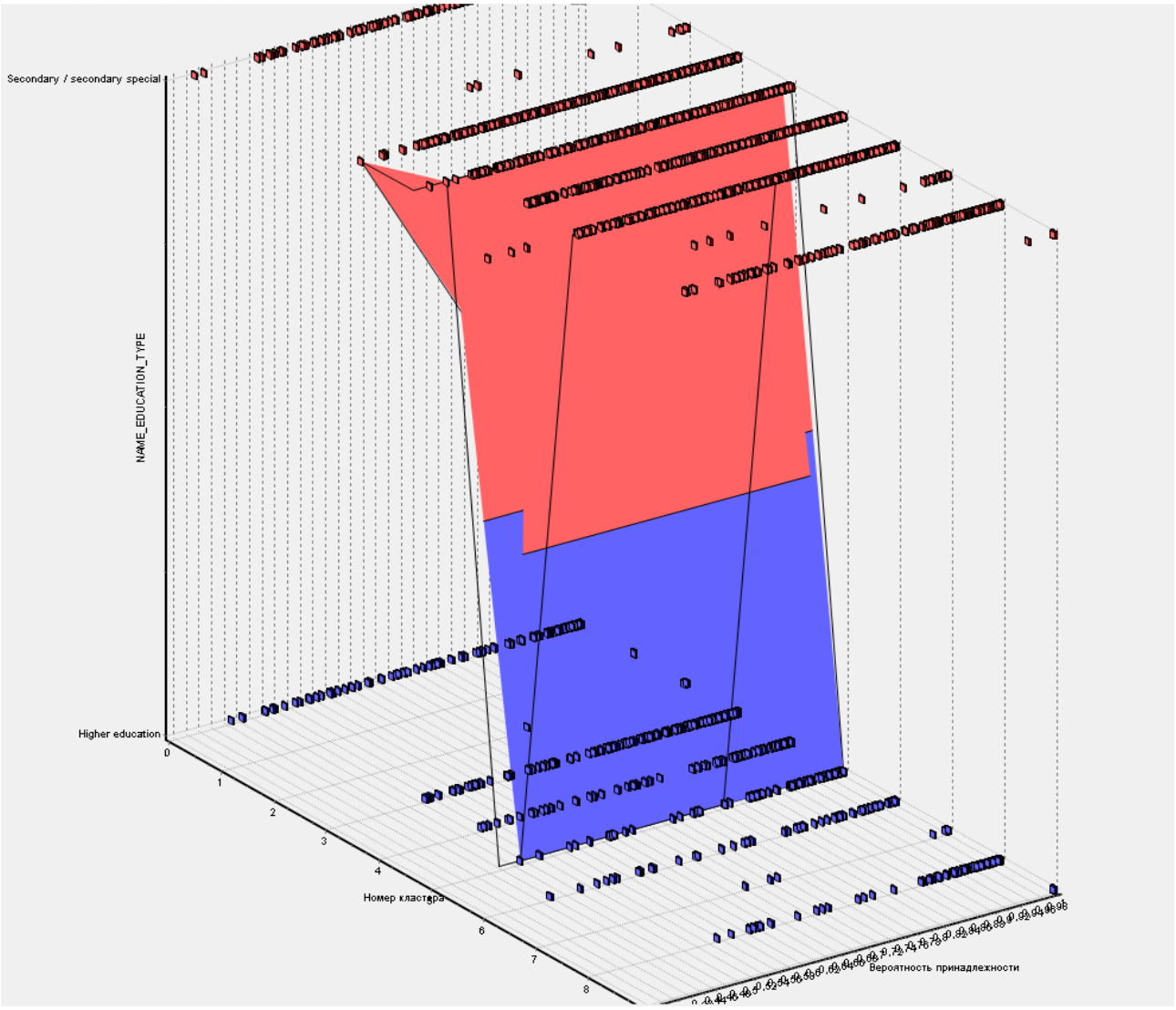


Рисунок 3.15 – Візуалізація характеристики “NAME_EDUCATION_TYPE”

Що стосується отриманих доходів “NAME_INCOME_TYPE”, то візуалізація наповненості кластерів клієнтів за даною характеристикою представлена на рисунку 3.16. Виявляється, що переважна кількість зловмисників отримують дохід від роботи або в результаті участі в комерційних об’єднаннях. Але цікавість викликає кластер під номером 1, куди попали ті шахраї, які отримують пенсії. З урахуванням визначення попередніх ознак виявляється, що даний кластер містить зловмисників, які знаходяться на пенсії, завершили середню школу, мають сім’ю, та можливо проблеми житлового характеру. Тобто, даний кластер є високо ризикованим щодо надання кредиту та його неповернення. Навіть, якщо клієнт і не має мотивації щодо цілеспрямованого шахрайства, то, можливо, різні фактори сприятимуть тому, що вони не зможуть повернути кошти банку.

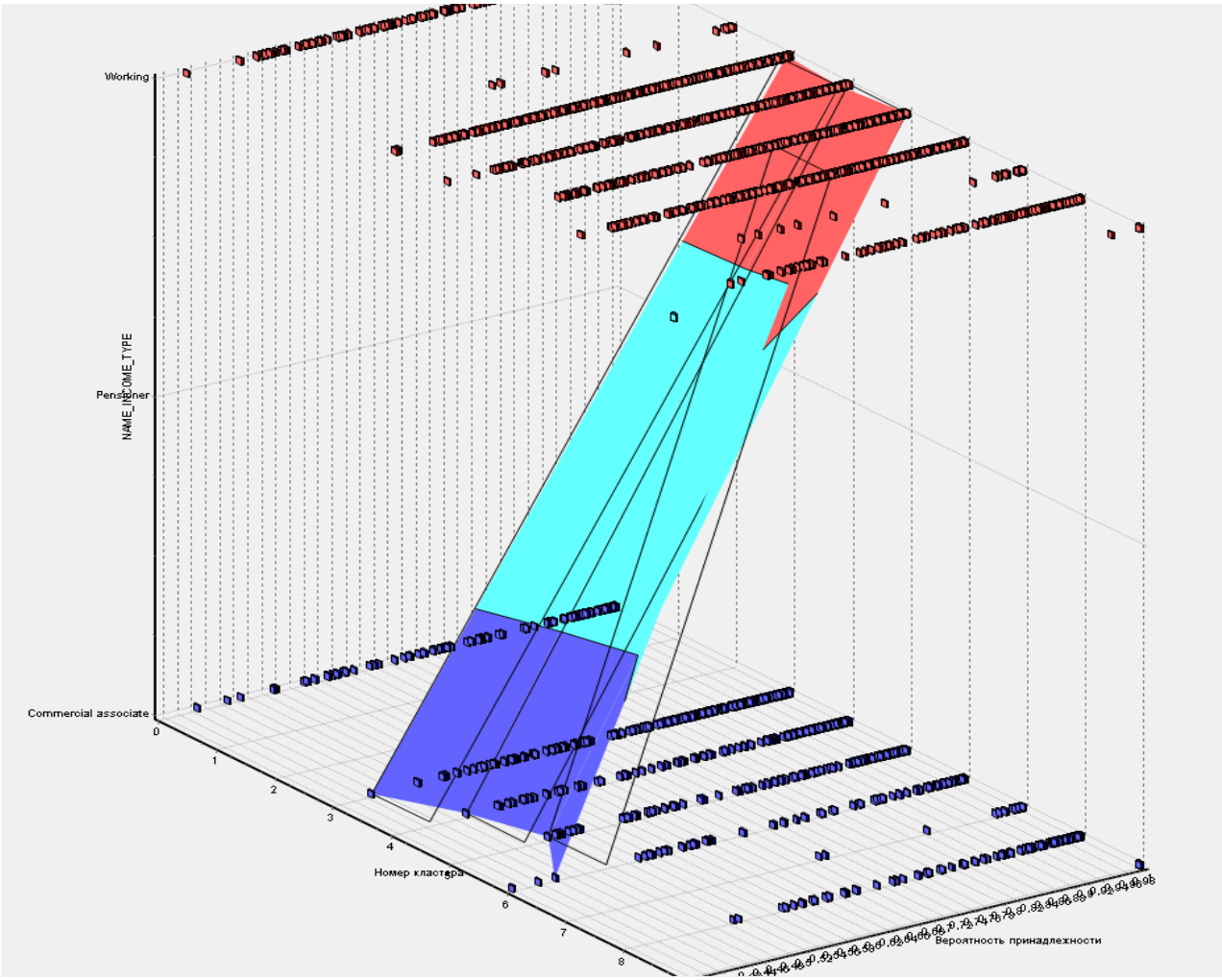


Рисунок 3.16 – Візуалізація характеристики “NAME_INCOME_TYPE”

Наступною характеристикою є “WEEKDAY_APPR_PROCESS_START”, для якої візуалізація наповненості кластерів клієнтів представлена на рисунку 3.17. Встановлено, що переважна кількість клієнтів звертається за кредитом у вівторок, а найменша кількість відповідає вихідним дням. На нашу думку, дана ознака можливо буде мало інформативною для формування кіберпрофілю у випадку таких злочинів як кібератака або соціальна інженерія. Щодо злочинів, пов’язаних з кредитним шахрайством, дана характеристика може дозволити тільки отримати додаткову оцінку завантаженості працівників банку на обробку заявок. Тобто дане знання може сприятиме формуванню додаткових організаційних заходів для підвищення уваги щодо неправильного прийняття рішення стосовно видачі кредиту.

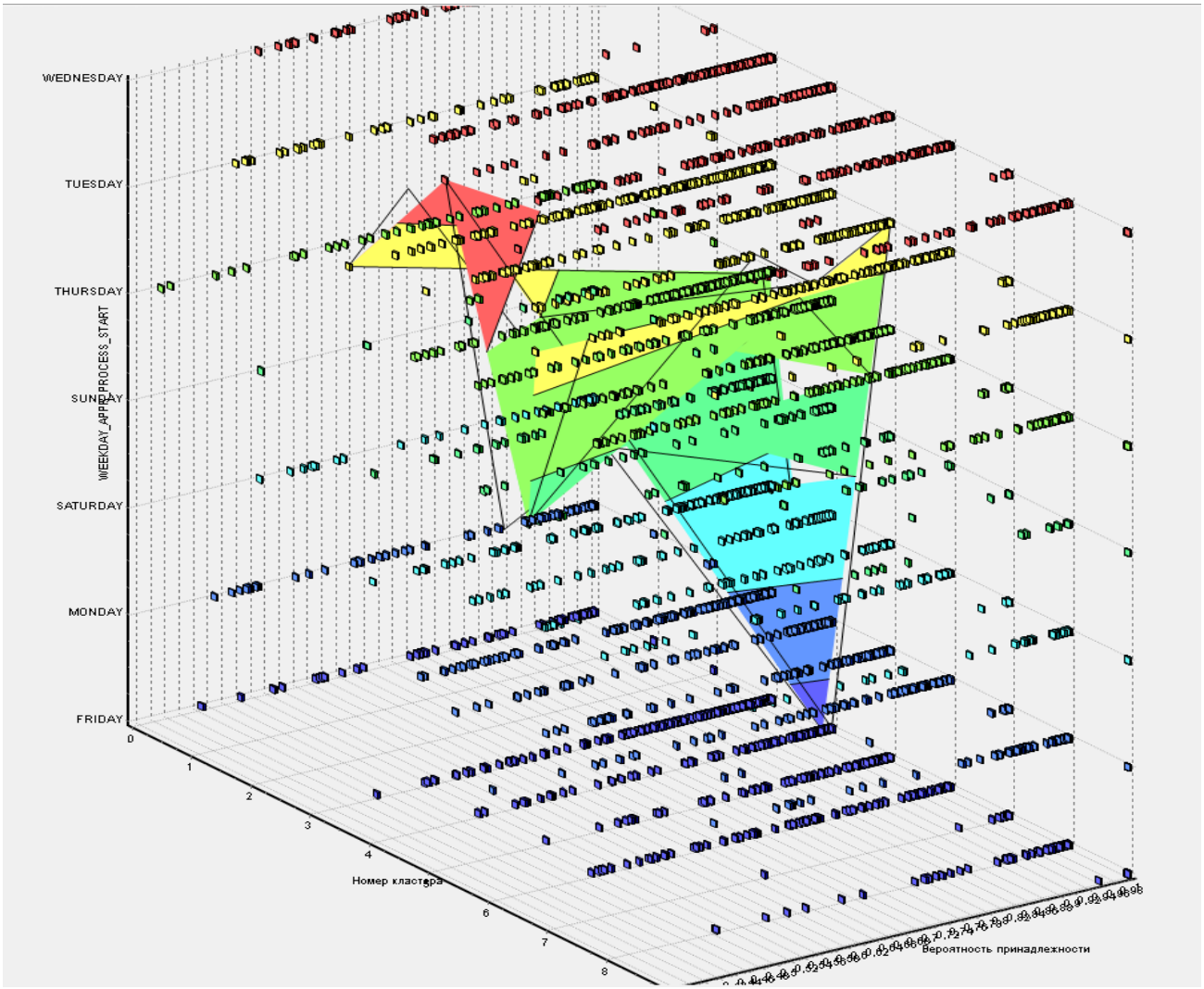


Рисунок 3.17 – Візуалізація характеристики “WEEKDAY_APPR_PROCESS_START”

В додатках представлені інші характеристики, які можна використовувати для формування та аналізу профілів кіберзлочинців. Для різних ситуацій та видів злочинів дані характеристики можуть бути замінені на ті, які є актуальними для даного випадку. Але запропонований в даній роботі підхід до профілювання кіберзлочинів є також ефективний у боротьбі з кіберзлочинністю для громадян, підприємств та державних установ поряд із програмним забезпеченням, математичними та технічними інструментами. У їх розвитку вкрай важливо враховувати різні характеристики, такі як поведінкові, географічні, психологічні та соціальні. Для їх формування профілів важливо застосовувати криміналістичні, слідчі, клінічні та статистичні методи, які сприятимуть виявленню кіберзагроз за різними аспектами їх виникнення.

Пункт 3.2 було виконано із використанням матеріалів публікацій виконавців [100].

3.3 Алгоритми розпізнавання поведінки кібершахраїв

Стрімке зростання кібершахрайств за останнє десятиліття пов’язане з різними чинниками, такими як масова цифровізація різних процесів у суспільстві, автоматизація бізнес-процесів суб’єктів господарювання, створення та впровадження розумних міст, комп’ютеризація державних структур тощо. Усе це створило сприятливі умови для розвитку та використання нових доступних комп’ютерних технологій, які використовуються злочинцями для викрадення приватної інформації, отримання доступу до акаунтів інших людей, зміни даних, пошкодження комп’ютерних мереж та виведення з ладу інформаційних систем. Хоча проблема боротьби з кіберзлочинністю не є головною для світу, її вирішення вкрай необхідно для суспільства.

Одним із інструментів протидії кібершахрайству є математичні методи та моделі, які використовуються для розробки алгоритмів розпізнавання поведінки кіберзлочинців. Існує безліч підходів, які можна застосувати для вирішення тієї чи іншої задачі протидії та виявлення шахрайств. Найбільш узагальненими є статистичні методи. Їх принциповими особливостями для дослідження даних є

усереднення характеристик вибірки. При дослідженні реальних процесів, наприклад, в банківській справі, зазначені характеристики є фіктивними величинами. Також статистика вивчає масові явища, тобто не поодинокі випадки, а сукупність фактів. При цьому кількісну розмірність показників неможливо досліджувати без їх якісної визначеності. Головним завданням статистичного дослідження є виявлення закономірностей, залежностей (взаємозалежностей) між явищами, перевірка гіпотези, розбиття матриці даних на підмножини (класи), аналіз залежності однієї величини від іншої, тощо.

Вивчаючи більш детально метод статистичного аналізу, наприклад авторами Nuha M., Mahmud S., Sattar A., визначаються ключові фактори зростання шахрайства в секторах електронного та мобільного банкінгу. Використовуючи кореляційний аналіз, автори виявили значну залежність між недостатньою обізнаністю та ймовірністю негативних результатів від шахрайства. Результати дослідження демонструють, що 86,3 % жертв електронного або мобільного банківського шахрайства раніше не знали про цей тип злочинності, в результаті чого зловмисники використовували тактику та емоційні маніпуляції для отримання конфіденційної інформації клієнтської бази замість процесу злому на основі кодування [101].

Статистичні методи дозволяють зробити тільки певні припущення в ситуаціях із кіберзлочинами. Більш сучасні системи протидії кіберзагрозам використовують інтелектуальний аналіз даних (Data Mining). Він є концепцією, мета якої полягає в аналізі даних різної природи, просіюванні величезних обсягів збережених даних, що можуть бути неточними, неповними, суперечливими, різнорідними [102]. Також інтелектуальний аналіз даних – це процес обробки значного масиву інформації, з метою виявлення в них латентних правил і закономірностей; процес виявлення в «сирих» даних: раніше невідомих, нетривіальних, практично корисних, доступних інтерпретації знань, необхідних для прийняття рішень [103].

Систематизовано підходи до методів інтелектуального аналізу даних, які використовувалися вченими у наукових працях для виявлення кібершахрайств у банках (таблиця 3.1).

Таблиця 3.1 – Підходи до методів інтелектуального аналізу даних для виявлення кібершахрайств у банках

№ з/п	П.І.Б. науковців	Методи інтелектуального аналізу
Вітчизняні дослідники та науковці		
1.	Яровенко Г.М., Сковронська А.І., Бояджян М.М.	метод дерева рішень (decision trees), нейронні мережі, логіт-регресія
2.	Syniavska O., Dekhtyar N., Deyneka O., Zhukova T., Syniavska O.	класична модель Лотки-Вольтерра з логістичним зростанням та динамічна модель Холлінга-Таннера
Зарубіжні дослідники та науковці		
3.	Lekha K. Chitra, Prakasam S.	k-means, Influenced Association Classifier, J48 Prediction tree
4.	Vinayakumar R., Alazab M., Soman K. P., Poornachandran P., Al-Nemrat A., Venkatraman S.	нейронна мережа (DNN), тип моделі глибинного навчання
5.	Nuha M., Mahmud S., Sattar A.	статистичний метод Data Mining (кореляційний аналіз)
6.	Kanimozhi V., Prem Jacob T.	метод штучного інтелекту
7.	Fedotova G. V., Gontar A. A., Titov V. A., Kurbanov A. K., Kuzmina E. V.	методи вертикального, горизонтального, фінансового, трендового, систематизації, аналогії, порівняння
8.	Akinbowale O. E., Klingelhöfer H. E., Zerihun M. F.	метод BSC (збалансування системи показників)
9.	Alshamasi S., Menai M.	cluster analysis

Із швидким розвитком інформаційних технологій з’явилися різновиди інтелектуального аналізу, такі як дерева рішень, нейронні мережі, логіт-регресія, система міркувань на основі аналогічних прикладів, метод штучного інтелекту, метод BSC (збалансування системи показників), еволюційне програмування, генетичні алгоритми, візуалізація багатовимірних даних, алгоритми обмеженого перебору, предметно-орієнтовні аналітичні системи, тощо [103].

Методи інтелектуального аналізу в банківській справі використовуються для виявлення фактів шахрайства з дебетовими та кредитними картками клієнтів, електронного та мобільного банкінгу; сегментація клієнтів для результативної маркетингової політики; прогнозування змін клієнтури, побудова моделей

прогнозування обсягів споживання відповідних послуг, тощо. Це великий клас традиційних підходів (перевірка гіпотез, факторний аналіз, кореляційний, канонічний, регресійний аналіз, кластеризація) та сучасні методи (дерева класифікації, багатомірне шкалювання, структурне моделювання, дискримінантний, логлінійний, дисперсійний аналіз, компоненти дисперсії, побудова класифікаційних, асоціативних правил). Для реалізації методів інтелектуального аналізу використовують різні версії статистичних пакетів, такі як Statistica, SPSS, SAS, STATGRAPICS, STADIA, Python, R, Deductor, тощо.

Alshamasi S., Menai M. для виявлення порушень стилю написання в документі, виявлення позиції зміни авторів вимагає декомпозиції тексту на його авторські компоненти. Найкращий метод для поставленої задачі є групування текстового документу в стилістично однорідні кластери, елементи об’єднуються включаючи схожі за стилем фрагменти тексту. Застосований метод кластеризації авторства всередині документу відіграє важливу роль в розслідуванні кіберзлочинів у банках, судовій лінгвістиці, тощо [104].

Результати кластеризації часто не піддаються змістової інтерпретації, не відповідають інтуїтивним очікуванням, набутому досвіду за даною проблематикою. Кластеризуючи показники бази даних слід аналізувати позитивні та негативні аспекти кластеризації, обрати максимально прийнятні алгоритми. Швидкий розвиток комп’ютерної технології породжує нові методи обробки інформації, наука про бази даних зростає, обсяги інформації невпинно збільшуються. З цих причин сучасним статистичним методам складно адекватно опрацювати значні масиви даних.

Використання нейромереж дає можливість виявити приховані взаємозв’язки, у процесі роботи системи виділити класи за подібністю об’єктів. Так, використання нейронних мереж дозволяє обробити вагомі масиви даних, використовувати алгоритми об’єднаних методів ієрархічної кластеризації з іншими методами. Нейронні мережі здатні знаходити розв’язки навіть за відсутності закономірностей та залежностей між перемінними, за відсутності апріорних відомостей про вибірку даних. Зазначимо, статистичний аналіз та

математичні методи поступають у адекватному вирішенні відповідних задач. Нейронні мережі відзначаються потенціальною швидкодією, яка формується на основі масового паралелізму обробки масиву даних.

Особлива привабливість нейронних мереж зумовлена здатністю давати точні прогнози, точність прогнозування значно вища відносно статистичного аналізу. Також, до переваг віднесемо можливість працювати з неповними даними, здатність до навчання (експерт вільний у виборі математичної моделі, адже її побудова відбувається адаптивно під час навчання), висока точність, реалізація нелінійних відображень, здатність системи до адаптації (реакція та пристосування до зміни навколишнього середовища),

Lekha K. Chitra, Prakasam S. досліджують дані про кіберзлочинність на основі методів інтелектуального аналізу даних, а саме, такі як K-Means, Influenced Association Classifier, J48 Prediction tree. Метою аналізу є розпізнавання закономірностей кіберзлочинів, щоб передбачити злочинність, передбачити злочинну діяльність і запобігти їй. Науково-методичний підхід алгоритму K-Means вибирає початкові центроїди, щоб класифікатор міг отримувати дані та формулювати прогнози кіберзлочинів за допомогою алгоритму J48. Об’єднання методів безсумнівно дасть покращений, об’єднаний і точний результат показників кіберзлочинності в банківському секторі, щоб подолати та запобігти кібератаці та прогнозування неплатежів, виявлення підроблених транзакцій тощо [105].

Vinayakumar R., Alazab M., Soman K. P., Poornachandran P., Al-Nemrat A., Venkatraman S. досліджують глибоку нейронну мережу (DNN), тип моделі глибинного навчання, для розробки гнучкої та ефективної IDS, що виявляє та класифікує непередбачувані кібератаки. Даний тип дослідження визначає найкращий алгоритм, який адекватно виявляє майбутні кібератаки. Комплексна оцінка експериментів DNN та різних класичних класифікаторів машинного навчання сповіщає на різних загальнодоступних наборах даних про зловмисне програмне забезпечення [106].

Kanimozhi V., Prem Jacob T. запропонований методичний підхід має виявити класифікацію ботнет-атаки, яка становить серйозну загрозу для фінансового сектора та банківських послуг. Використаний метод штучного інтелекту відіграє важливу роль у виявленні кібератак, системи виявлення вторгнень (IDS). Дослідження проводиться до реалістичного набору даних виявлення вторгнень кіберзахисту (CSE-CIC-IDS2018), створеного в 2018 році Канадським інститутом кібербезпеки (CIC) на AWS (веб-сервіси Amazon). Оцінка точності 99,97%, коефіцієнт помилкових позитивних результатів 0,001, тобто використаний метод штучного інтелекту виявлення атак ботнету є значущим, може бути запропонований для аналізу мережевого трафіку в реальному часі [107].

Syniavska O., Dekhtyar N., Deyneka O., Zhukova T., Syniavska O. запропонована математична модель присвячена питанню протидії кібератакам у сфері електронного банкінгу. В основі моделювання закладена класична модель Лотки-Вольтерра з логістичним зростанням та динамічна модель Холлінга-Таннера. На основі теорії біфуркації виділено типи фіксованих точок: сідло, стабільний вузол, стабільний вироджений вузол та лінії стабільних фіксованих точок, що малоймовірно зустрічаються в реальному житті. Розглянуту модель можна використовувати для теоретичних та емпіричних досліджень протидії кібератакам банківського сектору [108].

Fedotova G. V., Gontar A. A., Titov V. A., Kurbanov A. K., Kuzmina E. V. використовують методи вертикального, горизонтального, фінансового та трендового Data Mining масиву даних для оцінки динаміки та тренду розвитку кіберзлочинності в банківській сфері [109]. Проведено аналіз ситуації з кіберзлочинністю банківської системи, розглянуті механізми забезпечення безпеки особистих рахунків, розробка таргетів впровадження інформаційної безпеки платіжних систем банківської сфери. Методи інтелектуального аналізу даних, а саме, систематизації, аналогії, порівняння використані авторами при формуванні висновків та рекомендацій щодо розглянутих напрямів підвищення економічної безпеки інформаційних банківських систем.

Akinbowale O. E., Klingelhöfer H. E., Zerihun M. F. використовують метод BSC (збалансування системи показників) для аналізу впливу кіберзлочинності на банківський сектор. На основі аналізу, щоб запобігти значним збиткам від кіберзлочинності, запропонована система оповіщення. Споживачами можуть бути банки, клієнти шляхом ефективного впровадження та інтеграції технології великих даних у їхню систему для пом’якшення негативного впливу. кіберзлочинності [110].

У банківському секторі існує серйозна проблема, пов’язана із кібершахрайствами щодо здійснення кредитних операцій. Дана проблема була розглянута в п. 3.3, де було описано структуру вхідних даних та відповідні дії, які проводилися із їх очищенням та модернізацією. Оскільки методи інтелектуального аналізу є найбільш ефективними у боротьбі із кіберзлочинністю, то було розроблено алгоритми виявлення кіберзлочинної операції на основі побудови різних регресій, дерева рішень та нейронної мережі.

Перший алгоритм базується на побудові регресій. У випадку із шахрайствами доцільно застосувати саме логістичну регресію, але для наших даних виявилося неможливим її побудувати через незбіжність матриці. Тому було прийнято рішення щодо побудови узагальненої регресії, яка дозволить зробити прогноз не у бінарному вигляді, а в інтервалі від 0 до 1. За наявності значень, близьких до 1, вважається ознака відповідною зловживанню, в протилежному випадку, вона не вважається відповідною зловживанню. Регресію було побудовано з урахуванням також й фіктивних змінних, оскільки набір має велику кількість категоріальних змінних, які було перетворено на фіктивні. Результат узагальненої регресійної моделі представлений на рисунку 3.18. Рівень р-значущості для всіх змінних є меншим 0,05, тобто рівняння складатиметься із статистично значущих величин. Коефіцієнт детермінації дорівнює 0,775, що в принципі свідчить про непогану якість моделі.

OLS Regression Results

Dep. Variable:	y	R-squared:	0.775
Model:	OLS	Adj. R-squared:	0.775
Method:	Least Squares	F-statistic:	5335.
Date:	Thu, 08 Dec 2022	Prob (F-statistic):	0.00
Time:	20:13:28	Log-Likelihood:	4017.0
No. Observations:	193384	AIC:	-7782.
Df Residuals:	193258	BIC:	-6500.
Df Model:	125		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	1.3710	0.007	184.444	0.000	1.356	1.386
CNT_CHILDREN	-0.0739	0.002	-36.548	0.000	-0.078	-0.070
AMT_INCOME_TOTAL	2.234e-09	8.23e-10	2.714	0.007	6.2e-10	3.85e-09
AMT_CREDIT	1.43e-07	8.49e-09	16.835	0.000	1.26e-07	1.6e-07
AMT_ANNUITY	2.78e-07	6.29e-08	4.418	0.000	1.55e-07	4.01e-07
AMT_GOODS_PRICE	-1.731e-07	9.42e-09	-18.384	0.000	-1.92e-07	-1.55e-07
DAYS_EMPLOYED	6.704e-06	2.79e-07	24.047	0.000	6.16e-06	7.25e-06
CNT_FAM_MEMBERS	0.0617	0.002	34.713	0.000	0.058	0.065
REGION_RATING_CLIENT	-0.0231	0.003	-7.941	0.000	-0.029	-0.017
REGION_RATING_CLIENT_W_CITY	0.0249	0.003	8.452	0.000	0.019	0.031
HOUR_APPR_PROCESS_START	-0.0019	0.000	-10.478	0.000	-0.002	-0.002
NAME_CONTRACT_TYPE_Cash loans	-0.0487	0.004	-12.731	0.000	-0.056	-0.041
NAME_CONTRACT_TYPE_Revolving loans	-0.0777	0.004	-18.509	0.000	-0.086	-0.069
CODE_GENDER_F	-0.0629	0.002	-36.519	0.000	-0.066	-0.059
CODE_GENDER_M	-0.0434	0.002	-23.715	0.000	-0.047	-0.040
FLAG_OWN_CAR_N	-0.0601	0.002	-36.586	0.000	-0.063	-0.057
FLAG_OWN_CAR_Y	-0.0997	0.002	-51.133	0.000	-0.104	-0.096
FLAG_OWN_REALTY_N	-0.0872	0.002	-44.361	0.000	-0.091	-0.083
FLAG_OWN_REALTY_Y	-0.0681	0.002	-41.674	0.000	-0.071	-0.065
NAME_TYPE_SUITE_Children	-0.1091	0.008	-13.685	0.000	-0.125	-0.093
NAME_TYPE_SUITE_Family	-0.1125	0.003	-41.072	0.000	-0.118	-0.107
NAME_TYPE_SUITE_Group of people	-0.0740	0.025	-2.985	0.003	-0.123	-0.025
NAME_TYPE_SUITE_Other_A	-0.1194	0.013	-8.964	0.000	-0.145	-0.093
NAME_TYPE_SUITE_Other_B	-0.1112	0.010	-11.544	0.000	-0.130	-0.092
NAME_TYPE_SUITE_Spouse, partner	-0.1177	0.004	-27.509	0.000	-0.126	-0.109
NAME_TYPE_SUITE_Unaccompanied	-0.0684	0.002	-39.181	0.000	-0.072	-0.065
NAME_INCOME_TYPE_Businessman	-0.1427	0.084	-1.700	0.089	-0.307	0.022
NAME_INCOME_TYPE_Commercial associate	-0.1076	0.002	-53.949	0.000	-0.112	-0.104
NAME_INCOME_TYPE_Maternity leave	-0.1334	0.237	-0.563	0.574	-0.598	0.331
NAME_INCOME_TYPE_State servant	-0.0878	0.003	-25.912	0.000	-0.094	-0.081
NAME_INCOME_TYPE_Student	-0.1630	0.090	-1.817	0.069	-0.339	0.013
NAME_INCOME_TYPE_Working	-0.0690	0.002	-43.108	0.000	-0.072	-0.066
NAME_EDUCATION_TYPE_Academic degree	-0.1618	0.030	-5.481	0.000	-0.220	-0.104
NAME_EDUCATION_TYPE_Higher education	-0.1036	0.002	-47.031	0.000	-0.108	-0.099
NAME_EDUCATION_TYPE_Incomplete higher	-0.1149	0.004	-29.209	0.000	-0.123	-0.107
NAME_EDUCATION_TYPE_Lower secondary	-0.0900	0.009	-9.873	0.000	-0.108	-0.072
NAME_EDUCATION_TYPE_Secondary / secondary special	-0.0632	0.002	-38.518	0.000	-0.066	-0.060
NAME_FAMILY_STATUS_Civil marriage	-0.1603	0.003	-58.961	0.000	-0.166	-0.155
NAME_FAMILY_STATUS_Married	-0.1353	0.002	-77.191	0.000	-0.139	-0.132
NAME_FAMILY_STATUS_Separated	-0.1087	0.003	-33.060	0.000	-0.115	-0.102
NAME_FAMILY_STATUS_Single / not married	-0.0907	0.002	-36.351	0.000	-0.096	-0.086
NAME_FAMILY_STATUS_Widow	-0.1479	0.005	-30.694	0.000	-0.157	-0.138
NAME_HOUSING_TYPE_Co-op apartment	-0.0665	0.011	-5.883	0.000	-0.089	-0.044
NAME_HOUSING_TYPE_House / apartment	-0.0423	0.002	-23.078	0.000	-0.046	-0.039
NAME_HOUSING_TYPE_Municipal apartment	-0.0678	0.004	-17.497	0.000	-0.075	-0.060
NAME_HOUSING_TYPE_Office apartment	-0.0897	0.008	-10.726	0.000	-0.106	-0.073
NAME_HOUSING_TYPE_Rented apartment	-0.0568	0.007	-8.477	0.000	-0.070	-0.044
NAME_HOUSING_TYPE_With parents	-0.0604	0.004	-16.082	0.000	-0.068	-0.053
OCCUPATION_TYPE_Accountants	-0.2913	0.004	-75.091	0.000	-0.299	-0.284
OCCUPATION_TYPE_Cleaning staff	-0.2803	0.005	-53.581	0.000	-0.291	-0.270
OCCUPATION_TYPE_Cooking staff	-0.2834	0.005	-54.843	0.000	-0.294	-0.273
OCCUPATION_TYPE_Core staff	-0.2447	0.003	-79.875	0.000	-0.251	-0.239
OCCUPATION_TYPE_Drivers	-0.2386	0.003	-73.936	0.000	-0.245	-0.232
OCCUPATION_TYPE_HR staff	-0.2650	0.013	-20.519	0.000	-0.290	-0.240
OCCUPATION_TYPE_High skill tech staff	-0.2792	0.004	-78.292	0.000	-0.286	-0.272
OCCUPATION_TYPE_IT staff	-0.2872	0.013	-21.401	0.000	-0.313	-0.261
OCCUPATION_TYPE_Laborers	-0.2062	0.002	-97.241	0.000	-0.210	-0.202
OCCUPATION_TYPE_Low-skill Laborers	-0.2197	0.009	-24.926	0.000	-0.237	-0.202
OCCUPATION_TYPE_Managers	-0.2613	0.003	-89.404	0.000	-0.267	-0.256
OCCUPATION_TYPE_Medicine staff	-0.2749	0.005	-53.376	0.000	-0.285	-0.265
OCCUPATION_TYPE_Private service staff	-0.3049	0.007	-44.512	0.000	-0.318	-0.291
OCCUPATION_TYPE_Realty agents	-0.3057	0.012	-25.755	0.000	-0.329	-0.282
OCCUPATION_TYPE_Sales staff	-0.2516	0.003	-94.429	0.000	-0.257	-0.246
OCCUPATION_TYPE_Secretaries	-0.2671	0.009	-30.272	0.000	-0.284	-0.250
OCCUPATION_TYPE_Security staff	-0.2583	0.006	-46.931	0.000	-0.269	-0.247
OCCUPATION_TYPE_Waiters/barmen staff	-0.2767	0.010	-27.917	0.000	-0.296	-0.257

Рисунок 3.18 – Результати побудованої узагальненої регресії (початок)

ORGANIZATION_TYPE_Advertising	-0.3818	0.016	-23.670	0.000	-0.413	-0.350
ORGANIZATION_TYPE_Agriculture	-0.3875	0.015	-26.396	0.000	-0.416	-0.359
ORGANIZATION_TYPE_Bank	-0.3843	0.007	-56.184	0.000	-0.398	-0.371
ORGANIZATION_TYPE_Business Entity Type 1	-0.3897	0.005	-76.269	0.000	-0.400	-0.380
ORGANIZATION_TYPE_Business Entity Type 2	-0.3858	0.004	-97.820	0.000	-0.393	-0.378
ORGANIZATION_TYPE_Business Entity Type 3	-0.2964	0.002	-142.228	0.000	-0.301	-0.292
ORGANIZATION_TYPE_Cleaning	-0.3882	0.023	-16.797	0.000	-0.433	-0.343
ORGANIZATION_TYPE_Construction	-0.3707	0.005	-78.912	0.000	-0.380	-0.361
ORGANIZATION_TYPE_Culture	-0.3956	0.019	-21.133	0.000	-0.432	-0.359
ORGANIZATION_TYPE_Emergency	-0.4101	0.016	-24.973	0.000	-0.442	-0.378
ORGANIZATION_TYPE_Government	-0.3884	0.004	-87.467	0.000	-0.397	-0.380
ORGANIZATION_TYPE_Hotel	-0.3938	0.014	-28.001	0.000	-0.421	-0.366
ORGANIZATION_TYPE_Housing	-0.4130	0.007	-59.994	0.000	-0.426	-0.399
ORGANIZATION_TYPE_Industry: type 1	-0.4087	0.012	-35.176	0.000	-0.431	-0.386
ORGANIZATION_TYPE_Industry: type 10	-0.3994	0.031	-13.017	0.000	-0.460	-0.339
ORGANIZATION_TYPE_Industry: type 11	-0.4124	0.007	-57.910	0.000	-0.426	-0.398
ORGANIZATION_TYPE_Industry: type 12	-0.4215	0.018	-22.996	0.000	-0.457	-0.386
ORGANIZATION_TYPE_Industry: type 13	-0.3995	0.061	-6.522	0.000	-0.520	-0.279
ORGANIZATION_TYPE_Industry: type 2	-0.4377	0.015	-29.751	0.000	-0.467	-0.409
ORGANIZATION_TYPE_Industry: type 3	-0.4039	0.007	-55.707	0.000	-0.418	-0.390
ORGANIZATION_TYPE_Industry: type 4	-0.3951	0.013	-31.243	0.000	-0.420	-0.370
ORGANIZATION_TYPE_Industry: type 5	-0.4192	0.013	-31.296	0.000	-0.445	-0.393
ORGANIZATION_TYPE_Industry: type 6	-0.4425	0.038	-11.786	0.000	-0.516	-0.369
ORGANIZATION_TYPE_Industry: type 7	-0.4159	0.009	-43.851	0.000	-0.435	-0.397
ORGANIZATION_TYPE_Industry: type 8	-0.2662	0.075	-3.548	0.000	-0.413	-0.119
ORGANIZATION_TYPE_Industry: type 9	-0.4144	0.006	-65.382	0.000	-0.427	-0.402
ORGANIZATION_TYPE_Insurance	-0.4044	0.014	-29.124	0.000	-0.432	-0.377
ORGANIZATION_TYPE_Kindergarten	-0.3839	0.005	-78.271	0.000	-0.394	-0.374
ORGANIZATION_TYPE_Legal Services	-0.3731	0.018	-20.687	0.000	-0.408	-0.338
ORGANIZATION_TYPE_Medicine	-0.3740	0.005	-79.411	0.000	-0.383	-0.365
ORGANIZATION_TYPE_Military	-0.4046	0.008	-50.557	0.000	-0.420	-0.389
ORGANIZATION_TYPE_Mobile	-0.3518	0.018	-19.126	0.000	-0.388	-0.316
ORGANIZATION_TYPE_Other	-0.3811	0.004	-101.841	0.000	-0.388	-0.374
ORGANIZATION_TYPE_Police	-0.3974	0.008	-50.638	0.000	-0.413	-0.382
ORGANIZATION_TYPE_Postal	-0.4205	0.009	-47.057	0.000	-0.438	-0.403
ORGANIZATION_TYPE_Realtor	-0.3115	0.016	-19.238	0.000	-0.343	-0.280
ORGANIZATION_TYPE_Religion	-0.3998	0.053	-7.535	0.000	-0.504	-0.296
ORGANIZATION_TYPE_Restaurant	-0.3589	0.009	-39.170	0.000	-0.377	-0.341
ORGANIZATION_TYPE_School	-0.4010	0.005	-80.566	0.000	-0.411	-0.391
ORGANIZATION_TYPE_Security	-0.3829	0.008	-50.124	0.000	-0.398	-0.368
ORGANIZATION_TYPE_Security Ministries	-0.3957	0.009	-44.663	0.000	-0.413	-0.378
ORGANIZATION_TYPE_Self-employed	-0.3208	0.003	-125.543	0.000	-0.326	-0.316
ORGANIZATION_TYPE_Services	-0.3772	0.009	-42.413	0.000	-0.395	-0.360
ORGANIZATION_TYPE_Telecom	-0.4029	0.014	-28.669	0.000	-0.430	-0.375
ORGANIZATION_TYPE_Trade: type 1	-0.3748	0.018	-21.282	0.000	-0.409	-0.340
ORGANIZATION_TYPE_Trade: type 2	-0.3975	0.008	-51.227	0.000	-0.413	-0.382
ORGANIZATION_TYPE_Trade: type 3	-0.3609	0.006	-58.289	0.000	-0.373	-0.349
ORGANIZATION_TYPE_Trade: type 4	-0.4726	0.042	-11.258	0.000	-0.555	-0.390
ORGANIZATION_TYPE_Trade: type 5	-0.4704	0.045	-10.483	0.000	-0.558	-0.382
ORGANIZATION_TYPE_Trade: type 6	-0.3970	0.014	-28.382	0.000	-0.424	-0.370
ORGANIZATION_TYPE_Trade: type 7	-0.3725	0.004	-84.000	0.000	-0.381	-0.364
ORGANIZATION_TYPE_Transport: type 1	-0.3946	0.029	-13.679	0.000	-0.451	-0.338
ORGANIZATION_TYPE_Transport: type 2	-0.4077	0.008	-51.645	0.000	-0.423	-0.392
ORGANIZATION_TYPE_Transport: type 3	-0.3382	0.010	-32.387	0.000	-0.359	-0.318
ORGANIZATION_TYPE_Transport: type 4	-0.3816	0.005	-73.222	0.000	-0.392	-0.371
ORGANIZATION_TYPE_University	-0.3939	0.010	-40.649	0.000	-0.413	-0.375
HOUSETYPE_MODE_block of flats	-0.0534	0.004	-14.316	0.000	-0.061	-0.046
HOUSETYPE_MODE_specific housing	-0.0719	0.008	-9.044	0.000	-0.087	-0.056
HOUSETYPE_MODE_terraced house	-0.0794	0.009	-8.885	0.000	-0.097	-0.062

Omnibus:	85722.900	Durbin-Watson:	1.933
Prob(Omnibus):	0.000	Jarque-Bera (JB):	383793.968
Skew:	2.207	Prob(JB):	0.00
Kurtosis:	8.306	Cond. No.	4.42e+08

Рисунок 3.18 – Результати побудованої узагальненої регресії (продовження)

Оскільки модель регресії не є ефективною при побудові складних алгоритмів, у випадку, коли змінних дуже багато, де також не враховується їх нормальний розподіл, то було побудовано наступні моделі регресій, такі як LASSO, RIDGE та Elastic Net. Їх оцінки є менш зміщеними, що може дати кращі

результати в процесі прогнозування цільової змінної. Результати оцінок побудованих видів регресії представлені на рисунках 3.19-3.21. Значення коефіцієнту детермінації для LASSO регресії дорівнює 0,485, що говорить про не досить хорошу якість моделі. З урахуванням того, що оцінки є не такими зміщеними, як у випадку із узагальненою регресією, то не рекомендується застосовувати даний вид для алгоритму виявлення кіберзлочину.

Результат Elastic Net регресії (рис. 3.20) показує значення критерію детермінації, яке дорівнює 0,645. Тобто якість моделі є середньою і її результати можна брати до уваги в процесі протидії кіберзагроз. Найбільш ефективною виявилася RIDGE регресія, для якої коефіцієнт детермінації дорівнює 0,775 (рис. 3.21). І хоча його значення відповідає аналогічному для узагальненої регресії, у випадку для даних кібершахрайств цей вид регресії буде більш ефективним. Тому із запропонованих видів регресії обираємо RIDGE регресію.

```
[ 0.          -0.           0.          -0.          -0.          -0.
  0.00709414  0.          -0.          -0.          -0.           0.
 -0.          -0.17002758 -0.07068182 -0.07034849 -0.14391311 -0.12767708
 -0.11409848 -0.          -0.02288539 -0.          -0.          -0.
 -0.          -0.          -0.          -0.12148423 -0.          -0.04697268
 -0.          -0.06914145 -0.          -0.10276462 -0.          -0.
 -0.0052286  -0.05142318 -0.12352886 -0.03566322 -0.06989978 -0.00110099
 -0.          -0.          -0.          -0.          -0.          -0.
 -0.          -0.          -0.          -0.02688197 -0.          -0.
 -0.00278777 -0.          -0.04276375 -0.          -0.00402532 -0.
 -0.          -0.          -0.01437364 -0.          -0.          -0.
 -0.          -0.          -0.          -0.          -0.          -0.02625839
 -0.          -0.          -0.          -0.          -0.          -0.
 -0.          -0.          -0.          -0.          -0.          -0.
 -0.          -0.          -0.          -0.          -0.          -0.
 -0.          -0.          -0.          -0.          -0.          -0.
 -0.          -0.          -0.          -0.          -0.          -0.
 -0.          -0.          -0.          -0.          -0.          -0.
 -0.013564    -0.          -0.          -0.          -0.          -0.
 -0.          -0.          -0.          -0.          -0.          -0.
 -0.          -0.          -0.          -0.          -0.          -0.        ]
MSE train: 0.5145631, test: 0.5141511
R^2 train: 0.4854337, test: 0.4858315
```

Рисунок 3.19 – Результати побудованої LASSO регресії

```
[ 0.         -0.01195471  0.         -0.         -0.         -0.
  0.02756828  0.         -0.         -0.         -0.          0.
 -0.         -0.14838179 -0.07969526 -0.0889707  -0.14068873 -0.12803405
 -0.1151974  -0.         -0.0554265  -0.         -0.         -0.
 -0.01131085 -0.03925029 -0.         -0.13516669 -0.         -0.06805381
 -0.         -0.09783352 -0.         -0.12103528 -0.0247224  -0.
 -0.0524418  -0.08499625 -0.15366117 -0.07023511 -0.10681343 -0.03914577
 -0.         -0.01679104 -0.00547173 -0.         -0.         -0.
 -0.03612119 -0.02410465 -0.02056971 -0.09074713 -0.0542794  -0.
 -0.05793286 -0.         -0.10552381 -0.         -0.06719768 -0.03489841
 -0.00047365 -0.         -0.07605948 -0.         -0.02778745 -0.
 -0.         -0.         -0.         -0.         -0.01205588 -0.04883392
 -0.         -0.         -0.         -0.         -0.         -0.00404568
 -0.         -0.         -0.         -0.         -0.         -0.
 -0.         -0.         -0.         -0.         -0.         -0.
 -0.         -0.         -0.         -0.         -0.         -0.
 -0.0172396  -0.         -0.         -0.02055019 -0.         -0.
 -0.         -0.         -0.         -0.         -0.         -0.
 -0.03792866 -0.         -0.         -0.         -0.         -0.
 -0.         -0.         -0.         -0.00306903 -0.         -0.
 -0.         -0.         -0.         -0.         -0.         -0.        ]
MSE train: 0.355, test: 0.355
R^2 train: 0.645, test: 0.645
```

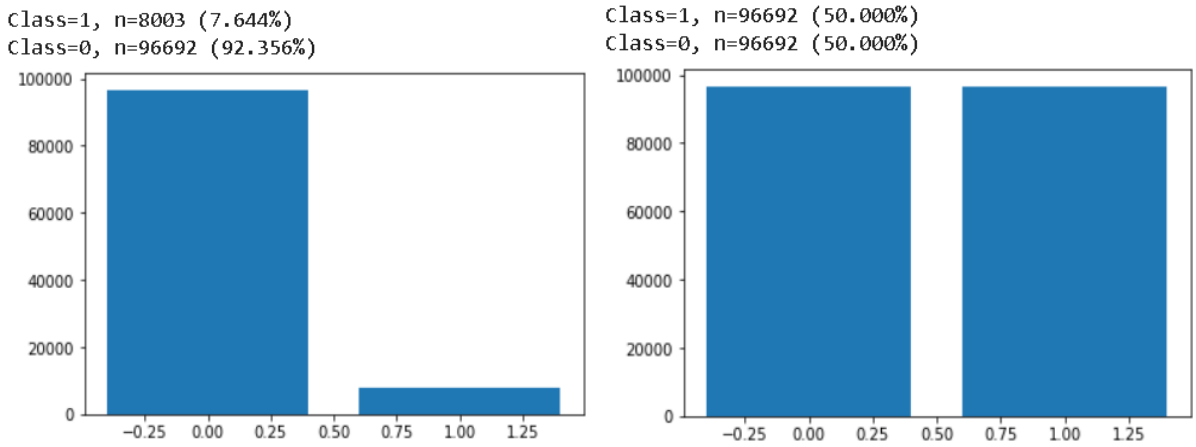
Рисунок 3.20 – Результати побудованої Elastic Net регресії

```
[ 0.         -0.0935649   0.00296911  0.11892204  0.00761869 -0.12913517
  0.02907017  0.10521608 -0.02273391  0.02370381 -0.01219976 -0.03006902
 -0.04096166 -0.06171671 -0.03961367 -0.05895927 -0.08629575 -0.06966483
 -0.06604109 -0.01495996 -0.05761171 -0.0056295  -0.00891418 -0.01353051
 -0.03252201 -0.05715769 -0.00186482 -0.08727753 -0.00061832 -0.03567371
 -0.00140046 -0.06880264 -0.00650444 -0.08127442 -0.03521828 -0.00921614
 -0.06114914 -0.08044876 -0.1354045  -0.04329814 -0.0566609  -0.03632243
 -0.00629059 -0.03384819 -0.02165712 -0.01174937 -0.0118305  -0.0197203
 -0.0961641  -0.06422306 -0.06753818 -0.13185843 -0.10355067 -0.02263505
 -0.1039079  -0.02503478 -0.15910371 -0.0281121  -0.13233491 -0.08087302
 -0.05384816 -0.02883609 -0.14721941 -0.03208359 -0.06368    -0.03264186
 -0.02671673 -0.02959126 -0.06487991 -0.08927476 -0.12119003 -0.23233962
 -0.01987878 -0.09207939 -0.02309125 -0.04150343 -0.02694944 -0.10920965
 -0.02940132 -0.06811247 -0.03740234 -0.01335678 -0.06399749 -0.02604187
 -0.00782066 -0.0322204  -0.0620455  -0.03387124 -0.03377689 -0.01207303
 -0.04952477 -0.00639443 -0.07446286 -0.03104605 -0.09803155 -0.02329821
 -0.11949942 -0.0584315  -0.02155258 -0.12947565 -0.05920128 -0.05335531
 -0.02226548 -0.00756368 -0.04304942 -0.10174997 -0.06590341 -0.05091359
 -0.18651761 -0.05025029 -0.03154031 -0.02348703 -0.05775551 -0.06606384
 -0.01226029 -0.01163781 -0.03195566 -0.10147148 -0.01543445 -0.05851367
 -0.03635182 -0.0854006  -0.04534433 -0.01753266 -0.00932793 -0.00966927]
MSE train: 0.225, test: 0.225
R^2 train: 0.775, test: 0.775
```

Рисунок 3.21 – Результати побудованої RIDGE регресії

В якості наступного алгоритму пропонуємо застосування дерева рішень. Оскільки маємо справу із бінарною цільовою змінною, то доцільно буде побудувати класифікаційне дерево рішень. Але спочатку треба провести аналіз збалансованості змінних, що є важливим для побудови даного виду моделей. Для збалансування масиву даних була застосована техніка передискретизації синтетичної меншості. Результати незбалансованого початкового набору даних та даних після збалансування представлені на рисунку 3.22.

На рисунку 3.22(а) чітко видно, що дані, які відповідають випадкам, ідентифікованим як кіберзлочин, дорівнюють всього 7,64%. Тобто набір даних не є збалансованим. Застосування техніки передискретизації синтетичної меншості дозволило отримати збалансований набір даних (рис. 3.22(б)). Також, побудова дерев рішень на різних видах збалансованої та незбалансованої вибірок згодом підтвердило, що ефективніше даний алгоритм будувати на основі саме збалансованого масиву даних.



а) б)

Рисунок 3.22 – Дані до (а) і після (b) методу передискретизації синтетичної меншості

Для побудови дерева рішень необхідно визначити його глибину, що сприятиме формуванню такої моделі, значення якої можна інтерпретувати та використати для моделювання. Тому проведено визначення точності поділу гілок дерева рішень для незбалансованих даних за допомогою тесту Джині та ентропії. Результати представлені на рисунку 3.23.

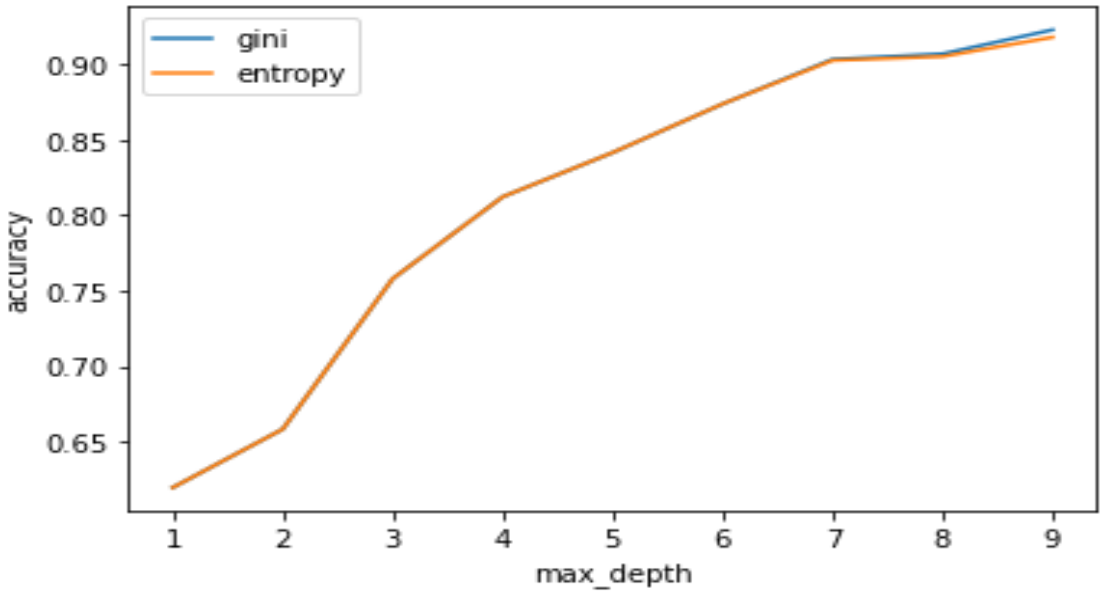


Рисунок 3.23 – Визначення точності поділу гілок дерева рішень за допомогою тесту Джині та ентропії

Дані рисунку 3.23 показують, що максимальної точності модель досягатиметься при глибині 9, але практика показує, що велике дерево рішень важко застосовувати для інтерпретації. Пожертвуємо часткою точності моделі з рахунок її спрощення. Тому обираємо глибину рівну 7 за умови точності 0,9, що є досить гарним показником. При цьому такий рівень досягається як із використанням ентропійного коефіцієнту, так і показника Джині. Для побудови дерева рішень обираємо показник ентропії. Результат його точності представлено на рисунку 3.24, а самої конфігурації моделі на рисунку 3.25.

```
Confusion Matrix:
[[19408   161]
 [ 3574 15534]]
Classification Report:
              precision    recall  f1-score   support

           0       0.84      0.99      0.91     19569
           1       0.99      0.81      0.89     19108

    accuracy                           0.90     38677
   macro avg       0.92      0.90      0.90     38677
weighted avg       0.92      0.90      0.90     38677

Accuracy: 0.9034309796519896
```

Рисунок 3.24 – Оцінки якості дерева рішень

Загальна якість моделі для класів «0» та «1» є високою і дорівнює 0,9034. Дерево рішень дасть правильний прогноз з імовірністю 90,34%. Точність для позитивних рангів коливається від 0,84 до 0,99, що вказує на високу ймовірність того, що модель робить багато точних оптимістичних прогнозів і меншу кількість неправильних позитивних класифікацій. Параметр чутливості для всіх класів становить від 0,81 до 0,99, що підтверджує високу здатність моделі правильно визначати позитивні ранги. Оскільки ми не отримали суттєво відмінних значень точності та повторного виклику, показник F1 має високі значення, які наближаються до 1, що вказує на гарне поєднання точності та запам’ятовування моделі. Таким чином, запропонована модель є дуже якісною.

Оскільки дана модель за показником точності перевищує регресійні моделі, то можна зробити висновок, що класифікаційне дерево рішень буде більш ефективним.

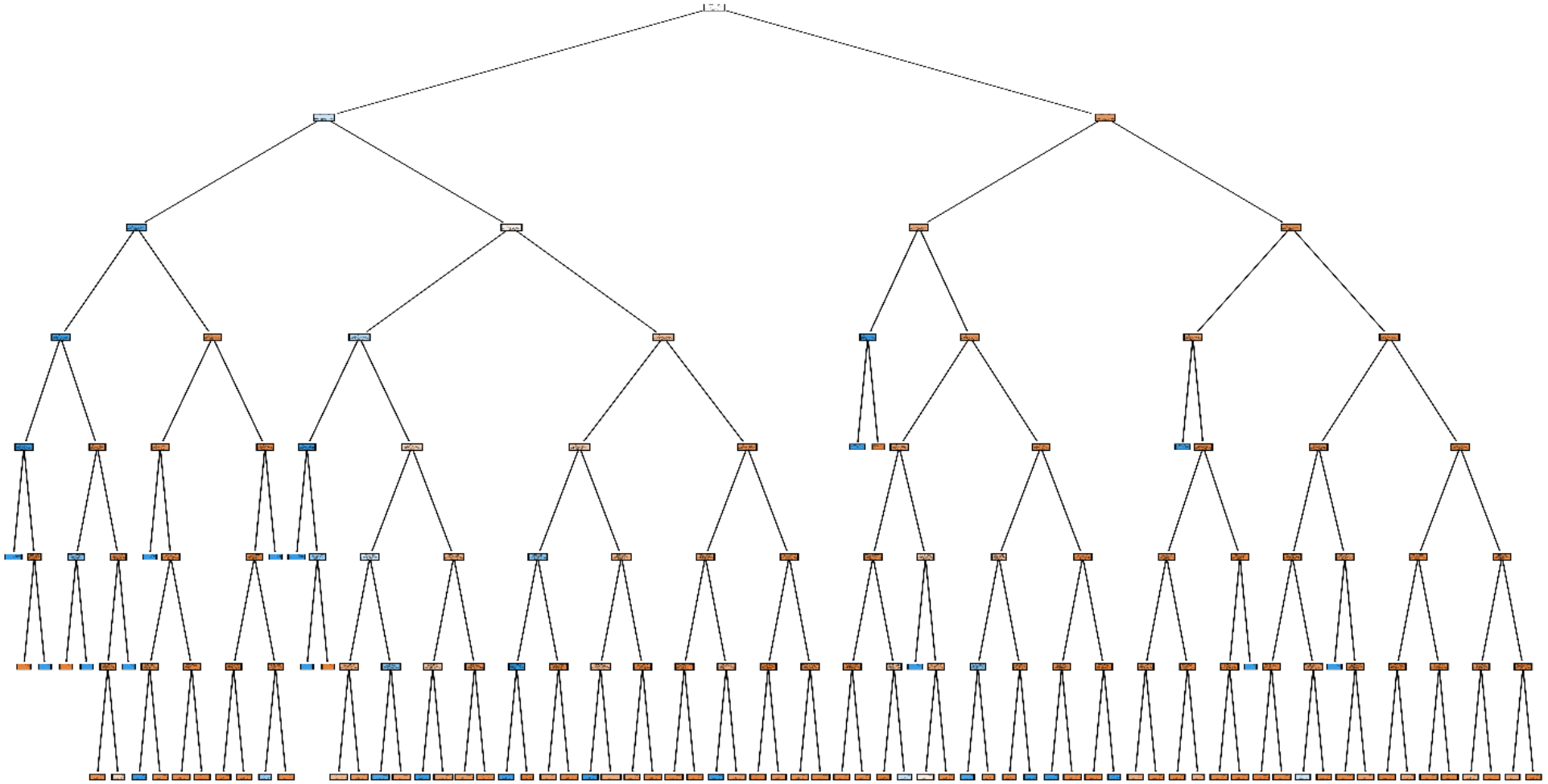


Рисунок 3.25 – Алгоритм розпізнавання поведінки кібершахраїв на основі моделі дерева прогнозних рішень

Визначимо третій алгоритм, який передбачає побудову нейронної мережі. Для даного дослідження було використано Deductor Academic, оскільки воно дозволяє здійснити візуалізацію нейронної мережі, що важко виконати, застосовуючи інші аналітичні пакети.

Математичну модель нейронної мережі з урахуванням вхідних та вихідних змінних щодо кібершахрайств з кредитними операціями можна представити наступним чином (формули 3.1-3.3):

$$h_1^{(2)} = f\left(w_{1_1}^{(1)} x_1 + w_{1_2}^{(1)} x_2 + \cdots + w_{1_{126}}^{(1)} x_{126} + b_1^{(1)}\right), \quad (3.1)$$

$$h_2^{(2)} = f\left(w_{2_1}^{(1)} x_1 + w_{2_2}^{(1)} x_2 + \cdots + w_{2_{126}}^{(1)} x_{126} + b_2^{(1)}\right), \quad (3.2)$$

$$y\left(\frac{p}{1-p}\right) = f\left(w_1^{(2)} h_1^{(2)} + w_2^{(2)} h_2^{(2)}\right) \quad (3.3)$$

де $f(\cdot)$ – активаційна функція вузла, в нашому випадку сигмоїдна (логістична) функція;

$h_1^{(2)}$ – вихід першого вузла у другому шарі нейронної мережі, входами у якій є вихід першого вузла, тобто $\left(w_{1_1}^{(1)} x_1 + w_{1_2}^{(1)} x_2 + \cdots + w_{1_{126}}^{(1)} x_{126}\right)$ та вільний член для даних першого шару $b_1^{(1)}$. Ці входи складаються та передаються в активаційну функцію для розрахунку виходу першого вузла. Інший вузол $h_2^{(2)}$ формується аналогічно;

y – вихід другого вузла у третьому шарі, в якому беруться зважені виходи вузлів другого шару $h_1^{(2)}, h_2^{(2)}$. Для кінцевого виходу p відповідає цільовій змінній, що дорівнює 0, $1-p$ – цільовій змінній, що дорівнює 1.

В якості активаційної функції для прихованих шарів та виходів застосовано сигмоїдальну (логістичну) функцію. Логістична функція для активації вихідних вузлів має вигляд (формула 3.4):

$$OUT = \frac{1}{1+\exp(-a \times net)}, \quad (3.4)$$

де OUT – виходи вузлів нейронної мережі у другому та третьому шарах, тобто $h_1^{(2)}, h_2^{(2)}$ та y;

net – сума вхідних сигналів, помножена на відповідні ваги для другого та третього шару, наприклад, $\left(w_{1_1}^{(1)} x_1 + w_{1_2}^{(1)} x_2 + \cdots + w_{1_{126}}^{(1)} x_{126} + b_1^{(1)}\right)$ для $h_1^{(1)}$ (див. формули 3.1-3.3);

a – ступінь крутизни логістичної функції.

Візуалізація отриманої нейронної мережі представлена на рисунку 3.26, де можна побачити, що на вході маємо 126 змінних та 3 шари. Третій шар відповідає прогнозованому значенню змінної, що сигналізує або випадок кіберзлочину, або його відсутності.

Представлена конфігурація нейронної мережі є на перший погляд дуже спрощеною. Перевіримо якість отриманої моделі, результати якої представлені на рисунку 3.27.

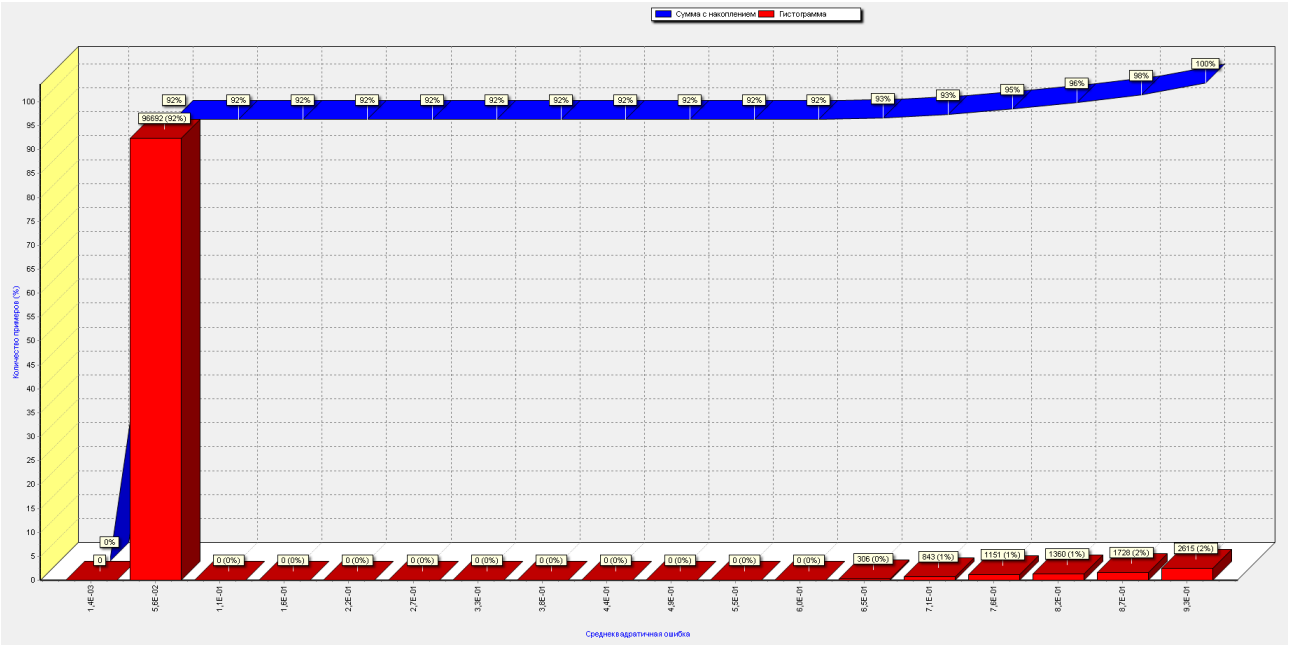


Рисунок 3.27 – Оцінка якості нейронної моделі

Графік на рисунку 3.27 показує, що середньоквадратична похибка практично дорівнює 0 для 92% прикладів. Навіть із накопиченням експериментів до 100% похибка не буде перевищувати 0,05. Отримані значення свідчать про достатньо високу якість алгоритму, побудованого на основі нейронної мережі.

Таким чином, методи виявлення та протидії кіберзагрозам є актуальними на сьогодні, особливо у контексті побудови відповідних алгоритмів розпізнавання поведінки кібершахраїв. У цьому контексті найбільш ефективними є алгоритми, засновані на математичних методах. У цьому дослідженні представлено алгоритми на основі регресійних моделей, класифікаційного дерева рішення та нейронної моделі. Отримані моделі демонструють досить непогані показники якості. Виявилося, що дерево рішень та нейронна мережа є більш точними в оцінці поведінки кібершахраїв. Саме тому пропонується їх використовувати на практиці в банківських установах для виявлення кіберзагроз на основі сформованих профілів кіберзлочинців.

Пункт 3.3 було виконано із використанням матеріалів публікацій виконавців [111, 112].

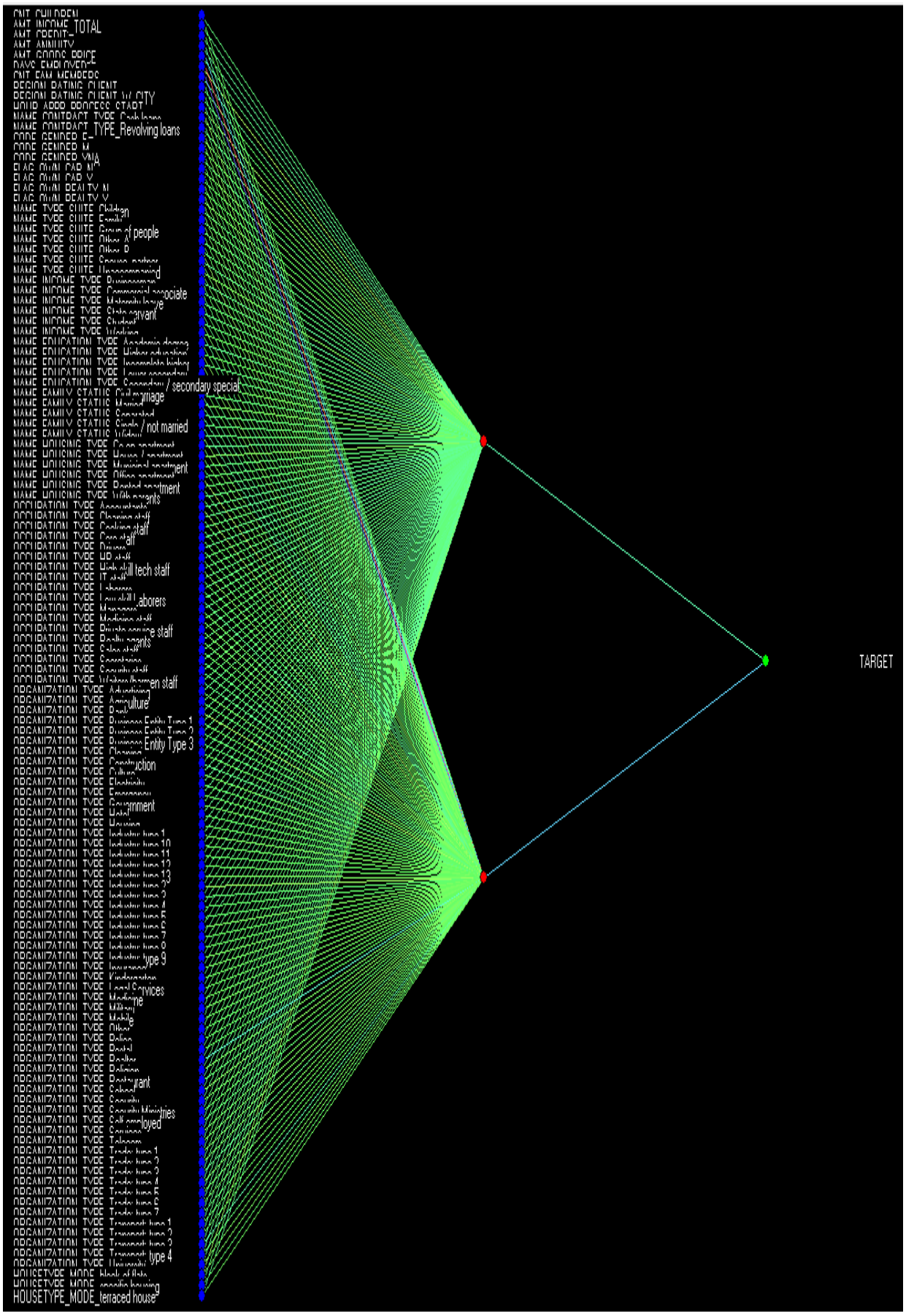


Рисунок 3.26 – Алгоритм розпізнавання поведінки кібершахраїв на основі нейронної моделі

4 МЕТОДИКА ПРОГНОЗУВАННЯ КІБЕРШАХРАЙСЬКИХ АТАК НА КОМП'ЮТЕРНІ СИСТЕМИ, МЕРЕЖЕВУ ТА ХМАРНУ ІНФРАСТРУКТУРУ ФІНАНСОВОЇ УСТАНОВИ

4.1 Розробка моделей прогнозування кібершахрайських атак

За останні два десятиліття Четверта промислова революція призвела до стрімкого зростання інформаційних та комунікаційних технологій у світі та їх активного впровадження у різні сфері життєдіяльності суспільства. З одного боку, це сприяло і сприяє виникненню позитивних тенденцій, таких як цифрова трансформація бізнесу, розвиток сфери Інтернету речей, економіки спільного користування, віртуалізації ІТ-інфраструктури, 3Д-маркетинг, поява та використання криптовалют, блокчейнів, штучного інтелекту, ай-трекінгу, тощо. З іншого боку, комп'ютеризація та цифровізація процесів призвела до появи такого негативного явища, як кіберзлочиннність, що супроводжується паралельним зростанням цифрової грамотності населення та зниженням вартості технологій для вчинення кіберзлочинів. Так, інсталяція шкідливого програмного забезпечення на темному веб-ринку вартує 1 долар, а персональні дані будь-якої людини можна отримати всього за 3 долара [113]. Тобто, будь-хто може стати кіберзлочинцем або отримати доступ до будь-яких конфіденційних даних за невелику ціну.

Про актуальність проблеми кіберзлочинів та кібершахрайств свідчить й інша статистика, яка показує динамічне зростання її негативних наслідків за останні роки. Так, середня вартість витоку даних у світі від кіберінцидентів у 2022 році склала 4,35 мільйонів доларів США, що збільшилась приблизно на 24.29% у порівнянні із 2014 роком (3,5 мільйонів доларів США) [83]. При цьому найбільш постраждалими є такі сектори, як охорона здоров'я (10,1 мільйонів доларів США), фінанси (5,97 мільйонів доларів США), фармацевтика (5,01 мільйонів доларів США), технології (4,97 мільйонів доларів США), енергетика (4,72 мільйони доларів США), послуги (4,7 мільйонів доларів США) та промисловість (4,47 мільйона доларів США) [84].

Те, що сьогодні проблема кіберзлочинності є значущою для світу, свідчать й намагання багатьох ІТ-компаній світу протидіяти їй шляхом створення відповідних рішень для кіберзахисту інформації та комп'ютерної інфраструктури та формування відповідно ринку. У 2022 році очікується дохід від кіберрішень та кіберпослуг у розмірі 159,84 мільярдів доларів США, що на 14,88% перевищує даний показник у 2021 році та на 91,68% у 2014 році [114]. При цьому прогнозується збільшення ринку кібербезпеки у 2027 році на 86,87% до 298,7 мільярдів доларів США [114]. За оцінками експертів зростатиме й ринок страхування від кіберінцидентів. У 2018 році його обсяг сягнув 4 мільярдів доларів США, у 2020 – 9 мільярдів доларів США, а у 2025 році його обсяг прогнозується рівним 20 мільярдів доларів США [115].

Боротьба із кіберзлочинністю є світовою проблемою, тому для її вирішення створено спеціальні організації, діяльність яких спрямована на формування механізму для забезпечення кібербезпеки. Серед них можна виділити Управління ООН з наркотиків і злочинності, Міжурядова група експертів відкритого складу з кіберзлочинності, Комісія із запобігання злочинності та кримінального правосуддя, Міжнародний союз електрозв'язку, Міжнародна організація кримінальної поліції, тощо. Також створюються регіональні та приватні організації, які займаються питаннями кіберзахисту.

Окрім інституційних механізмів світовими організаціями запроваджено ряд програм і ініціатив. У 2016 році країни-члени НАТО визнали кібербезпеку галуззю, якою повинен опікуватися Альянс на рівні із захистом на суші, повітрі та у морі, та прийняли оборонний мандат [116]. У 2021 році на саміті НАТО була запропонована та схвалена нова Комплексна політика кіберзахисту [116]. United Nations розробила програму «Кібербезпека та нові технології», спрямовану на розробку та посилення заходів боротьби із кібертероризмом для країн-членів та приватних компаній [117].

У зв'язку із війною, яку розпочала Російська Федерація проти України, багато країн стали впроваджувати посилені заходи щодо кібербезпеки. Наприклад, The White House в Інформаційному бюлетені виклав відповідні кроки

для приватних організацій для забезпечення протидії кібератакам, які можуть бути наслідками кібервійни [118]. National Cyber Security Centre підготував та опублікував нові вказівки, спрямовані на підтримку стійкості персоналу, якого повинні дотримуватися компанії в умовах кіберзагроз, ініційованих військовою агресією [119].

Таким чином, проблема боротьби із кіберзлочинністю є актуальною і інтерес до неї з часом тільки зростає, особливо в умовах світових пандемій та військових агресій. В даних умовах важливо приділяти увагу різним напрямкам її вирішення – інституційному, правовому, організаційному, методичному, тощо, що потребує системного підходу до їх дослідження і реалізації, як на практичному, так й на науковому рівнях.

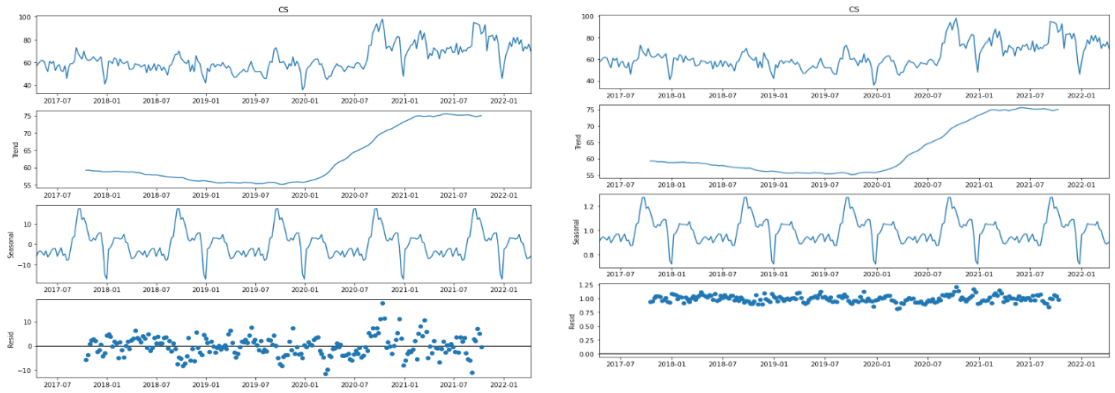
Можливості протидіяти кіберзлочинам передбачають необхідність прогнозувати потенційні кібератаки або кібершахрайства. Для вирішення даної проблеми найбільш ефективним є застосування математичних методів і моделей, які пропонуються науковцями різних світових дослідницьких шкіл. Серед них можна виділити традиційні економетричні методи дослідження, такі як регресійний аналіз [120], методи структурних рівнянь [121], VAR та VEC моделювання [122]. У дослідженнях набули популярності також методи нечітких множин [123], гравітаційне моделювання [124], Data mining [125], машинне навчання [126], штучний інтелект [127]. У даному досліджені в якості вхідних даних будуть використані інформаційні тренди найбільш популярних видів кіберзлочинів. Оскільки їх значення представлятимуть собою часові ряди, то для їх прогнозування доцільно використовувати саме економетричні методи, які є простими у реалізації та дають точні результати на коротко- та середньострокову перспективи.

Для дослідження і прогнозування тенденцій кіберзлочинів сформовано набір вхідних даних на основі запитів інструментарію Google Trends. Сюди увійшли найбільш популярні звернення Інтернет-користувачів до термінів “Кібератаки на комп’ютерні системи фінансової установи” (CS), “Кібератаки на мережеву інфраструктуру фінансової установи” (NI), “Кібератаки на хмарну

інфраструктуру фінансової установи” (CI) за період з 16.04.2017 по 10.04.2022 в розрізі потижневих рівнів.

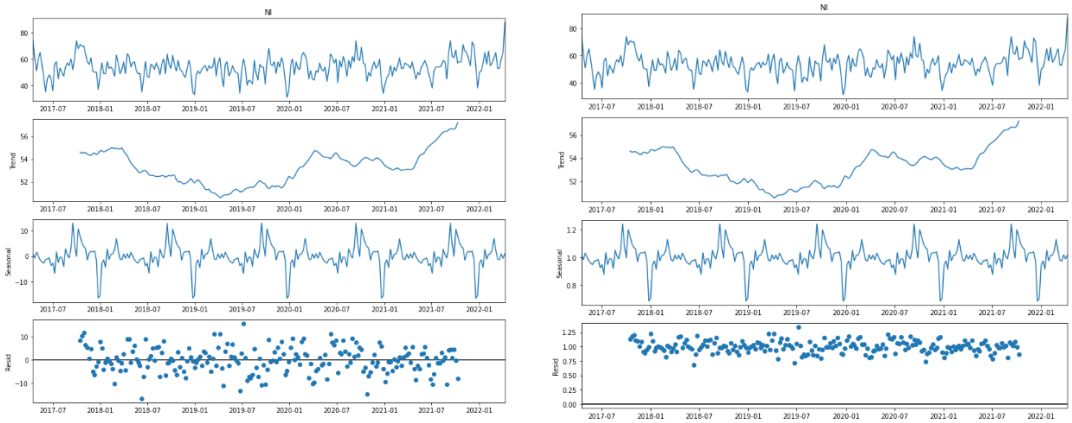
Цю інформацію було обрано, виходячи з наступних міркувань. Масові кібератаки, як правило, здійснюються по відношенню до суб’єктів економіки певної країни або країн. Віддзеркаленням цих подій є зростання зацікавленості Інтернет-користувачів у мережі щодо даних подій. Часовий розрив між реальним кіберзлочином та активністю в Інтернеті не може бути досить великим, оскільки реакція користувачів на значущі події у країні та світі є миттєвою. Офіційні джерела, які займаються збором, обробкою та публікацією статистичних даних, як правило, публікують її із значною затримкою у часі та в агрегованому вигляді. Тому у даному випадку інформаційні тренди, що відображають звернення Інтернет-користувачів, є швидким відкликом реальних подій. Відповідно, їх дослідження дозволить досить точно прогнозувати можливі кіберзлочини у світі.

Декомпозиція досліджуваних часових трендів запитів користувачів глобальної мережі з урахуванням сезонної, трендової та випадкової компонент для адитивної та мультиплікативної моделей представлена на рисунках 4.1-4.3.



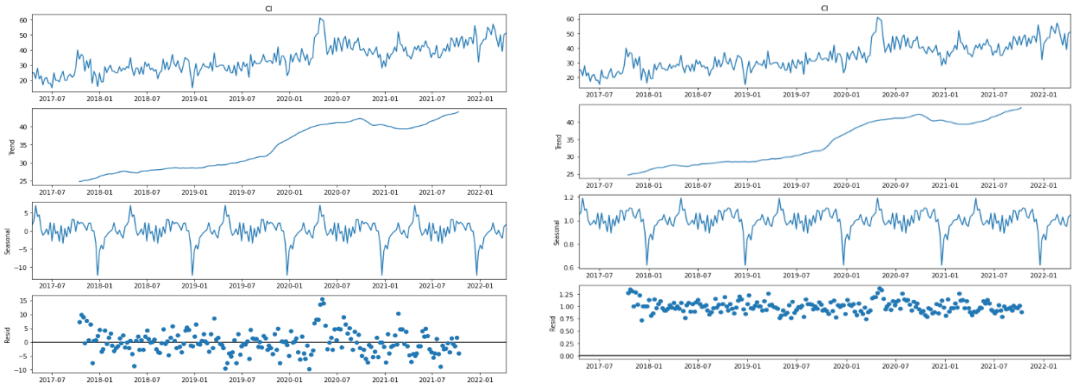
(**a**) (**b**)

Рисунок 4.1 – Декомпозиція (фактичні дані, трендова, сезонна та випадкова компоненти) часового ряду “Кібератаки на комп’ютерні системи фінансової установи”: (a) Адитивна модель; (b) Мультиплікативна модель



(**a**) (**b**)

Рисунок 4.2 – Декомпозиція (фактичні дані, трендова, сезонна та випадкова компоненти) часового ряду “Кібератаки на мережеву інфраструктуру фінансової установи”: (a) Адитивна модель; (b) Мультиплікативна модель



(**a**) (**b**)

Рисунок 4.3 – Декомпозиція (фактичні дані, трендова, сезонна та випадкова компоненти) часового ряду “Кібератаки на хмарну інфраструктуру фінансової установи”: (a) Адитивна модель; (b) Мультиплікативна модель

Аналіз декомпозиції ряду “Кібератаки на комп’ютерні системи фінансової установи” (рис. 4.1a-b) свідчить, що наявність трендової складової оцінити візуально досить складно, тому є необхідність застосування тестів для перевірки ряду на стаціонарність. Також потребує перевірці наявність викидів. Ряд містить сезонну компоненту та щільність розподілу залишків показує, що модель є адитивною. Аналіз декомпозиції “Кібератаки на мережеву інфраструктуру

фінансової установи” (рис. 4.2a-b) показує наявність сезонної компоненти, а щільність розподілу залишків свідчить на користь адитивного процесу. Щодо трендової компоненти, то візуальний аналіз не дозволяє зробити висновок про її відсутність чи наявність. На рисунках 4.3a-b представлена декомпозиція ряду “Кібератаки на хмарну інфраструктуру фінансової установи”, яка демонструє чітку наявність тренду, сезонної складової та відповідність адитивному процесу.

Тобто, досліджувані дані представляють собою часові ряди, моделювання яких можливе за допомогою моделей експоненційного згладжування або авторегресійних моделей в залежності від доведення стаціонарності чи нестаціонарності процесу.

Дане прогнозування інформаційних трендів показників кіберзлочинів передбачається здійснення наступних етапів.

Етап 1. Здійснення перевірки часових рядів на наявність та відсутність аномальних значень та проведення їх відповідного коректування. Для реалізації даного етапу буде застосовано статистичний метод Z-score. Z-score вимірює відстань між значенням спостереження та середнім значенням ряду за допомогою стандартних відхилень і розраховується за формулою (4.1):

$$z = (x - \mu)/\sigma\,, \tag{4.1}$$

де x – фактичне значення спостереження;

μ – середнє значення ряду;

σ – середньоквадратичне відхилення.

Розраховані значення Z-score порівнюють з пороговими (–3 та +3). Якщо одно із значень є більшим за +3 або меншим за –3, то дане спостереження є викидом.

Етап 2. Перевірка компонент сезонності для часових рядів CS, NI, CI, виконавши тест QS.

Сезонна стійкість виникає, коли процес є майже періодичним протягом сезону. У цьому випадку ми можемо думати, що середній рівень часового ряду x_t моделюється як:

$$x_t = S_t + w_t, \tag{4.2}$$

де S_t це сезонна складова, яка дещо змінюється від року до року відповідно до випадкового блукання:

$$S_t = S_{t-12} + v_t, \tag{4.3}$$

де w_t та v_t є некорельованими процесами білого шуму.

Для перевірки наявності компоненти сезонності в часових рядах CS, NI, CI пропонується використовувати тест QS та його застосування на базі пакету R. Ідея оцінки тесту QS базується на співвідношенні:

$$QS = n \cdot (n+2) \cdot \left(\frac{R_s^2}{n-s} + \frac{R_{2s}^2}{n-2s}\right), \tag{4.4}$$

де n – кількість спостережень у часовому ряду та s – періодичність даних (12 у цьому випадку з місячними даними);

R_s^2та R_{2s}^2 позначають автокореляції, отримані для відповідного часового ряду. Ця статистика приблизно відповідає розподілу $\chi 2$ з 2 ступенями свободи.

Для виконання тесту QS на сезонність у часовому ряді використовується функція:

$$qs(x, freq = NA, diff = T, residuals = F, autoarima = T), \tag{4.5}$$

де x - часові ряди;

freq - періодичність часового ряду;

diff – різницевий ряд;

residuals - залишки моделі;

autoarima - автоматичний.

Етап 3. Здійснення перевірки стаціонарності ряду шляхом застосування методу різниць середніх рівнів. Даний тест перевіряє гіпотезу про однорідність дисперсій частин часового ряду та гіпотезу про відсутність тренду. Застосування даного тесту є обґрунтованим для вхідних даних, оскільки на графіках трендів (рис. 4.1 – 4.3) можна побачити, що дані не є однорідними протягом всього періоду часу та мають перегин. Для його реалізації ряд необхідно розділити на дві частини з приблизно однаковою кількістю точок та обчислити їх дисперсію (4.6):

$$\sigma_1^2 = \frac{\sum_{t=1}^{n_1}(Y_{t_1} - \overline{Y_1})^2}{n_1 - 1};\ \sigma_2^2 = \frac{\sum_{t=1}^{n_2}(Y_{t_2} - \overline{Y_2})^2}{n_2 - 1}, \quad (4.6)$$

де σ_1^2, σ_2^2 – дисперсії двох частин часового ряду;

Y_{t_1}, Y_{t_2} – фактичні значення двох частин часового ряду;

$\overline{Y_1}$, $\overline{Y_2}$ – середнє значення двох частин часового ряду;

n_1, n_2 – кількість спостережень в 1-й та 2-й частинах часового ряду.

Перевірка гіпотези на однорідність ряду здійснюється за допомогою критерія Фішера (4.7):

$$F = \begin{cases} \sigma_1^2/\sigma_2^2\,,\ \sigma_1^2 > \sigma_2^2 \\ \sigma_2^2/\sigma_1^2\,,\ \sigma_2^2 > \sigma_1^2 \end{cases}, \quad (4.7)$$

де F – розраховане значення критерію Фішера. Якщо його значення менше табличного, визначеного для рівня значущості 0.05 та $(n_1 - 1)$, $(n_2 - 1)$ – ступенів вільності, то приймається гіпотеза про однорідність дисперсій, в протилежному випадку метод не дає відповіді на запитання про наявність чи відсутність тренду.

Перевірка гіпотези щодо відсутності тренду проводиться за допомогою критерію Стьюдента (4.8):

$$t = \frac{|\overline{Y_1} - \overline{Y_2}|}{\sqrt{\frac{(n_1-1)\cdot\sigma_1^2+(n_2-1)\cdot\sigma_2^2}{n_1+n_2-2}}\cdot\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}, \tag{4.8}$$

де t – розраховане значення критерію Ст'юдента. Якщо його значення менше табличного, визначеного для рівня значущості 0.05 та $(n_1 + n_2 - 2)$ – ступенів вільності, то приймається гіпотеза щодо відсутності тренду, в протилежному випадку тренд присутній.

Етап 4. В розділі 4 даного дослідження буде доведено, що аналізовані ряди є нестаціонарні, тому для прогнозування інформаційних тенденцій кіберзлочинів буде обрано моделі експоненційного згладжування.

Проста модель експоненційного згладжування має вигляд (4.9):

$$S_t = \alpha \cdot X_t + (1 - \alpha) \cdot S_{t-1}, \tag{4.9}$$

де S_t, S_{t-1} – експоненційно згладжене значення в момент часу t та $(t - 1)$ відповідно $(t = \overline{1, n})$;

α – параметр згладжування, який приймає значення від нуля (коли ігноруються усі поточні спостереження) до одиниці (коли повністю ігноруються усі попередні спостереження);

X_t – рівень часового ряду в момент часу t.

В даній роботі буде побудовано наступні різновиди моделей експоненційного згладжування:

- адитивна модель циклічності (4.10):

$$S_t = \alpha \cdot X_t + (1-\alpha) \cdot S_{t-1} + I_{t-p}, I_t = I_{t-p} + (1-\alpha) \cdot e_t, \quad (4.10)$$

де I_t, I_{t-p} – згладжений сезонний фактор у момент часу t та $t-p$ (довжина сезону);

e_t – залишки у момент часу t;

- тренд-циклічна адитивна модель з лінійним трендом (4.11):

$$S_t = LT_t + \alpha \cdot X_t + (1-\alpha) \cdot S_{t-1} + I_{t-p}, I_t = I_{t-p} + (1-\alpha) \cdot e_t, \quad (4.11)$$

де LT_t – лінійний тренд (значення в момент часу t);

- тренд-циклічна адитивна модель з експоненційним трендом (4.12):

$$S_t = ET_t + \alpha \cdot X_t + (1-\alpha) \cdot S_{t-1} + I_{t-p}, I_t = I_{t-p} + (1-\alpha) \cdot e_t, \quad (4.12)$$

де ET_t – експоненціальний тренд (значення в момент часу t);

- тренд-циклічна адитивна модель із затухаючим трендом (4.13):

$$S_t = DT_t + \alpha \cdot X_t + (1-\alpha) \cdot S_{t-1} + I_{t-p}, I_t = I_{t-p} + (1-\alpha) \cdot e_t, \quad (4.13)$$

де DT_t – затухаючий тренд (значення в момент часу t);

- мультиплікативна модель циклічності (14):

$$S_t = (\alpha \cdot X_t + (1-\alpha) \cdot S_{t-1}) \cdot I_{t-p}, I_t = I_{t-p} + \delta \cdot (1-\alpha) \cdot e_t / S_t, \quad (4.14)$$

де δ – сезонний параметр параметром згладжування, який зазначається лише для сезонних моделей;

- мультиплікативна тренд-циклічна модель з лінійним трендом (4.15):

$$S_t = LT_t \cdot (\alpha \cdot X_t + (1-\alpha) \cdot S_{t-1}) \cdot I_{t-p}, I_t = I_{t-p} + \delta \cdot (1-\alpha), \quad (4.15)$$

- мультиплікативна тренд-циклічна модель з експоненційним трендом (4.16):

$$S_t = ET_t \cdot (\alpha \cdot X_t + (1-\alpha) \cdot S_{t-1}) \cdot I_{t-p}, I_t = I_{t-p} + \delta \cdot (1-\alpha), \quad (4.16)$$

- мультиплікативна тренд-циклічна модель із затухаючим трендом (4.17):

$$S_t = DT_t \cdot (\alpha \cdot X_t + (1-\alpha) \cdot S_{t-1}) \cdot I_{t-p}, I_t = I_{t-p} + \delta \cdot (1-\alpha). \quad (4.17)$$

Хоча в результаті візуального аналізу початкових даних було доведено, що вони слідують адитивному процесу, буде побудовано також й мультиплікативні моделі експоненційного згладжування для математичного обґрунтування отриманих висновків.

Етап 5. На останньому етапі даного дослідження буде здійснено оцінку точності прогнозів показників «Кібератаки на комп’ютерні системи фінансової установи», «Кібератаки на мережеву інфраструктуру фінансової установи», «Кібератаки на хмарну інфраструктуру фінансової установи», розрахованих за побудованими моделями експоненційного згладжування. Для цього будуть розраховані помилки: «Mean Error», «Mean Absolute Error», «Sums of Squares», «Mean Square», «Mean Percentage Error» та «Mean Absolute Percentage Error».

На першому етапі запропонованої методології прогнозування інформаційних трендів кіберзлочинів було проаналізовано часові ряди на наявність аномальних значень. Для реалізації статистичного методу Z-score було використано мову програмування Python. В результаті для ряду показника «Кібератаки на комп'ютерні системи фінансової установи» виявлено одне аномальне значення, для «Кібератаки на мережеву інфраструктуру фінансової установи» – 5, для «Кібератаки на хмарну інфраструктуру фінансової установи» – 3. Виявлені значення були замінені на середньоарифметичні, узяті для спостережень, що є попереднім та наступним до аномального.

На другому етапі було застосовано QS тест, який було здійснено із використанням мови програмування R. В результаті встановлено, що значення циклічної компоненти для трьох рядів динаміки дорівнює 48, що також підтверджується візуалізацією сезонної складової на рисунках 4.1-4.3.

На третьому етапі побудовано автокореляційні функції часових рядів для проведення їх візуального аналізу на стаціонарність. Результати представлені на рисунках 4.4-4.6.

Аналізуючи отримані графіки, було зроблено попередній висновок, що ряди «Кібератаки на комп'ютерні системи фінансової установи» (рис. 4.4) та «Кібератаки на хмарну інфраструктуру фінансової установи» (рис. 4.6) є нестаціонарними, оскільки автокореляційні коефіцієнти для перших рівнів є статистично значущими. Що стосується ряду «Кібератаки на мережеву інфраструктуру фінансової установи», то однозначно не можна стверджувати, що ряд є стаціонарний чи нестаціонарний, оскільки значення автокореляційної функції для першого рівня дорівнює 0.5, що свідчить тільки про помітний рівень зв'язку і не дозволяє з впевненістю зробити висновок про стаціонарність. Тому було проведено тест різниць середніх рівнів із використанням програмного забезпечення MS Excel, результати якого представлені в таблиці 4.1.

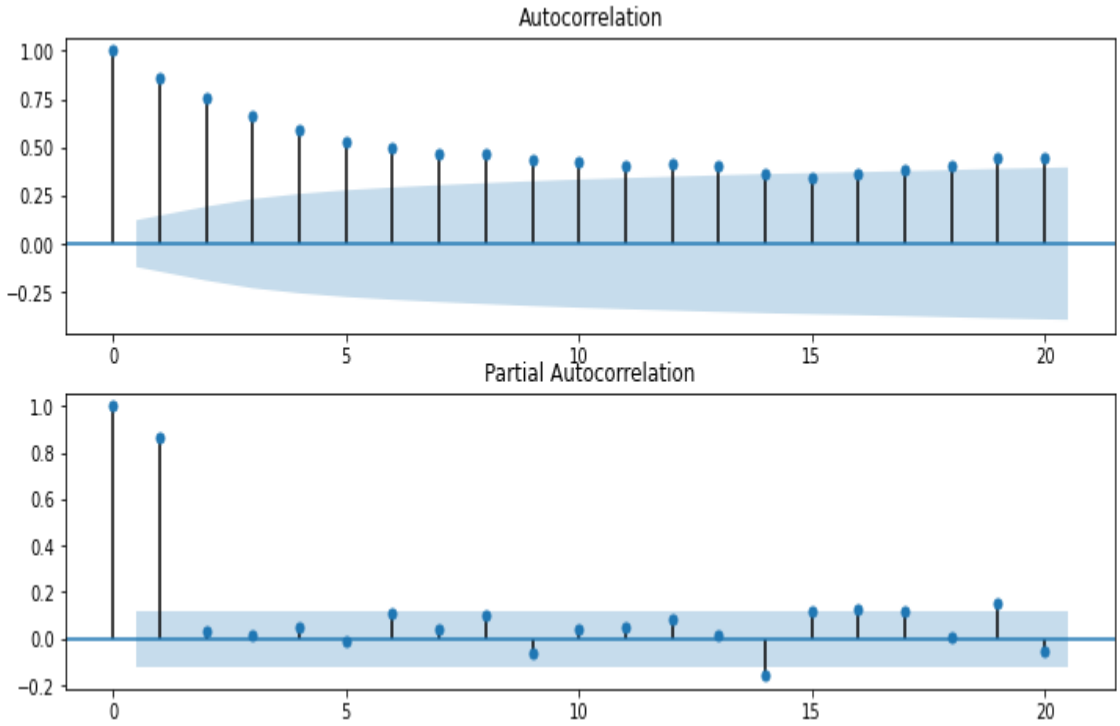


Рисунок 4.4 – Графіки автокореляційної функції та функції часткової автокореляції для показника «Кібератаки на комп'ютерні системи фінансової установи»

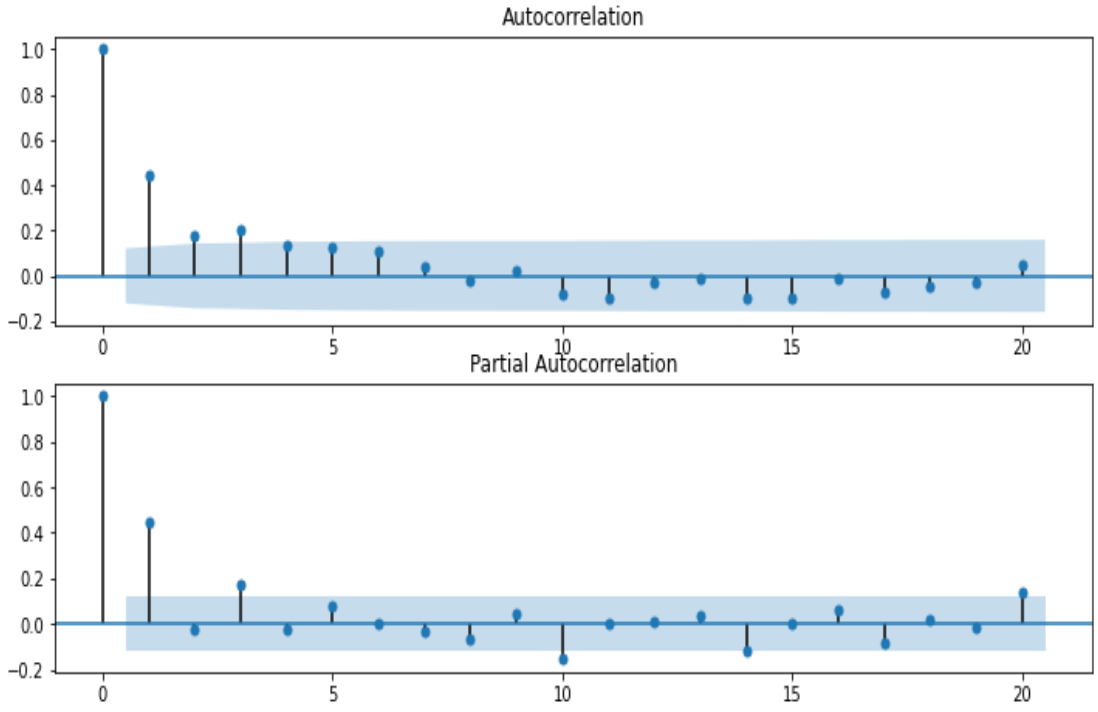


Рисунок 4.5 – Графіки автокореляційної функції та функції часткової автокореляції для показника «Кібератаки на мережеву інфраструктуру фінансової установи»

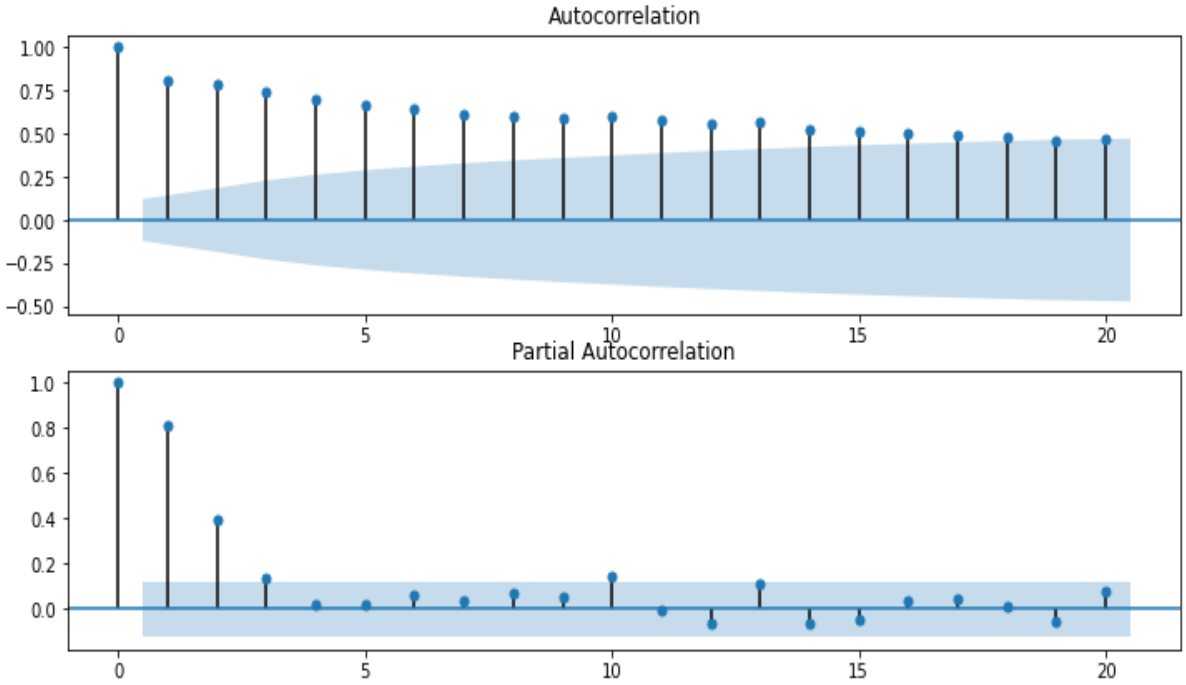


Рисунок 4.6 – Графіки автокореляційної функції та функції часткової автокореляції для показника «Кібератаки на хмарну інфраструктуру фінансової установи»

Таблиця 4.1 – Результати проведеного тесту різниць середніх рівнів

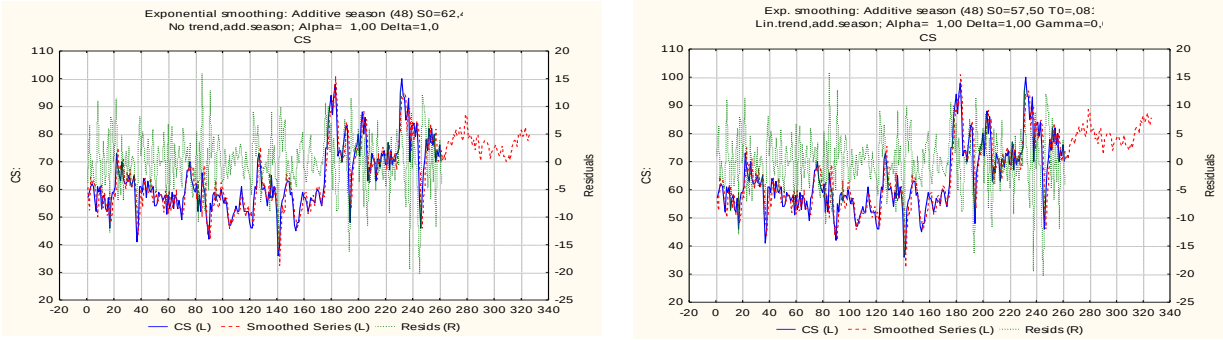
Критерії та висновки	CS	NI	CI
F розрахований	4.6901	1.1489	1.8905
F критичний	1.3374	1.3374	1.3374
Результат перевірки гіпотези на однорідність ряду	Гіпотезу однорідності відхилено	Гіпотезу однорідності прийнято	Гіпотезу однорідності відхилено
t розрахований	9.1187	2.3558	18.5668
t критичний	1.9692	1.9692	1.9692
Результат перевірки гіпотези щодо відсутності тенденції	Тенденція є	Тенденція є	Тенденція є

Результати проведеного тесту показують, що ряди «Кібератаки на комп’ютерні системи фінансової установи» та «Кібератаки на хмарну інфраструктуру фінансової установи» є неоднорідними та містять тренд. Для ряду «Кібератаки на мережеву інфраструктуру фінансової установи» було підтверджено наявність тренду, хоча він й виявився однорідним. Таким чином, можна застосувати клас моделей експоненційного згладжування для даних дослідження.

На четвертому етапі було побудовано моделі експоненційного згладжування для прогнозування інформаційних трендів запитів щодо кібератак на комп'ютерні системи, мережеву та хмарну інфраструктуру фінансової установи. З цією метою було використано інструментарій аналітичного пакету STATISTICA. Результати отриманих прогнозних моделей представлені у табл. 4.2.

Результати виявлених циклічних компонент розглянутих 3 часових рядів представлені в таблиці Е.1 Додатку Е.

Зобразимо результати моделювання на рисунках 4.7–4.9, як співвідношення фактичних, теоретичних та прогнозних рівнів показника «Кібератаки на комп'ютерні системи фінансової установи». Прогнозні значення відображають період з 16.04.2017 по 09.07.2023 рр.

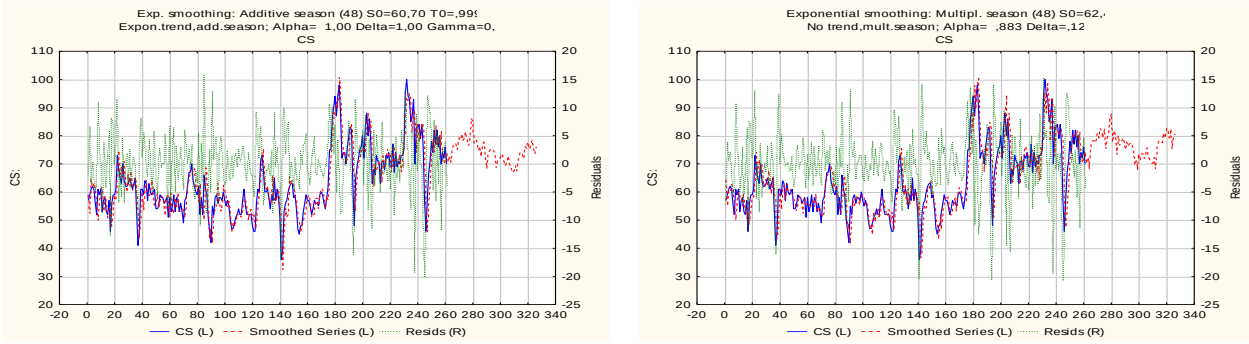


(a) (b)

Рисунок 4.7 – Співвідношення фактичних, теоретичних та прогнозних рівнів показника «Кібератаки на комп'ютерні системи фінансової установи»: (a) Адитивна модель циклічності; (b) Тренд-циклічна адитивна модель з лінійним трендом

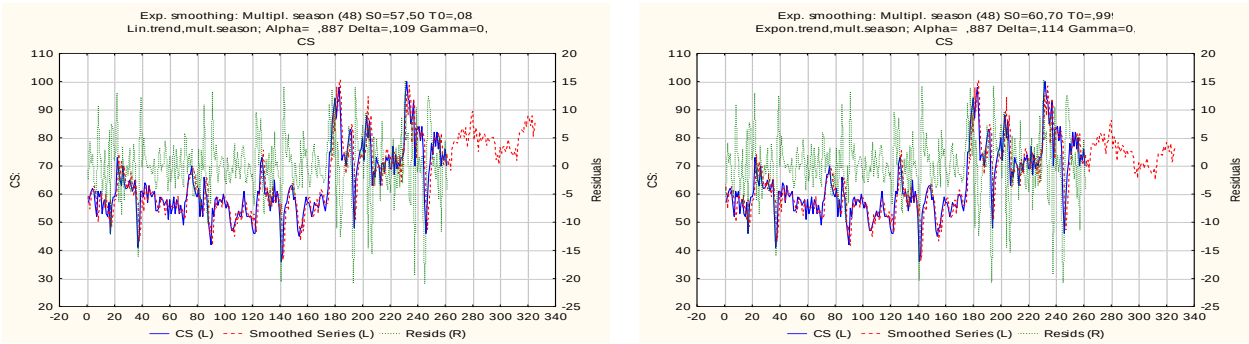
Таблиця 4.2 – Прогнозні моделі експоненціального згладжування інформаційних трендів запитів щодо кібератак на комп'ютерні системи, мережеву та хмарну інфраструктуру фінансової установи

Індикатор	Модель	Властивості моделі
Кібератаки на комп'ютерні системи фінансової установи	Модель 1	Additive season (48); S0=62,42; No trend; Alpha= 1,00; Delta=1,00
	Модель 2	Additive season (48); S0=57,50; T0=0,0815; Linear trend; Alpha= 1,00; Delta=1,00; Gamma=0,00
	Модель 3	Additive season (48); S0=60,70; T0=0,9991; Exponential trend; Alpha= 1,00; Delta=1,00; Gamma=0,00
	Модель 4	Multiplicative season (48); S0=62,42; No trend; Alpha=0,883; Delta=0,125
	Модель 5	Multiplicative season (48); S0=57,50; T0=0,0815; Linear trend; Alpha=0,887; Delta=0,109; Gamma=0,00
	Модель 6	Multiplicative season (48) S0=60,70; T0=0,9991; Exponential trend; Alpha=0,887; Delta=0,114; Gamma=0,00
Кібератаки на мережеву інфраструктуру фінансової установи	Модель 1	Additive season (48); S0=53,90; No trend; Alpha=0,569; Delta=0,00
	Модель 2	Additive season (48); S0=56,88; T0=-0,012; Linear trend; Alpha=0,564; Delta=0,00; Gamma=0,00
	Модель 3	Additive season (48); S0=58,86; T0=0,9984; Exponential trend; Alpha=0,573; Delta=0,00; Gamma=0,00
	Модель 4	Additive season (48); S0=74,05; T0=-0,728; Damped trend; Alpha=0,361; Delta=0,00; Phi=0,017
	Модель 5	Multiplicative season (48); S0=53,90; No trend; Alpha=0,518; Delta=0,00
	Модель 6	Multiplicative season (48); S0=56,88; T0=-0,012; Linear trend; Alpha=0,518; Delta=0,00; Gamma=0,00
	Модель 7	Multiplicative season (48); S0=58,86; T0=0,9984; Exponential trend; Alpha=0,527; Delta=0,00; Gamma=0,00
	Модель 8	Multiplicative season (48); S0=71,43; T0=-0,618; Damped trend; Alpha=0,328; Delta=0,00; Phi=0,020
Кібератаки на хмарну інфраструктуру фінансової установи	Модель 1	Additive season (48); S0=33,73; No trend; Alpha=0,763; Delta=0,00
	Модель 2	Additive season (48); S0=25,94; T0=0,0667; Linear trend; Alpha=0,756; Delta=0,00; Gamma=0,00
	Модель 3	Additive season (48); S0=27,21; T0=1,000; Exponential trend; Alpha=0,761; Delta=0,00; Gamma=0,00
	Модель 4	Multiplicative season (48); S0=33,73; No trend; Alpha=1,00; Delta=1,00
	Модель 5	Multiplicative season (48); S0=25,94; T0=0,0667; Linear trend; Alpha=1,00; Delta=1,00; Gamma=0,00
	Модель 6	Multiplicative season (48); S0=27,21; T0=1,000; Exponential trend; Alpha=0,815; Delta=0,00; Gamma=0,00



(a) (b)

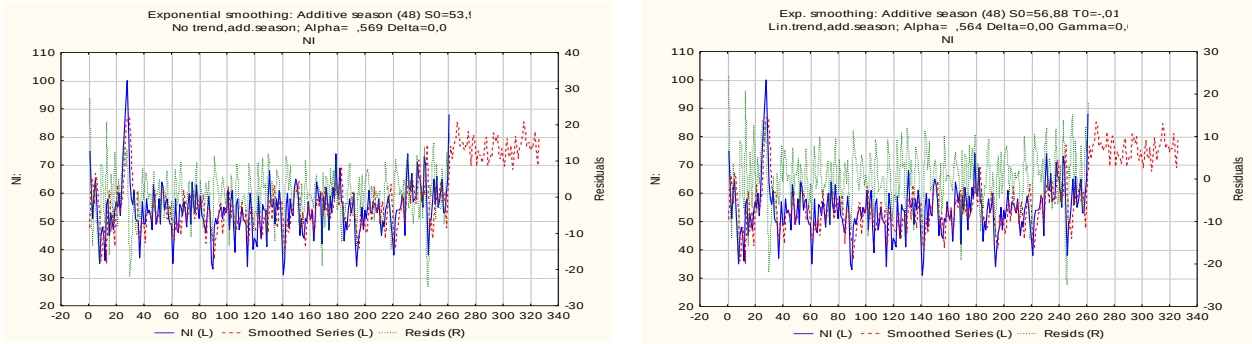
Рисунок 4.8 – Співвідношення фактичних, теоретичних та прогнозних рівнів показника «Кібератаки на комп'ютерні системи фінансової установи»: (a) Тренд-циклічна адитивна модель з експоненційним трендом; (b) Мультиплікативна модель циклічності



(a) (b)

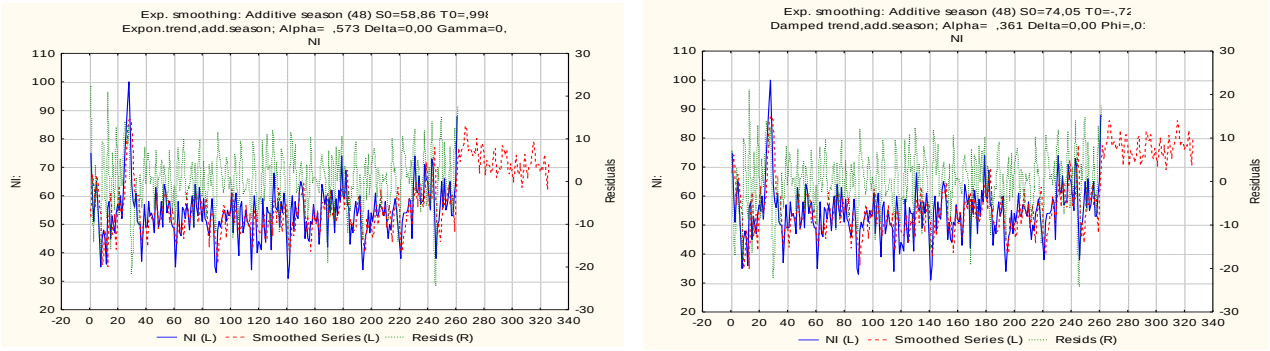
Рисунок 4.9 – Співвідношення фактичних, теоретичних та прогнозних рівнів показника «Кібератаки на комп'ютерні системи фінансової установи»: (a) Мультиплікативна тренд-циклічна модель з лінійним трендом; (b) Мультиплікативна тренд-циклічна модель з експоненційним трендом

Представимо результати експоненційного моделювання на рисунках 4.10–4.13, як співвідношення фактичних, теоретичних та прогнозних рівнів показника «Кібератаки на мережеву інфраструктуру фінансової установи». Прогнозні значення відображають період з 16.04.2017 по 09.07.2023 рр.



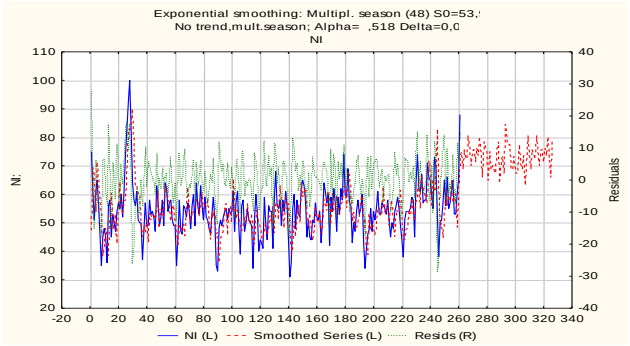
(a) (b)

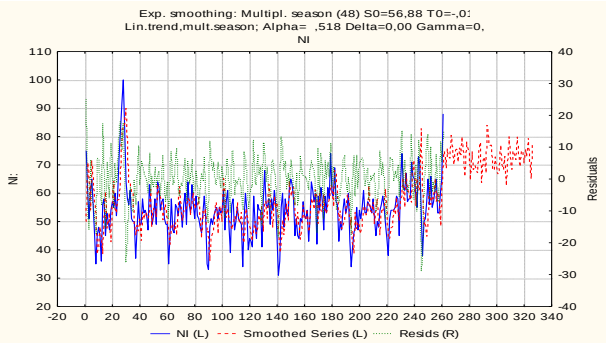
Рисунок 4.10 – Співвідношення фактичних, теоретичних та прогнозних рівнів показника «Кібератаки на мережеву інфраструктуру фінансової установи»: (a) Адитивна модель циклічності; (b) Тренд-циклічна адитивна модель з лінійним трендом



(a) (b)

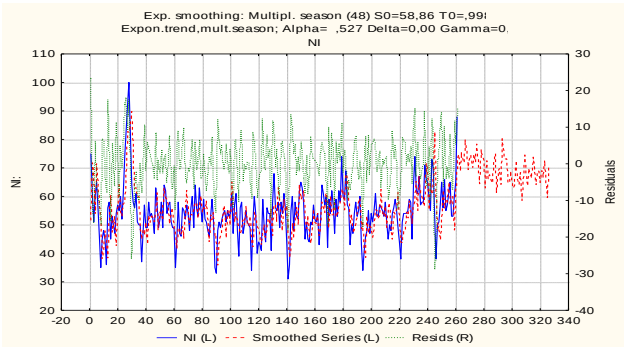
Рисунок 4.11 – Співвідношення фактичних, теоретичних та прогнозних рівнів показника «Кібератаки на мережеву інфраструктуру фінансової установи»: (a) Тренд-циклічна адитивна модель з експоненційним трендом; (b) Тренд-циклічна адитивна модель з затухаючим трендом

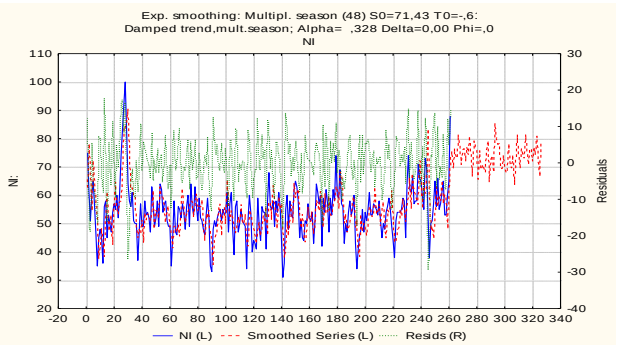




(a) (b)

Рисунок 4.12 – Співвідношення фактичних, теоретичних та прогнозних рівнів показника «Кібератаки на мережеву інфраструктуру фінансової установи»: (a) Мультиплікативна модель циклічності; (b) Мультиплікативна тренд-циклічна модель з лінійним трендом

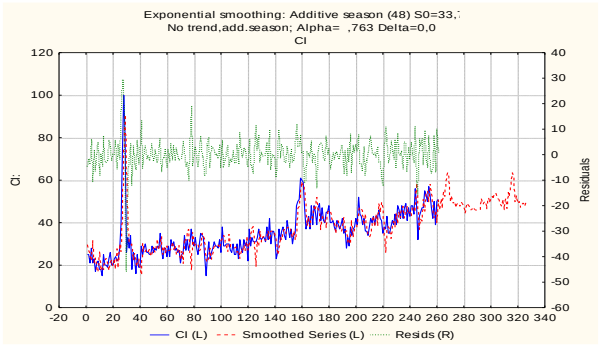


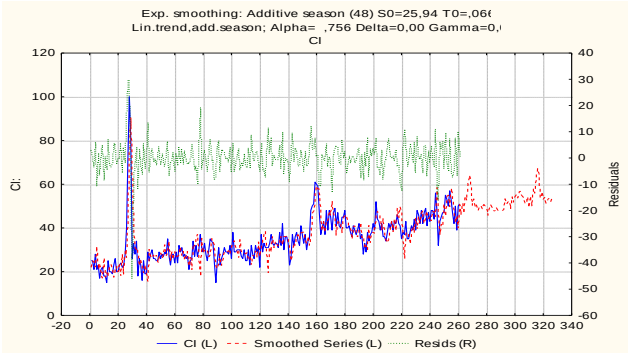


(a) (b)

Рисунок 4.13 – Співвідношення фактичних, теоретичних та прогнозних рівнів показника «Кібератаки на мережеву інфраструктуру фінансової установи»: (a) Мультиплікативна тренд-циклічна модель з експоненційним трендом; (b) Мультиплікативна тренд-циклічна модель з затухаючим трендом

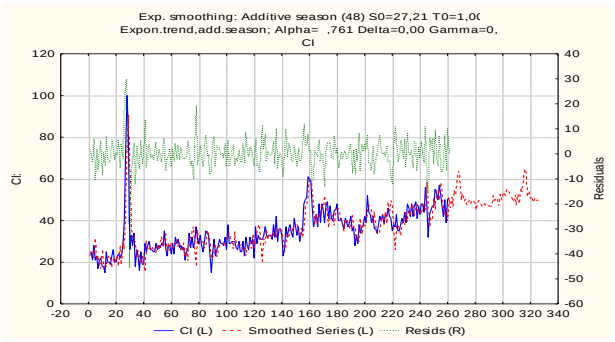
На рисунках 4.14-4.16 представлені результати моделювання співвідношення фактичних, теоретичних та прогнозних рівнів показника «Кібератаки на хмарну інфраструктуру фінансової установи». Прогнозні значення відображають період з 16.04.2017 по 09.07.2023 рр.

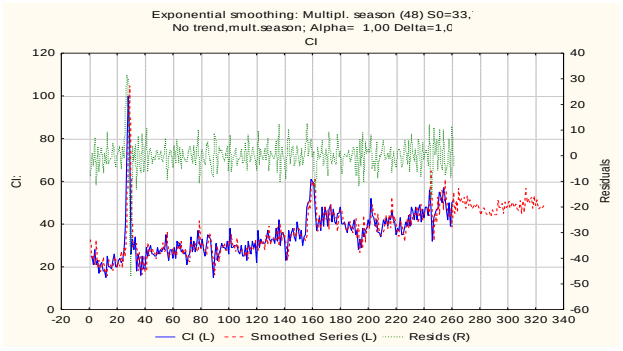




(a) (b)

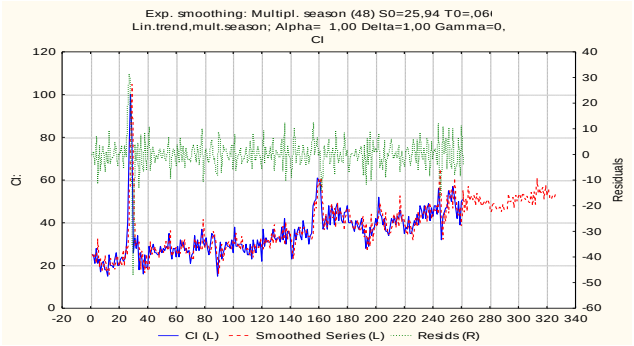
Рисунок 4.14 – Співвідношення фактичних, теоретичних та прогнозних рівнів показника «Кібератаки на хмарну інфраструктуру фінансової установи»: (a) Адитивна модель циклічності; (b) Тренд-циклічна адитивна модель з лінійним трендом



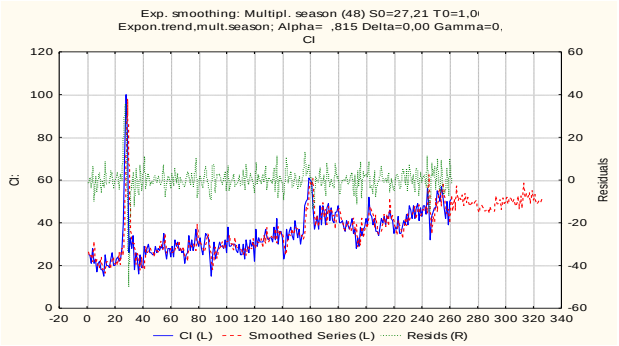


(a) (b)

Рисунок 4.15 – Співвідношення фактичних, теоретичних та прогнозних рівнів показника «Кібератаки на хмарну інфраструктуру фінансової установи»: (a) Тренд-циклічна адитивна модель з експоненційним трендом; (b) Мультиплікативна модель циклічності



(a)



(b)

Рисунок 4.16 – Співвідношення фактичних, теоретичних та прогнозних рівнів показника «Кібератаки на хмарну інфраструктуру фінансової установи»: (a) Мультиплікативна тренд-циклічна модель з лінійним трендом; (b) Мультиплікативна тренд-циклічна модель з експоненційним трендом

Розраховані прогнозні значення показників кібератак на комп'ютерні системи, мережеву та хмарну інфраструктуру фінансової установи за період з 17.04.2022 по 09.07.2023 систематизуємо у вигляді таблиці Е.2 та представимо у Додатку Е. Розроблення моделі прогнозування кібератак на комп'ютерні системи, мережеву та хмарну інфраструктуру фінансової установи вимагало проведення перевірки точності обчислених прогнозних рівнів. Тому на п'ятому етапі проведено аналіз наступного переліку показників: «Mean Error», «Mean Absolute Error», «Sums of Squares», «Mean Square», «Mean Percentage Error», «Mean Absolute Percentage Error» (таблиці 4.3 – 4.5 відповідно для трьох розглянутих напрямків кібершахрайських атак).

Таблиця 4.3 – Показники точності прогнозів показника «Кібератаки на комп'ютерні системи фінансової установи»

Назва помилки	Модель 1	Модель 2	Модель 3	Модель 4	Модель 5	Модель 6
Mean Error	0,0539	-0,0088*	0,1152	0,0229	-0,0472	0,0917
Mean Absolute Error	4,3026	4,2923*	4,2948	4,3488	4,3289	4,3471
Sums of Squares	8174,6881	8160,2652	8160,2379*	9340,3966	9300,4093	9312,8601
Mean Square	31,3206	31,2654	31,2653*	35,7870	35,6338	35,6815
Mean Percentage Error	-0,3187	-0,4186	-0,2200*	-0,4462	-0,5556	-0,3336
Mean Absolute Percentage Error	6,9939	6,9803	6,9776*	7,0286	6,9983	7,0197

* Найменші значення похибок виділені сірим кольором

Таблиця 4.4 – Показники точності прогнозів показника «Кібератаки на мережеву інфраструктуру фінансової установи»

Назва помилки	Модель 1	Модель 2	Модель 3	Модель 4	Модель 5	Модель 6	Модель 7	Модель 8
Mean Error	0,1481	0,1503	0,2695	0,0162*	0,0476	0,0507	0,1810	-0,0699
Mean Absolute Error	5,9178	5,9115	5,9130	5,8966*	5,9721	5,9706	5,9709	5,9698
Sums of Squares	14991,5	14869,8	14784,5	14545,7*	15997,2	15907,2	15828,7	15713,2
Mean Square	57,4386	56,9725	56,6455	55,7306*	61,2920	60,9472	60,6463	60,2038
Mean Percentage Error	-1,1104	-1,1033	-0,8663*	-1,2791	-1,3870	-1,3739	-1,1162	-1,5270
Mean Absolute Percentage Error	11,3017	11,2936	11,2782*	11,2887	11,3923	11,3905	11,3712	11,4027

* Найменші значення похибок виділені сірим кольором

Таблиця 4.5 – Показники точності прогнозів показника «Кібератаки на хмарну інфраструктуру фінансової установи»

Назва помилки	Модель1	Модель2	Модель3	Модель4	Модель5	Модель6
Mean Error	0,0812	0,0332*	0,0915	0,0430	0,0052*	0,0692
Mean Absolute Error	4,4419	4,4221	4,4229	4,6343	4,6063	4,3387*
Sums of Squares	10845,971	10828,033	10827,360*	11894,42	11833,363	11525,863
Mean Square	41,5554	41,4867	41,4841*	45,5725	45,3386	44,1604
Mean Percentage Error	-1,6617	-1,7840	-1,5980*	-1,6119	-1,7041	-1,6788
Mean Absolute Percentage Error	13,4378	13,3832	13,3663	13,9018	13,8019	12,9143*

* Найменші значення похибок виділені сірим кольором

Аналіз розрахованих показників точності дозволив обрати відповідні моделі для досліджуваних часових рядів. Для ряду «Кібератаки на комп'ютерні системи фінансової установи» найбільш точною за більшою кількістю показників виявилася модель 3 (табл. 4.3) – тренд-циклічна адитивна модель з експоненційним трендом. Тренд-циклічна адитивна модель з затухаючим трендом є точною для ряду «Кібератаки на мережеву інфраструктуру фінансової установи» (табл. 4.4). Тренд-циклічна адитивна модель з експоненційним трендом показала найкращі результати для «Кібератак на хмарну інфраструктуру фінансової установи» (табл. 4.5). Отримані результати також підтвердили, що досліджувані ряди слідують адитивному процесу та мають трендову і сезонну складові.

Досліджувана тема щодо прогнозування інформаційних трендів кіберзлочинів набуває актуальності у зв’язку із стрімким зростанням їх рівня за останнє десятиліття. Наслідки кіберзлочинності відчуваються по всьому світу, що пов’язано із збільшенням фінансових втрат від викрадення, втрати та відновлення персональної інформації, даних суб’єктів господарювання, урядових організації, тощо. Особливо відчутною дана проблема є в умовах війн та світових пандемій, оскільки вони формують сприятливі умови для кіберзлочинців та кібершахраїв. Саме тому їх попередження та завчасне виявлення є стратегічним завданням у боротьбі з цим явищем.

В роботі встановлено, що наукова спільнота активно досліджує проблематику кіберзлочинності. Вони приділяють увагу макроекономічним проблемам, а саме її впливу на макроекономічну стабільність, інноваційні можливості країни, її імідж, а також зростання тіньового сектору. Також науковці досліджують вплив інформаційних технологій на розвиток бізнесу, питання реорганізації бізнес-процесів в умовах впровадження хмарних технологій, умови зростання кіберризиків та заходи організації кібербезпеки. Актуальним є науковий напрямок, пов'язаний із питаннями кіберзлочинів по відношенню до користувачів інформаційних систем та комп’ютерних технологій, які можуть відбуватися через соціальні мережі, мобільні та Інтернет додатки. Досліджуються психологічні причини кіберзлочинів, мотивація злочинців, та інші фактори.

В роботі запропоновано методологію дослідження, яка передбачала дослідження початкового набору даних на наявність аномальних спостережень за допомогою Z-score, проведення QS-тест для виявлення періоду циклічності рядів, застосування тесту різниць середніх рівнів для доведення гіпотези щодо відсутності тренду, моделювання і прогнозування рядів динаміки на основі методу експоненційного згладжування та побудови адитивних та мультиплікативних моделей циклічності, тренд-циклічних з лінійним, експоненційним та затухаючим трендами, а також проведення оцінки якості побудованих моделей. Вхідними даними було обрано інформаційні тренди

запитів Google-користувачів щодо кібератак на комп’ютерні системи, мережеву та хмарну інфраструктуру фінансової установи. Вибір даних був пов'язаний із тими міркуваннями, що в Інтернет-мережі реакція на будь-яку подію є швидшою ніж у практичній діяльності, тому відповідне зростання запитів користувачів ідентифікується як відклик на кіберзлочини.

В результаті проведення декомпозиції обраних часових рядів встановлено, що вони слідують адитивному процесу, мають сезонну та трендову компоненти. Проведення аналізу рядів на предмет наявності аномальних спостережень дозволило встановити, що інформаційний тренд запитів щодо кібератак на комп’ютерні системи містить одне аномальне спостереження, тренд запитів щодо кібератак на мережеву інфраструктуру – 5, на хмарну інфраструктуру – 3. Їх значення були замінені на середньоарифметичні значення спостережень, що передують та слідують ним. Проведення QS-тест визначило, що період циклічності дорівнює 48 для всіх трьох рядів, що також підтверджується візуалізацією їх сезонної компоненти. За результатами тесту перевірки різниць середніх рівнів було виявлено, що інформаційні тренди запитів щодо кібератак на комп’ютерні системи та хмарну інфраструктуру мають неоднорідні дисперсії, а ряд запитів щодо кібератак на мережеву інфраструктуру має однорідні. Але дослідження значень критерія Стьюдента дозволило встановити, що ряди нестаціонарні та мають трендову складову, тому для їх моделювання та прогнозування можна використовувати моделі експоненційного згладжування. В результаті їх побудови та проведення оцінки якості визначено, що адитивна тренд-циклічна модель з експоненційним трендом гарно моделює та прогнозує ряди запитів щодо кібератак на комп’ютерні системи та хмарну інфраструктуру, а адитивна тренд-циклічна модель із затухаючим трендом – ряд запитів щодо кібератак на мережеву інфраструктуру.

Запропоновану в роботі методологію та результати прогнозування інформаційних трендів запитів користувачів щодо кіберзлочинів доцільно використовувати для удосконалення стратегії боротьби із кіберзлочинами як на рівні держави, так і на рівні фінансових установ. Дослідження подібних

тенденцій та їх прогнозування дозволить у майбутньому попереджати масові кібератаки, які сьогодні є поширеними в рамках ведення кібервійн та кібертерроризму. Отримання подібних прогнозів сприятиме активізації заходів кібербезпеки на прогнозовані періоди у більш активному режимі, а також швидше реагувати в ситуаціях, пов'язаних із кібератаками на різні об'єкти комп'ютерної та мережевої інфраструктури.

Пункт 4.1 було виконано із використанням матеріалів публікацій виконавців [128].

4.2 Розроблення ударно-хвильової моделі впливу кібершахрайських атак на рівень фінансової безпеки

Статистичну базу даних дослідження кібершахрайських атак на комп'ютерні системи, мережеву та хмарну інфраструктуру фінансової установи, а також рівня фінансової безпеки сформуємо на основі застосування інструментарію Google Trends. В якості індикаторів розглянемо кількість запитів інтернет-користувачів до доданих понять за період 16.04.2017 по 10.04.2022 в розрізі потижневих рівнів. Для проведення дескриптивного аналізу отриманих часових рядів, побудуємо графіки (рисунки 4.17-4.20) за допомогою програми Statistica.

Для опису інформаційних війн були обрані кількісні критерії: кількість запитів інтернет-користувачів до понять кібершахрайські атаки на комп'ютерні системи, мережеву та хмарну інфраструктуру фінансової установи. На основі рівнів обраних для дослідження часових рядів за допомогою щотижневих даних за проміжок часу з 16.04.2017 по 10.04.2022 були ідентифіковані «інформаційні бульбашки» та прояви їх розриву: computer system – 10.09.2017, 23.09.2018, 15.09.2019, 11.10.2020, 19.09.2021 (рисунок 4.18), cyber fraud – 26.11.2017, 25.11.2018, 01.12.2019, 12.07.2020, 09.01.2022 (рисунок 4.17), network infrastructure – 22.10.2017, 13.10.2019, 13.09.2020, 12.09.2021, 10.04.2022 (рисунок 4.19), cloud infrastructure – 22.10.2017, 26.04.2020, 21.02.2021, 12.12.2021 (рисунок 4.20).

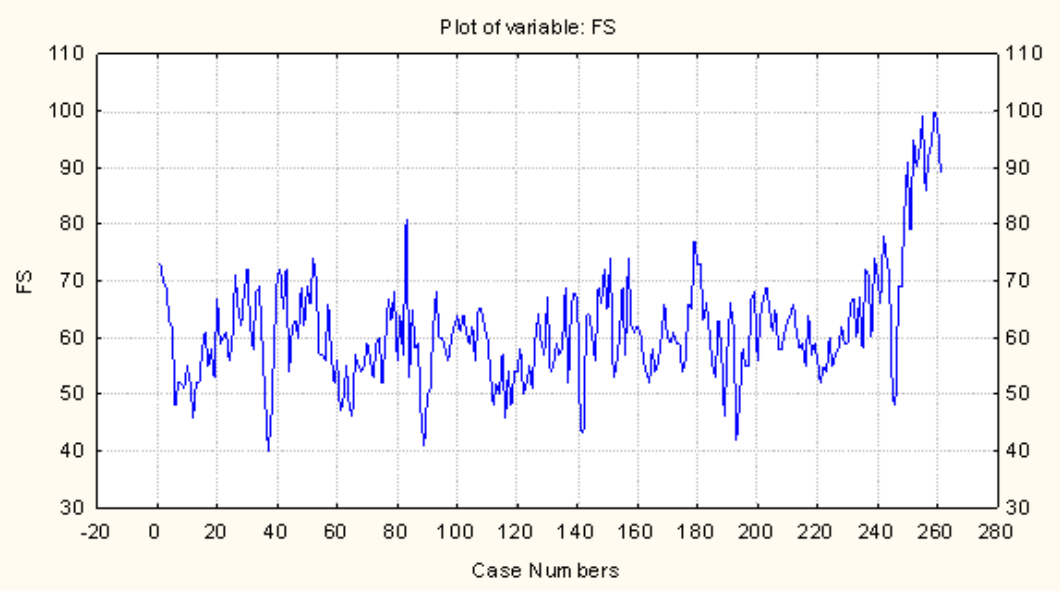


Рисунок 4.17 – Потижнева динаміка показника кількості запитів інтернет-користувачів до поняття фінансова безпека

Кожна із виділених дат виступає проявом розриву «інформаційної бульбашки», коли зазначені кількісні індикатори характеризують варіацію «маси», що має стрибкоподібний характер та визначає «енергію» відповідної ударної хвилі.

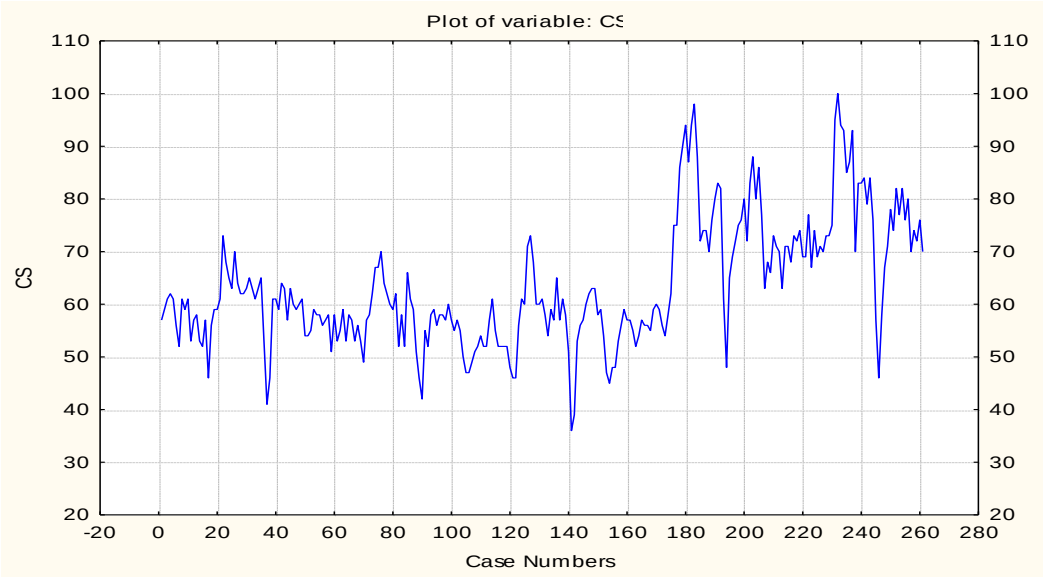


Рисунок 4.18 – Потижнева динаміка показника кількості запитів інтернет-користувачів до поняття кібернетичних атак на комп’ютерні системи фінансової установи

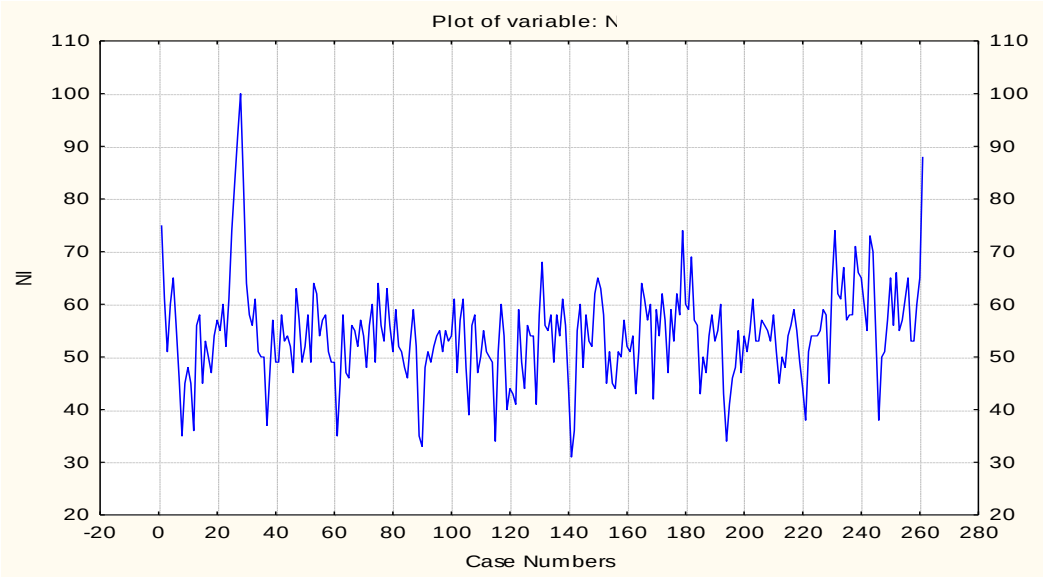


Рисунок 4.19 – Потижнева динаміка показника кількості запитів інтернет-користувачів до поняття кібернетичних атак на мережеву інфраструктуру фінансової установи

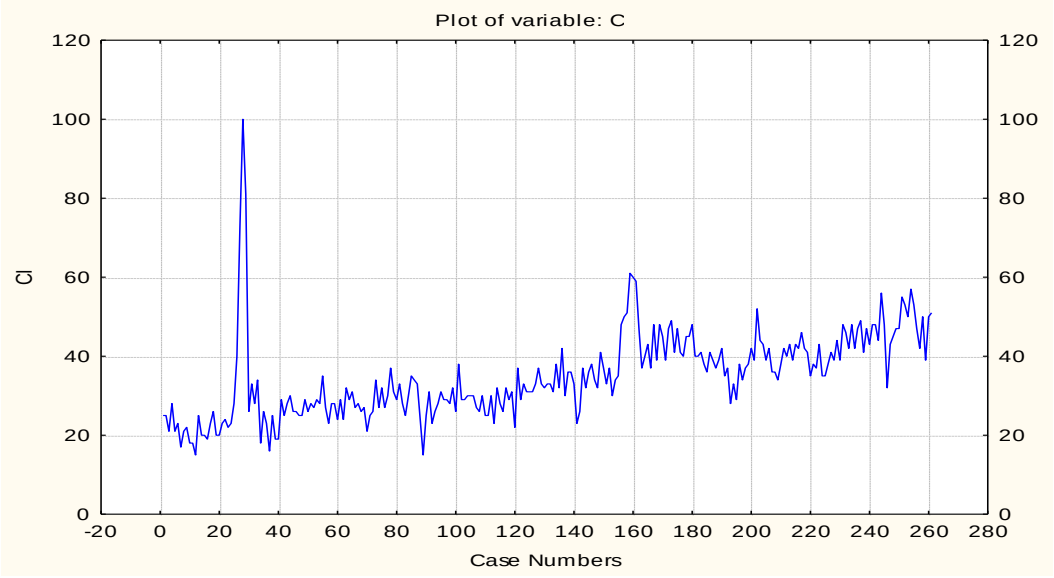


Рисунок 4.20 – Потижнева динаміка показника кількості запитів інтернет-користувачів до поняття кібернетичних атак на хмарну інфраструктуру фінансової установи

На рис. 4.17 – 4.20 представлена динаміка показників фінансової бузпеки та кібершахрайських атак на комп’ютерні системи, мережеву та хмарну

інфраструктуру фінансової установи, де зазначені вище дати чітко в розрізі кожного часового ряду ідентифікуються як аномальні рівні у вигляді певного несподіваного стрімкого стрибка з подальшим поверненням до попереднього рівня часового ряду.

Розглянемо теоретичну сутність моделі, яка виступає основною формалізації залежності фінансової безпеки від кібершахрайських атак на комп'ютерні системи, мережеву та хмарну інфраструктуру фінансової установи. Так, модель Сєдова-Тейлора для опису моделі ударної хвилі розповсюдження наслідків «інформаційних війн» (кібершахрайських атак на комп'ютерні системи, мережеву та хмарну інфраструктуру фінансової установи) на фінансову безпеку набуває вигляду:

$$\Delta F \quad (t) = \frac{C}{t^4} + k_1 \left(\frac{C}{t^4}\right)^{3/4} + k_2 \left(\frac{C}{t^4}\right)^{2/4} + k_3 \left(\frac{C}{t^4}\right)^{1/4} \qquad (4.18)$$

де C – енергія в початковий момент після розриву бульбашки, що в нашому випадку пропонується інтерпретувати як показник «інформаційних війн» (кібершахрайських атак на систем/мережевої/хмарної інфраструктури фінансової установи);

$\Delta F \quad (t)$ – як абсолютний приріст показника фінансової безпеки за період t;

t – період часу після розриву «інформаційної бульбашки» (кількість тижнів);

k_1, k_2, k_3 – характеристики середовищ поширення ударних хвиль наслідків «інформаційних війн» (впливу кібершахрайських атак на фінансову безпеку).

Нелінійна модель Сєдова-Тейлора як задача оптимізації нелінійного програмування набуває наступного вигляду:

$$\begin{cases} K_i \to min \\ K_i = \sum_{j=1}^{V} K_{ij}^{t_j} = \\ = \sum_{j=1}^{A} \left(\frac{e_{t_j}^i}{\tau_j^i} + a_1 \left(\frac{e_{t_j}^i}{\left(\tau_j^i\right)^2} \right)^{3/4} + a_2 \left(\frac{e_{t_j}^i}{\left(\tau_j^i\right)^3} \right)^{2/4} + a_3 \left(\frac{e_{t_j}^i}{\left(\tau_j^i\right)^4} \right)^{1/4} \right) \end{cases} \quad (4.19)$$

де K_i – розсіювання енергії ударної хвилі в і-му каналі поширення кібернетичних атак;

A – кількість інформаційних війн (кібернетичних атак);

$e_{t_j}^i$ – значення початкової енергії в момент часу t_j;

t_j – момент часу початку j-ої інформаційної війни в і-му напрямку здійснення кібернетичних атак;

τ_j^i - тривалість j-ої інформаційної війни в і-му напрямку здійснення кібернетичних атак.

Переходячи до практичного впровадження моделі Сєдова-Тейлора на прикладі залежності фінансової безпеки від кібершахрайських атак на комп’ютерні системи, мережеву та хмарну інфраструктуру фінансової установи, зазначимо, що починаючи із ідентифікованих дат розриву «інформаційної бульбашки» в розрізі розглянутих індикаторів ударна хвиля буде розповсюджуватися, однак може не дійти до фінансової установи і не завдати значних негативних наслідків чи дійшовши не призвести до суттєвого впливу. Початкові значення енергій розривів у моменти кризових явищ відображені в табл. 4.6-4.8 для кібершахрайських атак на комп’ютерні системи, мережеву та хмарну інфраструктуру фінансової установи.

Таблиця 4.6 – Початкові енергії розривів у моменти криз для показника computer cyber fraud (computer system)

data	financial security	cyber fraud (computer system)	ΔF	C/t4	(C/t4)^3/4	(C/t4)^2/4	(C/t4)^1/4
10.09.2017	60	73	1	0,0304	0,0728	0,1744	0,4176
23.09.2018	59	70	7	0,0292	0,0706	0,1707	0,4132
15.09.2019	64	73	4	0,0304	0,0728	0,1744	0,4176
11.10.2020	66	98	3	0,0408	0,0908	0,2020	0,4495
19.09.2021	67	100	1	0,0416	0,0922	0,2041	0,4518

Таблиця 4.7 – Початкові енергії розривів у моменти криз для показника cyber fraud (network infrastructure)

data	financial security	cyber fraud (network infrastructure)	ΔF	C/t4	(C/t4)^3/4	(C/t4)^2/4	(C/t4)^1/4
22.10.2017	62	100	-3	0,0416	0,0922	0,2041	0,4518
13.10.2019	54	68	-13	0,0283	0,0690	0,1683	0,4102
13.09.2020	77	74	12	0,0308	0,0736	0,1756	0,4190
12.09.2021	66	74	7	0,0308	0,0736	0,1756	0,4190
10.04.2022	89	88	-9	0,0367	0,0838	0,1914	0,4375

Таблиця 4.8 – Початкові енергії розривів у моменти криз для показника cyber fraud (cloud infrastructure)

data	financial security	cyber fraud (cloud infrastructure)	ΔF	C/t4	(C/t4)^3/4	(C/t4)^2/4	(C/t4)^1/4
22.10.2017	62	100	-3	0,0416	0,0922	0,2041	0,4518
26.04.2020	61	61	-1	0,0254	0,0636	0,1594	0,3992
21.02.2021	67	52	3	0,0217	0,0565	0,1472	0,3836
12.12.2021	71	56	-3	0,0233	0,0597	0,1527	0,3908

З метою побудови моделі Сєдова-Тейлора формалізації ударної хвилі наслідків розповсюдження інформаційних війн в процесі здійснення впливу кібершахрайських атак на комп’ютерні системи, мережеву та хмарну інфраструктуру фінансової установи на рівень фінансової безпеки виникає необхідність ідентифікації проміжку часу після моменту розриву «інформаційної бульбашки» (кількість тижнів) щодо її поширення. З метою розв’язання зазначеного питання необхідно провести автокореляційний аналіз

(рисунки 21-27), який дозволяє сформувати графіки корелограм нульових різниць для розглянутих часових рядів як фінансової безпеки, так і кібершахрайських атак на комп'ютерні системи, мережеву та хмарну інфраструктуру фінансової установи (рисунки 22-28).

Autocorrelation Function (Spreadsheet1 FS.
FS
(Standard errors are white-noise estimates)

Lag	Auto-Corr.	Std.Err.	Box & Ljung Q	p
1	0,68689	0,06154	124,567	0,00000
2	0,56569	0,06142	209,379	0,00000
3	0,45783	0,06130	265,149	0,00000
4	0,40837	0,06118	309,693	0,00000
5	0,33245	0,06106	339,329	0,00000
6	0,30150	0,06095	363,800	0,00000
7	0,27804	0,06083	384,693	0,00000
8	0,23711	0,06071	399,947	0,00000
9	0,16269	0,06059	407,157	0,00000
10	0,16454	0,06047	414,561	0,00000
11	0,13572	0,06034	419,619	0,00000
12	0,07335	0,06022	421,102	0,00000
13	0,03699	0,06010	421,481	0,00000
14	-0,01019	0,05998	421,510	0,00000
15	0,00102	0,05986	421,510	0,00000

Рисунок 4.21 – Значення автокорляційної функції нульових різниць часового ряду кібершахрайських атак

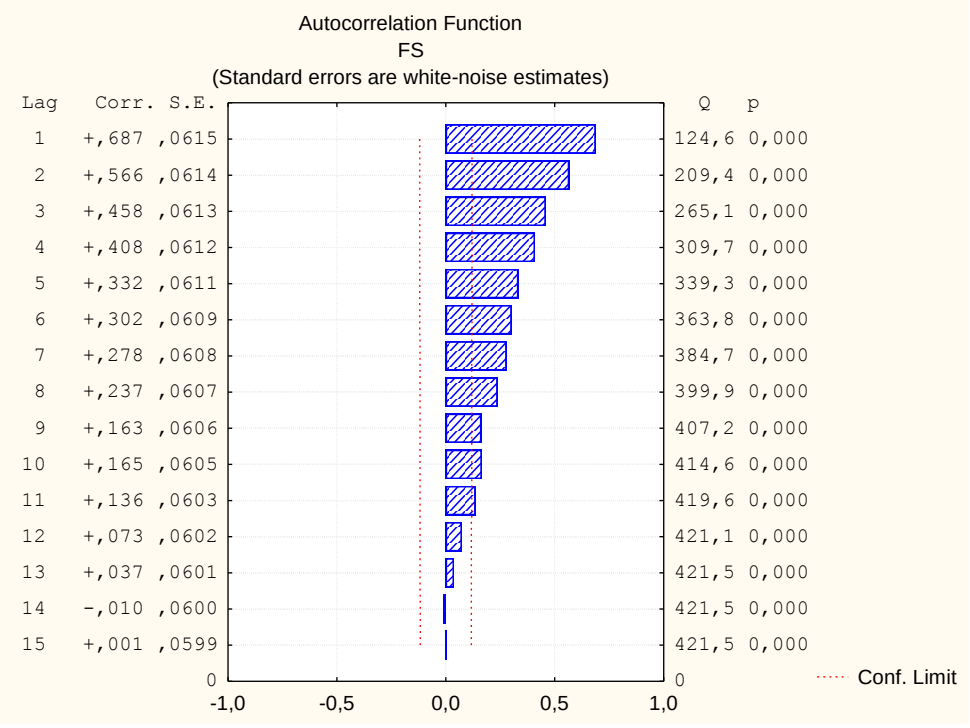


Рисунок 4.22 – Корелограма нульових різниць часового ряду кібершахрайських атак

Lag	Autocorrelation Function (Spreadsheet3.: CS (Standard errors are white-noise estimate Auto-Corr.	Std.Err.	Box & Ljung Q	p
1	0,86421	0,06154	197,18	0,00
2	0,75191	0,06142	347,02	0,00
3	0,65812	0,06130	462,26	0,00
4	0,58955	0,06118	555,09	0,00
5	0,52455	0,06106	628,87	0,00
6	0,48948	0,06095	693,37	0,00
7	0,46760	0,06083	752,46	0,00
8	0,46503	0,06071	811,13	0,00
9	0,43636	0,06059	863,00	0,00
10	0,41992	0,06047	911,22	0,00
11	0,40700	0,06034	956,71	0,00
12	0,40872	0,06022	1002,76	0,00
13	0,39974	0,06010	1046,99	0,00
14	0,35372	0,05998	1081,76	0,00
15	0,34255	0,05986	1114,50	0,00

Рисунок 4.23 – Значення автокорляційної функції нульових різниць часового ряду кібершахрайських атак на комп’ютерні системи

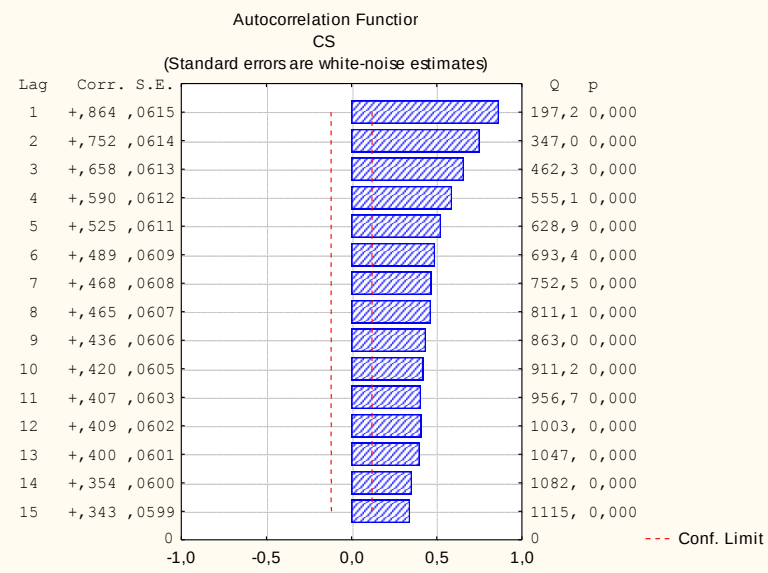


Рисунок 4.24 – Корелограма нульових різниць часового ряду кібершахрайських атак на комп’ютерні системи

Lag	Autocorrelation Function (Spreadsheet3.: NI (Standard errors are white-noise estimate Auto-Corr.	Std.Err.	Box & Ljung Q	p
1	0,53193	0,06154	74,703	0,00000
2	0,26924	0,06142	93,916	0,00000
3	0,24074	0,06130	109,336	0,00000
4	0,14934	0,06118	115,293	0,00000
5	0,11865	0,06106	119,068	0,00000
6	0,09744	0,06095	121,624	0,00000
7	0,03297	0,06083	121,918	0,00000
8	-0,03992	0,06071	122,350	0,00000
9	-0,02590	0,06059	122,533	0,00000
10	-0,10795	0,06047	125,721	0,00000
11	-0,11355	0,06034	129,261	0,00000
12	-0,04596	0,06022	129,843	0,00000
13	-0,02616	0,06010	130,033	0,00000
14	-0,09446	0,05998	132,512	0,00000
15	-0,10024	0,05986	135,317	0,00000

Рисунок 4.25 – Значення автокорляційної функції нульових різниць часового ряду кібершахрайських атак на мережеву інфраструктуру фінансової установи

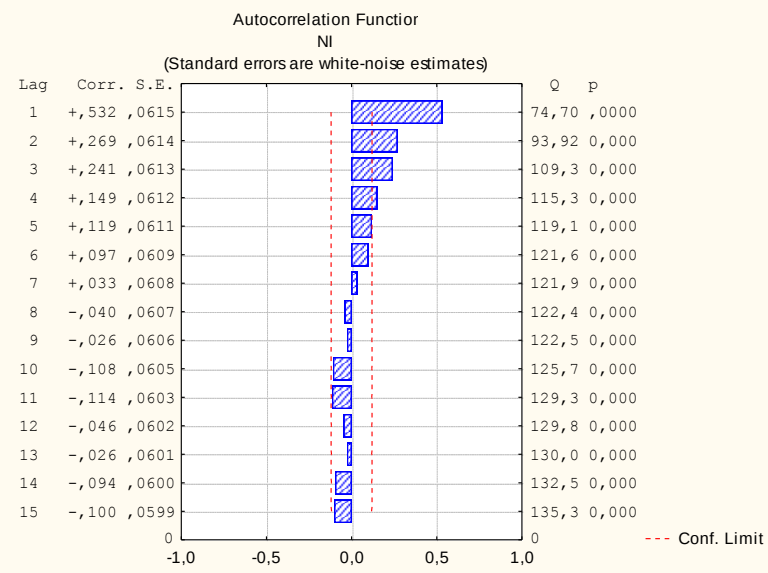


Рисунок 4.26 – Корелограма нульових різниць часового ряду кібершахрайських атак на мережеву інфраструктуру фінансової установи

Lag	Auto-Corr.	Std.Err.	Box & Ljung Q	p
1	0,77739	0,06154	159,551	0,00
2	0,62684	0,06142	263,691	0,00
3	0,50649	0,06130	331,944	0,00
4	0,44977	0,06118	385,977	0,00
5	0,39864	0,06106	428,589	0,00
6	0,37002	0,06095	465,446	0,00
7	0,33746	0,06083	496,222	0,00
8	0,31626	0,06071	523,360	0,00
9	0,30472	0,06059	548,654	0,00
10	0,32584	0,06047	577,691	0,00
11	0,30401	0,06034	603,069	0,00
12	0,28265	0,06022	625,094	0,00
13	0,30584	0,06010	650,985	0,00
14	0,29088	0,05998	674,500	0,00
15	0,27846	0,05986	696,137	0,00

Autocorrelation Function (Spreadsheet3.
CI
(Standard errors are white-noise estimate

Рисунок 4.27 – Значення автокорляційної функції нульових різниць часового ряду кібершахрайських атак на хмарну інфраструктуру фінансової установи

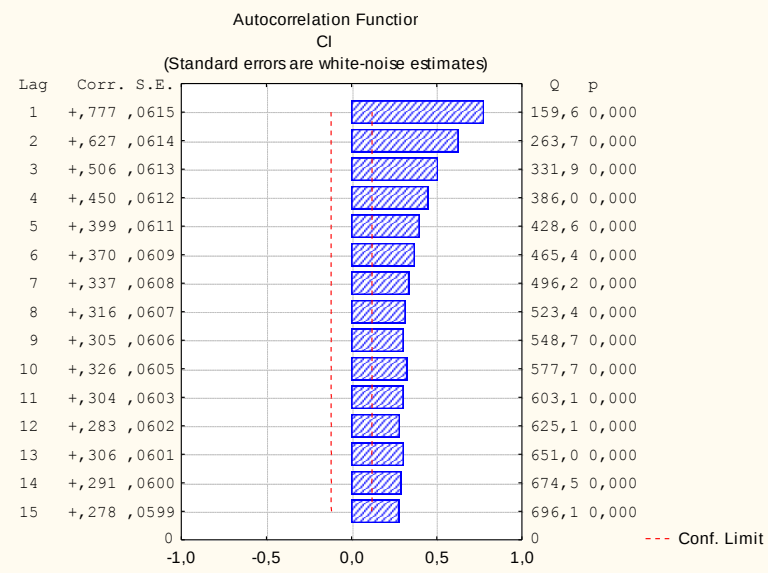


Рисунок 4.28 – Корелограма нульових різниць часового ряду кібершахрайських атак на хмарну інфраструктуру фінансової установи

Аналіз рисунків 4.21 і 4.22 дозволяє стверджувати, що для часового ряду кібершахрайських атак (нульові різниці) значення коефіцієнтів автокореляції для різних часових лагів є статистично значущими, постійно варіюються. Аналогічна

тенденція спостерігається і для корелограм кібершахрайських атак на комп’ютерні системи, та хмарну інфраструктуру фінансової установи. Винятком виступають лише значення коефіцієнтів автокореляції для мережевої інфраструктури, оскільки статистично значущими є лише перші три рівня, але вони мають коливальну тенденцію. Тому пропонується за часовий проміжок моделі Сєдова-Тейлора взяти тиждень (тобто 7 денний інтервал).

Переходячи до наступного кроку побудови моделі Сєдова-Тейлора для формалізації моделі ударної хвилі розповсюдження наслідків інформаційних війн у вигляді залежності фінансової безпеки від кібершахрайських атак на комп’ютерні системи, мережеву та хмарну інфраструктуру фінансової установи виникає необхідність обчислення коефіцієнтів економіко-математичної моделі розсіювання енергії ударних хвиль. З цією метою, базуючись на даних таблиць 1 – 3 в розрізі кібершахрайських атак на комп’ютерні системи, мережеву та хмарну інфраструктуру фінансової установи відповідно, які відображують початкові рівні енергії у кризові моменти часу, а також беручи до уваги обґрунтований на основі результатів проведення автокореляційного аналізу тижневий інтервал розповсюдження наслідків в розрізі фінансової безпеки після моменту розриву так званої «інформаційнох бульбашки», оптимізаційну модель загального вигляду (4.19) запишемо окремо в розрізі кожного напрямку кібершахрайських атак на основі наявних статистичних даних у вигляді формул (4.20) – (4.22). Крім того, для формування правої частини системи обмежень оптимізаційної задачі обчислимо значення абсолютних прирості поточного та попереднього рівнів відповідних часових рядів:

- для кібершахрайських атак на комп’ютерні системи фінансової установи:

$$\begin{cases} \frac{73}{7^4} + a_1\left(\frac{73}{7^4}\right)^{\frac{3}{4}} + a_2\left(\frac{73}{7^4}\right)^{\frac{2}{4}} + a_3\left(\frac{73}{7^4}\right)^{\frac{1}{4}} \to min \\ \frac{73}{7^4} + a_1\left(\frac{73}{7^4}\right)^{\frac{3}{4}} + a_2\left(\frac{73}{7^4}\right)^{\frac{2}{4}} + a_3\left(\frac{73}{7^4}\right)^{\frac{1}{4}} = 1 \\ \frac{70}{7^4} + a_1\left(\frac{70}{7^4}\right)^{\frac{3}{4}} + a_2\left(\frac{70}{7^4}\right)^{\frac{2}{4}} + a_3\left(\frac{70}{7^4}\right)^{\frac{1}{4}} = 7 \\ \frac{73}{7^4} + a_1\left(\frac{73}{7^4}\right)^{\frac{3}{4}} + a_2\left(\frac{73}{7^4}\right)^{\frac{2}{4}} + a_3\left(\frac{73}{7^4}\right)^{\frac{1}{4}} = 4 \\ \frac{98}{7^4} + a_1\left(\frac{98}{7^4}\right)^{\frac{3}{4}} + a_2\left(\frac{98}{7^4}\right)^{\frac{2}{4}} + a_3\left(\frac{98}{7^4}\right)^{\frac{1}{4}} = 3 \\ \frac{100}{7^4} + a_1\left(\frac{100}{7^4}\right)^{\frac{3}{4}} + a_2\left(\frac{100}{7^4}\right)^{\frac{2}{4}} + a_3\left(\frac{100}{7^4}\right)^{\frac{1}{4}} = 1 \end{cases} \quad (4.20)$$

- для кібершахрайських атак на мережеву інфраструктуру фінансової установи:

$$\begin{cases} \frac{100}{7^4} + a_1\left(\frac{100}{7^4}\right)^{\frac{3}{4}} + a_2\left(\frac{100}{7^4}\right)^{\frac{2}{4}} + a_3\left(\frac{100}{7^4}\right)^{\frac{1}{4}} \to min \\ \frac{100}{7^4} + a_1\left(\frac{100}{7^4}\right)^{\frac{3}{4}} + a_2\left(\frac{100}{7^4}\right)^{\frac{2}{4}} + a_3\left(\frac{100}{7^4}\right)^{\frac{1}{4}} = -3 \\ \frac{68}{7^4} + a_1\left(\frac{68}{7^4}\right)^{\frac{3}{4}} + a_2\left(\frac{68}{7^4}\right)^{\frac{2}{4}} + a_3\left(\frac{68}{7^4}\right)^{\frac{1}{4}} = -13 \\ \frac{74}{7^4} + a_1\left(\frac{74}{7^4}\right)^{\frac{3}{4}} + a_2\left(\frac{74}{7^4}\right)^{\frac{2}{4}} + a_3\left(\frac{74}{7^4}\right)^{\frac{1}{4}} = 12 \\ \frac{74}{7^4} + a_1\left(\frac{74}{7^4}\right)^{\frac{3}{4}} + a_2\left(\frac{74}{7^4}\right)^{\frac{2}{4}} + a_3\left(\frac{74}{7^4}\right)^{\frac{1}{4}} = 7 \\ \frac{88}{7^4} + a_1\left(\frac{88}{7^4}\right)^{\frac{3}{4}} + a_2\left(\frac{88}{7^4}\right)^{\frac{2}{4}} + a_3\left(\frac{88}{7^4}\right)^{\frac{1}{4}} = -9 \end{cases} \quad (4.21)$$

- кібершахрайських атак на хмарну інфраструктуру фінансової установи:

$$\begin{cases} \frac{100}{7^4} + a_1\left(\frac{100}{7^4}\right)^{\frac{3}{4}} + a_2\left(\frac{100}{7^4}\right)^{\frac{2}{4}} + a_3\left(\frac{100}{7^4}\right)^{\frac{1}{4}} \to min \\ \frac{100}{7^4} + a_1\left(\frac{100}{7^4}\right)^{\frac{3}{4}} + a_2\left(\frac{100}{7^4}\right)^{\frac{2}{4}} + a_3\left(\frac{100}{7^4}\right)^{\frac{1}{4}} = -3 \\ \frac{61}{7^4} + a_1\left(\frac{61}{7^4}\right)^{\frac{3}{4}} + a_2\left(\frac{61}{7^4}\right)^{\frac{2}{4}} + a_3\left(\frac{61}{7^4}\right)^{\frac{1}{4}} = -1 \\ \frac{52}{7^4} + a_1\left(\frac{52}{7^4}\right)^{\frac{3}{4}} + a_2\left(\frac{52}{7^4}\right)^{\frac{2}{4}} + a_3\left(\frac{52}{7^4}\right)^{\frac{1}{4}} = 3 \\ \frac{56}{7^4} + a_1\left(\frac{56}{7^4}\right)^{\frac{3}{4}} + a_2\left(\frac{56}{7^4}\right)^{\frac{2}{4}} + a_3\left(\frac{56}{7^4}\right)^{\frac{1}{4}} = -3 \end{cases} \quad (4.22)$$

З метою розв'язання оптимізаційних задач (3) – (5), які передбачають розрахунок коефіцієнтів моделі Сєдова-Тейлора формалізації моделі ударної хвилі поширення негативних наслідків інформаційних війн у вигляді залежності фінансової безпеки від кібершахрайських атак на комп'ютерні системи, мережеву та хмарну інфраструктуру фінансової установи, пропонується скористатись можливостями інструментарію MS Excel «Пошук рішення», зокрема методом ОПГ (метод узагальненого приведеного градієнта).

Таким чином, вирішення систем (4.20) – (4.22) шляхом пошуку відповідних коефіцієнтів надає можливість формалізувати шукану модель Сєдова-Тейлора для опису ударної хвилі поширення негативних наслідків інформаційних війн у вигляді залежності фінансової безпеки від кібершахрайських атак на комп'ютерні системи, мережеву та хмарну інфраструктуру фінансової установи в наступному вигляді:

- в розрізі кібершахрайських атак на комп'ютерні системи фінансової установи:

$$\Delta F \quad (t) = \frac{C}{t^4} - 0.7187 \cdot \left(\frac{C}{t^4}\right)^{\frac{3}{4}} + 0.0601 \cdot \left(\frac{C}{t^4}\right)^{\frac{2}{4}} + 0.0273 \cdot \left(\frac{C}{t^4}\right)^{1/4} \quad (4.23)$$

- у вигляді кібершахрайських атак на мережеву інфраструктуру фінансової установи:

$$\Delta F \quad (t) = \frac{C}{t^4} - 0.2510 \cdot \left(\frac{C}{t^4}\right)^{\frac{3}{4}} - 0.3076 \cdot \left(\frac{C}{t^4}\right)^{\frac{2}{4}} + 0.0994 \cdot \left(\frac{C}{t^4}\right)^{1/4} \qquad (4.24)$$

- у вигляді кібершахрайських атак на хмарну інфраструктуру фінансової установи:

$$\Delta F \quad (t) = \frac{C}{t^4} - 0.2214 \cdot \left(\frac{C}{t^4}\right)^{\frac{3}{4}} - 0.2784 \cdot \left(\frac{C}{t^4}\right)^{\frac{2}{4}} + 0.0829 \cdot \left(\frac{C}{t^4}\right)^{1/4} \qquad (4.25)$$

Візуалізація поширення наслідків ударної хвилі після моменту розриву «інформаційної бульбашки» набуває вигляду, представленому на рисунках 4.29 – 4.31 відповідно для кібершахрайських атак на комп'ютерні системи, мережеву та хмарну інфраструктуру фінансової установи.

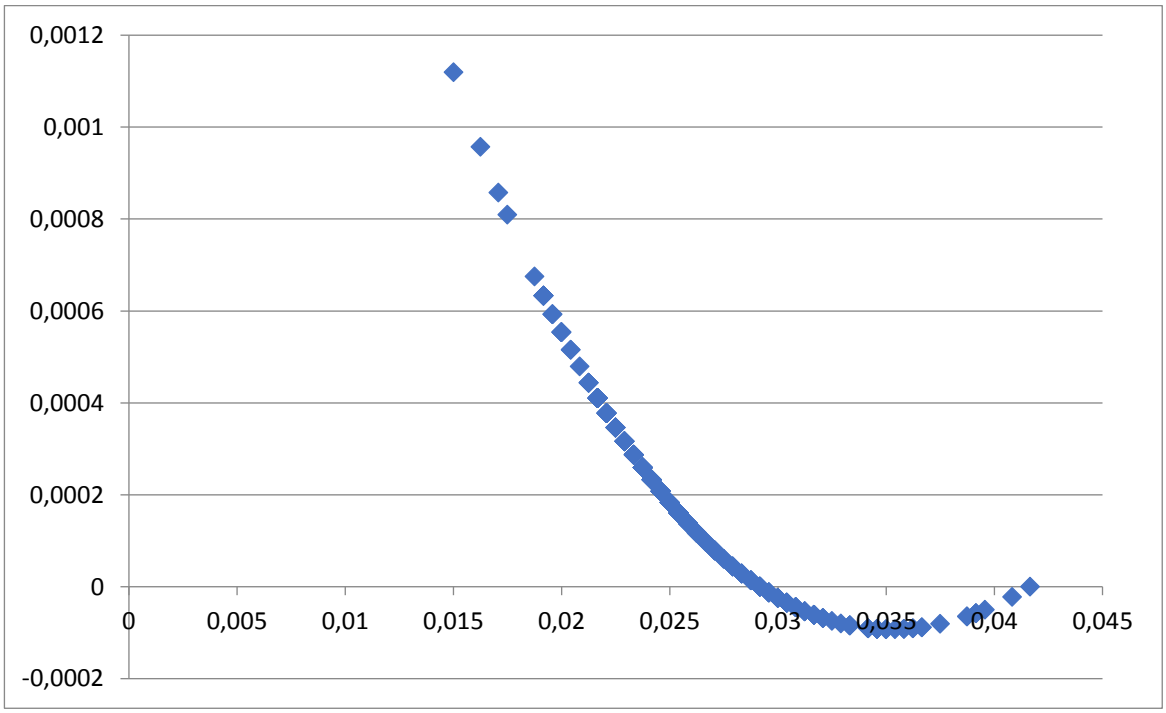


Рисунок 4.29 – Візуалізація поширення наслідків ударної хвилі після моменту розриву «інформаційної бульбашки» для впливу кібершахрайських атак на комп'ютерні системи фінансової установи на рівень фінансової безпеки

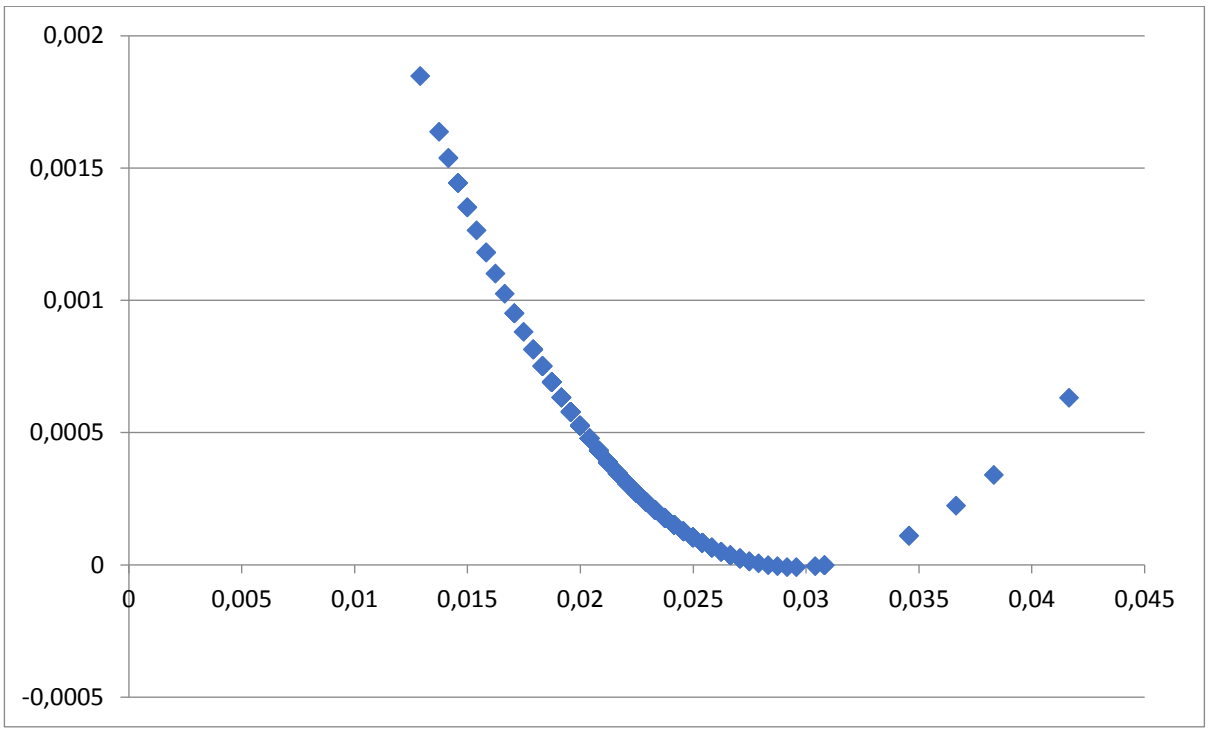


Рисунок 4.30 – Візуалізація поширення наслідків ударної хвилі після моменту розриву «інформаційної бульбашки» для впливу кібершахрайських атак на мережеву інфраструктуру фінансової установи на рівень фінансової безпеки

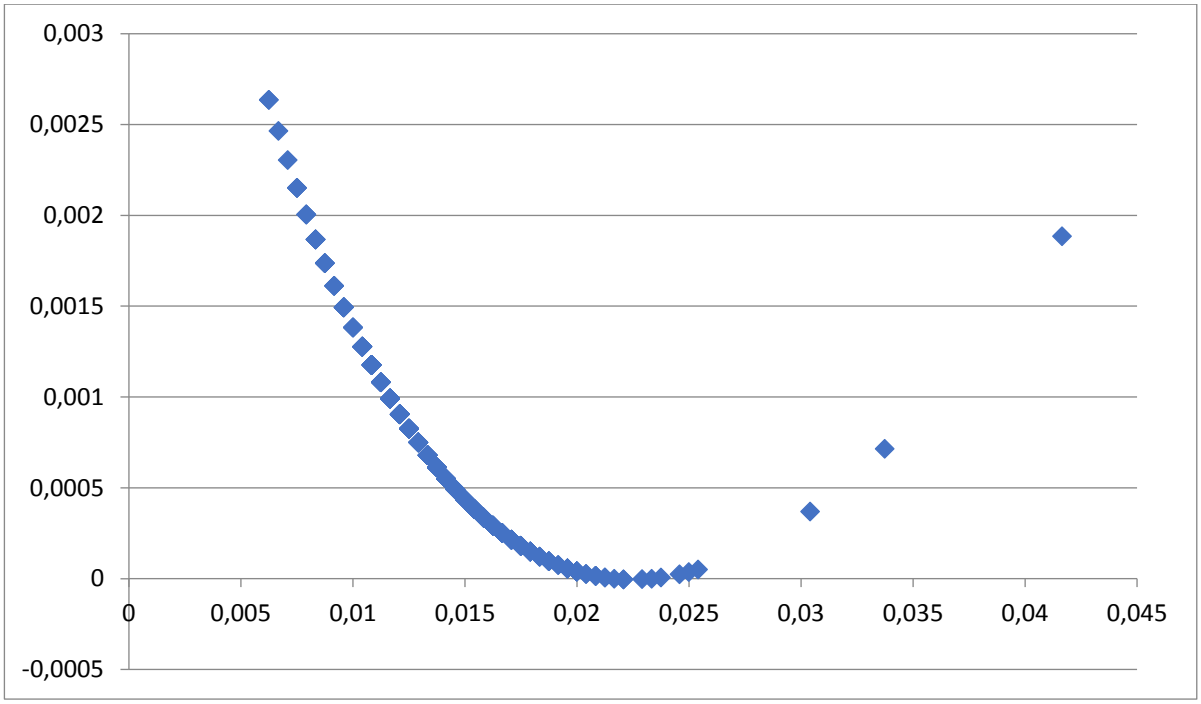


Рисунок 4.31 – Візуалізація поширення наслідків ударної хвилі після моменту розриву «інформаційної бульбашки» для впливу кібершахрайських атак на хмарну інфраструктуру фінансової установи на рівень фінансової безпеки

Аналіз рисунків 4.29 – 4.31 дозволяє констатувати наявність точки розриву «інформаційної бульбашки» (здійснення кібершахрайських атак) з подальшою адаптацією комп'ютерних систем, мережевої та хмарної інфраструктури фінансової установи і боротьбою з негативними наслідками поширення ударної хвилі на рівень фінансової безпеки, свідченням чого виступає зростаюча права гілка параболи.

З метою визначення критичних рівнів фінансової безпеки, кібершахрайських атак в розрізі комп'ютерних систем, мережевої та хмарної інфраструктури фінансової установи, перевищення яких супроводжується розривом «інформаційної бульбашки», визначимо такі рівні зазначених характеристик, за яких функція залежності фінансової безпеки від кібершахрайських атак має точку перегину. Такими значеннями виступають: 0,036, 0,030, 0,022. Для визначених рівнів обчислимо значення функції залежності фінансової безпеки від кібершахрайських атак другої похідної (рисунки 4.32 – 4.34).

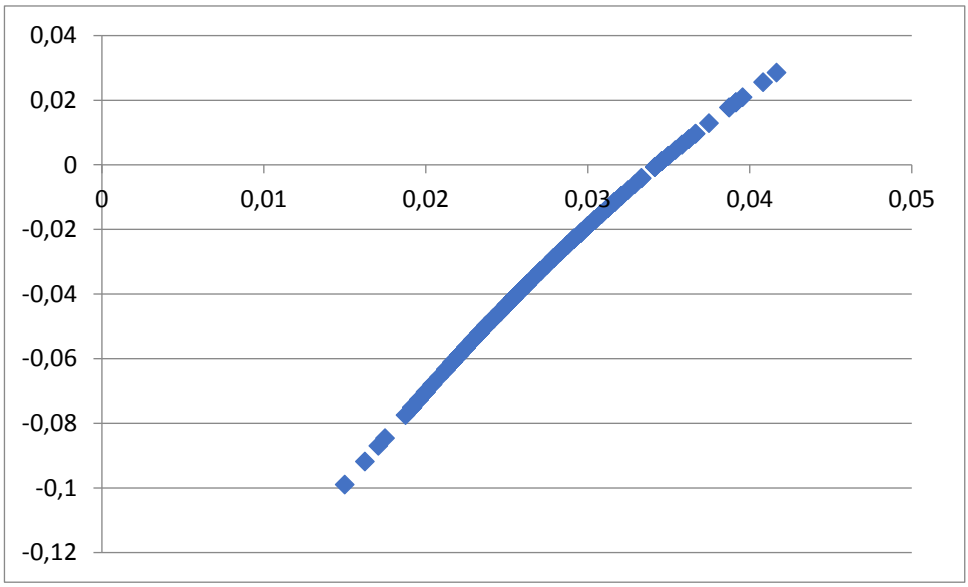


Рисунок 4.32 - Візуалізація функції першої похідної моделі Сєдова-Тейлора для опису ударної хвилі розповсюдження наслідків інформаційних війн у вигляді впливу кібершахрайських атак на комп'ютерні системи фінансової установи на рівень фінансової безпеки

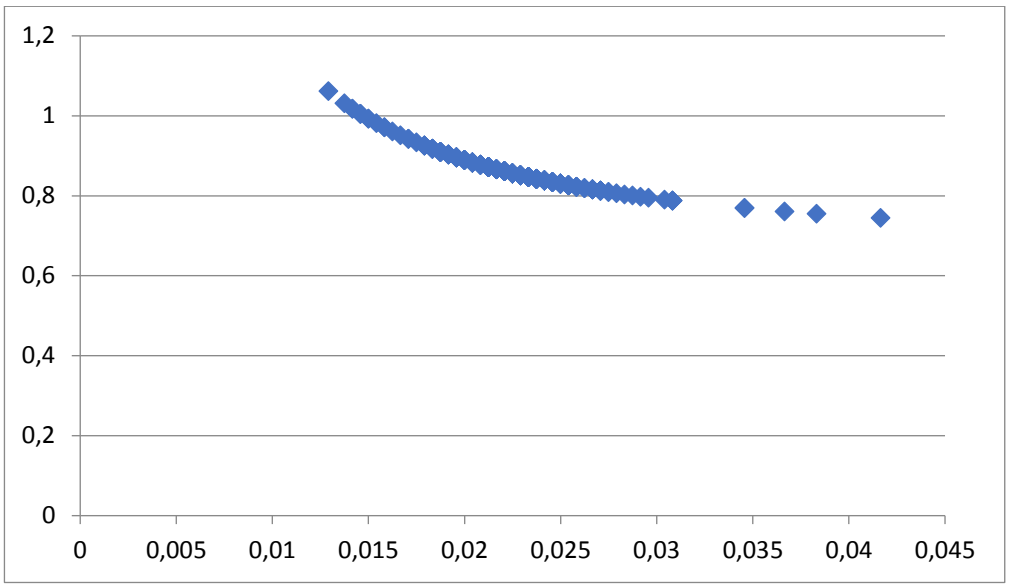


Рисунок 4.33 - Візуалізація функції першої похідної моделі Сєдова-Тейлора для опису ударної хвилі розповсюдження наслідків інформаційних війн у вигляді впливу кібершахрайських атак на мережеву інфраструктуру фінансової установи на рівень фінансової безпеки

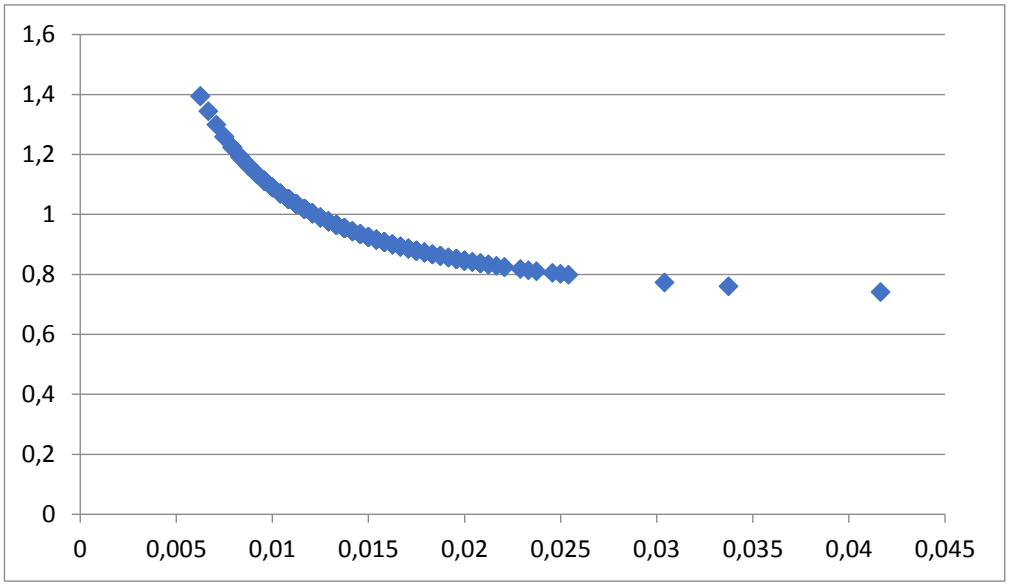


Рисунок 4.34 - Візуалізація функції першої похідної моделі Сєдова-Тейлора для опису ударної хвилі розповсюдження наслідків інформаційних війн у вигляді впливу кібершахрайських атак на хмарну інфраструктуру фінансової установи на рівень фінансової безпеки

Аналіз рисунків 4.32-4.34 дозволяє констатувати, що в момент розриву інформаційної бульбашки найбільший впливу кібершахрайських атак на рівень фінансової безпеки спостерігається в розрізі атак на комп'ютерні системи фінансової установи (0,797%), хоча в розрізі мережевої інфраструктури та хмарної інфраструктури фінансової установи рівень фінансової безпеки під впливом кібершахрайських атак відповідає рівням 82,86% та 78,82% відповідно. Після досягнення зазначених рівнів активність здійснення кібершахрайських атак знижується.

ВИСНОВКИ

У даному звіті представлені результати другого етапу науково-дослідної теми, який присвячений модернізації інструментарію протидії легалізації кримінальних доходів та кібершахрайствам. В рамках поставлених завдань було отримано наступні наукові результати.

Представлені у першому розділі наукові результати створюють передумови для здійснення аналізу джерел кібератак на прикладі ринку банківських послуг та пріоритезації найбільш поширених видів шахрайств. Серед основних результатів, що мають наукову новизну і практичну значущість, слід зазначити наступні:

- запропоновано концептуальну модель розробки прогнозних моделей кібератак, яка ідентифікує всі етапи процесу прогнозування; проведено статистичний аналіз базових статистик, декомпозиційний аналіз трендів, проведено перевірку на стаціонарність за допомогою розширеного тесту Дики-Фулера, на відповідність нормальному розподілу за допомогою тесту Харка-Бера. Це дозволило виявити: неоднорідність даних та встановити причини; відсутність трендової складової, наявність сезонності, вид зв'язків між компонентами часових рядів; доведення стаціонарності рядів; невідповідність нормальному розподілу. Отримані висновки дозволили здійснити процедури над даними та підготувати їх до наступних етапів багатошарового аналізу;
- проведено регресійний аналіз шляхом побудови об’єднаної регресії та регресій із випадковими та фіксованими ефектами для змінних “Mail Anti Virus”, “Kaspersky Anti-Spam” та “Intrusion Detection Spam”. Це дозволило визначити найбільш ефективну модель для прогнозування різних видів кібератак, якою виявилася модель об’єднаної регресі;
- проведено прогнозування на основі об’єднаної регресії та LSTM моделі на валідаційному наборі даних для різних країн світу. В результаті оцінок якості отриманих прогнозів було встановлено, що найбільш якісні прогнози генерує LSTM модель, не дивлячись на недостатній обсяг вхідних даних, але

об’єднана регресія більш якісно описує початкові дані, тобто перший вид моделювання доцільно використати для створення прогнозів, а другий – в рамках здійснення багатошарового аналізу.

Представлені у другому розділі наукові результати створюють передумови для розроблення мультисервісної моделі для комплексної оцінки та пріоритезації ризиків легалізації кримінальних доходів та кіберризиків. Серед основних результатів, що мають наукову новизну і практичну значущість, слід зазначити наступні:

- було проаналізовано наукові підходи щодо розуміння сутності поняття «протидія легалізації доходів, отриманих незаконним шляхом» в умовах діджиталізації суспільства. В результаті було встановлено, що проблематика протидії легалізації кримінальних коштів є досить актуальною в наукових колах і вона розглядається як комплекс заходів з попередження, виявлення та подальшого покарання злочинних дій, спрямованих на приховання чи маскування незаконного походження коштів або іншого майна;
- було запропоновано науково-методичний підхід до оцінювання ризиків конвергенції системи протидії фінансовим і кібершахрайствам на основі проведення сегментації країн із використанням кластерного аналізу за рівнем їх кібербезпеки, ризику відмивання кримінальних доходів, інтегрального рівня конвергенції систем протидії фінансовим і кібершахрайствам, побудові класифікаційної моделі дерева рішень оцінки ризиків конвергенції. В результаті було встановлено кластери країн, для яких було означено 9 груп ризику, що дозволяє оцінити можливості країн щодо спроможності та готовності систем їх фінансового моніторингу та кібербезпеки інтегруватися в єдину та комплексну систему фінансового кіберзахисту;
- було побудовано нейромережеву модель потенційної конвергенції систем фінансової та кібербезпеки, для чого було проведено статистичний та канонічний аналізи, застосовано метод головних компонент, побудовано автоматичну нейронну мережу та мережу на сітці, а також регресію. В результаті було встановлено, що важливими показниками в сфері кібербезпеки та протидії

фінансовим шахрайствам є фактори рівня цифрової трансформації і фактори, які характеризують легкості ведення бізнесу в країні, рівень споживчих цін та фінансової таємниці. Найбільш ефективною виявилася нейронна мережа, побудована на сітці, яка дозволяє прогнозувати рівень конвергенції систем фінансової та кібербезпеки в залежності від їх визначених ключових характеристик.

Представлені у третьому розділі наукові результати створюють передумови для розроблення алгоритмів розпізнавання поведінки кібершахраїв, кіберпрофілів сучасних фінансових злочинів та шахрайств. Серед основних результатів, що мають наукову новизну і практичну значущість, слід зазначити наступні:

- було проведено профілювання жертв кіберзлочинів на основі гендерного аналізу щодо використання ними пристроїв для доступу до Інтернету, їх активності в Інтернеті, відношення до ситуацій кібершахрайства, впевненості в ситуаціях із кібершахрайствами та способів захисту від кібершахрайства. В результаті було визначено, що саме чоловіки мають більшу схильність до того, щоб стати жертвою кібершахрайства, ніж жінки, а також було окреслено фактори, які можуть цьому сприяти;
- було проведено профілювання кіберзлочинів для випадків кібершахрайств із кредитними операціями, для чого було визначено найбільш ефективні підходи до виявлення їх характеристик, а також застосовано метод кластеризації «Очікування-максимізація» до набору вхідних даних щодо клієнтів банку і побудовано кіберпрофілі на основі 10 кластерів потенційних зловмисників. В результаті запропонованої методики профілювання встановлено п’ять найбільш значущих кластерів, які було сформовано під впливом таких характеристик, як сімейний стан, тип нерухомості, побутові умови, рівень освіти, тощо. Його використання не тільки дозволить виявити потенційних зловмисників, але й прийняти більш ефективне рішення щодо кредитування клієнтів, які мають статус потенційної загрози для банку;

- було розроблено алгоритми ідентифікації кіберзлочинців на основі методів інтелектуального аналізу даних, в результаті чого було побудовано об’єднану, LASSO, RIDGE, Elastic Net регресії, класифікаційне дерево рішення та нейронну мережу. Це дозволило визначити, що найбільш ефективними є алгоритми дерева рішення та нейронної мережі, які дозволятимуть з 90% рівнем упевненості ідентифікувати кіберзлочинця.

Представлені у четвертому розділі наукові результати створюють передумови для розроблення моделі прогнозування кібершахрайських атак на комп’ютерні системи, мережеву та хмарну інфраструктуру фінансової установи. Серед основних результатів, що мають наукову новизну і практичну значущість, слід зазначити наступні:

- було проведено дослідження інформаційних трендів трьох видів кібератак на комп’ютерні системи, мережеву та хмарну інфраструктуру фінансової установи, яке дозволило визначити інструмент моделювання та побудувати адитивні та мультиплікативні циклічні моделі експоненційного згладжування без тренду, з лінійним, експоненційним, затухаючим трендами та з урахуванням сезонної складової. Це дозволило виявити найкращі моделі для прогнозування конкретного виду кібератак та здійснити прогнози на короткострокову перспективу;
- було запропоновано ударно-хвильову модель впливу трьох видів кібератак на рівень фінансової безпеки, яка дозволила виявити моменти розриву інформаційної бульбашки, сформованої під найбільшим впливом кібершахрайських атак на рівень фінансової безпеки. Застосування даного підходу дозволяє прогнозувати пікові моменти кібератак, що сприяє формуванню комплексу превентивних дій в конкретних випадках та для конкретних видів кібератак.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. Ukraine cyber-attack: Russia to blame for hack, says Kyiv. *BBC* : website. URL: https://www.bbc.com/news/world-europe-59992531 (дата звернення 10.12.2022).

2. Ukraine's defence ministry and two banks targeted in cyberattack. *Euronews* : website. URL: https://www.euronews.com/my-europe/2022/02/15/ukraine-s-defence-ministry-and-two-banks-targeted-in-cyberattack (дата звернення 10.12.2022).

3. Report: Recent 10x Increase in Cyberattacks on Ukraine. *Krebsonsecurity* : website. URL: https://krebsonsecurity.com/2022/03/report-recent-10x-increase-in-cyberattacks-on-ukraine/ (дата звернення 10.12.2022).

4. UKRAINE: Timeline of Cyberattacks on critical infrastructure and civilian objects. *CyberPeace Institute* : website. URL: https://cyberpeaceinstitute.org/ukraine-timeline-of-cyberattacks/ (дата звернення 10.12.2022).

5. Mortgage Loan Fraud Connections with Other Financial Crime: An Evaluation of Suspicious Activity Reports Filed By Money Services Businesses, Securities and Futures Firms, Insurance Companies and Casinos. Office of Law *Enforcement Support Financial Crimes Enforcement Network* : website. URL: https://www.fincen.gov/sites/default/files/shared/mortgage_fraud.pdf (дата звернення 10.12.2022).

6. The connected defense: Elevating the fight against financial crime. Using 4IR technologies to prevent and detect the growing ecosystem of financial crime. *Deloitte Development LLC* : website. URL: https://www2.deloitte.com/content/dam/Deloitte/nl/Documents/financial-services/deloitte-nl-fsi-elevating-the-fight-against-financial-crime.pdf (дата звернення 10.12.2022).

7. Building a united front on financial crimes. *PwC* : website. URL: https://www.pwc.com/gx/en/financial-services/pdf/united-front-financial-crimes-2018-pwc.pdf (дата звернення 10.12.2022).

8. The Global Risk Report. *World Economic Forum* : website. URL: https://www3.weforum.org/docs/WEF_Global_Risk_Report_2020.pdf (дата звернення 10.12.2022).

9. Cost of a Data Breach Report 2022. *IBM Security* : website. URL: https://www.ibm.com/downloads/cas/3R8N1DZJ (дата звернення 10.12.2022).

10. Mclean M. Must-Know Cyber Attack Statistics and Trendshttps. *Embroker* : website. URL: www.embroker.com/blog/cyber-attack-statistics/ (дата звернення 10.12.2022).

11. Reports largest single day virus spike. *Abcnews* : website. URL: https://abcnews.go.com/Health/wireStory/latest-india-reports-largest-single-day-virus-spike-70826542 (дата звернення 10.12.2022).

12. Stacey P., Taylor R., Spanaki K. Emotional reactions and coping responses of employees to a cyber-attack: A case study. *International Journal of Information Management*. 2021, vol. 58, art. num. 102298. DOI: https://doi.org/10.1016/j.ijinfomgt.2020.102298.

13. Shandler R., Gomez M. A. The hidden threat of cyber-attacks–undermining public confidence in government. *Journal of Information Technology and Politics*. 2022. DOI: https://doi.org/10.1080/19331681.2022.2112796.

14. Lonsdale D. J. The Ethics of Cyber Attack: Pursuing Legitimate Security and the Common Good in Contemporary Conflict Scenarios. *Journal of Military Ethics*. 2020, vol. 19(1). P. 20–39. DOI: https://doi.org/10.1080/15027570.2020.1764694.

15. Bolpagni M. Cyber risk index: a socio-technical composite index for assessing risk of cyber attacks with negative outcome. *Quality and Quantity*. 2022, vol. 56(3). P. 1643–1659. DOI: https://doi.org/10.1007/s11135-021-01199-3.

16. Simons G., Danyk Y., Maliarchuk T. Hybrid war and cyber-attacks: creating legal and operational dilemmas. *Global Change, Peace and Security*. 2020, vol. 32(3). P. 337–342. DOI: https://doi.org/10.1080/14781158.2020.1732899.

17. Weaver G. A., Feddersen B., Marla L., Wei D., Rose A., Van Moer M. Estimating economic losses from cyber-attacks on shipping ports: An optimization-

based approach. *Transportation Research Part C: Emerging Technologies.* 2022, vol. 137, art. num. 103423. DOI: https://doi.org/10.1016/j.trc.2021.10342.

18. Leroy I. The relationship between cyber-attacks and dynamics of company stock: the role of reputation management. *International Journal of Electronic Security and Digital Forensics*. 2022, vol. 14(4). P. 309–317. DOI: https://doi.org/10.1504/IJESDF.2022.123891.

19. Akoto W. International trade and cyber conflict: Decomposing the effect of trade on state-sponsored cyber attacks. *Journal of Peace Research*. 2021, vol. 58(5). P. 1083–1097. DOI: https://doi.org/10.1177/0022343320964549.

20. Lallie H. S., Shepherd L. A., Nurse J. R.C., Erola A., Epiphaniou G., Maple C., Bellekens X. Cyber security in the age of COVID-19: A timeline and analysis of cyber-crime and cyber-attacks during the pandemic. *Computers and Security*. 2021, vol. 105, Art. Num. 102248. DOI: https://doi.org/10.1016/j.cose.2021.102248.

21. Definition and How to Run an F-Test. *Statistics How To* : website. URL: https://www.statisticshowto.com/probability-and-statistics/hypothesis-testing/f-test/ (дата звернення 10.12.2022).

22. Яровенко Г.М., Кобзенко В.В. Попередній аналіз і підготовка даних для прогнозування трендів кібератак. *Економіка та суспільство*. 2022, №45. DOI: https://doi.org/10.32782/2524-0072/2022-45-42.

23. Statsmodels. *Statsmodels* : website. URL: https://www.statsmodels.org/stable/index.html (дата звернення 10.12.2022).

24. Akaike Information Criterion | When & How to Use It (Example). *Scribbr* : website. URL: https://www.scribbr.com/statistics/akaike-information-criterion/ (дата звернення 10.12.2022).

25. How to Interpret Log-Likelihood Values (With Examples). *Statology* : website. URL: https://www.statology.org/interpret-log-likelihood/ (дата звернення 10.12.2022).

26. Residual Standard Deviation/Error: Guide for Beginners. *Quantifyinghealth* : website. URL: https://quantifyinghealth.com/residual-standard-deviation-error/ (дата звернення 10.12.2022).

27. How to Calculate Standardized Residuals in Python. *Statology* : website. URL: https://www.statology.org/standardized-residuals-python/ (дата звернення 10.12.2022).

28. The Breusch-Pagan Test: Definition & Example. *Statology* : website. URL: https://www.statology.org/breusch-pagan-test/ (дата звернення 10.12.2022).

29. Train-Test Split for Evaluating Machine Learning Algorithms. *Machinelearningmaster* : website. URL: https://machinelearningmastery.com/train-test-split-for-evaluating-machine-learning-algorithms/ (дата звернення 10.12.2022).

30. LabelEncoder. *Scikit-learn.org* : website. URL: https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.LabelEncoder.html (дата звернення 10.12.2022).

31. RMSE: Root Mean Square Error. *Statisticshowto.com* : website. URL: https://www.statisticshowto.com/probability-and-statistics/regression-analysis/rmse-root-mean-square-error/ (дата звернення 10.12.2022).

32. Sequence Classification with LSTM Recurrent Neural Networks in Python with Keras. *Machinelearningmaster* : website. URL: https://machinelearningmastery.com/sequence-classification-lstm-recurrent-neural-networks-python-keras/ (дата звернення 10.12.2022).

33. Кобзенко В.В. Моделювання та прогнозування трендів кібератак : робота на здобуття кваліфікаційного рівня магістр : спец. 051 - економіка / наук. кер. Г. М. Яровенко. Суми : СумДУ, 2022. 73 с.

34. Загальні тенденції тіньової економіки у 2021 році. *Міністерство економіки України* : веб-сайт. URL: https://www.me.gov.ua/Documents/Download?id=74e86de5-126a-4849-94d5-7d4ea048e4b8 (дата звернення: 06.12.2022).

35. Гордійчук М. Тіньова економіка: позитивні та негативні аспекти. *Траєкторія науки*. 2019, №5(3). С. 2001-2007. URL: https://pathofscience.org/index.php/ps/article/download/599/613 (дата звернення: 06.12.2022).

36. Живко З. Б., Родченко С. С., Висоцька І. Б. Вплив легалізації доходів, отриманих незаконним шляхом, на економічну безпеку. *Соціально-гуманітарний вісник*. 2021, №37. С. 48-51. URL: http://dspace.lvduvs.edu.ua/bitstream/1234567890/4195/1/вплив%20легалізації%20доходів.pdf (дата звернення: 06.12.2022).

37. High-Risk Jurisdictions subject to a Call for Action – 21 October 2022. *The Financial Action Task Force (FATF)* : website. URL: https://www.fatf-gafi.org/publications/high-risk-and-other-monitored-jurisdictions/documents/call-for-action-october-2022.html (date accessed: 06.12.2022).

38. Типологічне дослідження «Ризики використання суб'єктів з непрозорою структурою власності у схемах відмивання кримінальних доходів». *Державна служба фінансового моніторингу України :* веб-сайт. URL: https://fiu.gov.ua/assets/userfiles/411/Типолог%20ДСФМУ/2018%2012%2028_Typology2018_UA.pdf (дата звернення: 06.12.2022).

39. Типологічне дослідження «Актуальні методи і способи легалізації (відмивання) доходів, одержаних злочинним шляхом, та фінансування тероризму». *Державна служба фінансового моніторингу України :* веб-сайт. URL: https://fiu.gov.ua/assets/userfiles/411/Типолог%20ДСФМУ/2012%2027%2012_2012.pdf (дата звернення: 06.12.2022).

40. Типології легалізації (відмивання) доходів, одержаних злочинним шляхом «Використання готівки у схемах відмивання злочинних доходів». *Державна служба фінансового моніторингу України :* веб-сайт. URL: https://fiu.gov.ua/assets/userfiles/411/Типолог%20ДСФМУ/2011%2012%2029_gotivka.pdf (дата звернення: 06.12.2022).

41. Lemire K.A. Cryptocurrency and anti-money laundering enforcement. *Reuters* : website. URL: https://www.reuters.com/legal/transactional/cryptocurrency-anti-money-laundering-enforcement-2022-09-26 (date accessed: 06.12.2022).

42. Що таке NFT і як на ньому заробити. *МінфінМедіа :* веб-сайт. URL: https://minfin.com.ua/ua/invest/articles/scho-take-nft-i-yak-na-nomu-zarobyty/ (дата звернення: 06.12.2022).

43. PayPal в Україні під час війни спростив отримання міжнародних переказів на банківські картки. Як саме? *Економічна правда :* веб-сайт. URL: https://www.epravda.com.ua/publications/2022/06/16/688135 (дата звернення: 06.12.2022).

44. В Україні з'явилися перші криптомати для біткойнів. *Радіо Свобода :* веб-сайт. URL: https://www.radiosvoboda.org/a/28729154.html (дата звернення: 06.12.2022).

45. Dobrowolski Z., Sułkowski Ł. Implementing a Sustainable Model for Anti-Money Laundering in the United Nations Development Goals. *Sustainability*. 2020, vol. 12(1). 244. DOI: https://doi.org/10.3390/su12010244.

46. Ferwerda J., Kleemans E.R. Estimating Money Laundering Risks: An Application to Business Sectors in the Netherlands. *Eur J Crim Policy Res*. 2019, vol. 25. P. 45–62. URL: https://doi.org/10.1007/s10610-018-9391-4.

47. Salehi A., Ghazanfari M., Fathian M. Data mining techniques for anti money laundering. *International Journal of Applied Engineering Research*. 2017, vol. *12*(20). P. 10084–10094. DOI: https://doi.org/10.5120/ijca2016910953

48. Canhoto A.I. Leveraging machine learning in the global fight against money laundering and terrorism financing: An affordances perspective. *Journal of Business Research*. 2021, vol. 131. P. 441-452. DOI: https://doi.org/10.1016/j.jbusres.2020.10.012.

49. Vovk V., Zhezherun Y., Bilovodska O., Babenko V., Biriukova A. Financial Monitoring in the Bank as a Market Instrument in the Conditions of Innovative Development and Digitalization of Economy: Management and Legal Aspects of the Risk-Based Approach. *IJIEPR*. 2020, vol. 31(4). P. 559-570. URL: http://ijiepr.iust.ac.ir/article-1-1141-en.html.

50. Said Kh., Karimi D. Impact de la Digitalisation sur la Performance Bancaire dans la Prévention et la Lutte contre le Blanchiment de Capitaux. *African Scientific Journal*. 2022, vol. *3*(12). P. 461-476. DOI: https://doi.org/10.5281/zenodo.6874059.

51. Kobushko I., Tiutiunyk I., Kobushko I., Starinskyi M., Zavalna Z. The triadic approach to cash management: Communication, advocacy, and legal aspects. *Estudios*

De Economia Aplicada. 2021, vol. 39(7). DOI: https://doi.org/10.25115/eea.v39i7.5071.

52. Boiko A., Zwolińska-Ligaj M., Bozhenko V., Florczak E., Ovcharenko V. Readiness for implementing innovations in banking in advanced and emerging economies. *Journal of International Studies*. 2021, vol. 14(4). P. 236-250. DOI: https://doi.org/10.14254/2071-8330.2021/14-4/16.

53. Djalilov K., Hölscher J. Comparative analyses of the banking environment in transition countries. *Economic Annals*. 2016, vol. 61(208). P. 7-25. DOI: https://doi.org/10.2298/EKA1608007D.

54. Djalilov K., Hartwell C. Do social and environmental capabilities improve bank stability? evidence from transition countries. *Post-Communist Economies*. 2021, vol. 34(5). P.624-646. DOI: https://doi.org/10.1080/14631377.2021.1965359

55. Kuzior A., Kettler K., Rąb Ł. Digitalization of work and human resources processes as a way to create a sustainable and ethical organization. *Energies*. 2022, vol. 15(1). 172. DOI: https://doi.org/10.3390/en15010172.

56. Antonyuk N., Plikus I., Jammal M. Human Capital Quality Assurance under the Conditions of Digital Business Transformation and COVID-19 Impact. *Health Economics and Management Review*. 2021, vol. 2(3). P. 39-47. DOI: https://doi.org/10.21272/hem.2021.3-04.

57. Addo A., Senyo P.K. Digitalization and government corruption in developing countries: towards a framework and research agenda. *Academy of Management Proceedings*. 2020, vol. 1. DOI: https://doi.org/10.5465/AMBPP.2020.16765abstract.

58. Halis D. de C. Digitalization and Dissent in Legal Cultures. Chinese and Other Perspectives. Naveiñ Reet: *Nordic Journal of Law and Social Research (NNJLSR)*. 2019, vol. 9. P. 127-152. URL: https://tidsskrift.dk/nnjlsr/issue/download/8857/1189#page=129.

59. Mulyana Y. Digitalization of the court in the settlement of cases. *International Journal of Latin Notary*. 2021, vol. 1(2). P. 36-42. URL: https://i-latinnotary.notariat.unpas.ac.id/index.php/jurnal/article/view/6.

60. Wronka C. Money laundering through cryptocurrencies - analysis of the phenomenon and appropriate prevention measures. *Journal of Money Laundering Control*. 2022, vol. 25(1). P.79-94. DOI: https://doi.org/10.1108/JMLC-02-2021-0017.

61. Dupuis D., Gleason K. Money laundering with cryptocurrency: open doors and the regulatory dialectic. *Journal of Financial Crime*. 2020, vol. 28(1). P. 60-74. DOI: https://doi.org/10.1108/JFC-06-2020-0113.

62. Миненко С.В. Теоретичні засади до розуміння сутності поняття «протидія легалізації доходів отриманих незаконним шляхом» в умовах діджиталізації суспільства. *Modern aspects of science: 26- th volume of the international collective monograph*: / за ред. К. Недбаленка та ін.. : Publishing Group „Vědecká perspektiva“, 2022. С. 537-556.

63. Eling M., Schnell W. What do we know about cyber risk and cyber risk insurance? *The Journal of Risk Finance*. 2016, vol. 17(5). P. 474-491. DOI: https://doi.org/10.1108/JRF-09-2016-0122.

64. Кожедуб Ю. Аналіз документів з керування ризиком кібербезпеки. Information Technology and Security. 2017, vol. 5(1). P. 82-95. URL: http://nbuv.gov.ua/UJRN/inftech_2017_5_1_11.

65. Institute of Risk Management. *Theirm* : website. URL : https://www.theirm.org/knowledge-and resources/thought-leadership/cyber-risk (дата звернення 10.12.2022).

66. National Cybersecurity Index. NCSI : website. URL: https://ncsi.ega.ee/ (дата звернення 10.12.2022).

67. Basel AML Index Assessing Money Laundering Risks Around The World. Basel ALM Index : website. URL: https://index.baselgovernance.org/ (дата звернення 10.12.2022).

68. Bouveret A. Cyber Risk for the Financial Sector: A Framework for Quantitative Assessment. *IMF Working Paper*. 2018. URL: https://www.imf.org/en/Publications/WP/Issues/2018/06/22/Cyber-Risk-for-the-Financial-Sector-A-Framework-for-Quantitative-Assessment-45924 (дата звернення 10.12.2022).

69. Understanding Systemic Cyber Risk, Global Agenda Council on Risk & Resilience, White Paper, October 2022. *World Economic Forum* : website. URL: https://www3.weforum.org/docs/White_Paper_GAC_Cyber_Resilience_VERSION_2.pdf (дата звернення 10.12.2022).

70. Haro G. Shape from silhouette consensus and photo-consistency. URL: https://repositori.upf.edu/bitstream/handle/10230/35708/haro_icip14_shape.pdf;jsessionid=21B7F0CD85AB9435B87CD7AB8D338316?sequence=1

71. Яровенко Г.М., Рожкова М.С. Оцінка ризику конвергенції системи протидії відмивання грошей та кібербезпеки. *Економіка та суспільство*. 2022, № 45.

72. Кібербезпека у фінансовій сфері. *Risk-practice* : website. URL: https://risk-practice.ru/magazine/112/eau_112_659/ (дата звернення 10.12.2022).

73. Мельничук Я.О., Кравченко С.М. Аналіз даних та візуалізація за допомогою мови Python. URL: https://conf.ztu.edu.ua/wp-content/uploads/2017/11/54.pdf .

74. NumPy. *Wikipedia* : website. URL: https://ru.wikipedia.org/wiki/NumPy (дата звернення 10.12.2022).

75. Global Cybersecurity Index. *ITU* : website. URL: https://www.itu.int/en/ITU-D/Cybersecurity/Pages/global-cybersecurity-index.aspx (дата звернення 10.12.2022).

76. Networked Readiness Index. *Wikipedia* : website. URL: https://en.wikipedia.org/wiki/Networked_Readiness_Index#:~:text=The%20Networked%20Readiness%20Index%20is,by%20information%20and%20communications%20technology (дата звернення 10.12.2022).

77. Візуалізація даних. *Oracle* : website. URL: https://www.oracle.com/ru/business-analytics/what-is-data-visualization/ (дата звернення 10.12.2022).

78. Гістограми. *Sixsigmaonline* : website. URL: http://sixsigmaonline.ru/baza-znanij/gistogrammy-chto-ehto-kak-postroit-kak-predstavit-dannye-kak-provesti-analiz (дата звернення 10.12.2022).

79. Matplotlib: Наукова графіка в Python. *Pythonworld* : website. URL: https://pythonworld.ru/novosti-mira-python/scientific-graphics-in-python.html (дата звернення 10.12.2022).

80. Яровенко Г. М. Канонічний аналіз взаємозв'язку інформаційної безпеки та соціо-економіко-політичного розвитку країни. *Науковий вісник Ужгородського національного університету. Серія: Міжнародні економічні відносини та світове господарство*. 2020, № 31. С. 160–167.

81. Principal component analysis. *Royalsocietypublishing* : website. URL: https://royalsocietypublishing.org/doi/10.1098/rsta.2015.0202 (дата звернення 10.12.2022).

82. Svitlychna A. O. Modelling the potential convergence of the cybersecurity system and combating money laundering : bachelor's qualification work : specialty 051 – economics / head H. Yarovenko. Sumy : Sumy State University, 2022. 56 p.

83. Average cost of a data breach worldwide from 2014 to 2022 (in a million U.S. dollars). *Statista* : website. URL: https://www.statista.com/statistics/987474/global-average-cost-data-breach/ (дата звернення 10.12.2022).

84. Average cost of a data breach worldwide from May 2020 to March 2022, by industry (in a million U.S. dollars). *Statista* : website. URL: https://www.statista.com/statistics/387861/cost-data-breach-by-industry/ (дата звернення 10.12.2022).

85. UAE victims of cybercrime lose $746m a year. *The National* : website. URL: https://www.thenationalnews.com/business/technology/2021/08/13/uae-victims-of-cybercrime-lose-746m-a-year/ (дата звернення 10.12.2022).

86. Bhardwaj G., Bawa R. K. Machine learning techniques based exploration of various types of crimes in India. *Indian Journal of Computer Science and Engineering*. 2022, vol. 13(4). P. 1293–1307. DOI: https://doi.org/110.21817/indjcse/2022/v13i4/221304142.

87. Syeda R. Z., Chishti M. A., Baba A. I., Wu F. Detecting Covid-19 chaos driven phishing/malicious URL attacks by a fuzzy logic and data mining based

intelligence system. *Egyptian Informatics Journal*. 2022, vol. 23(2). P. 197–214. DOI: https://doi.org/10.1016/j.eij.2021.12.003.

88. Monika, Bhat A. Automatic Twitter Crime Prediction Using Hybrid Wavelet Convolutional Neural Network with World Cup Optimization. *International Journal of Pattern Recognition and Artificial Intelligence*. 2022, vol. 36(5). DOI: https://doi.org/10.1142/S0218001422590054.

89. Gomathi C., Jayasri K. Rain Drop Service and Biometric Verification Based Blockchain Technology for Securing the Bank Transactions from Cyber Crimes Using Weighted Fair Blockchain (WFB) Algorith. *Cybernetics and Systems*. 2022. DOI: https://doi.org/10.1080/01969722.2022.2103229.

90. Dupont B., Holt T. The Human Factor of Cybercrime. *Social Science Computer Review*. 2022, vol. 40(4). P. 860–864. DOI: https://doi.org/10.1177/08944393211011584.

91. Lazarus S., Button M., Kapend R. Exploring the value of feminist theory in understanding digital crimes: Gender and cybercrime types. *Howard Journal of Crime and Justice*. 2022, vol. 61(3). P. 381–398. DOI: https://doi.org/10.1111/hojo.12485.

92. Connolly A. Y., Borrion H. Reducing Ransomware Crime: Analysis of Victims' Payment Decisions. *Computers and Security*. 2022, vol. 119. DOI: https://doi.org/10.1016/j.cose.2022.102760.

93. Witsenboer J. W. A., Sijtsma K., Scheele F. Measuring cyber secure behavior of elementary and high school students in the Netherlands. *Computers and Education*. 2022, vol. 186. DOI: https://doi.org/10.1016/j.compedu.2022.104536.

94. Drury B., Drury S.M., Rahman Md A., Ihsan U. A social network of crime: A review of the use of social networks for crime and the detection of crime. *Online Social Networks and Media*. 2022, vol. 30. DOI: https://doi.org/10.1016/j.osnem.2022.100211.

95. Lee Yi Y., Gan C. L., Liew T. W. Phishing victimization among Malaysian young adults: cyber routine activities theory and attitude in information sharing online. *Journal of Adult Protection*. 2022, vol. 24(3-4). P. 179–194. DOI: https://doi.org/10.1108/JAP-06-2022-0011.

96. Special Eurobarometer 499 : Europeans' attitudes towards cyber security (cybercrime). *Data.europa.eu* : website. URL: https://data.europa.eu/data/datasets/s2249_92_2_499_eng?locale=en (дата звернення 10.12.2022).

97. Glossary on gender-related terms. *OSCE.org* : website. URL: https://www.osce.org/files/f/documents/1/2/26397.pdf (дата звернення 10.12.2022)

98. Gender Impact Assessment: Gender Mainstreaming Toolkit. *EIGE.europa.eu* : website. URL: https://eige.europa.eu/sites/default/files/mh0416171enn.pdf (дата звернення 10.12.2022).

99. Cyber crime categories that were reported most often in 2021, by number of victims. *Statista* : website. URL: https://www.statista.com/statistics/184083/commonly-reported-types-of-cyber-crime/ (дата звернення 10.12.2022).

100. Yarovenko H., Rymar V. Development of modern cyber fraud profiles. *Міжнародний науковий журнал «Грааль науки», 2022. : за матеріалами V Міжнародної науково-практичної конференції «An integrated approach to science modernization: methods, models and multidisciplinarity»*, що проводилася 23 грудня 2022 року ГО «Європейська наукова платформа» (Вінниця, Україна) та ТОВ «International Centre Corporative Management» (Відень, Австрія). P. 267-268. DOI: https://doi.org/10.36074/grail-of-science.23.12.2022.40.

101. Nuha M., Mahmud S., Sattar A. (2021). A case study and fraud rate prediction in e-banking systems using machine learning and data mining. Soft *Computing Techniques and Applications*. 2022. P. 71-83. DOI: https://doi.org/10.1007/978-981-15-7394-1_6.

102. Ланде Д. В., Субач І. Ю., Бояринова Ю. Є. Основи теорії і практики інтелектуального аналізу даних у сфері кібербезпеки : навчальний посібник. Київ : КПІ ім. Ігоря Сікорського, 2018. 300 с.

103. Дюк В., Самойленко А. Data Mining: учебный курс (+CD). СПб: Изд. Питер, 2001. 368 с. URL: https://www.azstat.org/Kitweb/zipfiles/11337.pdf (дата звернення 10.12.2022).

104. Alshamasi S., Menai M. Ensemble-based clustering for writing style change detection in multi-authored textual documents. *Paper presented at the CEUR Workshop Proceedings*. 2022, art. no. 3180. P. 2357-2374.

105. Lekha K. C., Prakasam S. Data mining techniques in detecting and predicting cyber crimes in banking sector. Paper presented at *the 2017 International Conference on Energy, Communication, Data Analytics and Soft Computing, ICECDS*. 2017. P. 1639–1643. DOI: https://doi.org/10.1109/ICECDS.2017.8389725.

106. Vinayakumar R., Alazab M., Soman K. P., Poornachandran P., Al-Nemrat A., Venkatraman S. Deep learning approach for intelligent intrusion detection system. *IEEE Access*. 2019, vol. 7. P. 41525–41550. DOI: https://doi.org/10.1109/ACCESS.2019.2895334.

107. Kanimozhi V., Prem Jacob T. Artificial intelligence based network intrusion detection with hyper-parameter optimization tuning on the realistic cyber dataset CSE-CIC-IDS2018 using cloud computing. Paper presented at the *Proceedings of the 2019 IEEE International Conference on Communication and Signal Processing, ICCSP*. 2019. P. 33-36. DOI: https://doi.org/10.1109/ICCSP.2019.8698029.

108. Syniavska O., Dekhtyar N., Deyneka O., Zhukova T., Syniavska O. Modeling the process of counteracting fraud in e-banking. Paper presented at *the CEUR Workshop Proceedings*. 2019, vol. 2422. P. 100–110.

109. Fedotova G. V., Gontar A. A., Titov V. A., Kurbanov A. K., Kuzmina E. V. Increasing the economic security of information banking systems. In book: *Ubiquitous Computing and the Internet of Things: Prerequisites for the Development of ICT*. 2019. P. 1153-1161. DOI: https://doi.org/10.1007/978-3-030-13397-9_118.

110. Akinbowale O. E., Klingelhöfer H. E., Zerihun M. F. Analysis of cyber-crime effects on the banking sector using the balanced score card: A survey of literature. *Journal of Financial Crime*. 2020, vol. 27(3). P. 945-958. DOI: https://doi.org/10.1108/JFC-03-2020-0037.

111. Yarovenko H. Development of algorithms for recognizing the cyber fraudsters' behaviour. *Міжнародний науковий журнал «Грааль науки», 2022. : за матеріалами V Міжнародної науково-практичної конференції «An integrated approach to science modernization: methods, models and multidisciplinarity»*, що проводилася 23 грудня 2022 року ГО «Європейська наукова платформа» (Вінниця, Україна) та ТОВ «International Centre Corporative Management» (Відень, Австрія). P. 265-266. DOI: https://doi.org/10.36074/grail-of-science.23.12.2022.39.

112. Яровенко Г.М. Жертва кіберзлочинів: ознаки та методи виявлення. Здобутки та досягнення прикладних та фундаментальних наук XXI століття: матеріали IV Міжнародної наукової конференції, м. Вінниця, 16 грудня, 2022 р. / Міжнародний центр наукових досліджень. – Вінниця: Європейська наукова платформа, 2022. С. 204-205. https://doi.org/10.36074/mcnd-16.12.2022.

113. Vojinovic I. (2022). More Than 70 Cybercrime Statistics - A $6 Trillion Problem. *Dataprot* : website. URL: https://dataprot.net/statistics/cybercrime-statistics/ (дата звернення 10.12.2022).

114. Cybersecurity. *Statista* : website. URL: https://www.statista.com/outlook/tmo/cybersecurity/worldwide (дата звернення 10.12.2022).

115. Estimated value of cyber insurance premiums written worldwide in 2018, 2020 and 2025. *Statista* : website. URL: https://www.statista.com/statistics/976526/global-cyber-insurance-market-size/ (дата звернення 10.12.2022).

116. Cyber defence. *NATO* : website. URL: https://www.nato.int/cps/en/natohq/topics_78170.htm (дата звернення 10.12.2022).

117. Cybersecurity. *United Nations* : website. URL: https://www.un.org/counterterrorism/cct/programme-projects/cybersecurity (дата звернення 10.12.2022).

118. FACT SHEET: Act Now to Protect Against Potential Cyberattacks. *The White House* : website. URL: https://www.whitehouse.gov/briefing-room/statements-

releases/2022/03/21/fact-sheet-act-now-to-protect-against-potential-cyberattacks/ (дата звернення 10.12.2022).

119. Maintaining a sustainable strengthened cyber security posture. *National Cyber Security Centre* : website. URL: https://www.ncsc.gov.uk/guidance/maintaining-a-sustainable-strengthened-cyber-security-posture (дата звернення 10.12.2022).

120. Leonov S., Frolov S., Plastun V. Potential of institutional investors and stock market development as an alternative to households' savings allocation in banks. *Economic Annals-XXI*. 2014, vol. 11-12. P. 65-68. URL: http://soskin.info/en/material/1/about-journal.html.

121. Brychko M., Bilan Y., Lyeonov S., Mentel G. Trust crisis in the financial sector and macroeconomic stability: A structural equation modelling approach. *Economic Research-Ekonomska Istrazivanja*. 2021, vol. 34(1). P. 828-855. DOI: https://doi.org/10.1080/1331677X.2020.1804970.

122. Tiutiunyk I. V., Zolkover A. O., Lyeonov S. V., Ryabushka L. B. The impact of economic shadowing on social development: challenges for macroeconomic stability. *Naukovyi Visnyk Natsionalnoho Hirnychoho Universytetu*. 2022, vol. 1. P. 183-191. DOI: https://doi.org/10.33271/nvngu/2022-1/183.

123. Sarwar M., Akram M., Shahzadi S. Bipolar fuzzy soft information applied to hypergraphs. *Soft Computing*. 2021, vol. 25(5). P. 3417-3439. DOI: https://doi.org/10.1007/s00500-021-05610-x.

124. Lyeonov S., Żurakowska-Sawa J., Kuzmenko O., Koibichuk V. Gravitational and intellectual data analysis to assess the money laundering risk of financial institutions. *Journal of International Studies*. 2020, vol. 13(4). P. 259-272. DOI: https://doi.org/10.14254/2071-8330.2020/13-4/18.

125. Kuzmenko O., Šuleř P., Lyeonov S., Judrupa I., Boiko A. Data mining and bifurcation analysis of the risk of money laundering with the involvement of financial institutions. *Journal of International Studies*. 2020, vol. 13(3). P. 332-339. DOI: https://doi.org/10.14254/2071-8330.2020/13-3/22.

126. Sivakumar P., Jayabalaguru V., Ramsugumar R., Kalaisriram S. Real Time Crime Detection Using Deep Learning Algorithm. In *2021 International Conference on System, Computation, Automation and Networking, ICSCAN 2021*. 2021. DOI: https://doi.org/10.1109/ICSCAN53069.2021.9526393.

127. Obeid H., Hillani F, Fakih R., Mozannar K. Artificial Intelligence: Serving American Security and Chinese Ambitions. *Financial Markets, Institutions and Risks*. 2020, vol. 4(3). P. 42-52. DOI: https://doi.org/10.21272/fmir.4(3).42-52.2020.

128. Kuzior, A., Brożek, P., Kuzmenko, O., Yarovenko, H., Vasilyeva, T. (2022). Countering cybercrime risks in financial institutions: forecasting information trends. *Journal of Risk Financial Management*. 2022, vol. 15. Art. no. 613. DOI: https://doi.org/10.3390/jrfm15120613.

129. Kuzmenko O.V., Kubálek J., Bozhenko V.V., Kushneryov O.S., Vida I. An approach to managing innovation to protect financial sector against cybercrime | Podejście do zarządzania innowacjami w celu ochrony sektora finansowego przed cyberprzestępczością. *Polish Journal of Management Studies*. 2021, 24(2). P. 276–291.

130. Lieonov S., Hlawiczka R., Boiko A., Mynenko S., Garai-Fodor M. Structural modelling for assessing the effectiveness of system for countering legalization of illicit money. *Journal of International Studies*. 2022, 15(3). P. 215-233. DOI: https://doi.org/10.14254/2071-8330.2022/15-3/15.

131. Kuzior A., Vasylieva T., Kuzmenko O., Koibichuk V., Brożek P. Global digital convergence: impact of cyber security, business transparency, economic transformation and AML efficiency. *Journal of Open Innovation: Technology, Market, and Complexity*. 2022, vol. 8. P. 195. DOI: https://doi.org/10.3390/joitmc8040195.

132. Vasilyeva T.A., Kuzmenko O.V., Stoyanets N.V., Artyukhov A.E., Bozhenko V.V. The depiction of cybercrime victims using data mining techniques. *Naukovyi Visnyk Natsionalnoho Hirnychoho Universytetu*. 2022, vol. 5. P. 174 – 178.

133. Kuzmenko O., Yarovenko H., Perkhun L. Assessing the maturity of the current global system for combating financial and cyber fraud. *STATISTICS IN TRANSITION new series and STATYSTYKA UKRAïNY. Joint Special Issue: A New*

Role for Statistics. 2022, Vol. 23(5). pp. 1–XX, DOI: https://doi.org/10.2478/stattrans-2022-xxx.

134. Kuzior A., Krawczyk D., Brożek P., Pakhnenko O., Vasylieva T., Lyeonov S. Resilience of smart cities to the consequences of the COVID-19 pandemic in the context of sustainable development. *Sustainability (Switzerland)*. 2022, vol. 14(19). DOI: https://doi.org/10.3390/su141912645.

135. Яровенко Г.М., Кочережченко Р.Д. Аналіз та моделювання соціально-економічного розвитку країн з урахуванням рівня їх кібербезпеки. *Вісник СумДУ. Серія «Економіка»*. 2022, № 1. С. 53-62. DOI: https://doi.org/10.21272/1817-9215.2022.1-5.

136. Кузьменко О.В., Яровенко Г.М., Скринька Л.О. Аналіз математичних моделей протидії банківським кібершахрайствам. *Вісник СумДУ. Серія «Економіка»*. 2022, № 2. С. 111-120 DOI: https://doi.org/10.21272/1817-9215.2022.2-13.

137. Яровенко Г.М., Ліцман М.А. Аналіз і прогнозування впливу рівня цифровізації країни на її економічний розвиток. *Вісник СумДУ. Серія «Економіка»*. 2021, № 4. С. 203-214. DOI: https://doi.org/10.21272/1817-9215.2021.4-24.

138. Кузьменко О.В., Бойко А.О, Доценко Т.В. Ризик легалізації коштів клієнтом банку від азартних ігор, що проводяться в мережі інтернет: підходи до вимірювання. *Вісник СумДУ. Серія «Економіка»*. 2022, № 3. С. 31-41. DOI: https://doi.org/10.21272/1817-9215.2022.3-3.

139. Yarovenko H., Rogkova M. Dynamic and bibliometric analysis of terms identifying the combating financial and cyber fraud system. *Financial Markets, Institutions and Risks*. 2022, 6(3). P. 93-104. DOI: https://doi.org/10.21272/fmir.6(3).93-104.2022.

140. Яровенко Г.М., Колотіліна О.В. Оцінювання взаємозалежності між втратою довіри до публічної влади та макроекономічною стабільністю України. *Міжнародний науковий журнал "Інтернаука". Серія: "Економічні науки"*. 2022, №11. DOI: https://doi.org/10.25313/2520-2294-2022-11-8437.

ДОДАТКИ

Додаток А

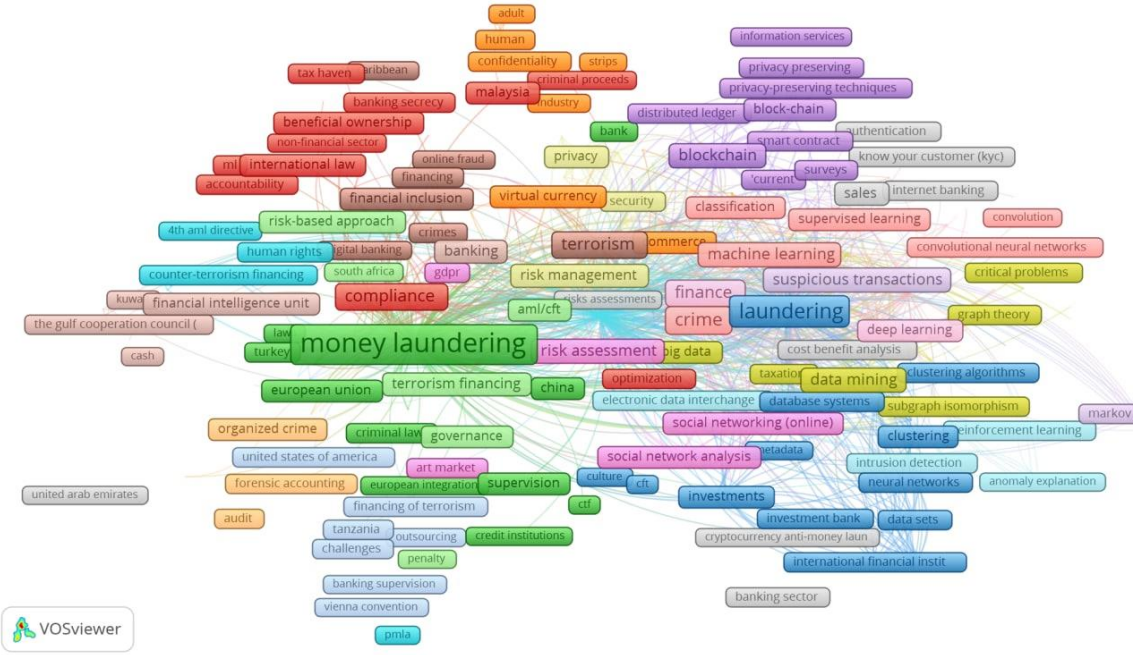


Рисунок А.1 – Наукова бібліографія поняття «anti-money laundering» (протидія легалізації доходів, отриманих незаконним шляхом) за 2012-2022 рр.

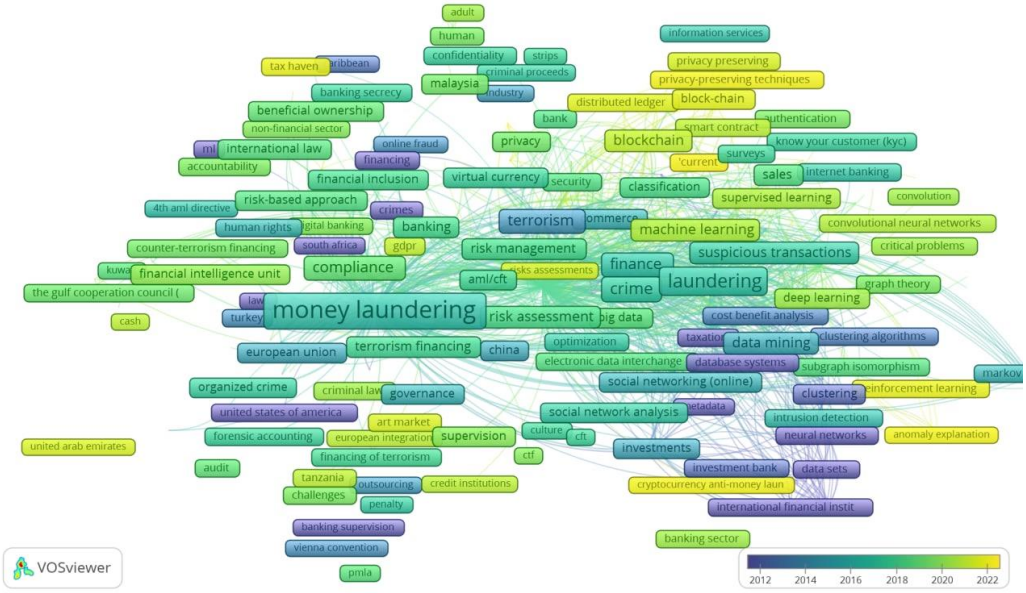


Рисунок А.2 – Візуалізація часового виміру досліджень стосовно протидії легалізації доходів, отриманих незаконним шляхом, опублікованих у виданнях, що індексуються наукометричною базою даних Scopus у 2012-2022 роках

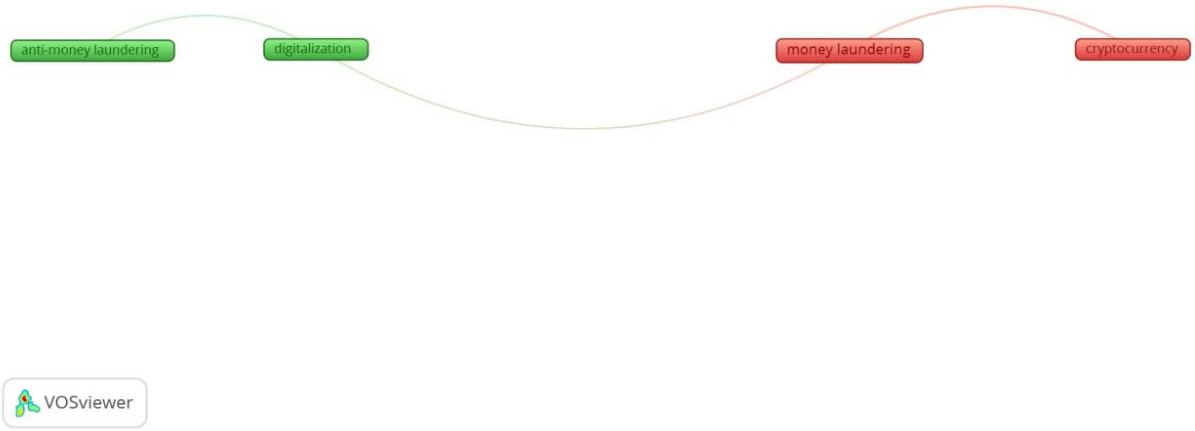


Рисунок А.3 – Візуалізація взаємозв’язку протидії легалізації доходів, отриманих незаконним шляхом з цифровізацією відповідно до досліджень, опублікованих у виданнях, що індексуються наукометричною базою даних Scopus у 2012-2022 роках

Додаток Б

Гістограми розподілу змінних

```
df.hist('PSI')
plt.ylabel("Frequency")
```

```
Text(0, 0.5, 'Frequency')
```

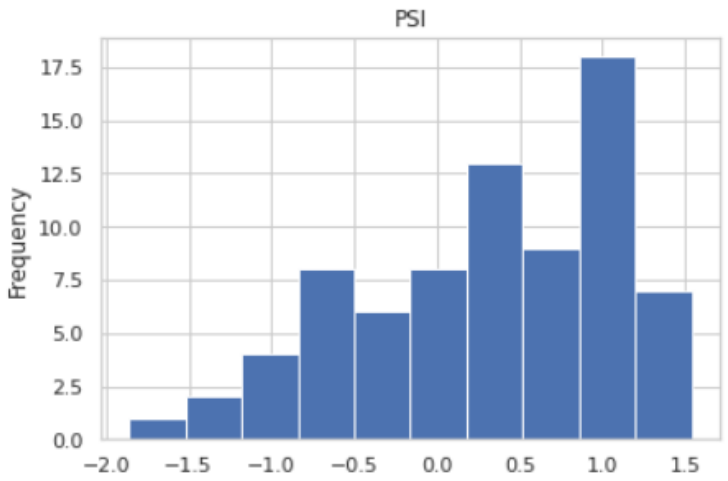


Рисунок Б.1 – Гістограма розподілу показника «PSI»

```
df.hist('EDB')
plt.ylabel("Frequency")
```

```
Text(0, 0.5, 'Frequency')
```

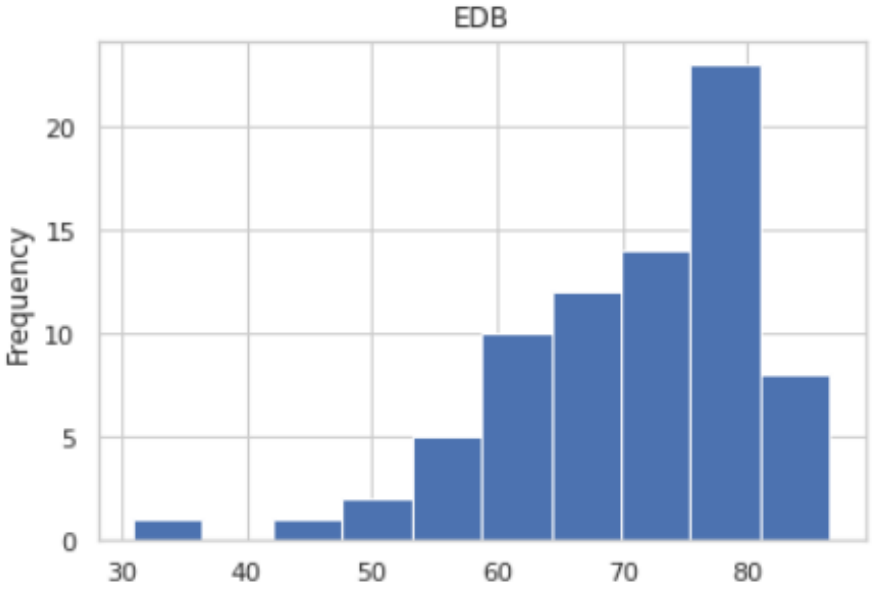


Рисунок Б.2 – Гістограма розподілу показника «EDB»

```
df.hist('CI')
plt.ylabel("Frequency")
```

```
Text(0, 0.5, 'Frequency')
```

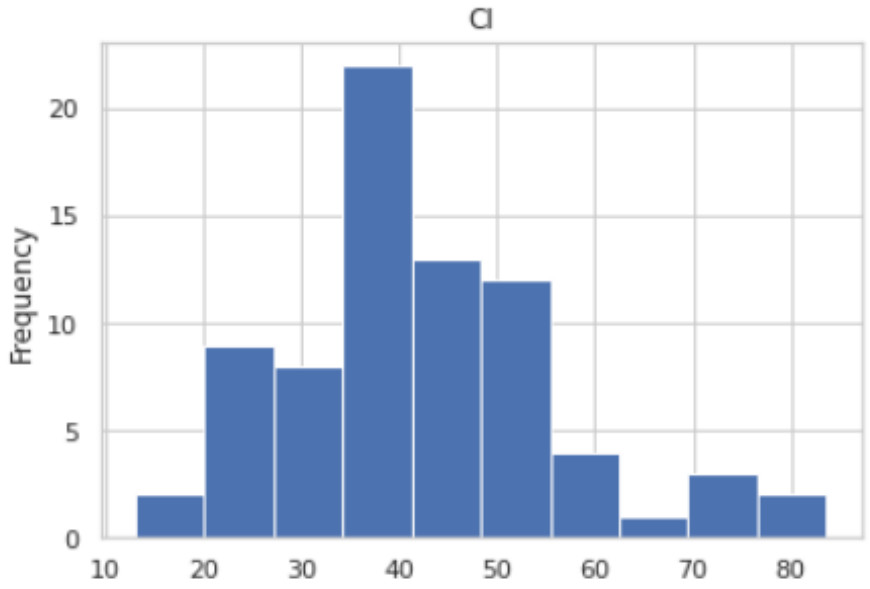


Рисунок Б.3 – Гістограма розподілу показника «CI»

```
df.hist('CPI')
plt.ylabel("Frequency")
```

```
Text(0, 0.5, 'Frequency')
```

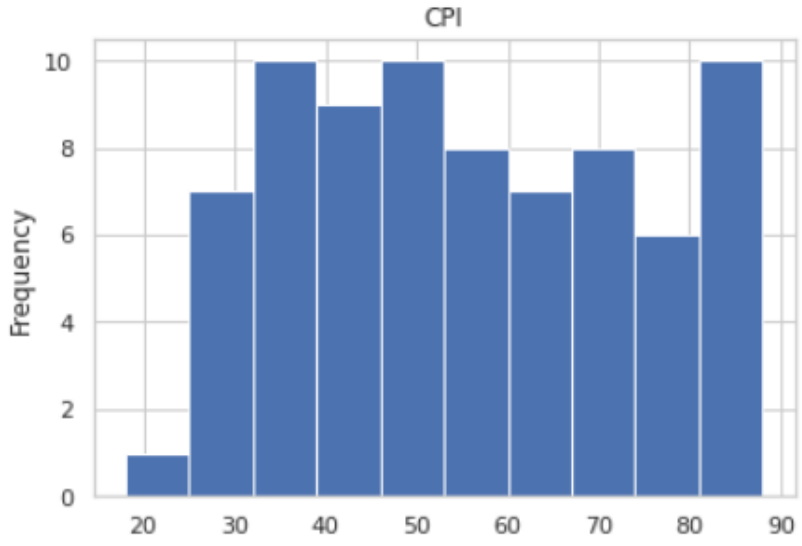


Рисунок Б.4 – Гістограма розподілу показника «CPI»

```
df.hist('GTI')
plt.ylabel("Frequency")
```

```
Text(0, 0.5, 'Frequency')
```

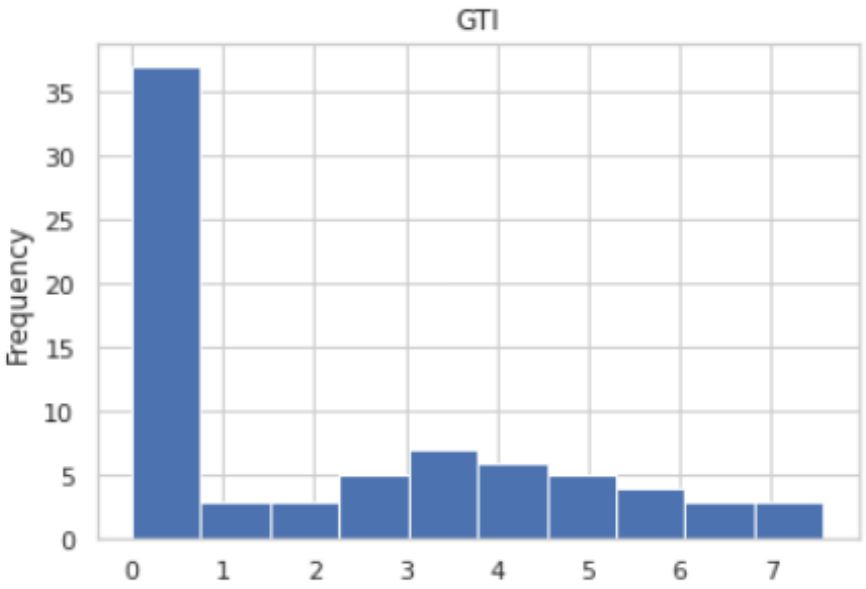


Рисунок Б.5 – Гістограма розподілу показника «GTI»

Додаток В

Результати канонічного аналізу

N=76	Canonical Analysis Summary (Convergensy.sta) Canonical R: .75906 Chi²(7)=60.519 p=0.0000	
	Left Set	Right Set
No. of variables	1	7
Variance extracted	100.000%	21.8918%
Total redundancy	57.6170%	12.6134%
Variables: 1	ICTDI	PSI
2		GEI
3		EDB
4		CI
5		CPI
6		GTI
7		FCI

Рисунок В.1 - Результати канонічного аналізу.

N=76	Canonical Analysis Summary (Convergensy.sta) Canonical R: .71093 Chi²(7)=49.636 p=0.0000	
	Left Set	Right Set
No. of variables	1	7
Variance extracted	100.000%	15.8106%
Total redundancy	50.5428%	7.99112%
Variables: 1	NRI	PSI
2		GEI
3		EDB
4		CI
5		CPI
6		GTI
7		FCI

Рисунок В.2 - Результати канонічного аналізу.

N=76	Canonical Analysis Summary (Convergensy.sta) Canonical R: .74559 Chi²(7)=57.225 p=0.0000	
	Left Set	Right Set
No. of variables	1	7
Variance extracted	100.000%	42.1733%
Total redundancy	55.5897%	23.4440%
Variables: 1	NCSI	PSI
2		GEI
3		EDB
4		CI
5		CPI
6		GTI
7		FCI

Рисунок В.3 - Результати канонічного аналізу.

N=76	Canonical Analysis Summary (Convergensy.sta) Canonical R: .95472 Chi²(7)=170.94 p=0.0000	
	Left Set	Right Set
No. of variables	1	7
Variance extracted	100.000%	43.6225%
Total redundancy	91.1491%	39.7615%
Variables: 1	DDL	PSI
2		GEI
3		EDB
4		CI
5		CPI
6		GTI
7		FCI

Рисунок В.4 - Результати канонічного аналізу.

Додаток Г

Нейромережева сітка

```
{'cv': 5,
 'error_score': nan,
 'estimator__activation': 'relu',
 'estimator__alpha': 0.0001,
 'estimator__batch_size': 'auto',
 'estimator__beta_1': 0.9,
 'estimator__beta_2': 0.999,
 'estimator__early_stopping': False,
 'estimator__epsilon': 1e-08,
 'estimator__hidden_layer_sizes': (45, 45, 45),
 'estimator__learning_rate': 'constant',
 'estimator__learning_rate_init': 0.001,
 'estimator__max_fun': 15000,
 'estimator__max_iter': 500,
 'estimator__momentum': 0.9,
 'estimator__n_iter_no_change': 10,
 'estimator__nesterovs_momentum': True,
 'estimator__power_t': 0.5,
 'estimator__random_state': None,
 'estimator__shuffle': True,
 'estimator__solver': 'adam',
 'estimator__tol': 0.0001,
 'estimator__validation_fraction': 0.1,
 'estimator__verbose': False,
 'estimator__warm_start': False,
 'estimator': MLPRegressor(activation='relu', alpha=0.0001, batch_size='auto', beta_1=0.9,
              beta_2=0.999, early_stopping=False, epsilon=1e-08,
              hidden_layer_sizes=(45, 45, 45), learning_rate='constant',
              learning_rate_init=0.001, max_fun=15000, max_iter=500,
              momentum=0.9, n_iter_no_change=10, nesterovs_momentum=True,
              power_t=0.5, random_state=None, shuffle=True, solver='adam',
              tol=0.0001, validation_fraction=0.1, verbose=False,
              warm_start=False),
 'iid': 'deprecated',
 'n_jobs': -1,
 'param_grid': {'hidden_layer_sizes': [(40, 40, 40),
   (35, 35, 35),
   (30, 30, 30)],
  'max_iter': [100, 500],
  'activation': ['tanh', 'relu'],
  'solver': ['sgd', 'adam'],
  'alpha': [0.0001, 0.05],
  'learning_rate': ['constant', 'adaptive']},
 'pre_dispatch': '2*n_jobs',
 'refit': True,
 'return_train_score': False,
 'scoring': None,
 'verbose': 0}
```

Рисунок Г.1 – Характеристика параметрів нейромережевої сітки

Додаток Д

Результати кластерного аналізу і формування кіберпрофілів

Рисунок Д.1 – Формування кіберпрофілів (продовження)

Поля	Показатели	4	3	1	7	9	5	6	0	8	2	Итого
		9311 (37,5%)	4593 (18,5%)	2990 (12,0%)	2930 (11,8%)	2544 (10,2%)	964 (3,9%)	567 (2,3%)	477 (1,9%)	417 (1,7%)	32 (0,1%)	
REG_REGION_NOT_WORK_REGION	Значимость	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	98,7%	100,0%
	Распределение											
	False	9311 (100,0%)	4593 (100,0%)	2990 (100,0%)	2930 (100,0%)	2494 (98,0%)	81 (8,4%)	501 (88,4%)	477 (100,0%)	33 (7,9%)	27 (84,4%)	23437 (94,4%)
	True	0 (0,0%)	0 (0,0%)	0 (0,0%)	0 (0,0%)	50 (2,0%)	883 (91,6%)	66 (11,6%)	0 (0,0%)	384 (92,1%)	5 (15,6%)	1388 (5,6%)
REG_REGION_NOT_LIVE_REGION	Значимость	100,0%	100,0%	100,0%	100,0%	60,6%	100,0%	99,8%	99,6%	100,0%	100,0%	100,0%
	Распределение											
	False	9311 (100,0%)	4593 (100,0%)	2974 (99,5%)	2930 (100,0%)	2494 (98,0%)	671 (69,6%)	567 (100,0%)	477 (100,0%)	348 (83,5%)	27 (84,4%)	24392 (98,3%)
	True	0 (0,0%)	0 (0,0%)	16 (0,5%)	0 (0,0%)	50 (2,0%)	293 (30,4%)	0 (0,0%)	0 (0,0%)	69 (16,5%)	5 (15,6%)	433 (1,7%)
9.0 HOUR_APPR_PROCESS_START	Значимость	100,0%	100,0%	100,0%	99,5%	98,0%	100,0%	99,9%	99,0%	100,0%	5,9%	100,0%
	Доверительный интервал											
	Среднее	11,63590756	12,16011878	11,26407873	11,98231769	11,63663342	12,6421292	11,34744268	12,1815286	13,13329225	11,75	11,79733845
	Стандартн. откл.	3,241909449	3,354384019	3,087345393	3,372272868	3,319035696	3,212653968	3,048070058	3,212693949	3,006106313	3,610021893	3,280111194
	Стандартн. ошиб.	0,03359717155	0,04949536801	0,05646113759	0,06230012215	0,06580416305	0,1034726171	0,1280069104	0,1470991509	0,1472096595	0,6381677401	0,02081823649
9.0 REGION_RATING_CLIENT_W_CITY	Значимость	100,0%	100,0%	32,4%	100,0%	100,0%	100,0%	100,0%	99,0%	100,0%	80,0%	100,0%
	Доверительный интервал											
	Среднее	2,214799699	2,040278685	2,138795987	2,068600683	2,201650943	1,995850622	2,220458554	2,075471698	1,683453237	2,25	2,134823766
	Стандартн. откл.	0,4747547958	0,4994658391	0,4892281169	0,4996452353	0,4967662103	0,5094979535	0,4516218922	0,4963797909	0,6205918146	0,508000508	0,4999091464
	Стандартн. ошиб.	0,00492069043	0,00736983165	0,008946966572	0,009230557672	0,009849030771	0,01640982414	0,01896633672	0,02272766934	0,03039051191	0,08980265101	0,003172827449
9.0 REGION_RATING_CLIENT	Значимость	100,0%	100,0%	38,9%	100,0%	100,0%	100,0%	100,0%	98,9%	100,0%	83,3%	100,0%
	Доверительный интервал											
	Среднее	2,224250886	2,075332027	2,158528428	2,108191126	2,211477987	1,997925311	2,222222222	2,094339623	1,695443645	2,28125	2,153635448
	Стандартн. откл.	0,4755870531	0,5154273114	0,4958366394	0,5175985904	0,4954471611	0,5095106387	0,4527107205	0,5036728901	0,624656342	0,5226714875	0,5050516003
	Стандартн. ошиб.	0,00492869405	0,007605349987	0,009067822727	0,009562231964	0,009822878918	0,01641023271	0,01901206321	0,02306159741	0,03058955229	0,09239613829	0,003205465618
CNT_FAM_MEMBERS	Значимость	100,0%	99,9%	100,0%	99,9%	100,0%	98,9%	99,8%	40,6%	100,0%	29,8%	100,0%
	Доверительный интервал											
	Среднее	09.02.2000 23:29:32	04.02.2000 9:14:27	02.01.2000 15:31:04	07.02.2000 15:30:21	08.02.2000 20:03:24	03.02.2000 13:39:51	09.02.2000 14:53:18	06.02.2000 13:21:16	01.01.2000 18:18:29	08.02.2000 4:30:00	05.02.2000 20:43:24
	Стандартн. откл.	29дн. 16:33:21	26дн. 19:15:16	18дн. 01:30:20	28дн. 02:37:39	28дн. 08:48:27	27дн. 09:47:25	27дн. 22:45:33	28дн. 03:33:51	24дн. 15:54:03	34дн. 09:04:03	28дн. 00:41:05
	Стандартн. ошиб.	07:23:04	09:29:29	07:55:40	12:27:47	13:29:52	21:11:09	1дн. 04:10:09	1дн. 06:55:55	1дн. 04:59:07	6дн. 01:51:09	04:16:09

Рисунок Д.2 – Формування кіберпрофілів (продовження)

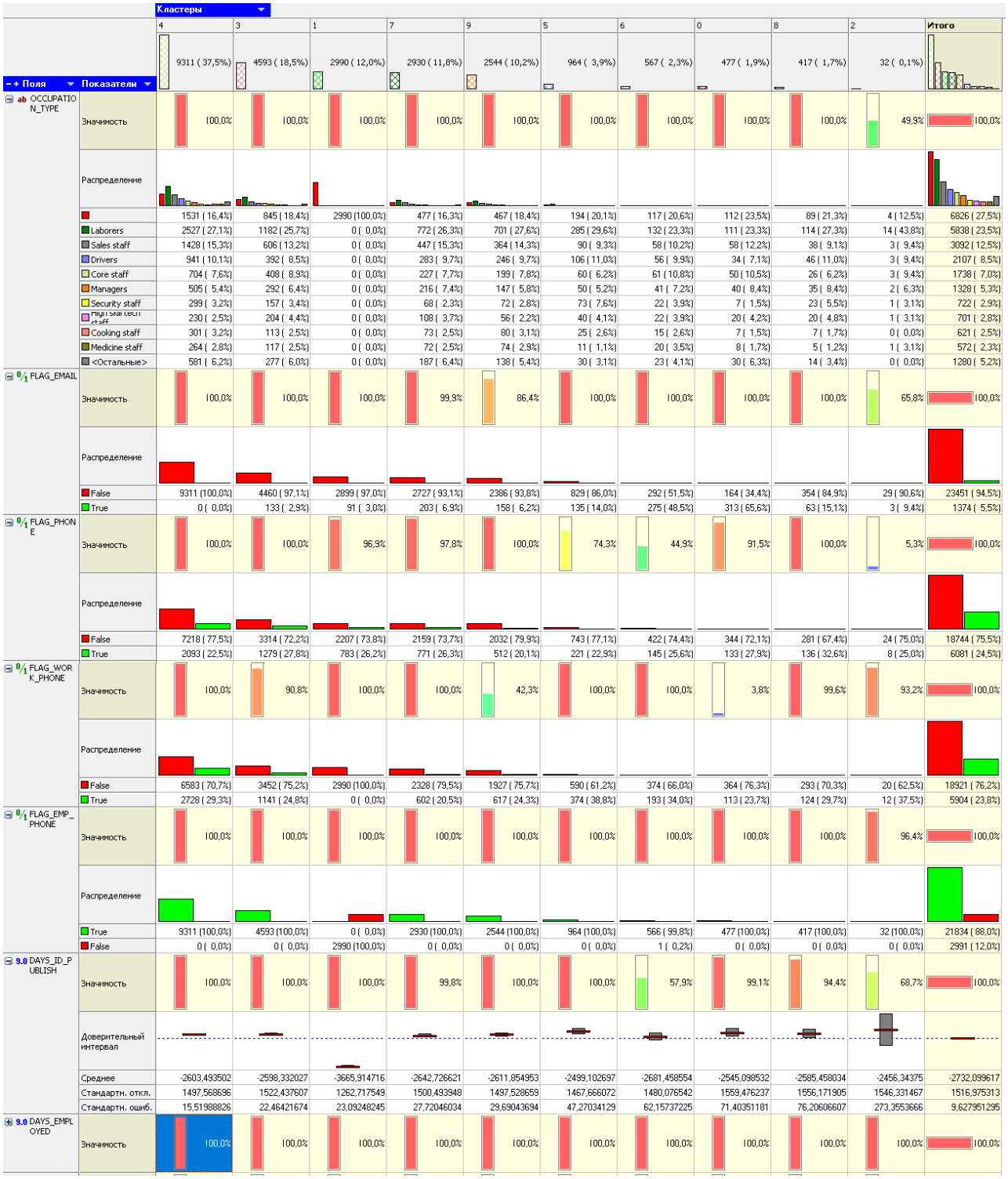


Рисунок Д.3 – Формування кіберпрофілів (продовження)

+ Поля	Показатели	Кластеры 4	3	1	7	9	5	6	0	8	2	Итого
		9311 (37,5%)	4593 (18,5%)	2990 (12,0%)	2930 (11,8%)	2544 (10,2%)	964 (3,9%)	567 (2,3%)	477 (1,9%)	417 (1,7%)	32 (0,1%)	
9.0 DAYS_EMPLOYED	Значимость	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%
	Доверительный интервал											
	Среднее	-1578,735796	-2052,988025	365243	-1875,920819	-1844,283412	-1151,283195	-4510,641975	-2093,8826	-1464,009592	-1088,09375	42394,67545
	Стандартн. откл.	1352,952523	2119,172125	0	1821,671059	1897,177045	1055,036468	3818,340369	2301,716142	1435,06535	871,0591758	119484,6343
	Стандартн. ошиб.	14,02117448	31,26928926	0	33,65395802	37,6139816	33,98043661	160,3552228	105,3883425	70,27545252	153,9829625	758,3460516
9.0 DAYS_BIRTH	Значимость	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	99,3%	100,0%	100,0%	100,0%	100,0%
	Доверительный интервал											
	Среднее	-13778,27666	-14320,0614	-21441,499	-14144,72526	-14025,77634	-13286,38278	-15306,46914	-13482,37526	-14149,57554	-12508,5	-14884,82808
	Стандартн. откл.	3480,155285	3619,295161	2388,188281	3471,81797	3554,2711	3426,997579	3696,991431	3392,249512	3602,402471	3105,534144	4192,844583
	Стандартн. ошиб.	36,06620603	53,40424498	43,67500553	64,13914061	70,4680083	110,37616	155,2590464	155,3204353	176,410408	548,9860632	26,61118021
REGION_POPULATION_RELATIVE	Значимость	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	98,0%	100,0%
	Распределение											
0.035792000000000004		622 (6,7%)	85 (1,9%)	135 (4,5%)	61 (2,1%)	169 (6,6%)	30 (3,1%)	29 (5,1%)	17 (3,6%)	5 (1,2%)	0 (0,0%)	1153 (4,6%)
0.025164		340 (3,7%)	266 (5,8%)	121 (4,0%)	154 (5,3%)	95 (3,7%)	23 (2,4%)	24 (4,2%)	21 (4,4%)	4 (1,0%)	0 (0,0%)	1048 (4,2%)
0.030755		413 (4,4%)	102 (2,2%)	132 (4,4%)	71 (2,4%)	126 (5,0%)	48 (5,0%)	26 (4,6%)	22 (4,6%)	8 (1,9%)	2 (6,3%)	950 (3,8%)
0.028663		305 (3,3%)	171 (3,7%)	150 (5,0%)	91 (3,1%)	92 (3,6%)	41 (4,3%)	47 (8,3%)	28 (5,9%)	11 (2,6%)	1 (3,1%)	937 (3,8%)
0.031329		300 (3,2%)	254 (5,5%)	109 (3,6%)	145 (4,9%)	76 (3,0%)	11 (1,1%)	10 (1,8%)	8 (1,7%)	8 (1,9%)	0 (0,0%)	921 (3,7%)
0.020246		316 (3,4%)	167 (3,6%)	117 (3,9%)	100 (3,4%)	102 (4,0%)	11 (1,1%)	13 (2,3%)	16 (3,4%)	9 (2,2%)	1 (3,1%)	852 (3,4%)
0.020713		304 (3,3%)	185 (4,0%)	108 (3,6%)	131 (4,5%)	83 (3,3%)	8 (0,8%)	10 (1,8%)	11 (2,3%)	7 (1,7%)	0 (0,0%)	847 (3,4%)
0.018029		314 (3,4%)	128 (2,8%)	87 (2,9%)	96 (3,3%)	74 (2,9%)	10 (1,0%)	15 (2,6%)	11 (2,3%)	1 (0,2%)	5 (15,6%)	741 (3,0%)
0.026392000000000002		168 (1,8%)	215 (4,7%)	90 (3,0%)	138 (4,7%)	57 (2,2%)	7 (0,7%)	13 (2,3%)	23 (4,8%)	10 (2,4%)	1 (3,1%)	722 (2,9%)
0.04622		124 (1,3%)	150 (3,3%)	48 (1,6%)	94 (3,2%)	42 (1,7%)	81 (8,4%)	5 (0,9%)	12 (2,5%)	107 (25,7%)	0 (0,0%)	663 (2,7%)
	<Остальные>	6105 (65,6%)	2870 (62,5%)	1893 (63,3%)	1849 (63,1%)	1628 (64,0%)	694 (72,0%)	375 (66,1%)	308 (64,6%)	247 (59,2%)	22 (68,8%)	15991 (64,4%)
AMT_GOODS_PRICE	Значимость	100,0%	100,0%	100,0%	100,0%	100,0%	99,7%	96,4%	100,0%	91,8%	85,7%	100,0%
	Распределение											
450000.0		2105 (22,6%)	124 (2,7%)	1018 (34,0%)	2929 (100,0%)	2291 (90,1%)	324 (33,6%)	179 (31,6%)	61 (12,8%)	136 (32,6%)	5 (15,6%)	9172 (36,9%)
225000.0		1111 (11,9%)	632 (13,8%)	298 (10,0%)	0 (0,0%)	11 (0,4%)	91 (9,4%)	42 (7,4%)	74 (15,5%)	29 (7,0%)	6 (18,8%)	2294 (9,2%)
675000.0		970 (10,4%)	636 (13,8%)	333 (11,1%)	0 (0,0%)	0 (0,0%)	114 (11,8%)	53 (9,3%)	45 (9,4%)	53 (12,7%)	0 (0,0%)	2204 (8,9%)
900000.0		504 (5,4%)	347 (7,6%)	73 (2,4%)	0 (0,0%)	0 (0,0%)	52 (5,4%)	24 (4,2%)	19 (4,0%)	28 (6,7%)	1 (3,1%)	1048 (4,2%)
180000.0		438 (4,7%)	253 (5,5%)	58 (1,9%)	0 (0,0%)	7 (0,3%)	35 (3,6%)	15 (2,6%)	24 (5,0%)	22 (5,3%)	2 (6,3%)	854 (3,4%)
270000.0		405 (4,3%)	251 (5,5%)	75 (2,5%)	0 (0,0%)	11 (0,4%)	34 (3,5%)	23 (4,1%)	16 (3,4%)	14 (3,4%)	1 (3,1%)	830 (3,3%)
135000.0		287 (3,1%)	129 (2,8%)	51 (1,7%)	0 (0,0%)	2 (0,1%)	28 (2,9%)	12 (2,1%)	15 (3,1%)	9 (2,2%)	1 (3,1%)	534 (2,2%)
454500.0		232 (2,5%)	124 (2,7%)	99 (3,3%)	0 (0,0%)	6 (0,2%)	19 (2,0%)	20 (3,5%)	16 (3,4%)	7 (1,7%)	1 (3,1%)	524 (2,1%)
1125000.0		207 (2,2%)	185 (4,0%)	63 (2,1%)	0 (0,0%)	8 (0,3%)	19 (2,0%)	13 (2,3%)	11 (2,3%)	10 (2,4%)	0 (0,0%)	516 (2,1%)
360000.0		227 (2,4%)	122 (2,7%)	67 (2,2%)	0 (0,0%)	7 (0,3%)	17 (1,8%)	16 (2,8%)	18 (3,8%)	11 (2,6%)	2 (6,3%)	487 (2,0%)
	<Остальные>	2825 (30,3%)	1790 (39,0%)	855 (28,6%)	1 (0,0%)	201 (7,9%)	231 (24,0%)	170 (30,0%)	178 (37,3%)	98 (23,5%)	13 (40,6%)	6362 (25,6%)
AMT_CREDIT	Значимость	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	36,3%	90,4%	100,0%
	Распределение											
450000.0		168 (1,8%)	247 (5,4%)	660 (22,1%)	1407 (48,0%)	2544 (100,0%)	138 (14,3%)	49 (8,6%)	80 (16,8%)	92 (22,1%)	2 (6,3%)	5387 (21,7%)
545040.0		375 (4,0%)	1 (0,0%)	45 (1,5%)	251 (8,6%)	0 (0,0%)	42 (4,4%)	22 (3,9%)	5 (1,0%)	4 (1,0%)	1 (3,1%)	746 (3,0%)
675000.0		301 (3,2%)	189 (4,1%)	95 (3,2%)	0 (0,0%)	0 (0,0%)	32 (3,3%)	16 (2,8%)	25 (5,2%)	10 (2,4%)	0 (0,0%)	668 (2,7%)
225000.0		276 (3,0%)	201 (4,4%)	62 (2,1%)	0 (0,0%)	0 (0,0%)	24 (2,5%)	11 (1,9%)	15 (3,1%)	7 (1,7%)	1 (3,1%)	597 (2,4%)
180000.0		311 (3,3%)	189 (4,1%)	23 (0,8%)	3 (0,1%)	0 (0,0%)	22 (2,3%)	7 (1,2%)	17 (3,6%)	18 (4,3%)	2 (6,3%)	592 (2,4%)
521280.0		228 (2,4%)	4 (0,1%)	28 (0,9%)	174 (5,9%)	0 (0,0%)	26 (2,7%)	16 (2,8%)	2 (0,4%)	9 (2,2%)	0 (0,0%)	487 (2,0%)
270000.0		222 (2,4%)	155 (3,4%)	29 (1,0%)	1 (0,0%)	0 (0,0%)	15 (1,6%)	9 (1,6%)	7 (1,5%)	9 (2,2%)	0 (0,0%)	447 (1,8%)
640080.0		206 (2,2%)	1 (0,0%)	17 (0,6%)	149 (5,1%)	0 (0,0%)	17 (1,8%)	10 (1,8%)	4 (0,8%)	7 (1,7%)	1 (3,1%)	412 (1,7%)
808650.0		157 (1,7%)	122 (2,7%)	74 (2,5%)	0 (0,0%)	0 (0,0%)	13 (1,3%)	9 (1,6%)	0 (0,0%)	14 (3,4%)	0 (0,0%)	389 (1,6%)
284400.0		196 (2,1%)	103 (2,2%)	51 (1,7%)	0 (0,0%)	0 (0,0%)	13 (1,3%)	5 (0,9%)	11 (2,3%)	3 (0,7%)	1 (3,1%)	383 (1,5%)
	<Остальные>	6871 (73,8%)	3381 (73,6%)	1906 (63,7%)	945 (32,3%)	0 (0,0%)	622 (64,5%)	413 (72,8%)	311 (65,2%)	244 (58,5%)	24 (75,0%)	14717 (59,3%)

Рисунок Д.4 – Формування кіберпрофілів (продовження)

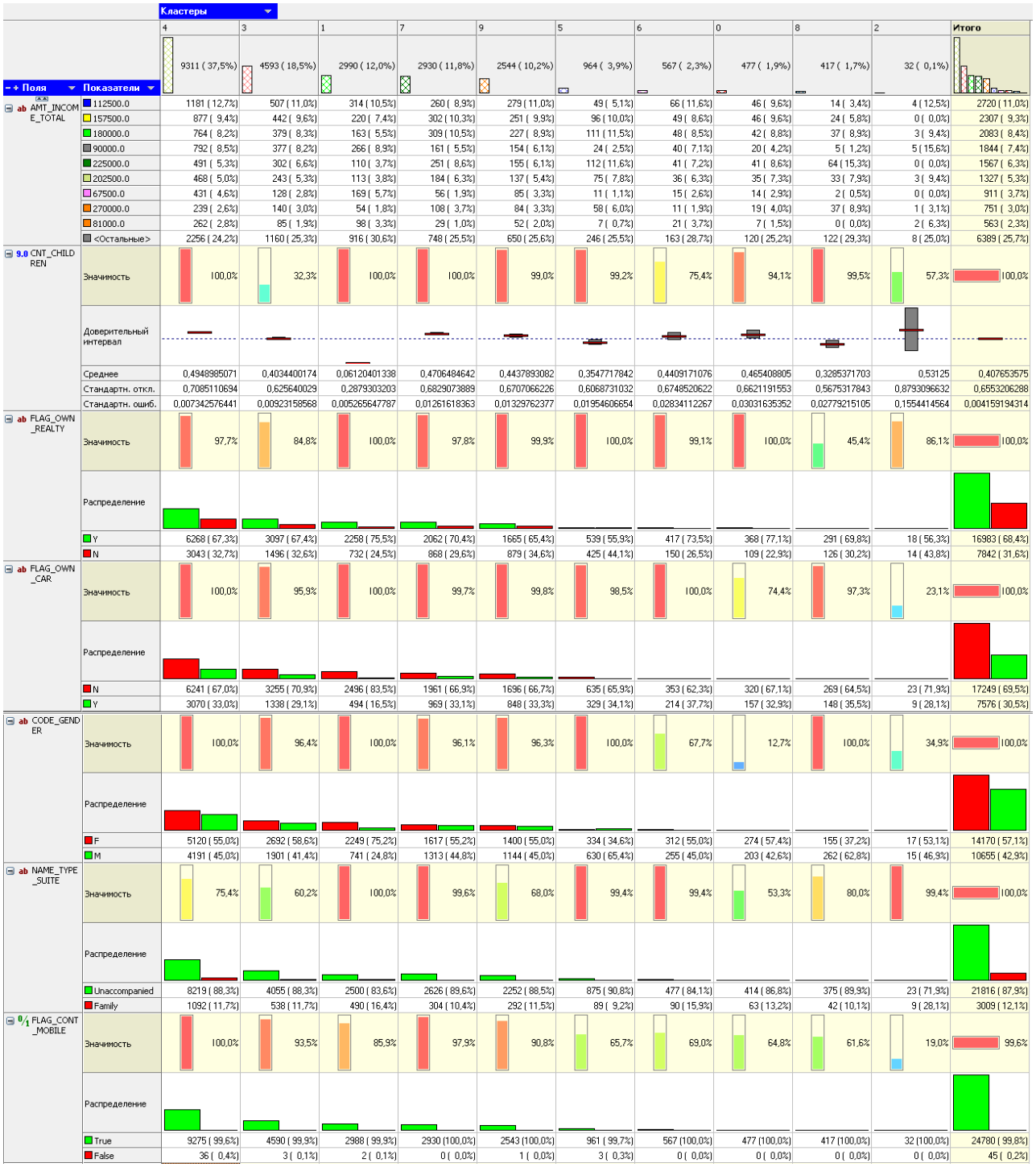


Рисунок Д.5 – Формування кіберпрофілів (завершення)

Додаток Е

Розрахунки для прогнозних моделей інформаційних трендів кібератак

Таблиця Е.1 – Значення циклічних складових інформаційних трендів запитів щодо кібератак на комп’ютерні системи, мережеву та хмарну інфраструктуру фінансової установи

Спостере-ження	CS	NI	CI	Спостере-ження	CS	NI	CI
1	-1,63689	-6,11530	-4,06617	25	-1,84210	1,41283	4,52446
2	-6,18898	-4,03717	-3,40992	26	0,44540	3,27533	2,42862
3	-0,69939	-2,39134	-3,92033	27	1,36207	9,50033	9,33279
4	-0,00148	-3,16738	-3,75367	28	0,86623	8,00866	13,63696
5	1,21207	5,62428	1,33487	29	-0,99210	0,76699	11,32029
6	0,92561	4,35866	-1,19117	30	3,42457	2,01283	-0,19221
7	0,85269	2,42637	-0,24846	31	2,66623	2,08783	2,77029
8	-1,15773	4,69720	-0,05054	32	5,04123	-0,52884	-0,73804
9	-5,46502	-2,29238	-3,58700	33	2,61623	1,20449	0,16196
10	-3,00668	-3,75592	1,90779	34	3,56623	0,55449	-2,74221
11	-2,29835	-0,02676	-0,69638	35	5,89540	6,08366	-1,06304
12	-2,07439	-1,82363	0,97550	36	3,00790	0,03783	-0,61721
13	-0,42856	-6,36530	2,64217	37	1,10373	-6,08717	-2,17554
14	-1,09523	-2,40697	0,99633	38	3,82457	3,71699	-0,27971
15	-3,49627	3,50449	3,92342	39	10,70373	4,72116	-1,03388
16	-4,42335	-1,83405	4,02237	40	9,60373	0,85449	-1,97138
17	-2,62648	-3,15176	1,88175	41	2,10790	-6,19551	-5,48804
18	-0,80877	1,03574	1,06404	42	3,20790	-0,80801	-0,96721
19	-4,70981	-7,51113	-1,30575	43	3,33290	-2,43301	-0,86721
20	-4,57439	-3,53197	-2,12867	44	-0,32127	-2,45801	-2,14638
21	-6,48064	4,44720	1,03279	45	1,40790	-5,09134	-3,67138
22	-5,88689	0,91595	-3,83700	46	-2,22960	-3,02884	-3,78804
23	-3,93898	-3,70905	1,33487	47	1,89436	2,86908	-2,35783
24	-5,81398	0,52533	-3,27971	48	3,12873	4,10866	-3,71721

Таблиця Е.2 – Прогнозований рівень кібератак на комп’ютерні системи, мережу та хмарну інфраструктуру фінансової установи

Дата	CS	NI	CI	Дата	CS	NI	CI
17.04.2022	70	75	46	04.12.2022	80	77	51
24.04.2022	72	69	51	11.12.2022	77	78	51
01.05.2022	68	76	47	18.12.2022	72	69	47
08.05.2022	75	74	55	25.12.2022	74	69	53
15.05.2022	78	72	52	01.01.2023	71	73	50
22.05.2022	79	81	59	08.01.2023	77	70	52
29.05.2022	79	76	64	15.01.2023	74	67	54
05.06.2022	78	70	61	22.01.2023	77	69	52
12.06.2022	82	75	50	29.01.2023	77	78	55
19.06.2022	80	76	53	05.02.2023	75	70	55
26.06.2022	84	71	49	12.02.2023	77	69	53
03.07.2022	83	76	50	19.02.2023	79	74	52
10.07.2022	80	74	47	26.02.2023	76	63	50
17.07.2022	84	80	49	05.03.2023	75	69	49
24.07.2022	77	73	50	12.03.2023	74	81	52
31.07.2022	81	66	48	19.03.2023	74	75	47
07.08.2022	82	78	50	26.03.2023	76	69	53
14.08.2022	88	76	49	02.04.2023	72	76	48
21.08.2022	90	75	48	09.04.2023	79	74	56
28.08.2022	79	65	45	16.04.2023	82	72	54
04.09.2022	84	75	49	23.04.2023	83	81	61
11.09.2022	82	69	50	30.04.2023	83	76	65
18.09.2022	75	70	48	07.05.2023	82	70	63
25.09.2022	82	69	47	14.05.2023	86	75	51
02.10.2022	79	67	47	21.05.2023	84	76	54
09.10.2022	79	75	48	28.05.2023	88	71	51
16.10.2022	82	78	47	04.06.2023	87	76	52
23.10.2022	81	64	47	11.06.2023	84	74	49
30.10.2022	77	72	47	18.06.2023	88	80	50
06.11.2022	80	73	47	25.06.2023	80	73	51
13.11.2022	79	67	47	02.07.2023	85	66	49
20.11.2022	80	85	52	09.07.2023	86	78	51
27.11.2022	80	79	50				