МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Сумський державний університет
Навчально-науковий інститут бізнесу, економіки та менеджменту
Кафедра економічної кібернетики

«До захисту допущено»

Завідувач кафедри

_____ Віталія КОЙБІЧУК
(підпис) (Ім'я та ПРІЗВИЩЕ)

_____ 2023 р.

# КВІЛІФІКАЦІЙНА РОБОТА

**на здобуття освітнього ступеня** бакалавр
(бакалавр / магістр)

зі спеціальності 051 «Економіка»,
(код та назва)

освітньо-професійної програми Економічна кібернетика
(освітньо-професійної / освітньо-наукової) (назва програми)

на тему: Прогнозування інформаційних трендів кіберзлочинів

Здобувачки групи ЕК-91а Солярової Катерини Геннадіївни
(шифр групи) (прізвище, ім'я, по-батькові)

Кваліфікаційна робота містить результати власних досліджень. Використання ідей, результатів і текстів інших авторів мають посилання на відповідне джерело.

Керівник доцентка, д. е. н., Ганна Яровенко _____
(посада, науковий ступінь, вчене звання, Ім'я та ПРІЗВИЩЕ) (підпис)

Суми – 2023

Міністерство освіти і науки України
Сумський державний університет
Навчально-науковий інститут бізнесу, економіки та менеджменту
Кафедра економічної кібернетики

ЗАТВЕРДЖУЮ
Завідувач кафедри
к.е.н., доцент
_____Віталія КОЙБІЧУК
“___” _____ 2023 р.

ЗАВДАННЯ НА КВАЛІФІКАЦІЙНУ БАКАЛАВРСЬКУ РОБОТУ

(спеціальність 051 Економіка «Економічна кібернетика», «Бізнес аналітика»)

студенту 4 курсу, групи ЕК-91а

Соляровій Катерині Геннадіївні
(прізвище, ім’я, по батькові студента)

1. Тема роботи Прогнозування інформаційних трендів кіберзлочинів затверджена наказом по університету від «23» травня 2023 року № 0554-VI

2. Термін подання студентом закінченої роботи «16» червня 2023 року

3. Мета кваліфікаційної роботи полягає в розробці прогнозних моделей інформаційних трендів кібератак.

4. Об’єкт дослідження — інформація, що ідентифікує кількість запитів користувачів глобальної мережі, що є свідченням реакцій на масовість кіберзлочинів.

5. Предмет дослідження статистичний, аналітичний та програмний інструментарій для прогнозування інформаційних трендів кібератак.

6. Кваліфікаційна робота виконується на матеріалах аналітичної системи Google Trends.

7. Орієнтовний план кваліфікаційної роботи, терміни подання розділів керівникові та зміст завдань для виконання поставленої мети

Розділ 1 Аналіз інформаційних трендів кібератак «23» травня 2023 року

У розділі 1 необхідно охарактеризувати проблематику кібератак та напрямки боротьби з ними, проаналізувати соціальну інженерію, DoS-атаки та атаки на пароль, провести первинний статистичний аналіз трендів кібератак.

Розділ 2 Статистичні тести інформаційних трендів кібератак «3» червня 2023 року

У розділі 2 необхідно перевірити ряди на відповідність нормальному закону розподілу, перевірити ряди на стаціонарність, провести перевірку сезонної компоненти.

Розділ 3 Побудова моделей та прогнозів «10» червня 2023 року

У розділі 3 необхідно охарактеризувати сутність математичного апарату авторегресійних моделей, описати результати побудови авторегресійних моделей, зпрогнозувати інформаційні тренди кібератак, верифікувати якість отриманих прогнозів.

8. Консультації з роботи:

| Розділ | Прізвище, ініціали та посада консультанта | Підпис, дата | |
|---|---|---|---|
| | | завдання видав | завдання прийняв |
| 1 | Яровенко Г.М., доцент | Яровенко Г.М. 03.04.23 | Солярова К.Г. 03.04.23 |
| 2 | Яровенко Г.М., доцент | Яровенко Г.М. 23.05.23 | Солярова К.Г. 23.05.23 |
| 3 | Яровенко Г.М., доцент | Яровенко Г.М. 03.06.23 | Солярова К.Г. 03.06.23 |

9. Дата видачі завдання: «10» квітня 2023 року

Керівник кваліфікаційної роботи _____ Г. М. Яровенко
( підпис) (ініціали, прізвище)

Завдання до виконання одержав _____ К. Г. Солярова
(підпис) (ініціали, прізвище)

# АНОТАЦІЯ

кваліфікаційної роботи бакалавра на тему

«ПРОГНОЗУВАННЯ ІНФОРМАЦІЙНИХ ТРЕНДІВ КІБЕРЗЛОЧИНІВ»

студентки Солярової Катерини Геннадіївни
(прізвище, ім’я, по батькові)

Актуальність теми, обраної для дослідження, визначається тим, що дозволяє зрозуміти, які нові загрози можуть з'явитися, які різновиди кібератак можуть стати популярними та які сектори чи організації можуть стати об'єктами нападу. Це допомагає підготуватися до майбутніх загроз, розробити відповідні заходи безпеки та зменшити ризики вразливості.

Мета кваліфікаційної роботи полягає в розробці прогнозних моделей інформаційних трендів кібератак.

Об’єктом дослідження є інформація, що ідентифікує кількість запитів користувачів глобальної мережі, що є свідченням реакцій на масовість кіберзлочинів.

Предметом дослідження є статистичний, аналітичний та програмний інструментарій для прогнозування інформаційних трендів кібератак.

Задачами дослідження є:

1) охарактеризувати проблематику кібератак та напрямки боротьби з ними;

2) проаналізувати соціальну інженерію, DoS-атаки та атаки на пароль;

3) провести первинний статистичний аналіз трендів кібератак;

4) перевірити ряди на відповідність нормальному закону розподілу;

5) перевірити ряди на стаціонарність;

6) провести перевірку сезонної компоненти;

7) охарактеризувати сутність математичного апарату авторегресійних моделей;

8) описати результати побудови авторегресійних моделей;

9) зпрогнозувати інформаційні тренди кібератак;

10) верифікувати якість отриманих прогнозів.

Для досягнення поставленої мети та задач дослідження були використані такі методи дослідження: аналіз та інтеграція, узагальнення, деталізація, обґрунтування, порівнювання та систематизація, за допомогою яких були зроблені загальні висновки, методи статистичного аналізу для проведення розрахунків.

Інформаційною базою кваліфікаційної роботи є Google Trends.

Основний науковий результат кваліфікаційної роботи полягає у такому: були створені моделі та перевірені на наявність сезонної компоненти, вибрано відповідну, що дає змогу отримати прогноз щодо тренду кібератак на майбутні періоди.

Одержані результати можуть бути використані міжнародними організаціями для прийняття належних заходів щодо захисту даних та забезпечення національної безпеки, оскільки їх інформація потенційно піддається ризику видалення та перехоплення. Результати було реалізовано в рамках науково-дослідної роботи та опубліковано в 1 тезах конференції, 1 статті.

Ключові слова: ARIMA-модель, вразливість, економіка, кібератака, прогнозування, SARIMA-модель.

Зміст кваліфікаційної роботи викладено на 68 сторінках. Список використаних джерел із 39 найменувань, розміщений на 4 сторінках. Робота містить 1 таблицю, 54 рисунків, а також 2 додатки, розміщених на 9 сторінках.

Рік виконання кваліфікаційної роботи – 2023 рік.

Рік захисту роботи – 2023 рік.

ЗМІСТ

# ВСТУП

Кожного дня збільшується кількість постраждалих від кібератак. Вони привертають дуже велику увагу, особливо, коли про них повідомляють у засобах масової інформації.

При розробці програмних продуктів, основним питанням постає питання кібербезпеки, тому що автоматизовані системи управління технологічних процесів часто знаходяться під загрозою різного роду вірусів та кібератак.

Для того, аби не постраждати від кібератаки потрібно завчасно подбати про свою безпеку. Для цього рекомендується використовувати надійне програмне забезпечення та розробити ефективну кіберстратегію, яка допоможе захистити особисту та корпоративну інформацію від зловмисників.

Мета дослідження полягає в розробці прогнозних моделей інформаційних трендів кібератак.

Актуальність теми, обраної для дослідження, визначається тим, що дозволяє зрозуміти, які нові загрози можуть з'явитися, які різновиди кібератак можуть стати популярними та які сектори чи організації можуть стати об'єктами нападу. Це допомагає підготуватися до майбутніх загроз, розробити відповідні заходи безпеки та зменшити ризики вразливості.

Об’єктом дослідження є інформація, що ідентифікує кількість запитів користувачів глобальної мережі, що є свідченням реакцій на масовість кіберзлочинів.

Предметом дослідження є статистичний, аналітичний та програмний інструментарій для прогнозування інформаційних трендів кібератак.

Для того, щоб отримати певні результати, потрібно виконати наступні задачі роботи:

— охарактеризувати проблематику кібератак та напрямки боротьби з ними;
— проаналізувати соціальну інженерію, DoS-атаки та атаки на пароль;
— провести первинний статистичний аналіз трендів кібератак;
— перевірити ряди на відповідність нормальному закону розподілу;
— перевірити ряди на стаціонарність;
— провести перевірку сезонної компоненти;
— охарактеризувати сутність математичного апарату авторегресійних моделей;
— описати результати побудови авторегресійних моделей;
— зпрогнозувати інформаційні тренди кібератак;
— верифікувати якість отриманих прогнозів.

Для дослідження було використано: набір даних (261 спостереження та 3 змінних), на основі цих даних проводився аналіз, будувалась модель та робилися прогнози. Для здійснення розрахунків та візуального зображення результатів було використано документацію по мові програмування Python.

Результати роботи було опубліковано в 1 тезах конференції, 1 статті та виконано в рамках держбюджетної науково-дослідної роботи 0121U109559 «Національна безпека через конвергенцію систем фінансового моніторингу та кібербезпеки: інтелектуальне моделювання механізмів регулювання фінансового ринку».

# РОЗДІЛ 1 АНАЛІЗ ІНФОРМАЦІЙНИХ ТРЕНДІВ КІБЕРАТАК

## 1.1 Характеристика проблематики кібератак та напрямки боротьби з ними

Під поняттям кібератака розуміються умисні дії у кіберпросторі з використанням електронної комунікації для досягнення однієї або кількох з наступних цілей: порушення повної безпеки, конфіденційності різної інформації, доступності електронної інформації, що зберігаються або передаються у комунікаційних та/або технічних системах. [1] Часто кібератаки здійснюються однією особою або групою осіб.

Якщо аналізувати кібератаки із соціальної сторони, то кількість конфліктів у таких сферах, як релігія, політика, соціологія, трудова збільшується. Також з кожним днем розвиваються нанотехнологій, і тому кількість кіберзлочинів також зростає.

Існують такі види кібератак:

- Розподілені атаки на відмову в обслуговуванні (DDoS-атаки)
- Шкідливе програмне забезпечення
- SQL-ін’єкції
- Фішинг
- Бот-мережі
- Міжсайтові сценарії (XSS)
- Зловмисні програми з вимогою викупу [2]

Аби убезпечити свої особисті та корпоративні дані, необхідно завчасно подбати про те, як запобігти кібератакам.

Для цього можна скористатися такими порадами:

- Почати інвестувати в надійну систему кібербезпеки;

- Найняти IT-адміністраторів, які будуть слідкувати за всіма мережами в компанії та моніторити інформацію, якщо виникне загроза;
- Застосуйте двофакторну або багатофакторну систему автентифікації. Це надасть впевненості, що учасники, які мають доступ до системи або власний обліковий запис являються працівниками або зацікавленими сторонами компанії.
- Проводьте інформаційні тренінги для працівників компанії, де буде йтися про можливі види кібератак та як захистити себе від них, а також радьте, що робити, якщо відбулось порушення безпеки даних.
- Залучайте сторонніх фахівців із кібербезпеки, для того, щоб вони допомагали внутрішньому IT-відділу контролювати корпоративні системи та мережі. [2]

## 1.2. Аналіз соціальної інженерії, DoS-атак та атак на паролі

Соціальна інженерія — використовується для широкого спектру зловмисних дій, що здійснюються через взаємодію людей. Він використовує психологічні маніпуляції, щоб обманом змусити користувачів зробити помилки безпеки або надати конфіденційну інформацію. [3]

Атаки соціальної інженерії мають велику кількість різних форм і можуть бути здійснені там, де задіяна взаємодія людей.

Можна виділити п’ять найпоширеніших форм цифрової соціальної інженерії.

1. Цькування — це використання фальшивих обіцянок, щоб розчулити людину жадібністю або зацікавити її. Найбільш серйозною формою цькування є фізичні носії для розповсюдження зловмисного програмного забезпечення. До таких носіїв відносять флеш-накопичувачі, які заражені шкідливим

програмним забезпеченням, їх залишають на відкритому просторі, аби жертва точно його помітила.

2. Відлякувальні програми — вони передбачають залякування жертв фіктивними погрозами та помилковими тривогами.

3. Претекстування — таким видом атаки зловмисник отримує інформацію за допомогою серії неправдивої інформації, кажучи, що йому необхідна конфіденційна інформація від жертви для виконання критично важливих завдань.

4. Фішингове шахрайство — це листи, відправлені електронною поштою та текстовими повідомленнями, спрямовані на те, щоб викликати у жертв відчуття терміновості, цікавості чи страху.

5. Кві-про-кво — цей метод використовується для маніпулювання людьми з метою отримання їх особистої інформації, такої як паролі, соціальні номери, номери кредитних карток та інші конфіденційні дані [4].

Слідуючи наступним крокам, можна захиститись від атак соціальної інженерії:

I крок — перевіряйте адресу відправника листа та доменне ім'я сайту, на якому плануєте вводити особисті дані.

II крок — уникайте підозрілих ресурсів та не завантажуйте невідомі програми.

III крок — створюйте різні паролі для доступу до особистої та корпоративної пошти, соціальних мереж і банківських додатків.

IV крок — антивірусні програмні забезпечення дозволять швидко виявляти загрози, відсутність ліцензійних файлів, спаму та небажаних програм [5].

DoS атаки — це спроба зробити сервіс, мережу або сайт недоступним для легітимних користувачів, завантажуючи його запитами або трафіком, що перевищує максимальні можливості, які ця система може обробляти. Ці атаки

можуть бути здійснені з одного або кількох джерел, які спробують зламати мережу, сайт або сервіс запитами, щоб перевантажити його та вивести із ладу.

Існують такі види DoS-атак, як:

1. Атака з переповненням мережі: здійснюється за допомогою відправлення великої кількості запитів до мережевого протоколу з метою перевантаження його ресурсів.

2. SYN-флуд атака: надсилання великої кількості запитів на встановлення з'єднання (SYN-пакети) до цільового сервера, не завершуючи їх, призводить до виснаження ресурсів сервера та недоступності новим клієнтам.

3. Ампліфікаційна атака: використання слабко налаштованих серверів або протоколів, що надсилають відповіді більшого розміру, ніж початковий запит. Як результат, після одного запиту зловмисник може створити велику навантаження на цільовий ресурс.

4. DNS-атака: це атака, в якій зловмисники спробують змінити або перехопити DNS запити, що використовуються для перетворення доменних імен на IP-адреси. Це може призвести до перенаправлення користувачів на небажані та шкідливі веб-сайти, перехоплення їхньої комунікації або зловживання даними.

5. HTTP-атака: ця атака використовується для перевантаження веб-сайту шляхом відправлення великої кількості запитів до сервера [6].

Для захисту від DoS-атак можуть бути використані такі методи, як фільтрація трафіку, розподілені системи захисту, обмеження ресурсів, моніторинг трафіку та інші техніки.

Останній вид кібератак, який розглянемо — Password Attack. Це процес спроби отримати доступ до облікового запису, перебираючи можливі варіанти паролів. Така атака може бути зроблена ручним чином або за допомогою спеціальних програм, які автоматизують процес перебору паролів.

До основних методів атак на паролі відносять такі:

1. Словникова атака — це метод, при якому програма використовує список зі слів або фраз для перебирання паролів. Цей метод є менш часовим, ніж брутфорс, оскільки перебирання відбувається на основі відомих слів.

2. Брутфорс — це метод, при якому програма перебирає всі можливі комбінації символів для створення пароля. Цей метод може зайняти дуже багато часу, але при правильному налаштуванні програми може бути досить ефективним.

3. Гібридна атакА — це метод, при якому програма комбінує словникову атаку з брутфорсом. Наприклад, програма може додавати до слова деякі символи для створення нових комбінацій [7].

Аби створити надійний захист для свого облікового запису від атак на паролі, користувачі можуть використовувати довгі та складні паролі, які складаються з комбінації букв верхнього та нижнього регістрів, цифр та спеціальних символів. Також рекомендується використовувати різні паролі для різних облікових записів та періодично змінювати паролі. Двофакторна аутентифікація також може використовуватися, як додатковий захист облікового запису від атак на паролі.

## 1.3 Первинний статистичний аналіз трендів кібератак

Для того, щоб побудувати моделі та спрогнозувати тренди кібератак, будуть використані такі вхідні дані: соціальна інженерія, DoS-атаки, атаки на паролі. Дані взяті за період з 28.01.2018 по 22.01.2023.

Соціальна інженерія — це техніки та методи маніпулювання людьми з метою отримання неправомірного доступу до конфіденційної інформації, ресурсів, або виконання певних дій [8].

DoS-атаки — це кібератаки, які спрямовані на перевантаження серверів або мережевих ресурсів з метою заборонити доступ користувачам до них [9].

Атаки на паролі — це спроба зламати або отримати доступ до облікового запису користувача шляхом зламу пароля або використання інших методів для визначення пароля [10].

Проведення статистичного аналіз даних буде за допомогою мови програмування Python з використанням бібліотеки Pandas, яка містить велику кількість функцій для маніпулювання даними.

Для початку підключимо бібліотеки NumPy та Pandas за допомогою методу import().

```
import numpy as np
import pandas as pd
```

Рисунок 1.1 — Підключення бібліотеки Pandas

NumPy — Python-бібліотека, що дозволяє працювати з багатовимірними масивами, містить різноманітні похідні об'єкти та має широкий набір процедур для швидких операцій над масивами [11].

Бібліотека Pandas — це інструмент для роботи з даними в мові програмування Python, що надає широкий набір функцій для роботи з табличними даними, а також можливості для читання та запису даних з різних джерел. [12]

Наступник кроком імпортуємо вхідні дані [13], які представлені в Додатку А таблиця 1.1. Результат відображений на рисунку 1.2. За рисунком 1.2 бачимо, що таблиця даних складається з 261 рядка (спостереження) та 3 стовбці (факторові ознаки).

```
df=pd.read_excel('Attacks.xlsx',parse_dates=['Date'], index_col='Date', header=0)
```

```
df
```

| Date | SE | PA | DoS |
|---|---|---|---|
| 2018-01-28 | 57 | 31 | 30 |
| 2018-02-04 | 54 | 41 | 52 |
| 2018-02-11 | 49 | 31 | 23 |
| 2018-02-18 | 48 | 37 | 58 |
| 2018-02-25 | 45 | 37 | 65 |
| ... | ... | ... | ... |
| 2022-12-25 | 55 | 64 | 47 |
| 2023-01-01 | 64 | 65 | 32 |
| 2023-01-08 | 75 | 60 | 75 |
| 2023-01-15 | 78 | 73 | 43 |
| 2023-01-22 | 83 | 64 | 51 |

261 rows × 3 columns

Рисунок 1.2 — Імпортування вхідних даних

За допомогою метода describe() отримаємо коротку статистику основних характеристик числових ознак. До статистики входять такі характеристики, як кількість значень вибірки, середнє значення, стандартне відхилення, мінімальне значення, медіана, 0,25 та 0,75 чверті, максимальне значення. Результат відображений на рисунку 1.3.

```
df.describe()
```

| | SE | PA | DoS |
|---|---|---|---|
| count | 261.000000 | 261.000000 | 261.000000 |
| mean | 63.724138 | 50.210728 | 47.275862 |
| std | 12.976627 | 18.056685 | 20.811252 |
| min | 30.000000 | 16.000000 | 0.000000 |
| 25% | 54.000000 | 37.000000 | 31.000000 |
| 50% | 63.000000 | 46.000000 | 48.000000 |
| 75% | 72.000000 | 61.000000 | 61.000000 |
| max | 100.000000 | 100.000000 | 100.000000 |

Рисунок 1.3 — Статистика основних характеристик числових ознак

За результатами метода describe(), що представлені на рисунку 1.3 можна зробити такі висновки:

- середнє значення соціальної інженерії – 63.7, стандартне відхилення – 12.9, мінімальне значення – 30 та максимальне – 100;
- середнє значення атак на паролі – 50.2, стандартне відхилення – 18.1, мінімальне значення – 16.0 та максимальне – 100;
- середнє значення DoS-атак – 47.3, стандартне відхилення – 20.8, мінімальне значення – 0.0 та максимальне – 100.

# РОЗДІЛ 2 СТАТИСТИЧНІ ТЕСТИ ІНФОРМАЦІЙНИХ ТРЕНДІВ КІБЕРАТАК

## 2.1 Перевірка рядів на відповідність нормальному закону розподілу

Перевірка рядів на відповідність нормальному закону розподілу — це статистичний аналіз даних, який визначає, наскільки точно ряди даних можна описати за допомогою нормального розподілу [14].

Для перевірки рядів на відповідність нормальному закону розподілу потрібно візуалізувати дані. За допомогою графіків, діаграм та описових статистик можна виявити наявність можливих трендів. Для цього використовується ряд таких бібліотек (рисунок 2.1) мови програмування Python.

```
import matplotlib.pyplot as plt
import matplotlib
import numpy as np
import pandas as pd
%matplotlib inline
import itertools
import warnings
plt.style.use('ggplot')
```

Рисунок 2.1 — Підключення бібліотеки для візуалізації даних

Візуалізувавши дані, можна наближено оцінити тип розподілу за описовими статистичними параметрами, таких як асиметрія, ексцес, статистика Жак-Бера та гістограмами. Для того, щоб подивитись на розподіл числової змінної можна відобразити дані у вигляді гістограми, використавши метод hist () бібліотеки Pandas (рисунок 2.2 – рисунок 2.4).

Гістограма — це графічний метод відображення розподілу числової змінної. Вона складається з прямокутників, які відповідають інтервалам значень числової змінної або діапазону значень, і висота кожного

прямокутника відповідає кількості спостережень, що потрапляють у відповідний інтервал або діапазон.

```
df1['DoS'].hist()
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7fa0d0a7d8e0>
```

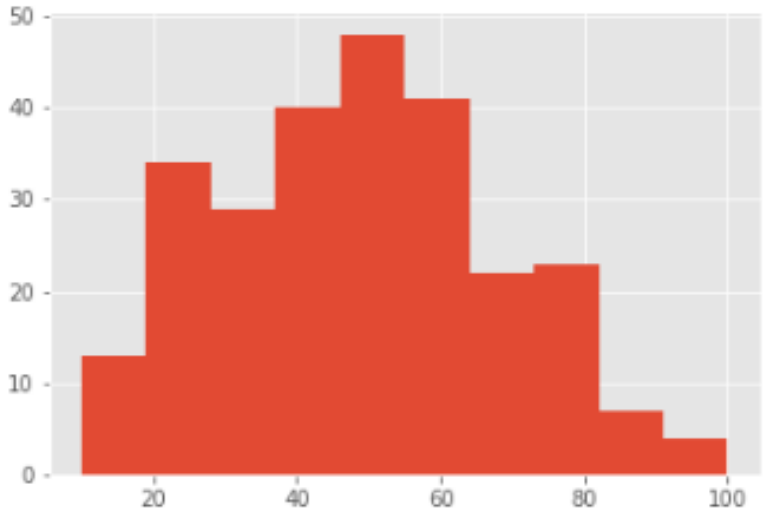



Рисунок 2.2 — Візуалізація даних ряду “DoS-атаки” у вигляді гістограми

```
df1['SE'].hist()
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f5d3c9478b0>
```

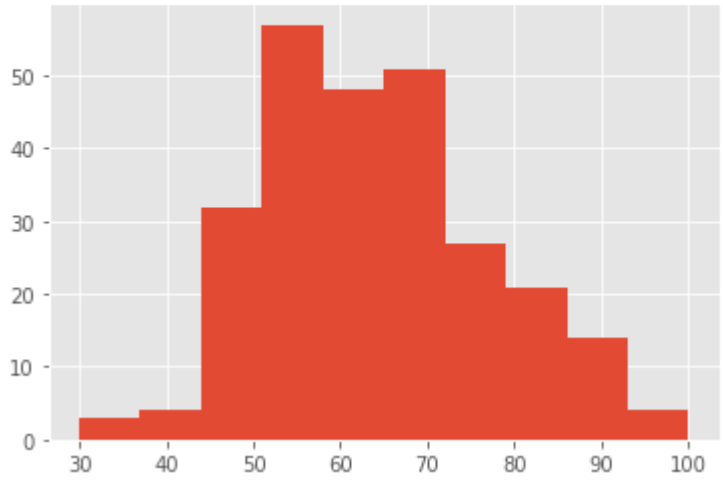



Рисунок 2.3 — Візуалізація даних ряду “Соціальна інженерія” у вигляді гістограми

```
df1['PA'].hist()
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f56f7f63400>
```

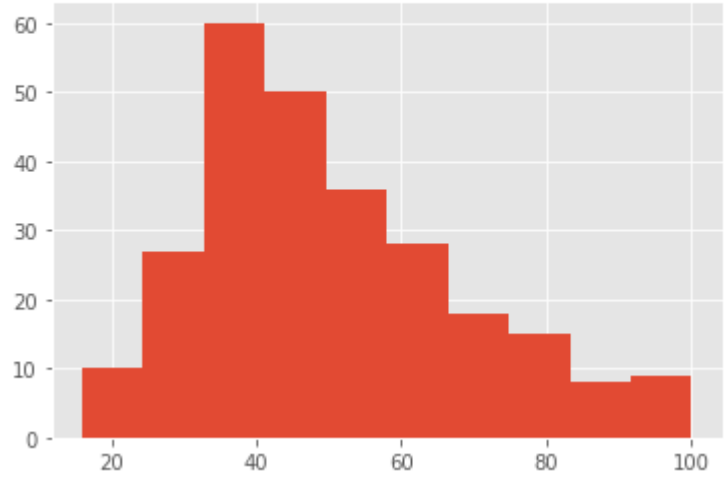



Рисунок 2.4 — Візуалізація даних “Атаки на паролі” у вигляді гістограми

Аналізуючи гістограми, бачимо, що DoS-атаки гістограма більше відповідає “дзвіноподібній кривій”; соціальна інженерія гістограма більше відповідає двопіковому типу; атаки на паролі гістограма не відповідає заданим критеріям, можливо треба виконати ряд дій для зниження розмаху вибірки. Це означає, що дані розподілено ненормально. Отже, можна стверджувати, що закон на нормальність підтверджується тільки для DoS-атаки гістограми.

У вигляді графіку зобразимо наші дані за допомогою метода plot(). Для цього треба визначити розмір графіка. У вигляді кортежу передамо необхідні дані параметру figsize() (рисунок 2.5 – рисунок 2.7).

```
df1=pd.read_excel('Attacks1.xlsx',parse_dates=['Date'], index_col='Date', header=0)
```

```
df1['DoS'].plot(figsize=(15,5))
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7fa0d0b3d370>
```

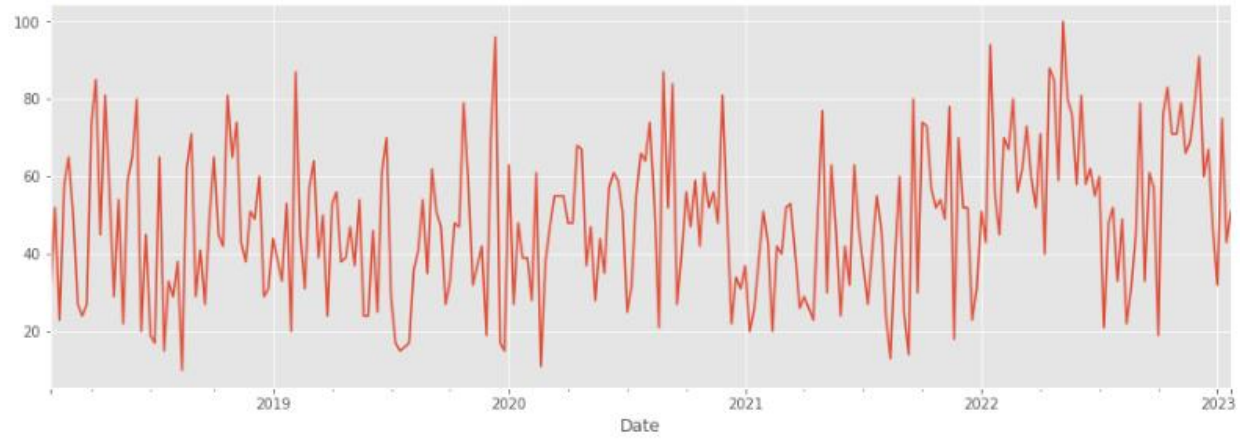



Рисунок 2.5 — Графік ряду “DoS-атаки”

```
df1=pd.read_excel('Attacks1.xlsx',parse_dates=['Date'], index_col='Date', header=0)
```

```
df1['SE'].plot(figsize=(15,5))
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f5d3ba4c100>
```

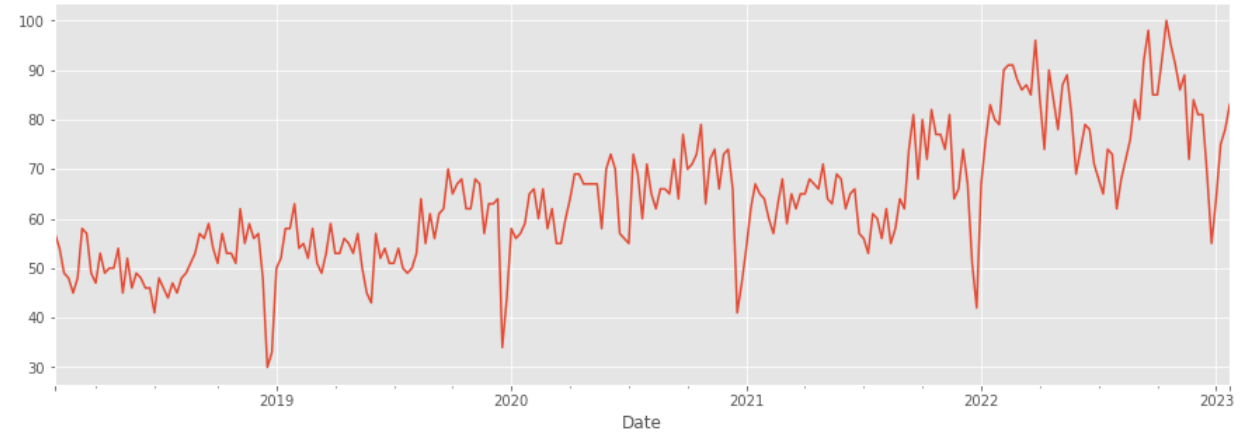



Рисунок 2.6 — Графік ряду “Соціальна інженерія”

```
df1=pd.read_excel('Attacks1.xlsx',parse_dates=['Date'], index_col='Date', header=0)

df1['PA'].plot(figsize=(15,5))

<matplotlib.axes._subplots.AxesSubplot at 0x7f56f7a773a0>
```

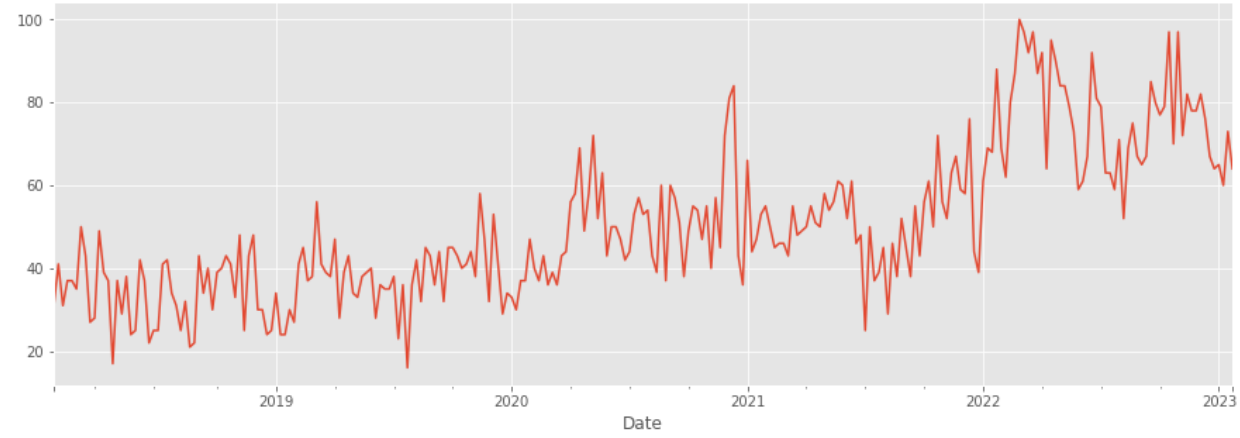



Рисунок 2.7 — Графік ряду “Атаки на паролі”

Аналізуючи графіки можна сказати, що обрані дані є часовими рядами. Ряд “DoS-атаки” не мають чітко вираженої тенденції та сезонності (рисунок 2.5) і скоріше за все ряд є стаціонарним. Ряд “Соціальна інженерія” має сезонність та тенденцію (рисунок 2.6) і тому, можливо, він є нестаціонарним. Ряд “Атаки на паролі” показує ймовірну наявність тенденції (рисунок 2.7). Ці висновки потребують подальшого тестування та перевірки.

Нормальний закон розподілу — це статистичний закон, що характеризує розподіл випадкової величини. Він може бути отриманий як сума великої кількості незалежних випадкових величин, які слабо впливають на всю суму. [15]

Для того, аби визначити нормальність розподілу, використаємо критерій Харке-Бера. Цей критерій визначає, чи мають вибрані дані ексцес та асиметрію, що відповідають нормальному закону розподілу. Критерій Харке-Бера завжди має позитивне число, і чим далі це число від нуля, тим більше видно, що вибірка не підпорядковується нормальному закону розподілу [16].

Для того, щоб виконати тест Jarque-Bera на Python, спочатку підключимо бібліотеку Scipy (рисунок 2.8) та використаємо функцію jarque_bera (рисунок 2.9 – рисунок 2.11).

```
import scipy.stats as stats
```

Рисунок 2.8 — Підключення бібліотеки Scipy

```
from statsmodels.iolib.table import SimpleTable
row =  [u'JB', u'p-value', u'skew', u'kurtosis']
jb_test = sm.stats.stattools.jarque_bera(df1['DoS'])
a = np.vstack([jb_test])
itog = SimpleTable(a, row)
print(itog)
```

```
================================================================================
        JB              p-value                skew             kurtosis
--------------------------------------------------------------------------------
6.237777236522556 0.044206271171204674 0.22044426538861728 2.3842029434680088
--------------------------------------------------------------------------------
```

Рисунок 2.9 — Тест Харке-Бера для ряду “DoS-атаки”

```
from statsmodels.iolib.table import SimpleTable
row =  [u'JB', u'p-value', u'skew', u'kurtosis']
jb_test = sm.stats.stattools.jarque_bera(df1['SE'])
a = np.vstack([jb_test])
itog = SimpleTable(a, row)
print(itog)
```

```
================================================================================
        JB              p-value                skew             kurtosis
--------------------------------------------------------------------------------
7.400825642940537 0.024713322174212866 0.4070274799663118 2.8663953087875598
--------------------------------------------------------------------------------
```

Рисунок 2.10 — Тест Харке-Бера для ряду “Соціальна інженерія”

```
from statsmodels.iolib.table import SimpleTable
row =  [u'JB', u'p-value', u'skew', u'kurtosis']
jb_test = sm.stats.stattools.jarque_bera(df1['PA'])
a = np.vstack([jb_test])
itog = SimpleTable(a, row)
print(itog)
```

```
=============================================================================
       JB               p-value                skew            kurtosis
-----------------------------------------------------------------------------
22.1747107019934 1.5304627091639918e-05 0.7135326532471882 2.94961678201348
-----------------------------------------------------------------------------
```

Рисунок 2.11 — Тест Харке-Бера для ряду “Атаки на паролі”

За результатом тесту Харке-Бера бачимо, що для DoS-атак значення p-value становить 0,0442, це означає, що ми не маємо достатніх доказів для відхилення нульової гіпотези. Відповідно, є потреба у здійсненні логарифмування ряду. Але візуальний аналіз гістограми розподілу для даного ряду та перевірка висновку шляхом логарифмування початкових даних дозволили дійти висновку у відсутності доцільності здійснення даної процедури.

Для ряду “Соціальна інженерія”значення p-value становить 0,0247, це також свідчить про відхилення нульової гіпотези про нормальний розподіл. Здійснення процедури логарифмування дозволило підвищити p-значення до 0.4139. Результат представлено на рисунку 2.12.

```
=============================================================================
        JB              p-value              skew            kurtosis
-----------------------------------------------------------------------------
1.7644896209829573 0.41385284694984936 -0.175727203772701 3.1968037565387335
-----------------------------------------------------------------------------
```

Рисунок 2.12 — Логарифмування для ряду “Соціальна інженерія”

Для ряду “Атаки на паролі” значення p-value було отримано 1.5305, тому нульова гіпотеза про нормальний розподілу підтверджується. Але візуальний аналіз гістограми дозволив виявити різку зміну у значеннях із середини 2021

року, тому цей факт вимагав здійснення процедури логарифмування початкових даних. Результат представлено на рисунку 2.13.

```
================================================================================
        JB             p-value               skew              kurtosis
--------------------------------------------------------------------------------
1.229105977478851 0.5408826219006762 -0.10817449173259704 2.7426784955314125
--------------------------------------------------------------------------------
```

Рисунок 2.13 — Логарифмування для ряду “Атаки на паролі”

За допомогою графіка коробки та вуса можемо побачити асиметрію даних (рисунок 2.14 – рисунок 2.16).

```
plt.boxplot(df1['DoS'])
plt.show()
```

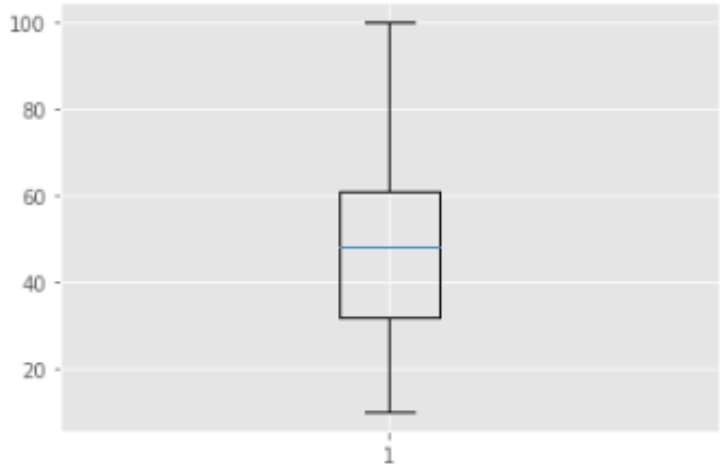



Рисунок 2.14 — Графік коробки та вуса для ряду “DoS-атаки”

```
plt.boxplot(df1['SE'])
plt.show()
```

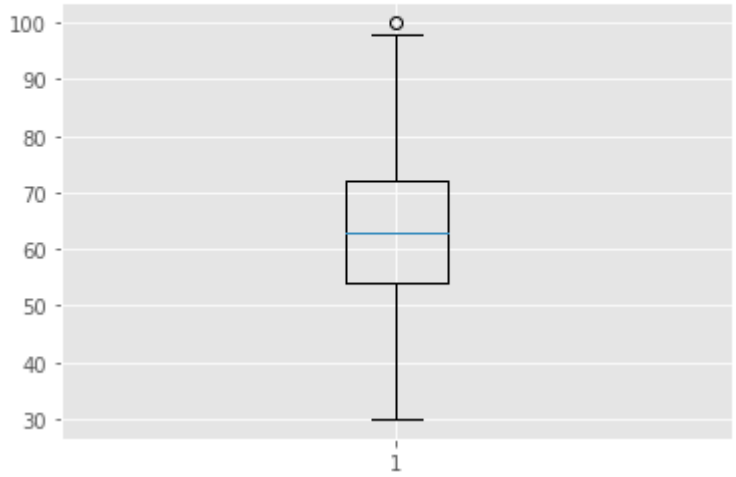



Рисунок 2.15 — Графік коробки та вуса для ряду “Соціальна інженерія”

```
plt.boxplot(df1['PA'])
plt.show()
```

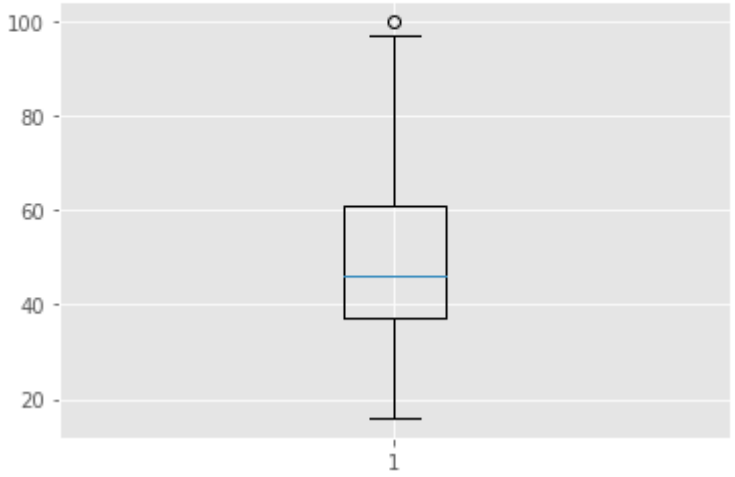



Рисунок 2.16 — Графік коробки та вуса для ряду “Атаки на паролі”

На рисунку 2.14 для ряду “DoS-атаки” присутня правостороння асиметрія, правий хвіст розподілу є довшим.

На графіку 2.15 для ряду “Соціальна інженерія” присутня правостороння асиметрія, можемо стверджувати, що правий хвіст розподілу є довшим.

На графіку 2.16 для ряду “Атаки на паролі” присутня лівостороння асиметрія, можемо стверджувати, що лівий хвіст розподілу є довшим.

## 2.2 Перевірка на стаціонарність рядів

Перевірка на стаціонарність рядів є важливим етапом в аналізі часових рядів. Стаціонарність означає, що статистичні властивості ряду не змінюються з часом, такі як середнє значення, дисперсія та кореляції [17].

Для того, що перевірити ряд на стаціонарність проведемо розширений тест Дики-Фулера, інша назва «ADF тест».

Розширений тест Дики-Фулера — це статистичний тест, який використовується для перевірки наявності одиничних коренів у часових рядах і тим самим визначає стаціонарність ряду [18]. Тест є розширенням базового тесту Дики-Фулера і дозволяє враховувати наявність лінійних трендів та автокореляції у ряді.

Для проведення ADF тесту скористаємося функцією adfuller() модуля statsmodels (рисунок 2.17 – рисунок 2.19).

```
dftest = adfuller(df1['DoS'].dropna(), autolag = 'AIC')
print("1. ADF : ",dftest[0])
print("2. P-Value : ", dftest[1])
print("3. Num Of Lags : ", dftest[2])
print("4. Num Of Observations Used For ADF Regression and Critical Values Calculation :", dftest[3])
print("5. Critical Values :")
for key, val in dftest[4].items():
  print("\t",key, ": ", val)
if dftest[0]> dftest[4]['5%']:
    print ('The series has unit roots and is not stationary')
else:
    print ('The series does not have unit roots and is stationary')
```

```
1. ADF :  -4.08327222645947
2. P-Value :  0.0010320578515232027
3. Num Of Lags :  5
4. Num Of Observations Used For ADF Regression and Critical Values Calculation : 255
5. Critical Values :
         1% :  -3.4562572510874396
         5% :  -2.8729420379793598
         10% :  -2.5728461399461744
The series does not have unit roots and is stationary
```

Рисунок 2.17 — Розширений тест Дики-Фулера для ряду “DoS-атаки”

```
dftest = adfuller(df1['SE'].dropna(), autolag = 'AIC')
print("1. ADF : ",dftest[0])
print("2. P-Value : ", dftest[1])
print("3. Num Of Lags : ", dftest[2])
print("4. Num Of Observations Used For ADF Regression and Critical Values Calculation :", dftest[3])
print("5. Critical Values :")
for key, val in dftest[4].items():
  print("\t",key, ": ", val)
if dftest[0]> dftest[4]['5%']:
    print ('The series has unit roots and is not stationary')
else:
    print ('The series does not have unit roots and is stationary')
```

```
1. ADF :  -2.8747719095682465
2. P-Value :  0.04836131656535359
3. Num Of Lags :  3
4. Num Of Observations Used For ADF Regression and Critical Values Calculation : 257
5. Critical Values :
         1% :  -3.4560535712549925
         5% :  -2.8728527662442334
         10% :  -2.5727985212493754
The series does not have unit roots and is stationary
```

Рисунок 2.18 — Розширений тест Дики-Фулера для ряду “Соціальна інженерія”

```
dftest = adfuller(df1['Lag PA'].dropna(), autolag = 'AIC')
print("1. ADF : ",dftest[0])
print("2. P-Value : ", dftest[1])
print("3. Num Of Lags : ", dftest[2])
print("4. Num Of Observations Used For ADF Regression and Critical Values Calculation :", dftest[3])
print("5. Critical Values :")
for key, val in dftest[4].items():
  print("\t",key, ": ", val)
if dftest[0]> dftest[4]['5%']:
    print ('The series has unit roots and is not stationary')
else:
    print ('The series does not have unit roots and is stationary')
```

```
1. ADF :  -13.003226838875209
2. P-Value :  2.660405064278572e-24
3. Num Of Lags :  3
4. Num Of Observations Used For ADF Regression and Critical Values Calculation : 256
5. Critical Values :
         1% :  -3.4561550092339512
         5% :  -2.8728972266578676
         10% :  -2.5728222369384763
The series does not have unit roots and is stationary
```

Рисунок 2.19 — Розширений тест Дики-Фулера для ряду “Атаки на паролі”

Проведений тест показав, що для усіх рядів критичними значеннями є: 1%: -3.4563; 5%: -2.8729; 10%: -2.5728. Для DoS-атак визначено результати тесту: ADF: -4.0833; P-value: 0.0010. Отримані значення дозволили дійти висновку, що цей досліджуваний ряд є стаціонарним та не має одиничних коренів, тому для побудови авторегресійної моделі потреба у його інтегруванні відпадає.

Для ряду “Соціальна інженерія” значення тесту Дики-Фулера є наступними: ADF: -2.8748; P-value: 0.0484. Результати показали, що ряд є також стаціонарним та не має одиничних коренів, тому немає потреби у його інтегруванні.

Для ряду “Атаки на паролі” результати є наступними: ADF: -1.9839; P-value: 0.2937. Це свідчить про те, що він є нестаціонарним та має одиничні корені. Тому для даного випадку є необхідність у здійсненні процедури інтегрування, що сприятиме перетворенню ряду у стаціонарний. Дані висновки було підтверджено повторним проведенням тестування (рисунок 2.20), але вже після процедури інтегрування. В результаті було отримано:

ADF: -13.0032; P-value: 0.0000. Тобто, перетворення сприяло створенню стаціонарного ряду.

```
dftest = adfuller(df1['Lag PA'].dropna(), autolag = 'AIC')
print("1. ADF : ",dftest[0])
print("2. P-Value : ", dftest[1])
print("3. Num Of Lags : ", dftest[2])
print("4. Num Of Observations Used For ADF Regression and Critical Values Calculation :", dftest[3])
print("5. Critical Values :")
for key, val in dftest[4].items():
  print("\t",key, ": ", val)
if dftest[0]> dftest[4]['5%']:
    print ('The series has unit roots and is not stationary')
else:
    print ('The series does not have unit roots and is stationary')
```

```
1. ADF :  -13.003226838875209
2. P-Value :  2.660405064278572e-24
3. Num Of Lags :  3
4. Num Of Observations Used For ADF Regression and Critical Values Calculation : 256
5. Critical Values :
         1% :  -3.4561550092339512
         5% :  -2.8728972266578676
         10% :  -2.5728222369384763
The series does not have unit roots and is stationary
```

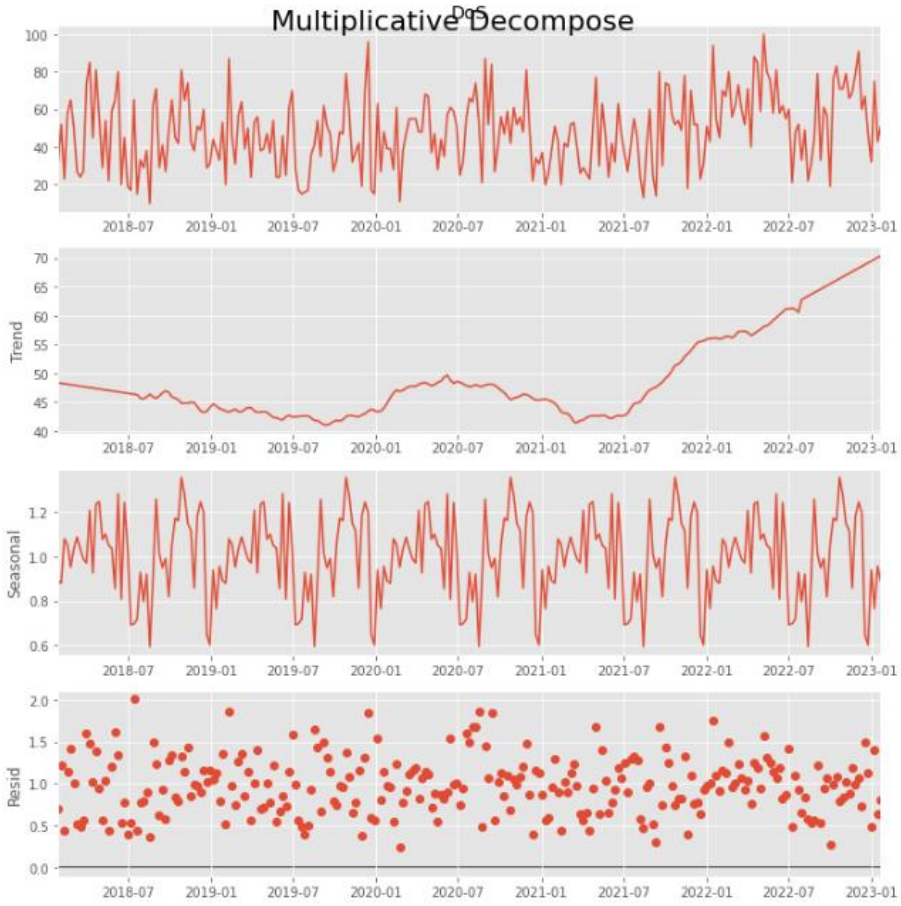
Рисунок 2.20 — Повторний розширений тест Дики-Фулера для ряду “Атаки на паролі”

## 2.3 Перевірка сезонної компоненти

Сезонна компонента — це регулярні коливання, які повторюються в часовому ряді з фіксованою періодичністю. Вона відображає систематичні зміни, що повторюються з однаковою або приблизно однаковою інтенсивністю та залежать від певного періоду [19]. Сезонні коливання можуть мати різну тривалість: денні, тижневі, місячні або річні. Вони можуть бути регулярними або нерегулярними, симетричними або асиметричними.

Врахування сезонної компоненти в аналізі часових рядів є важливим, оскільки дозволяє виявити та врахувати систематичні зміни, пов'язані зі сезонністю, при моделюванні та прогнозуванні.

Для того, аби розкласти часовий ряд на складові, розглянемо його як мультиплікативну та адитивну комбінацію, та залишки (рисунок 2.21 – рисунок 2.23).



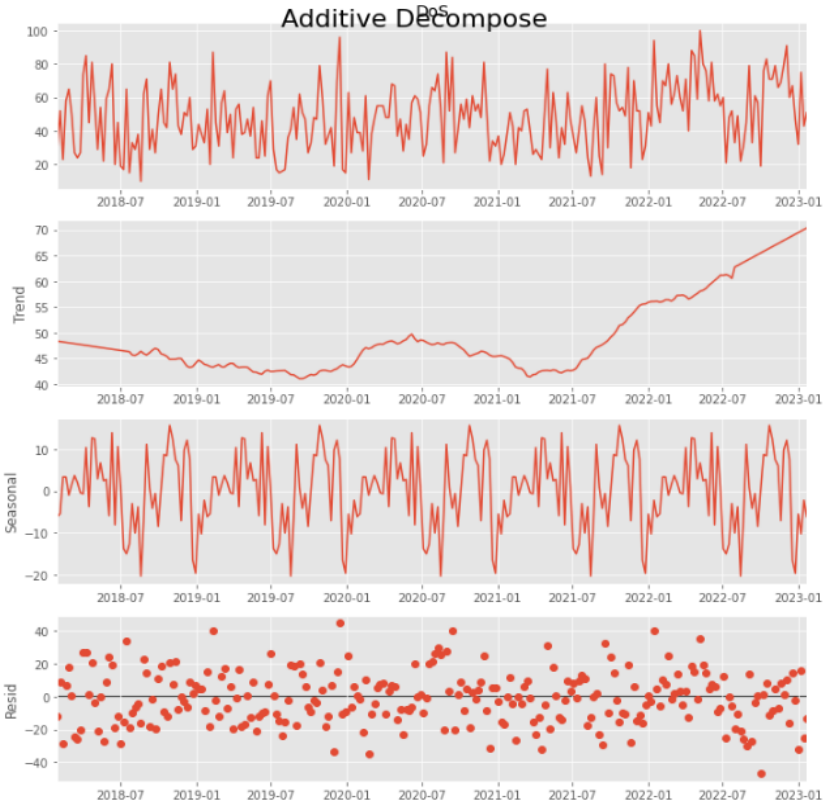



Рисунок 2.21 — Перевірка сезонної компоненти для ряду “DoS-атаки”

В результаті проведеного аналізу було встановлено, що ряд “DoS-атаки” скоріше за все не містить сезонності, але містить авторегресійну та складову ковзного середнього, тобто в результаті буде побудовано ARMA-модель.

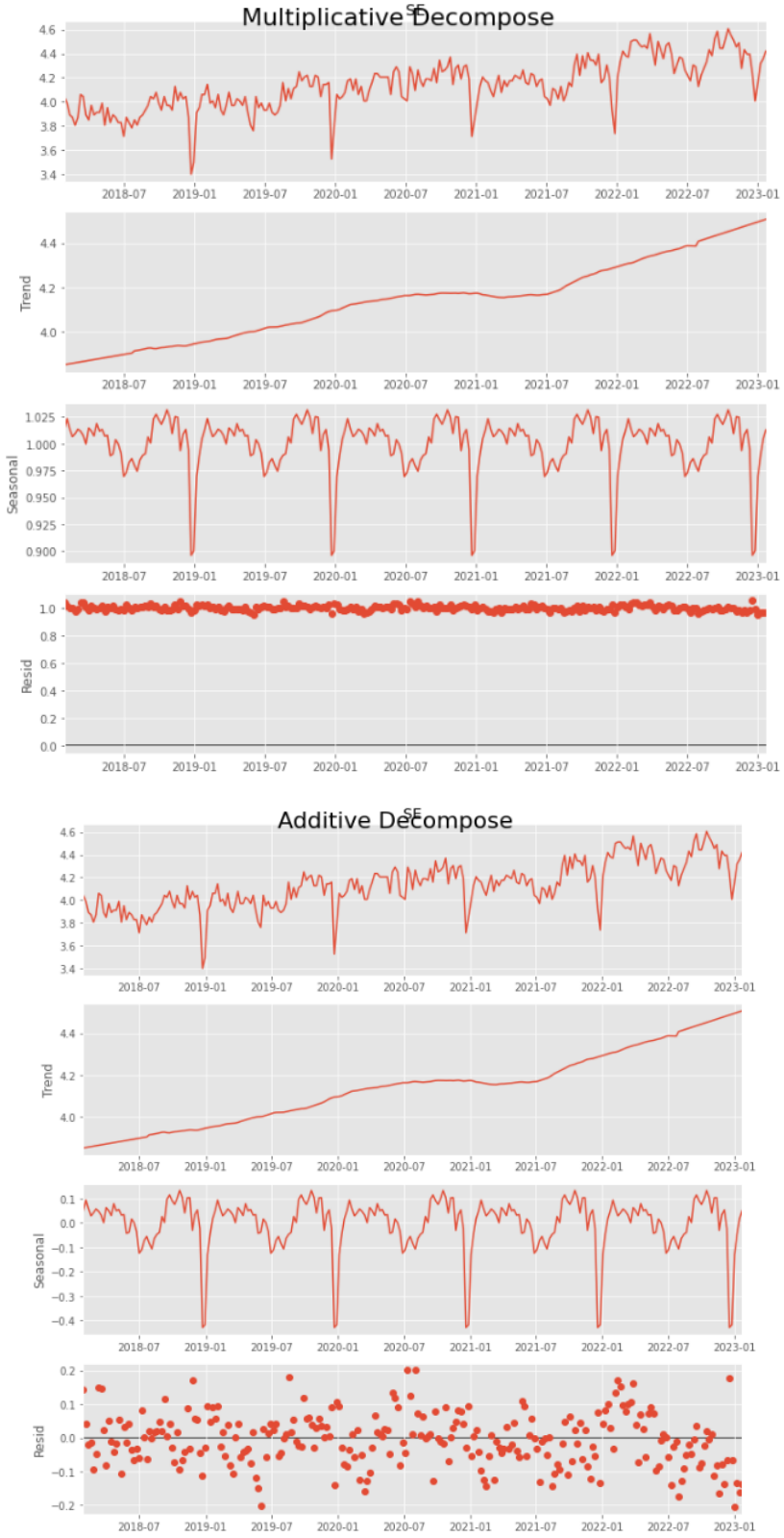



Рисунок 2. 22 — Перевірка сезонної компоненти для ряду “Соціальна інженерія”

Візуалізація представлених графіків (рисунок 2.20) дозволяє говорити про наявність чи відсутність сезонної компоненти. Автокореляційні функції ряду “Соціальна інженерія” дозволили виявити сезонний компонент із лагом 52, а також авторегресійний процес, тому пропонується модель SARIMA.

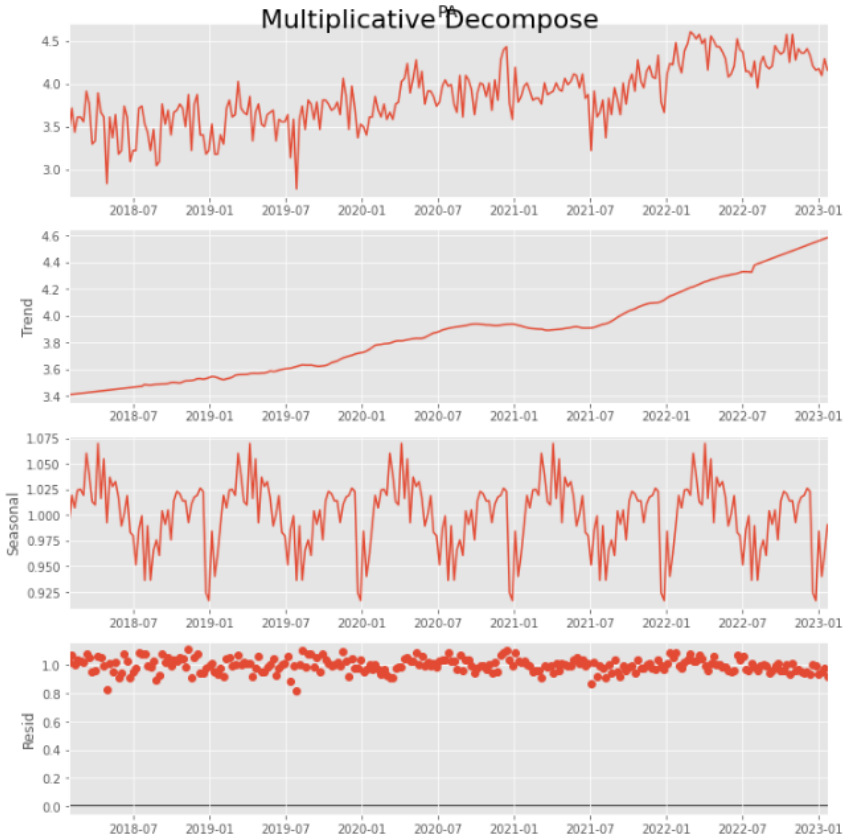



Рисунок 2. 23 — Перевірка сезонної компоненти для ряду “Атаки на паролі”

Аналіз для ряду “Атаки на паролі” також підтвердив наявність сезонності та авторегресійного процесу, що потребує побудову SARIMA-моделі (рисунок 2.23).

Результати даного розділу було опубліковано в [Яровенко Г.М., Солярова К.Г. Прогнозування інформаційних трендів кібератак як інструмент протидії вразливостей в економіці. Економіка та суспільство, 2023. №51. - прийнято до друку].

# РОЗДІЛ 3 ПОБУДОВА МОДЕЛЕЙ ТА ПРОГНОЗІВ

## 3.1 Охарактеризувати сутність математичного апарату авторегресійних моделей

ARIMA (Autoregressive Integrated Moving Average) — це статистична модель, яка використовується для аналізу та прогнозування часових рядів. ARIMA поєднує в собі авторегресійну (AR), модель з рухомим середнім (MA) та інтегровану (I) складові.

ARIMA модель може бути використана для прогнозування майбутніх значень часового ряду, а також для аналізу та виявлення складових ряду, таких як тренд, сезонність та шум. Вона широко застосовується в економічному прогнозуванні, фінансовому аналізі, кліматичних дослідженнях та інших галузях, де важливо аналізувати та прогнозувати часові ряди даних [20].

SARIMA модель — це узагальнення ARIMA-моделі на тимчасові ряди, в яких сезонна компонента яскраво виражена.

SARIMA модель складається за допомогою метода додавання додаткових членів до ARIMA моделі. Сезонна частина моделі включає в себе компоненти, які мають походження від несезонних компонентів моделі, але включають зрушення назад сезонного періоду [21].

Загальний вигляд SARIMA модель матиме такий: SARIMA(p, d, q)(P, D, Q)m [22].

Щоб використати SARIMA модель, необхідно зробити три пункти:

1. Визначити модель.

2. Підігнати визначену модель.

3. Зробити прогноз за допомогою відповідної моделі [23].

## 3.2 Результати побудови математичного апарату авторегресійних моделей

Для побудови SARIMA моделі використаємо необхідні бібліотеки (рисунок 3.1).

```
from statsmodels.tsa.statespace.sarimax import SARIMAX
from sklearn.metrics import mean_squared_error, r2_score, mean_absolute_error
from sklearn.metrics import median_absolute_error, mean_squared_log_error
```

Рисунок 3.1 — Підключення бібліотек

Щоб визначити сезонний авторегресійний порядок і порядок сезонної змінної середньої, треба побудувати корелограми ACF та PACF. Для цього використаємо функції plot_acf() [24] та plot_pacf() [25] бібліотеки Statsmodels (рисунок 3.2)

```
fig = plt.figure(figsize=(14,6))
ax1 = fig.add_subplot(211)
fig = sm.graphics.tsa.plot_acf(df1['DoS'].dropna(),lags=52,ax=ax1)
ax2 = fig.add_subplot(212)
fig = sm.graphics.tsa.plot_pacf(df1['DoS'].dropna(),lags=52,ax=ax2)
```

Рисунок 3.2 — Бібліотеки Statsmodels

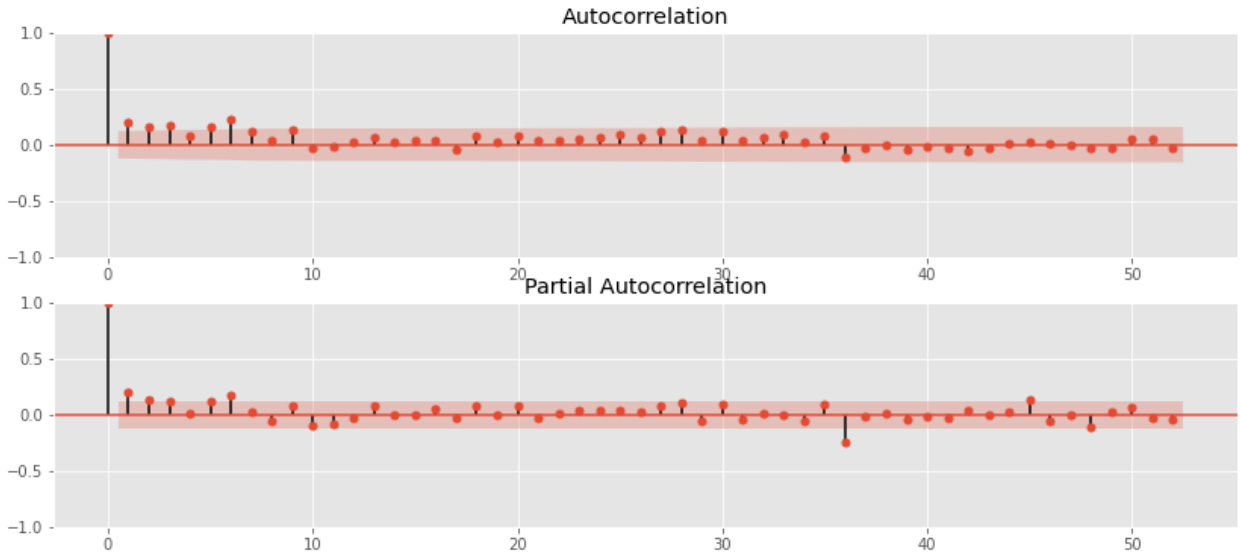



Рисунок 3.3 — Корелограми для ряду “DoS-атаки”

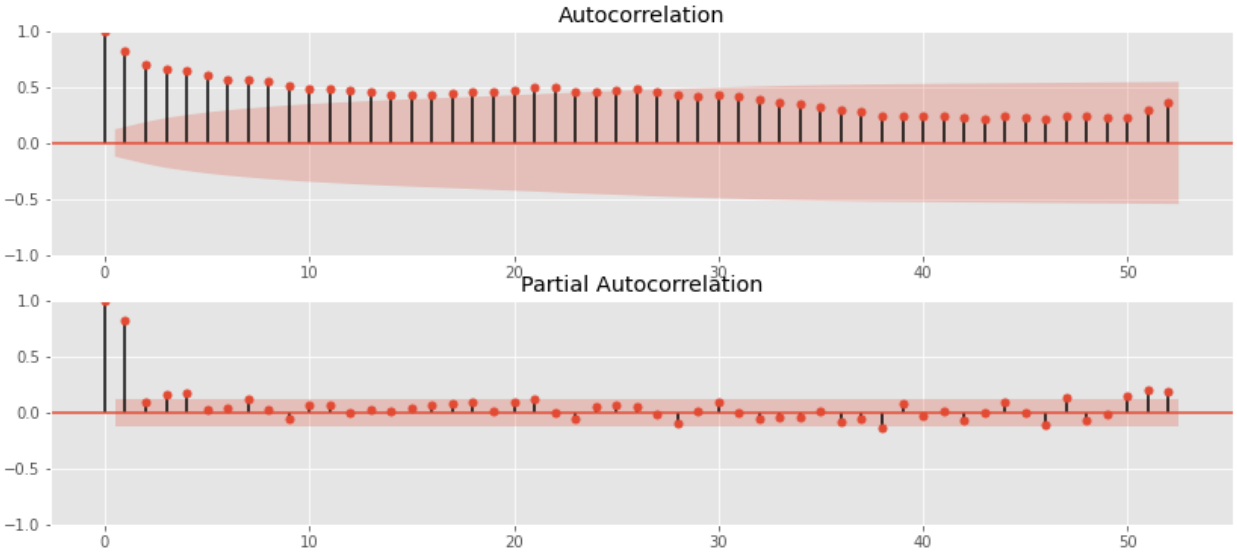



Рисунок 3.4 — Корелограми для ряду “Соціальна інженерія”

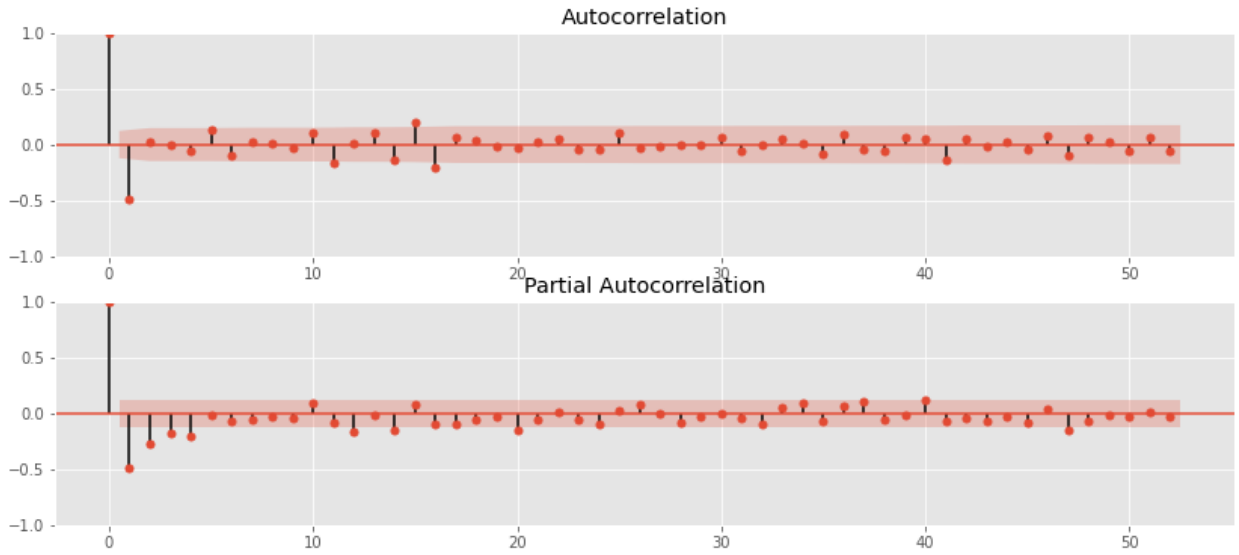



Рисунок 3.5 — Корелограми для ряду “Атаки на паролі”

За формою автокореляційної функції та часткової автокореляційної функції видно, що дані ряду “DoS-атаки” містять авторегресійний процес і процес ковзної середнє.

Для ряду “Соціальна інженерія” на графіку автокореляції (рисунок 3.4) бачимо поступове зменшення, це говорить про наявність автокореляційного процесу. Для часткової автокореляції перше значення є значущим, інші незначущі.

Для ряду “Атаки на паролі” можемо припустити, що можливо є авторегресійний процес і ковзне середнє.

Наступним кроком буде побудова SARIMA моделі (рисунок 3.6 – рисунок 3.8).

SARIMAX Results

| | | | |
|---|---|---|---|
| Dep. Variable: | DoS | No. Observations: | 261 |
| Model: | SARIMAX(3, 0, 3) | Log Likelihood | -1137.549 |
| Date: | Wed, 01 Feb 2023 | AIC | 2289.098 |
| Time: | 18:34:18 | BIC | 2314.049 |
| Sample: | 01-28-2018 - 01-22-2023 | HQIC | 2299.127 |
| Covariance Type: | opg | | |

| | coef | std err | z | P>\|z\| | [0.025 | 0.975] |
|---|---|---|---|---|---|---|
| ar.L1 | -0.4065 | 0.037 | -10.995 | 0.000 | -0.479 | -0.334 |
| ar.L2 | 0.4490 | 0.030 | 14.811 | 0.000 | 0.390 | 0.508 |
| ar.L3 | 0.9571 | 0.041 | 23.078 | 0.000 | 0.876 | 1.038 |
| ma.L1 | 0.5176 | 0.044 | 11.850 | 0.000 | 0.432 | 0.603 |
| ma.L2 | -0.3562 | 0.043 | -8.200 | 0.000 | -0.441 | -0.271 |
| ma.L3 | -0.9166 | 0.054 | -17.036 | 0.000 | -1.022 | -0.811 |
| sigma2 | 347.2051 | 36.900 | 9.409 | 0.000 | 274.883 | 419.528 |

| | | | |
|---|---|---|---|
| Ljung-Box (L1) (Q): | 0.85 | Jarque-Bera (JB): | 3.77 |
| Prob(Q): | 0.36 | Prob(JB): | 0.15 |
| Heteroskedasticity (H): | 1.12 | Skew: | 0.17 |
| Prob(H) (two-sided): | 0.61 | Kurtosis: | 2.52 |

Рисунок 3.6 — Побудова SARIMA моделі для DoS Attacks

SARIMAX Results

| | | | |
|---|---|---|---|
| Dep. Variable: | SE | No. Observations: | 261 |
| Model: | SARIMAX(1, 0, 0)x(1, 0, [1], 52) | Log Likelihood | 196.982 |
| Date: | Wed, 01 Feb 2023 | AIC | -385.965 |
| Time: | 18:59:51 | BIC | -371.707 |
| Sample: | 01-28-2018 - 01-22-2023 | HQIC | -380.233 |
| Covariance Type: | opg | | |

| | coef | std err | z | P>\|z\| | [0.025 | 0.975] |
|---|---|---|---|---|---|---|
| ar.L1 | 0.9999 | 0.000 | 6768.409 | 0.000 | 1.000 | 1.000 |
| ar.S.L52 | 0.7462 | 0.110 | 6.776 | 0.000 | 0.530 | 0.962 |
| ma.S.L52 | -0.4378 | 0.152 | -2.880 | 0.004 | -0.736 | -0.140 |
| sigma2 | 0.0119 | 0.001 | 12.168 | 0.000 | 0.010 | 0.014 |

| | | | |
|---|---|---|---|
| Ljung-Box (L1) (Q): | 26.01 | Jarque-Bera (JB): | 38.07 |
| Prob(Q): | 0.00 | Prob(JB): | 0.00 |
| Heteroskedasticity (H): | 0.99 | Skew: | -0.25 |
| Prob(H) (two-sided): | 0.98 | Kurtosis: | 4.80 |

Рисунок 3.7 — Побудова SARIMA моделі для Social Engeneering

SARIMAX Results

| | | | |
|---|---|---|---|
| Dep. Variable: | PA | No. Observations: | 261 |
| Model: | SARIMAX(4, 1, 0)x(1, 1, 0, 52) | Log Likelihood | -20.531 |
| Date: | Wed, 01 Feb 2023 | AIC | 53.061 |
| Time: | 18:13:15 | BIC | 73.087 |
| Sample: | 01-28-2018 - 01-22-2023 | HQIC | 61.159 |
| Covariance Type: | opg | | |

| | coef | std err | z | P>\|z\| | [0.025 | 0.975] |
|---|---|---|---|---|---|---|
| ar.L1 | -0.7478 | 0.068 | -10.989 | 0.000 | -0.881 | -0.614 |
| ar.L2 | -0.4684 | 0.070 | -6.675 | 0.000 | -0.606 | -0.331 |
| ar.L3 | -0.3102 | 0.071 | -4.372 | 0.000 | -0.449 | -0.171 |
| ar.L4 | -0.2429 | 0.061 | -3.976 | 0.000 | -0.363 | -0.123 |
| ar.S.L52 | -0.5677 | 0.055 | -10.240 | 0.000 | -0.676 | -0.459 |
| sigma2 | 0.0645 | 0.007 | 9.449 | 0.000 | 0.051 | 0.078 |

| | | | |
|---|---|---|---|
| Ljung-Box (L1) (Q): | 0.02 | Jarque-Bera (JB): | 0.21 |
| Prob(Q): | 0.88 | Prob(JB): | 0.90 |
| Heteroskedasticity (H): | 0.50 | Skew: | 0.05 |
| Prob(H) (two-sided): | 0.00 | Kurtosis: | 3.11 |

Рисунок 3.8 — Побудова SARIMA моделі для Password Attacks

З урахуванням отриманих результатів для всіх проведених тестів було побудовано серію моделей для кожного виду інформаційних трендів кібератак та обрано ту, яка є найкращою за інформаційними показниками AIC [26], BIC [27] та HQIC [28].

Аналізуючи побудовані моделі можна зробити такі висновки.

Спершу проаналізуємо отримані результати для ряду “DoS-атаки”. На рисунку 3.6 можна побачити, що було отримано ARMA-модель, яка містить авторегресію 3-го порядку та ковзне середнє 3-го порядку. Кожна складова моделі є статистично значущою, оскільки визначені для них p-значення менші 0,05. Ймовірність для Люнга-Бокса вище 0,05, тому ми не можемо відхилити гіпотезу, що помилки є білим шумом [29]. Значення p-статистики для гетероскедастичності також вище 0,05, що свідчить про гомоскедастичність залишків. Виходячи із того, що результати даної моделі було обрано за найкращими значеннями інформаційних критеріїв, то можна сказати, що її

оцінки є статистично значущими, а залишки некорельовані та із постійною дисперсією. Дана модель є ефективною для прогнозування.

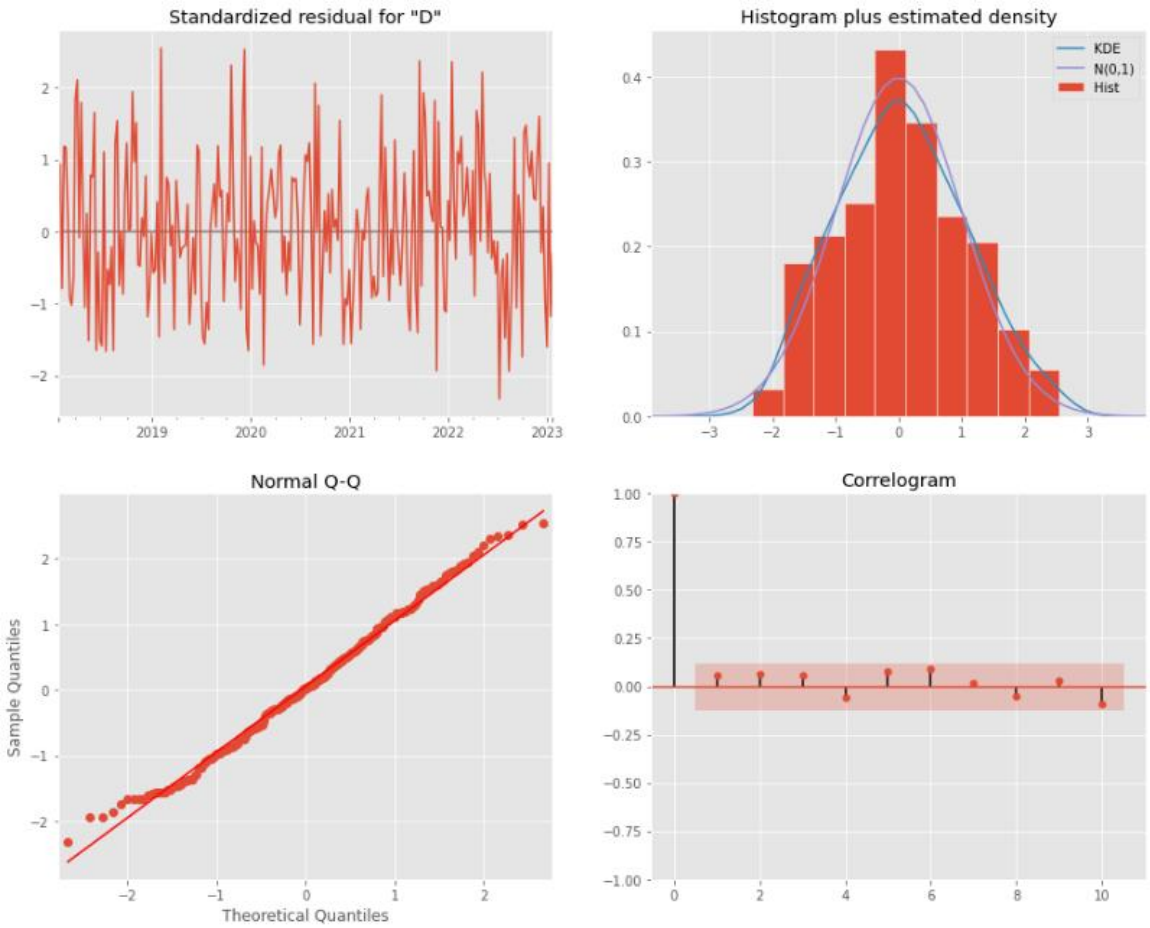



Рисунок 3.9 — Графіки для ряду“DoS-атаки”

Аналізуючи гістограму (рисунок 3.9), бачимо що дані нормально розподілені. За графіком Normal Q-Q спостерігаємо невелику невідповідність з одного краю нашого ряду. За графіком автокореляції бачимо, що значення не виходять за рамки.

Проведемо аналіз результатів для ряду “Соціальна інженерія”. Рисунок 3.7 показує, що було отримано модель, яка містить авторегресію 1-го порядку, сезонну компоненту із лагом 52, для якої існує ковзна середня 1-го порядку. Статистичну значущість кожної складової моделі підтверджує p-значення, яке є меншим ніж 0,05. Ймовірність для Люнга-Бокса нижче 0,05, тому ми

відхиляємо гіпотезу, що помилки є білим шумом. Але розрахунок даного параметру для кожного спостереження дозволив виявити, що на це впливає наявність сезонної складової. Для всіх інших состережень автокореляція відсутня. P-статистика для гетероскедастичності вище 0,05, що свідчить про гомоскедастичність залишків. Склад моделі було обрано за найкращими значеннями інформаційних критеріїв, то можна зробити висновок, що модель має статистично значущі оцінки її параметрів, гомоскедастичні залишки, але присутня автокореляція між ними, що може впливати на певну зміщеність оцінок. Оскільки застосування різних комбінацій не дозволило покращити параметри моделі, то верифікація результатів прогнозування дозволить прийняти остаточне рішення щодо її якості.

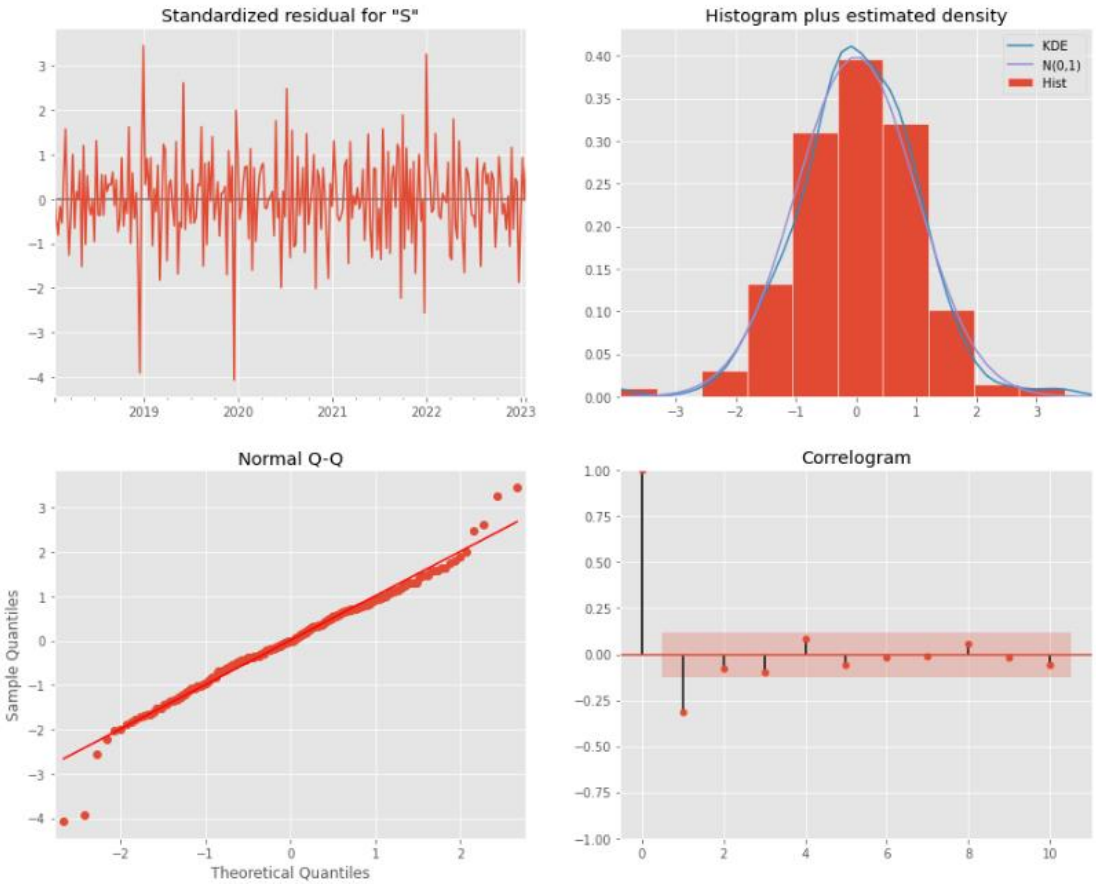



Рисунок 3.10 — Графіки для ряду “Соціальна інженерія”

Аналізуючи гістограму (рисунок 3.10), бачимо що дані нормально розподілені. За графіком Normal Q-Q бачимо, що невідповідність тільки по краях нашого ряду. За графіком автокореляції бачимо, що тільки перше значення виходить за рамки, але воно є меншим за 0.05. Це говорить про невелику автокореляцію, але вона може бути викликана сезонним компонентом.

Проаналізуємо результати для ряду “Атаки на паролі користувачів”. На рисунку 3.8 можна побачити, що було отримано модель, яка містить авторегресію 4-го порядку, сезонну складову із лагом 52. При цьому враховується також й порядок інтеграції 1. Всі складові моделі є статистично значущими (p-значення менші 0,05). Ймовірність для Люнга-Бокса вище 0,05, тому гіпотеза, що помилки є білим шумом, не відхиляється. P-статистика для гетероскедастичності є менше ніж 0,05, тому ми відхиляємо гіпотезу про гомоскедастичність залишків. Результати даної моделі обиралися за найкращими значеннями інформаційних критеріїв, то вона має статистично значущі оцінки, некорельовані, але гетероскедастичні залишки. Її ефективність для прогнозування буде підтверджено або відхилено шляхом визначення показників оцінки якості прогнозів.

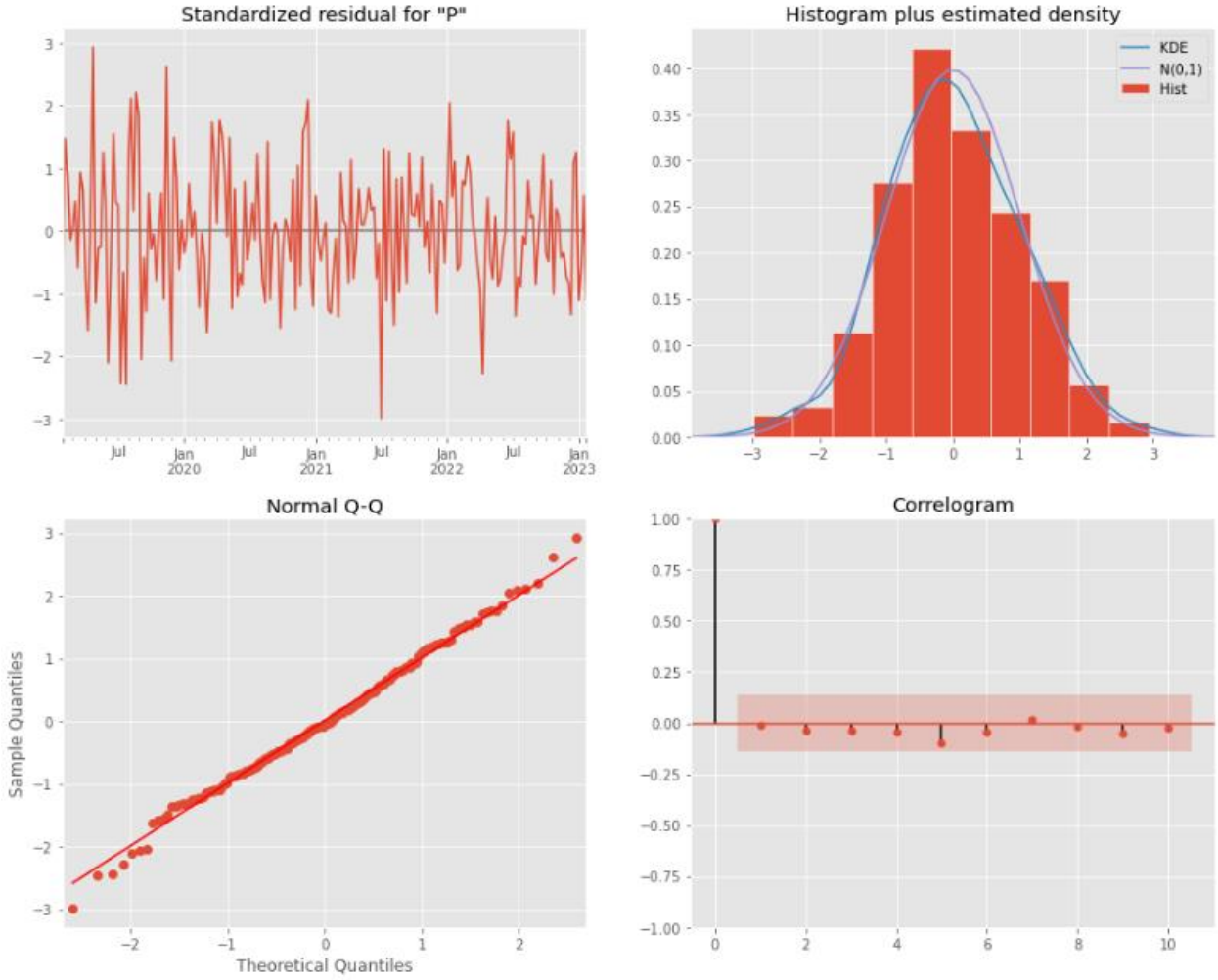



Рисунок 3.11 — Графіки для ряду “Атаки на паролі”

На рисунку 3.11 можемо побачити, що більшість точок розташовані вздовж прямої лінії з дрібними відхиленнями з кожного краю. На підставі цього графіку, можемо зробити обґрунтоване припущення, що ці дані демонструють нормальний розподіл.

Окремо було розраховано значення Люнга-Бокса для кожного ряду кібератак (рисунок 3.12 – рисунок 3.14).

```
from pandas import DataFrame

q_test = sm.tsa.stattools.acf(test_model.resid, qstat=True)
print(DataFrame({'Q-stat':q_test[1], 'p-value':q_test[2]}))
```

```
      Q-stat   p-value
0   0.937144  0.333013
1   1.919657  0.382959
2   2.805200  0.422645
3   3.591171  0.464152
4   5.035234  0.411595
5   6.974324  0.323228
6   7.012760  0.427552
7   7.695952  0.463720
8   8.265744  0.507603
9  10.161542  0.426437
10 14.143284  0.225166
11 15.085441  0.236797
12 15.086267  0.301998
13 16.942806  0.259239
14 17.080509  0.314074
15 17.388322  0.360909
16 21.666311  0.197895
17 22.419506  0.213892
18 23.188023  0.229162
19 23.188922  0.279614
20 23.374098  0.324361
21 24.013055  0.346545
22 24.015770  0.402936
23 24.023395  0.460260
```

Рисунок 3.12 — Значення Люнга-Бокса для ряду “DoS-атаки”

```
from pandas import DataFrame

q_test = sm.tsa.stattools.acf(test_model.resid, qstat=True)
print(DataFrame({'Q-stat':q_test[1], 'p-value':q_test[2]}))
```

```
      Q-stat   p-value
0   1.019376  0.312667
1   1.319461  0.516991
2   1.440712  0.696020
3   1.440713  0.837089
4   1.443738  0.919470
5   1.772374  0.939398
6   1.779706  0.971022
7   1.907873  0.983707
8   1.947293  0.992274
9   2.005836  0.996295
10  2.043730  0.998336
11  2.054975  0.999317
12  2.061560  0.999732
13  2.132370  0.999877
14  2.712837  0.999786
15  2.983669  0.999837
16  3.243081  0.999879
17  3.388964  0.999930
18  3.432788  0.999968
19  3.505124  0.999985
20  3.526009  0.999993
21  3.578204  0.999997
22  3.794998  0.999998
23  3.850194  0.999999
```

Рисунок 3.13 — Значення Люнга-Бокса для ряду “Соціальна інженерія”

```
from pandas import DataFrame

q_test = sm.tsa.stattools.acf(test_model.resid, qstat=True)
print(DataFrame({'Q-stat':q_test[1], 'p-value':q_test[2]}))
```

```
      Q-stat   p-value
0   0.287819  0.591622
1   1.623965  0.443977
2   2.131589  0.545548
3   2.285092  0.683485
4   2.336765  0.800852
5   2.362868  0.883488
6   2.401826  0.934306
7   3.080339  0.929215
8   3.324653  0.950022
9   5.396284  0.863184
10  8.967369  0.624904
11  9.062562  0.697578
12 11.514405  0.567807
13 13.827010  0.462677
14 14.875750  0.460404
15 16.467463  0.420838
16 16.617889  0.480531
17 17.283213  0.503721
18 19.546146  0.422338
19 21.166811  0.387373
20 22.241745  0.385696
21 23.652624  0.365698
22 24.414904  0.381140
23 26.473501  0.329585
```

Рисунок 3.14 — Значення Люнга-Бокса для ряду “Атаки на паролі”

Для кожного ряду p-значення більше за 0,05, тобто залишки є білим шумом і автокореляція відсутня.

## 3.3 Прогнозування інформаційних трендів кібератак

Прогнозування інформаційних трендів кібератак — це процес виявлення та передбачення змін і розвитку кібератак на основі аналізу наявних даних і трендів у кібербезпеці [30]. Цей процес включає в себе використання різноманітних методів і технік для аналізу минулих подій, виявлення патернів

та встановлення зв'язків, що допомагають передбачити майбутні напрямки кібератак [31].

Для прогнозування набори даних було поділено на тестову та верифікаційну вибірки. Безпосередньо його результати представлено на рисунку 3.16 – рисунку 3.18.

Для того, щоб провести прогнозування використаємо функцію prediction() (рисунок 3.15). За допомогою цієї функції можна передбачити мітки значень даних на основі навченої моделі [32].

```
prediction.plot(legend = True)
test['DoS'].plot(legend = True)
#df1['PA_Forecast']=arima_model.predict(start='30-01-2022',end='22-01-2023')
#df1[['Lag PA','PA_Forecast']].plot(figsize=(12,8))
```

Рисунок 3.15 — Використання функції prediction()

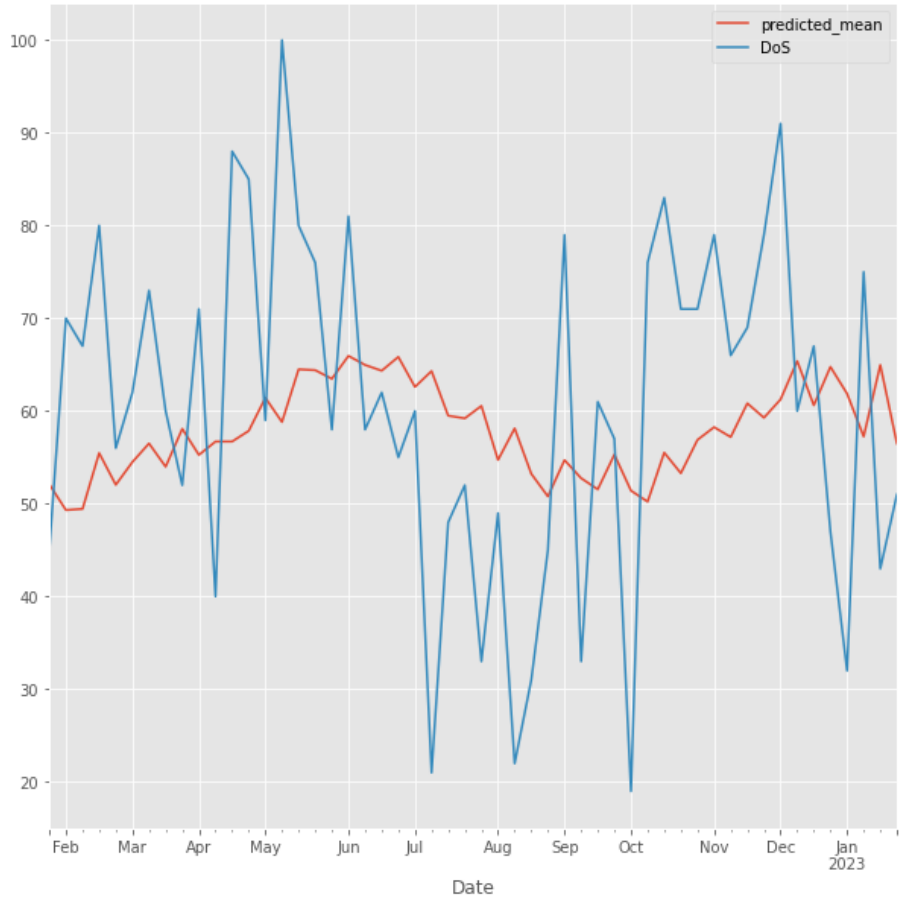



Рисунок 3.16 — Прогноз для ряду “DoS-атаки”

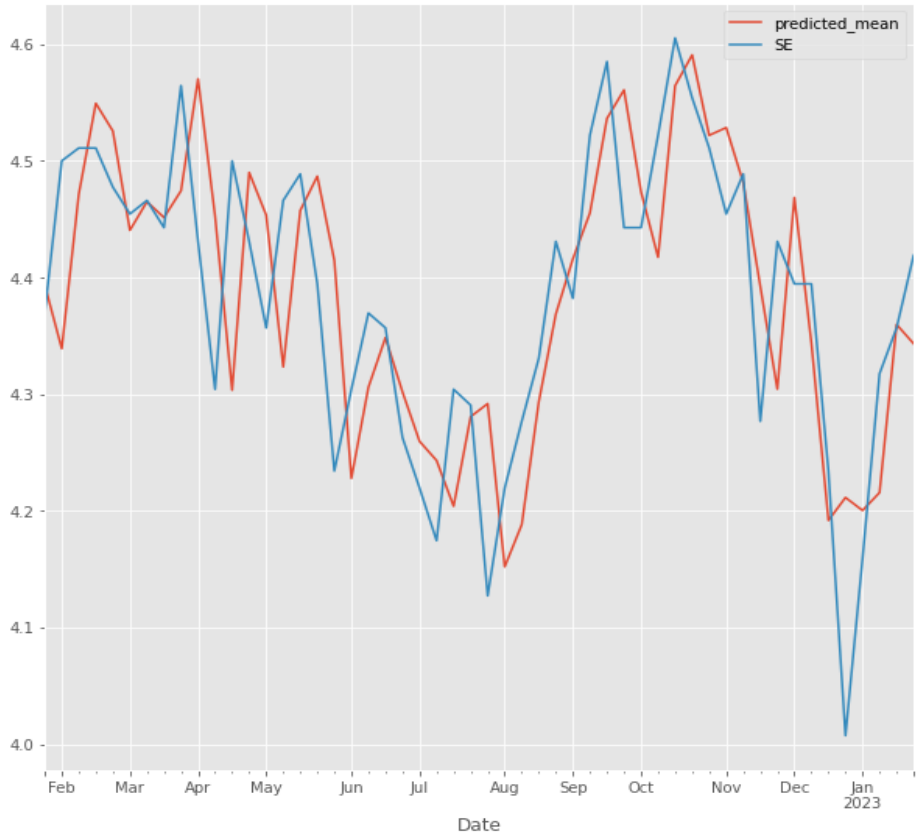



Рисунок 3.17 — Прогноз для ряду “Соціальна інженерія”

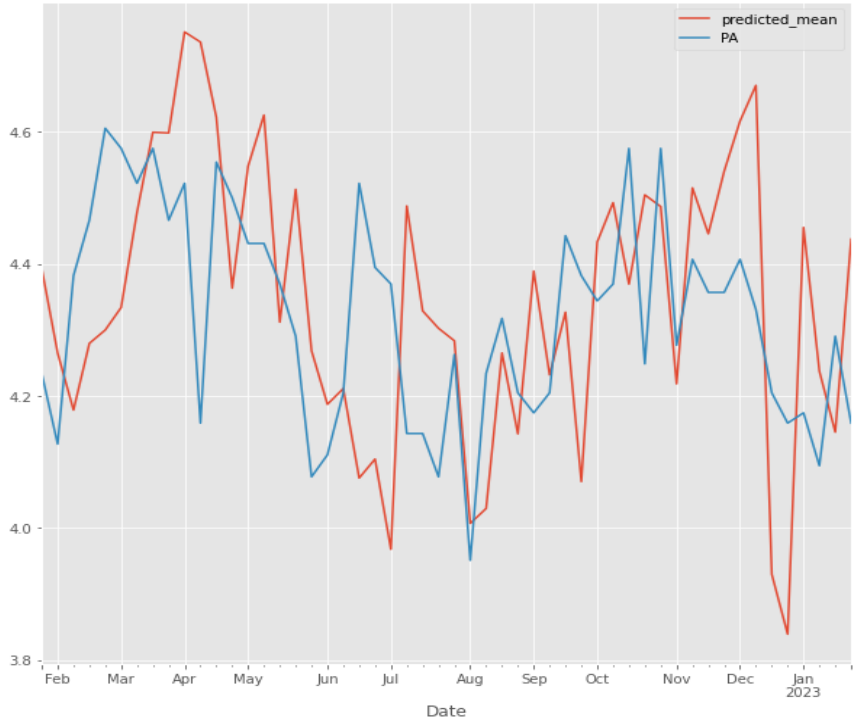



Рисунок 3.18 — Прогноз для ряду “Атаки на паролі”

Рисунок 3.16 показує, що прогнозна модель ряду “DoS-атаки” за своєю формою нагадує ковзну середню, але це є очікувано, оскільки вона містить її компонент.

Візуалізація результатів прогнозування ряду “Соціальна інженерія” на рисунку 3.17, показує досить гарні результати співпадіння модельованих та фактичних значень, які також враховують й сезонну компоненту.

Результати прогнозування ряду “Атаки на паролі користувачів” представлені на рисунку 3.18, де чітко можна побачити, що у більшості випадків модель видає правильні результати.

## 3.4 Верифікація якості прогнозів

Верифікація якості прогнозів — це процес оцінки та перевірки точності та надійності прогнозів, зроблених моделями або методами прогнозування. Його мета полягає в тому, щоб визначити, наскільки добре прогнози відповідають фактичним спостереженням або подіям [33].

Розрахуємо середню абсолютну похибку прогнозу (MAE), середня квадратична помилка (MSE), середню абсолютну відсоткову помилку (MAPE) та квадратний корінь середньої квадратичної помилки (RMSE) (рисунок 3.19 – рисунок 3.21).

MAE — вимірює середню абсолютну величину помилок прогнозування, не залежно від їхнього напрямку [34].

MSE — обчислює середнє значення квадратів помилок прогнозування, що дає більшу вагу великим відхиленням [35].

MAPE — вимірює відсоткову величину середньої абсолютної помилки, дозволяючи оцінити відносну точність прогнозів [36].

RMSE — квадратним коренем MSE і вимірює середнє квадратичне значення помилок прогнозування, що дозволяє оцінити дисперсію помилок [37].

```
from sklearn.metrics import mean_squared_error
def RMSE(actual,prediction):
    rmse = np.sqrt(mean_squared_error(actual,prediction))
    return rmse
print(RMSE(df1['DoS_Forecast'].iloc[size:], df1['DoS'].iloc[size:]))
```

```
19.24211012357412
```

```
from sklearn.metrics import mean_squared_error, r2_score, mean_absolute_error, mean_absolute_percentage_error
print('MAE: ', mean_absolute_error(df1['DoS_Forecast'].iloc[size:], df1['DoS'].iloc[size:]))
print('MSE: ', mean_squared_error(df1['DoS_Forecast'].iloc[size:], df1['DoS'].iloc[size:]))
print('MAPE:', mean_absolute_percentage_error(df1['DoS_Forecast'].iloc[size:], df1['DoS'].iloc[size:]))
```

```
MAE:  16.12908019257232
MSE:  370.2588020077536
MAPE: 0.28086147206713924
```

Рисунок 3.19 — Значення MAE, MSE, MAPE, RMSE для ряду “DoS-атаки”

```
from sklearn.metrics import mean_squared_error
def RMSE(actual,prediction):
    rmse = np.sqrt(mean_squared_error(actual,prediction))
    return rmse
print(RMSE(df1['SE_Forecast'].iloc[size:], df1['SE'].iloc[size:]))
```

```
0.08871235588781153
```

```
from sklearn.metrics import mean_squared_error, r2_score, mean_absolute_error, mean_absolute_percentage_error
print('MAE: ', mean_absolute_error(df1['SE_Forecast'].iloc[size:], df1['SE'].iloc[size:]))
print('MSE: ', mean_squared_error(df1['SE_Forecast'].iloc[size:], df1['SE'].iloc[size:]))
print('MAPE:', mean_absolute_percentage_error(df1['SE_Forecast'].iloc[size:], df1['SE'].iloc[size:]))
```

```
MAE:  0.0720229034503203
MSE:  0.007869882087165728
MAPE: 0.016494660325596586
```

Рисунок 3.20 — Значення MAE, MSE, MAPE, RMSE для ряду “Соціальна інженерія”

```
from sklearn.metrics import mean_squared_error
def RMSE(actual,prediction):
    rmse = np.sqrt(mean_squared_error(actual,prediction))
    return rmse
print(RMSE(df1['PA_Forecast'].iloc[size:], df1['PA'].iloc[size:]))
```

```
0.21630363950925016
```

```
from sklearn.metrics import mean_squared_error, r2_score, mean_absolute_error, mean_absolute_percentage_error
print('MAE: ', mean_absolute_error(df1['PA_Forecast'].iloc[size:], df1['PA'].iloc[size:]))
print('MSE: ', mean_squared_error(df1['PA_Forecast'].iloc[size:], df1['PA'].iloc[size:]))
print('MAPE:', mean_absolute_percentage_error(df1['PA_Forecast'].iloc[size:], df1['PA'].iloc[size:]))
```

```
MAE:  0.18184375287824484
MSE:  0.046787264464947645
MAPE: 0.04202609906564069
```

Рисунок 3.21 — Значення MAE, MSE, MAPE, RMSE для ряду “Атаки на паролі”

У таблиці 1 наведено розраховані показники якості прогнозів, які свідчать про можливості застосування побудованих моделей на практиці. Верифікація прогнозних результатів для ряду “DoS-атаки” свідчить, що рівень моделі є між добрим та задовільним, на що вказують значення показників оцінки якості прогнозів (Таблиця 3.1). Оскільки обрана прогнозна модель підтвердила всі тести, то її можна використовувати для прогнозування DoS-атак.

Таблиця 3.1 – Результати верифікації прогнозів

| Показники верифікації | DoS-атаки | Соціальна інженерія | Атаки на паролі користувачів |
|---|---|---|---|
| RMSE | 19,2421 | 0,0887 | 0,2163 |
| MAE | 16,1291 | 0,0720 | 0,1818 |
| MSE | 370,2588 | 0,0079 | 0,0468 |
| MAPE | 28,0861% | 1,6595% | 4,2026% |

Для ряду “Соціальна інженерія” значення оцінок якості прогнозів (Таблиця 1) знаходяться на високому рівні та підтверджують, що прогноз є високої якості. Хоча побудована модель не пройшла тест на автокорельованість залишків, але при цьому вона видає гарні результати

прогнозів, то приходимо до висновку щодо доцільності її застосування для прогнозування кіберзлочинів, пов’язаних із соціальною інженерією.

Розрахунок показників якості прогнозів для ряду “Атаки на паролі” (Таблиця 1) показує, що прогноз можна віднести до високоякісних, оскільки всі показники наближаються до нульових значень, а MAPE є меншим ніж 5%. Оскільки модель не пройшла перевірку на гетероскедастичність, то її оцінки мають завищені значення, що також може вплинути на результативність. Тому модель потребує доопрацювання в майбутньому.

Результати даного розділу було опубліковано в [Яровенко Г.М., Солярова К.Г. Прогнозування інформаційних трендів кібератак як інструмент протидії вразливостей в економіці. Економіка та суспільство, 2023. №51. - прийнято до друку].

# ВИСНОВКИ

Проблема кіберзлочинності є актуальною у наш час, оскільки її масовість та масштабність може впливати на розвиток економіки в країні шляхом формування вразливостей за рахунок її тінізації, відмивання доходів, отриманих злочинним шляхом, підтримки Даркнету, тощо. Це питання потребує систематичного дослідження та розробки відповідних превентивних заходів, оскільки характер злочинів, об’єкти та інструментарій їх здійснення постійно змінюються. Тому розробка прогнозних моделей інформаційних трендів кібератак є актуальною темою.

Ця робота базувалася на дослідженні емпіричних даних, отриманих на основі даних Google Trends, оскільки ця інформація є свідченням реакцій користувачів глобальної мережі на масовість кіберзлочинів. Було обрано три види найпоширених видів кібератак, пов’язаних із соціальною інженерією, DoS-атаками та атаками на паролі користувачів. Процес прогнозування передбачив здійснення тестів Харке-Бера, Дики-Фулера, аналіз гістограм розподілу, декомпозиції часового ряду та автокореляційних функцій для підтвердження чи відхилення гіпотез про нормальність, стаціонарність, наявність сезонної компоненти та вибір структури моделі. В результаті було побудовано для ряду DoS-атакам ARMA-модель, яка містить процеси авторегресії та ковзного середнього 3-го порядку. Тестування залишків та якості прогнозів даної моделі дозволили встановити, що її якість є задовільною при статистичній значущості параметрів, відсутності автокореляції та гетероскедастичності залишків. В цілому, її використання дозволить зробити прогноз середньої якості.

Для ряду “Соціальна інженерія” побудовано SARIMA-модель, яка містить авторегресійний процес 1-го порядку, сезонну компоненту та ковзну середню 1-го порядку для неї. Візуалізація прогнозів та оцінка їх якості

показала, що модель демонструє гарні результати. Але тест Люнга-Бокса підтвердив наявність автокореляції залишків. Оскільки розраховане значення для прогнозних спостережень не виявила її, то на даний результат вплинула наявність сезонної компонент із значним лагом, що потребує збільшення вибірки дослідження. Не дивлячись на даний нюанс, модель можна використовувати для прогнозування кіберзлочинів, пов’язаних із соціальною інженерією. Для ряду “Атаки на паролі користувачів” було побудовано SARIMA-модель із авторегресійним процесом та сезонною складовою. При цьому ряд було проінтегровано, оскільки його значення є нестаціонарними. Хоча верифікація прогнозів показала високу якість моделі, але її залишки є гетероскедастичними, що вплинуло на завищення оцінок її параметрів. Тому модель потребує проведення модифікації.

Отримані результати цього дослідження можна використовувати для удосконалення стратегічних планів країни щодо формування комплексу превентивних заходів для попередження кіберзагроз. Також це потребуватиме створення динамічної бази статистичних даних на основі відкритих та закритих джерел офіційних даних, які стосуються різних видів масових кіберзлочинів.

# СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. NIST Computer Security Resource Center | CSRC. URL: https://CyberAttack-Glossary|CSRC (дата звернення: 15.05.2023).

2. What Is a Cyberattack? - Most Common Types. URL: https://www.cisco.com/c/en/us/products/security/common-cyberattacks.html (дата звернення: 15. 05.2023).

3. Cybersecurity & Threat Intelligence Services | Webroot. URL: https://What-is-Social-Engineering? (дата звернення: 15.05.2023).

4. Low-Code DevOps & Testing Tools for Salesforce & Other SaaS – Copado. URL: https://12-Types-of-Social-Engineering-Attacks-to-Look-Out-For (дата звернення: 17.05.2023).

5. Блог Portmone.com. URL: https://blog.portmone.com.ua/information-posts-uk/sotsialna-inzheneriya (дата звернення: 18.05.2023).

6. Минин Андрій. KING SERVERS. Які існують види DDoS-атак і як від них захиститись. URL: https://www.kingservers.com/blog/ddos-atak-i-kak-ot-nikh-zashchititsia/ (дата звернення: 20.05.2023).

7. Identity Management Institute. URL: https://7-hacking-password-attack-methods (дата звернення: 23.05.2023).

8. Мельниченко А.А. Соціальна інженерія як фактор забезпечення стійкого розвитку соціальних систем. Вісник НТУУ "КПІ". Політологія. Соціологія. Право: збірник наукових праць, 2012, №1 (13). С. 73 – 78.

9. Краснобрижий І.В. Види та методики реалізації dos та ddos атак на державні автоматизовані системи, а також можливі шляхи боротьби з ними / І.В. Краснобрижий // Економічна та інформаційна безпека: проблеми та перспективи: матеріали Всеукр. наук.-практ. конф. (м. Дніпро, 14 квітня 2017 р.). – Дніпро: ДДУВС, 2017. – С. 89-94.

10. Ракович, Д. О. Рефлексивний аналіз сценаріїв атак соціальної інженерії : дипломний проект бакалавра : 125 Кібербезпека / Ракович Дарина Олександрівна. – Київ, 2021. – 45 с.

11. Парненко В. С. Мова програмування Python. NumPy в Python. Інформатика. Презентація до практики 9, 2023. URL: https://ela.kpi.ua/handle/123456789/56657.

12. Pandas. Python Data Analysis Library. URL: https://pandas.pydata.org/.

13. Google Trends. URL: https://trends.google.com/trends?geo=&hl=uk (дата звернення: 26.05.2023).

14. Теорія імовірностей та математична статистика. Курс лекцій. / Уклад.: Т.А.Ліхоузова – К.: КПІ ім. Ігоря Сікорського, 2018. – 300 с.

15. Методичні вказівки до практичних робіт з курсу “Теорія ймовірностей та математична статистика”. Тема 4 “Нормальний розподіл” / укладачі: О. С. Мазманішвілі, О. А. Шовкопляс. – Суми: Сумський Державний Університет, 2012. – 36 с.

16. R. S. Tsay, Analysis of Financial Time Series (third edition). Hoboken: John Wiley & Sons, Inc. 2010.

17. Лук’яненко І. Г., В. М. Жук. АНАЛІЗ ЧАСОВИХ РЯДІВ Побудова ARIMA, ARCH/GARCH моделей з використанням пакета E.Views 6.0. – Київ: НаУКМА, 2013. – 188 с.

18. Розширений тест Дікі-Фуллера. URL: https://dickey-fuller-test (дата звернення: 01.06.2023).

19. Прогнозування та аналіз часових рядів. Методичні вказівки до практичних занять та самостійної роботи студентів спеціальності 051 «Економіка» освітня програма «Економічна кібернетика», «Економічна аналітика» / Укл.: Юрченко М. Є. – Чернігів: ЧНТУ, 2018. – 88 с.

20. Box, G.E.P., Jenkins, G.M., Reinsel, G.C., Ljung, G.M.. Time Series Analysis: Forecasting and Control (5th Edition), 2015. Hoboken, NJ: John Wiley and Sons Inc. ISBN 978-1-118-67502-1. URL: https://Time_Series_Analysis.

21. І. Фармага, В. Гадомський. Система проведення аналізу, дослідження та передбачення подій у послідовностях даних дискретного часу. Комп'ютерні системи проектування. Теорія і практика, 2021, Випуск 3, №1. С. 61 – 66.

22. Гайдей, Р. В. Аналіз часових рядів статистичними методами Data Mining : дипломна робота магістра / Р. В. Гайдей. – Одеса, 2018. – 88 с.

23. Shumway, R. H., & Stoffer, D. S. Time series analysis and its applications: with R examples, 3th ed. Springer, 2015.

24. Statsmodels 0.14.0. Statsmodels.graphics.tsaplots.plot_acf. URL: https://statsmodels.graphics.tsaplots.plot_acf.

25. Statsmodels 0.14.0. Statsmodels.graphics.tsaplots.plot_pacf. URL: https://statsmodels.graphics.tsaplots.plot_pacf.

26. Joseph E. Cavanaugh. Unifying the derivations for the Akaike and corrected Akaike information criteria. Statistics & Probability Letters. 1997. Vol. 33, Issue 2. P. 201 – 208.

27. Economy-Pedia. Інформаційний критерій Байєса. URL: https://bayesian-information-criterion.

28. James R Knaub. Heteroscedasticity and homoscedasticity. 2007. URL: https://HETEROSCEDASTICITY_AND_HOMOSCEDASTICITY.

29. Statistics How To. Ljung Box Test: Definition. URL: https://www.statisticshowto.com/ljung-box-test.

30. Яренко А. В. Систематизація кількісних методів прогнозування кон'юнктури ринку в маркетингових дослідженнях / А. В. Яренко // Вісник Київського національного університету технологій та дизайну. - 2015. - № 3 (87) : Серія "Економічні науки". - С. 11-18.

31. Конспект лекцій з дисципліни «Прогнозування» (для студентів 3 курсу денної і 4 курсу заочної форм навчання напряму підготовки 6.030504 «Економіка підприємства» 0501 «Економіка і підприємництво» спеціальності ЕП) / Авт.: Світлична Т.І., Дріль Н.В.; Харк. нац. акад. міськ. госп-ва. – Х: ХНАМГ, 2010. – 112 с.

32. RDocumentation. Prediction function. URL: https://prediction.

33. Політологія. Проблема якості і верифікації прогнозів - Філософія глобальних проблем сучасності: Навч.-метод. посібник. URL: http://politics.ellib.org.ua/pages-3051.html.

34. IBM - Deutschland | IBM. IBM Documentation. URL: https://Mean_Absolute_Error.

35. GeeksforGeeks. Python | Mean Squared Error. URL: https://python-mean-squared-error.

36. Oracle Help Center. IPM Insights Metrics. URL: https://insights_metrics_MAPE.

37. Statistics How To. RMSE: Root Mean Square Error. URL: https://rmse-root-mean-square-error.

38. Яровенко Г.М., Солярова К.Г. Прогнозування інформаційних трендів кібератак як інструмент протидії вразливостей в економіці. Економіка та суспільство, 2023. №51. - прийнято до друку.

39. Солярова К., Яровенко Г. Прогнозування інформаційних трендів кіберзлочинів. Виклики кібербезпеки індустрії фінансових послуг: Матеріали міжнародної науково-практичної конференції (Суми, 02 червня 2023), 2023. С.71-73

# ДОДАТКИ

Додаток А

## SUMMARY

Soliarova K. G. Forecasting information trends cyberattacks. - Bachelor's qualifying work. Sumy State University, Sumy, 2023

The main aspects of cyberattacks, their essence and varieties are investigated, statistical data and current trends in the development of cyber security policy were analyzed. Information trends of cyberattacks are predicted using the construction of autoregressive models. The main goal of the study was to develop predictive models of cyber-attack information trends.

Key words: ARIMA model, vulnerability, economy, cyber attack, forecasting, SARIMA model.

## АНОТАЦІЯ

Солярова К. Г. Прогнозування інформаційних трендів кіберзлочинів. Кваліфікаційна робота бакалавра. Сумський лердавний університет, Суми, 2023 р.

У роботі було досліджено основні аспекти кібератак, їх сутність та різновиди, були проаналізовані статистичні дані та актуальні напрямки розвитку політики кібербезпеки. Було зпрогнозовано інформаційні тренди кібератак за допомогою побудови авторегресійних моделей. Основна мета дослідження полягала в розробці прогнозних моделей інформаційних трендів кібератак.

Ключові слова: ARIMA-модель, вразливість, економіка, кібератака, прогнозування, SARIMA-модель.

# Додаток Б

## Вхідні дані

| 1 | Date | SE | PA | DoS |
|---|---|---|---|---|
| 2 | 28.01.2018 | 57 | 31 | 30 |
| 3 | 04.02.2018 | 54 | 41 | 52 |
| 4 | 11.02.2018 | 49 | 31 | 23 |
| 5 | 18.02.2018 | 48 | 37 | 58 |
| 6 | 25.02.2018 | 45 | 37 | 65 |
| 7 | 04.03.2018 | 48 | 35 | 50 |
| 8 | 11.03.2018 | 58 | 50 | 27 |
| 9 | 18.03.2018 | 57 | 43 | 24 |
| 10 | 25.03.2018 | 49 | 27 | 27 |
| 11 | 01.04.2018 | 47 | 28 | 74 |
| 12 | 08.04.2018 | 53 | 49 | 85 |
| 13 | 15.04.2018 | 49 | 39 | 45 |
| 14 | 22.04.2018 | 50 | 37 | 81 |
| 15 | 29.04.2018 | 50 | 17 | 56 |
| 16 | 06.05.2018 | 54 | 37 | 29 |
| 17 | 13.05.2018 | 45 | 29 | 54 |
| 18 | 20.05.2018 | 52 | 38 | 22 |
| 19 | 27.05.2018 | 46 | 24 | 59 |
| 20 | 03.06.2018 | 49 | 25 | 65 |
| 21 | 10.06.2018 | 48 | 42 | 80 |
| 22 | 17.06.2018 | 46 | 37 | 20 |
| 23 | 24.06.2018 | 46 | 22 | 45 |
| 24 | 01.07.2018 | 41 | 25 | 19 |
| 25 | 08.07.2018 | 48 | 25 | 17 |
| 26 | 15.07.2018 | 46 | 41 | 65 |
| 27 | 22.07.2018 | 44 | 42 | 15 |
| 28 | 29.07.2018 | 47 | 34 | 33 |
| 29 | 05.08.2018 | 45 | 31 | 29 |
| 30 | 12.08.2018 | 48 | 25 | 38 |
| 31 | 19.08.2018 | 49 | 32 | 10 |
| 32 | 26.08.2018 | 51 | 21 | 62 |
| 33 | 02.09.2018 | 53 | 22 | 71 |
| 34 | 09.09.2018 | 57 | 43 | 29 |

Рисунок Б.1 — Вхідні дані

| 1 | Date | SE | PA | DoS |
|---|---|---|---|---|
| 35 | 16.09.2018 | 56 | 34 | 41 |
| 36 | 23.09.2018 | 59 | 40 | 27 |
| 37 | 30.09.2018 | 54 | 30 | 49 |
| 38 | 07.10.2018 | 51 | 39 | 65 |
| 39 | 14.10.2018 | 57 | 40 | 45 |
| 40 | 21.10.2018 | 53 | 43 | 42 |
| 41 | 28.10.2018 | 53 | 41 | 81 |
| 42 | 04.11.2018 | 51 | 33 | 65 |
| 43 | 11.11.2018 | 62 | 48 | 74 |
| 44 | 18.11.2018 | 55 | 25 | 43 |
| 45 | 25.11.2018 | 59 | 43 | 38 |
| 46 | 02.12.2018 | 56 | 48 | 51 |
| 47 | 09.12.2018 | 57 | 30 | 49 |
| 48 | 16.12.2018 | 48 | 30 | 60 |
| 49 | 23.12.2018 | 30 | 24 | 29 |
| 50 | 30.12.2018 | 33 | 25 | 31 |
| 51 | 06.01.2019 | 50 | 34 | 44 |
| 52 | 13.01.2019 | 52 | 24 | 38,5 |
| 53 | 20.01.2019 | 58 | 24 | 33 |
| 54 | 27.01.2019 | 58 | 30 | 53 |
| 55 | 03.02.2019 | 63 | 27 | 20 |
| 56 | 10.02.2019 | 54 | 41 | 87 |
| 57 | 17.02.2019 | 55 | 45 | 45 |
| 58 | 24.02.2019 | 52 | 37 | 31 |
| 59 | 03.03.2019 | 58 | 38 | 57 |
| 60 | 10.03.2019 | 51 | 56 | 64 |
| 61 | 17.03.2019 | 49 | 41 | 39 |
| 62 | 24.03.2019 | 53 | 39 | 50 |
| 63 | 31.03.2019 | 59 | 38 | 24 |
| 64 | 07.04.2019 | 53 | 47 | 53 |
| 65 | 14.04.2019 | 53 | 28 | 56 |
| 66 | 21.04.2019 | 56 | 39 | 38 |
| 67 | 28.04.2019 | 55 | 43 | 39 |

Рисунок Б.2 — Продовження рисунку Б.1

| 1 | Date | SE | PA | DoS |
|---|---|---|---|---|
| 68 | 05.05.2019 | 53 | 34 | 47 |
| 69 | 12.05.2019 | 57 | 33 | 37 |
| 70 | 19.05.2019 | 50 | 38 | 54 |
| 71 | 26.05.2019 | 45 | 39 | 24 |
| 72 | 02.06.2019 | 43 | 40 | 24 |
| 73 | 09.06.2019 | 57 | 28 | 46 |
| 74 | 16.06.2019 | 52 | 36 | 25 |
| 75 | 23.06.2019 | 54 | 35 | 61 |
| 76 | 30.06.2019 | 51 | 35 | 70 |
| 77 | 07.07.2019 | 51 | 38 | 29 |
| 78 | 14.07.2019 | 54 | 23 | 17 |
| 79 | 21.07.2019 | 50 | 36 | 15 |
| 80 | 28.07.2019 | 49 | 16 | 16 |
| 81 | 04.08.2019 | 50 | 36 | 17 |
| 82 | 11.08.2019 | 53 | 42 | 36 |
| 83 | 18.08.2019 | 64 | 32 | 41 |
| 84 | 25.08.2019 | 55 | 45 | 54 |
| 85 | 01.09.2019 | 61 | 43 | 35 |
| 86 | 08.09.2019 | 56 | 36 | 62 |
| 87 | 15.09.2019 | 61 | 44 | 51 |
| 88 | 22.09.2019 | 62 | 32 | 47 |
| 89 | 29.09.2019 | 70 | 45 | 27 |
| 90 | 06.10.2019 | 65 | 45 | 33 |
| 91 | 13.10.2019 | 67 | 43 | 48 |
| 92 | 20.10.2019 | 68 | 40 | 47 |
| 93 | 27.10.2019 | 62 | 41 | 79 |
| 94 | 03.11.2019 | 62 | 44 | 59 |
| 95 | 10.11.2019 | 68 | 38 | 32 |
| 96 | 17.11.2019 | 67 | 58 | 37 |
| 97 | 24.11.2019 | 57 | 47 | 42 |
| 98 | 01.12.2019 | 63 | 32 | 19 |
| 99 | 08.12.2019 | 63 | 53 | 70 |
| 100 | 15.12.2019 | 64 | 41 | 96 |

Рисунок Б.3 — Продовження рисунку Б.1

| 1 | Date | SE | PA | DoS |
|---|---|---|---|---|
| 101 | 22.12.2019 | 34 | 29 | 17 |
| 102 | 29.12.2019 | 44 | 34 | 15 |
| 103 | 05.01.2020 | 58 | 33 | 63 |
| 104 | 12.01.2020 | 56 | 30 | 27 |
| 105 | 19.01.2020 | 57 | 37 | 48 |
| 106 | 26.01.2020 | 59 | 37 | 39 |
| 107 | 02.02.2020 | 65 | 47 | 39 |
| 108 | 09.02.2020 | 66 | 40 | 28 |
| 109 | 16.02.2020 | 60 | 37 | 61 |
| 110 | 23.02.2020 | 66 | 43 | 11 |
| 111 | 01.03.2020 | 58 | 36 | 38 |
| 112 | 08.03.2020 | 62 | 39 | 47 |
| 113 | 15.03.2020 | 55 | 36 | 55 |
| 114 | 22.03.2020 | 55 | 43 | 55 |
| 115 | 29.03.2020 | 60 | 44 | 55 |
| 116 | 05.04.2020 | 64 | 56 | 48 |
| 117 | 12.04.2020 | 69 | 58 | 48 |
| 118 | 19.04.2020 | 69 | 69 | 68 |
| 119 | 26.04.2020 | 67 | 49 | 67 |
| 120 | 03.05.2020 | 67 | 58 | 37 |
| 121 | 10.05.2020 | 67 | 72 | 47 |
| 122 | 17.05.2020 | 67 | 52 | 28 |
| 123 | 24.05.2020 | 58 | 63 | 44 |
| 124 | 31.05.2020 | 70 | 43 | 35 |
| 125 | 07.06.2020 | 73 | 50 | 57 |
| 126 | 14.06.2020 | 70 | 50 | 61 |
| 127 | 21.06.2020 | 57 | 47 | 59 |
| 128 | 28.06.2020 | 56 | 42 | 51 |
| 129 | 05.07.2020 | 55 | 44 | 25 |
| 130 | 12.07.2020 | 73 | 53 | 32 |
| 131 | 19.07.2020 | 69 | 57 | 55 |
| 132 | 26.07.2020 | 60 | 53 | 66 |
| 133 | 02.08.2020 | 71 | 54 | 64 |

Рисунок Б.4 — Продовження рисунку Б.1

| 1 | Date | SE | PA | DoS |
|---|---|---|---|---|
| 134 | 09.08.2020 | 65 | 43 | 74 |
| 135 | 16.08.2020 | 62 | 39 | 53 |
| 136 | 23.08.2020 | 66 | 60 | 21 |
| 137 | 30.08.2020 | 66 | 37 | 87 |
| 138 | 06.09.2020 | 65 | 60 | 52 |
| 139 | 13.09.2020 | 72 | 57 | 84 |
| 140 | 20.09.2020 | 64 | 51 | 27 |
| 141 | 27.09.2020 | 77 | 38 | 40 |
| 142 | 04.10.2020 | 70 | 49 | 56 |
| 143 | 11.10.2020 | 71 | 55 | 47 |
| 144 | 18.10.2020 | 73 | 54 | 59 |
| 145 | 25.10.2020 | 79 | 47 | 42 |
| 146 | 01.11.2020 | 63 | 55 | 61 |
| 147 | 08.11.2020 | 72 | 40 | 52 |
| 148 | 15.11.2020 | 74 | 57 | 56 |
| 149 | 22.11.2020 | 66 | 45 | 48 |
| 150 | 29.11.2020 | 73 | 72 | 81 |
| 151 | 06.12.2020 | 74 | 81 | 50 |
| 152 | 13.12.2020 | 66 | 84 | 22 |
| 153 | 20.12.2020 | 41 | 43 | 34 |
| 154 | 27.12.2020 | 47 | 36 | 31 |
| 155 | 03.01.2021 | 54 | 66 | 37 |
| 156 | 10.01.2021 | 62 | 44 | 20 |
| 157 | 17.01.2021 | 67 | 47 | 26 |
| 158 | 24.01.2021 | 65 | 53 | 38,5 |
| 159 | 31.01.2021 | 64 | 55 | 51 |
| 160 | 07.02.2021 | 60 | 50 | 43 |
| 161 | 14.02.2021 | 57 | 45 | 20 |
| 162 | 21.02.2021 | 63 | 46 | 42 |
| 163 | 28.02.2021 | 68 | 46 | 40 |
| 164 | 07.03.2021 | 59 | 43 | 52 |
| 165 | 14.03.2021 | 65 | 55 | 53 |
| 166 | 21.03.2021 | 62 | 48 | 40 |

Рисунок Б.5 — Продовження рисунку Б.1

| 1 | Date | SE | PA | DoS |
|---|---|---|---|---|
| 167 | 28.03.2021 | 65 | 49 | 26 |
| 168 | 04.04.2021 | 65 | 50 | 29 |
| 169 | 11.04.2021 | 68 | 55 | 26 |
| 170 | 18.04.2021 | 67 | 51 | 23 |
| 171 | 25.04.2021 | 66 | 50 | 50 |
| 172 | 02.05.2021 | 71 | 58 | 77 |
| 173 | 09.05.2021 | 64 | 54 | 30 |
| 174 | 16.05.2021 | 63 | 56 | 63 |
| 175 | 23.05.2021 | 69 | 61 | 46 |
| 176 | 30.05.2021 | 68 | 60 | 24 |
| 177 | 06.06.2021 | 62 | 52 | 42 |
| 178 | 13.06.2021 | 65 | 61 | 32 |
| 179 | 20.06.2021 | 66 | 46 | 63 |
| 180 | 27.06.2021 | 57 | 48 | 47 |
| 181 | 04.07.2021 | 56 | 25 | 37 |
| 182 | 11.07.2021 | 53 | 50 | 27 |
| 183 | 18.07.2021 | 61 | 37 | 41 |
| 184 | 25.07.2021 | 60 | 39 | 55 |
| 185 | 01.08.2021 | 56 | 45 | 46 |
| 186 | 08.08.2021 | 62 | 29 | 24 |
| 187 | 15.08.2021 | 55 | 46 | 13 |
| 188 | 22.08.2021 | 58 | 38 | 40 |
| 189 | 29.08.2021 | 64 | 52 | 60 |
| 190 | 05.09.2021 | 62 | 45 | 25 |
| 191 | 12.09.2021 | 74 | 38 | 14 |
| 192 | 19.09.2021 | 81 | 55 | 80 |
| 193 | 26.09.2021 | 68 | 43 | 30 |
| 194 | 03.10.2021 | 80 | 56 | 74 |
| 195 | 10.10.2021 | 72 | 61 | 73 |
| 196 | 17.10.2021 | 82 | 50 | 57 |
| 197 | 24.10.2021 | 77 | 72 | 52 |
| 198 | 31.10.2021 | 77 | 56 | 54 |
| 199 | 07.11.2021 | 74 | 52 | 49 |

Рисунок Б.6 — Продовження рисунку Б.1

| 1 | Date | SE | PA | DoS |
|---|---|---|---|---|
| 200 | 14.11.2021 | 81 | 63 | 78 |
| 201 | 21.11.2021 | 64 | 67 | 18 |
| 202 | 28.11.2021 | 66 | 59 | 70 |
| 203 | 05.12.2021 | 74 | 58 | 52 |
| 204 | 12.12.2021 | 67 | 76 | 52 |
| 205 | 19.12.2021 | 51 | 44 | 23 |
| 206 | 26.12.2021 | 42 | 39 | 31 |
| 207 | 02.01.2022 | 67 | 61 | 51 |
| 208 | 09.01.2022 | 76 | 69 | 43 |
| 209 | 16.01.2022 | 83 | 68 | 94 |
| 210 | 23.01.2022 | 80 | 88 | 55 |
| 211 | 30.01.2022 | 79 | 69 | 45 |
| 212 | 06.02.2022 | 90 | 62 | 70 |
| 213 | 13.02.2022 | 91 | 80 | 67 |
| 214 | 20.02.2022 | 91 | 87 | 80 |
| 215 | 27.02.2022 | 88 | 100 | 56 |
| 216 | 06.03.2022 | 86 | 97 | 62 |
| 217 | 13.03.2022 | 87 | 92 | 73 |
| 218 | 20.03.2022 | 85 | 97 | 60 |
| 219 | 27.03.2022 | 96 | 87 | 52 |
| 220 | 03.04.2022 | 84 | 92 | 71 |
| 221 | 10.04.2022 | 74 | 64 | 40 |
| 222 | 17.04.2022 | 90 | 95 | 88 |
| 223 | 24.04.2022 | 84 | 90 | 85 |
| 224 | 01.05.2022 | 78 | 84 | 59 |
| 225 | 08.05.2022 | 87 | 84 | 100 |
| 226 | 15.05.2022 | 89 | 79 | 80 |
| 227 | 22.05.2022 | 81 | 73 | 76 |
| 228 | 29.05.2022 | 69 | 59 | 58 |
| 229 | 05.06.2022 | 74 | 61 | 81 |
| 230 | 12.06.2022 | 79 | 67 | 58 |
| 231 | 19.06.2022 | 78 | 92 | 62 |
| 232 | 26.06.2022 | 71 | 81 | 55 |

Рисунок Б.7 — Продовження рисунку Б.1

| 1 | Date | SE | PA | DoS |
|---|---|---|---|---|
| 233 | 03.07.2022 | 68 | 79 | 60 |
| 234 | 10.07.2022 | 65 | 63 | 21 |
| 235 | 17.07.2022 | 74 | 63 | 48 |
| 236 | 24.07.2022 | 73 | 59 | 52 |
| 237 | 31.07.2022 | 62 | 71 | 33 |
| 238 | 07.08.2022 | 68 | 52 | 49 |
| 239 | 14.08.2022 | 72 | 69 | 22 |
| 240 | 21.08.2022 | 76 | 75 | 31 |
| 241 | 28.08.2022 | 84 | 67 | 45 |
| 242 | 04.09.2022 | 80 | 65 | 79 |
| 243 | 11.09.2022 | 92 | 67 | 33 |
| 244 | 18.09.2022 | 98 | 85 | 61 |
| 245 | 25.09.2022 | 85 | 80 | 57 |
| 246 | 02.10.2022 | 85 | 77 | 19 |
| 247 | 09.10.2022 | 92 | 79 | 76 |
| 248 | 16.10.2022 | 100 | 97 | 83 |
| 249 | 23.10.2022 | 95 | 70 | 71 |
| 250 | 30.10.2022 | 91 | 97 | 71 |
| 251 | 06.11.2022 | 86 | 72 | 79 |
| 252 | 13.11.2022 | 89 | 82 | 66 |
| 253 | 20.11.2022 | 72 | 78 | 69 |
| 254 | 27.11.2022 | 84 | 78 | 79 |
| 255 | 04.12.2022 | 81 | 82 | 91 |
| 256 | 11.12.2022 | 81 | 76 | 60 |
| 257 | 18.12.2022 | 69 | 67 | 67 |
| 258 | 25.12.2022 | 55 | 64 | 47 |
| 259 | 01.01.2023 | 64 | 65 | 32 |
| 260 | 08.01.2023 | 75 | 60 | 75 |
| 261 | 15.01.2023 | 78 | 73 | 43 |
| 262 | 22.01.2023 | 83 | 64 | 51 |

Рисунок Б.8 — Продовження рисунку Б.1