Міністерство освіти і науки України
Київський національний університет імені Тараса Шевченка

Факультет інформаційних технологій
Кафедра кібербезпеки та захисту інформації

ПОЯСНЮВАЛЬНА ЗАПИСКА
кваліфікаційної роботи магістра

галузь знань *12 Інформаційні технології*
(шифр і назва галузі знань)

спеціальність *125 Кібербезпека*
(код і назва спеціальності)

освітній ступень *магістр*

освітньо-наукова програма *Кібербезпека*
(назва освітньої програми)

на тему: «Модель оцінки рівня кібербезпеки організації»

Виконавець: студентка ІІ курсу, групи КБм-21

(підпис) **Катерина Моклякова** (Ім'я, ПРІЗВИЩЕ)

	Ім'я, ПРІЗВИЩЕ	**Підпис**
Науковий керівник	Тетяна Бабенко	
Нормоконтроль	Сергій ДАКОВ	

Київ 2023

Міністерство освіти і науки України
Київський національний університет імені Тараса Шевченка

Факультет інформаційних технологій
Кафедра кібербезпеки та захисту інформації

ЗАТВЕРДЖЕНО:
В.о. завідувача кафедри
кібербезпеки
та захисту інформації
_____ Сергій ТОЛЮПА
«24» жовтня 2022 р.

ЗАВДАННЯ
на виконання кваліфікаційної роботи

спеціальності *125 Кібербезпека*
(код і назва спеціальності)

освітній ступень *магістр*

Здобувача(ки) КБм-21 Моклякової Катерини Павлівни
(група) (прізвище ім'я по-батькові)

Тема дипломної роботи Модель оцінки рівня кібербезпеки організації

1. ПІДСТАВИ ДЛЯ ПРОВЕДЕННЯ РОБОТИ

Рішення засідання кафедри кібербезпеки та захисту інформації факультету інформаційних технологій протокол № 3 від 20.10.2022

2. МЕТА ТА ВИХІДНІ ДАНІ ДЛЯ ПРОВЕДЕННЯ РОБІТ

Об'єкт досліджень Процес оцінки рівня кібербезпеки організації.

Предмет досліджень Моделі оцінки рівня кібербезпеки організації

Мета Розробка моделі оцінки рівня кібербезпеки організації.

Вихідні дані для проведення роботи Модель оцінки рівня кібербезпеки організації

3. ОЧІКУВАНІ НАУКОВІ РЕЗУЛЬТАТИ

Наукова новизна Вперше запропонована модель оцінки рівня кібербезпеки рівня організацій, СУІБ яких має гетерогенну природу

Практична цінність	Можливість використання запропонованої моделі як елемента СУІБ організації та в якості інструментарію при проведенні аудиту ІС

4. ВИМОГИ ДО РЕЗУЛЬТАТІВ ВИКОНАННЯ РОБОТИ

Робота виконана у повному обсязі відповідно до теми.

5. ЕТАПИ ВИКОНАННЯ РОБОТИ

Найменування етапів робіт	Строки виконання робіт (початок-кінець)
Розробка плану для досягнення мети роботи	24.10.2022 – 23.01.2023
Аналіз літературних джерел	24.01.2023 – 28.02.2023
Розробка моделі оцінки рівня кібербезпеки організації	28.02.2023 – 05.05.2023
Оформлення і друк пояснювальної записки	05.05.2023 – 19.05.2023

6. РЕАЛІЗАЦІЯ РЕЗУЛЬТАТІВ ТА ЕФЕКТИВНІСТЬ

Економічний ефект	Зниження імовірності реалізації загроз. Усунення економічного дублювання
Соціальний ефект	Покращення процесу управління кібербезпекою організації.

7. ДОДАТКОВІ ВИМОГИ

Завдання видала _____ Тетяна Бабенко
(підпис) (Ім'я, ПРІЗВИЩЕ)

Завдання прийняла до виконання _____ Катерина Моклякова
(підпис) (Ім'я, ПРІЗВИЩЕ)

Дата видачі завдання: 24.10.2022 р.
Термін подання дипломної роботи до ЕК 19.05.2023 р.

РЕФЕРАТ

Пояснювальна записка до дипломної роботи «Моделі оцінки рівня кібербезпеки організації» складається зі вступу, основної частини, що містить 3 розділи, висновків і списку літератури та джерел. Загальний обсяг роботи – 75 сторінок. Робота містить 19 рисунків, 2 таблиці. Список використаних джерел включає 60 джерел.

Об’єкт дослідження – процес оцінки рівня кібербезпеки організації.

Мета роботи – розробка моделі оцінки рівня кібербезпеки організації.

Предмет - моделі оцінки рівня кібербезпеки організації.

Актуальність: моделі оцінки дозволяють організаціям діяти проактивно, а не реактивний в галузі кібербезпеки, що допомагає визначити вразливі місця, стимулює впровадження ефективних засобів контролю та швидке реагування на нові загрози, окрім цього підвищує «видимість» . будь-яка система управління не є гомогенною, різні елементи спрямовані на різні технічні процеси, які вимагають відповідність різним стандартам, що ускладнює процес комплексної оцінки рівня кібербезпеки організації.

Практичною цінністю є можливість використання запропонованої моделі як елемента СУІБ організації та в якості інструментарію при проведенні аудиту ІС.

Наукова новизна: вперше запропонована модель оцінки рівня кібербезпеки рівня організацій, СУІБ яких має гетерогенну природу.

Ключові слова: модель оцінки зрілості кібербезпеки, оцінка зрілості кібербезпеки, бібліотека оцінки кібербезпеки, ефективність кібербезпеки.

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ ТА СКОРОЧЕНЬ

IT	–	Інформаційні технології
ІБ	–	Інформаційна безпека
CIA	–	Конфіденційність, цілісність, доступність
COBIT	–	Цілі управління інформаційними та суміжними технологіями
SIEM	–	Системи управління інформацією та подіями безпеки
CIS	–	Критичні засоби контролю безпеки
GDPR	–	Загальний регламент захисту даних
ISO	–	International Organization for Standardization
NIST	–	Національний інститут стандартів і технологій
PCI DSS	–	Стандарт безпеки даних платіжних карток
HIPAA	–	Закон про переносимість та підзвітність медичного страхування
CMMC	–	Сертифікація моделі зрілості кібербезпеки
CMMI	–	Модель зрілості можливостей
C2M2	–	Модель зрілості потенціалу кібербезпеки
SEI	–	Інститут розробки програмного забезпечення
EDA	–	Дослідницький аналіз даних
TF	–	Періодичність терміну
DIF	–	зворотна частота документа

ЗМІСТ

ВСТУП

Концепція інформаційної безпеки з акцентом на кібербезпеку існує вже наукодостатньо довгий час. Однак загрози в кіберпросторі розвиваються, і це вимагає від компаній адаптації, модернізації та інвестицій в кібербезпеку організації. Аби впевнитися в тому, що в організації прикладається достатньо зусиль і стратегія кібербезпеки є адекватною, компанії необіхдно проводити різного роду оцінювання як для перевірки себе і визначення вектору розвитку кібербезпеки, так і для того, щоб підтвердити відповідність стандартам і мати змогу вести бізнес з іншими компаніями, державою, чи в різних локаціях.

Процес оцінки рівня кібербезпеки організації є тестуванням та/або оцінкою управлінських, операційних і технічних засобів контролю безпеки в інформаційній системі для визначення ступеня, в якому засоби контролю реалізовані правильно, працюють за призначенням і дають бажаний результат щодо виконання вимог безпеки для системи [1].

Актуальність класифікаційної роботи полягає в тому, що природа кіберзагроз, що постійно розвиваються, і дедалі складніша технологічна екосистема вимагають проактивного підходу до кібербезпеки. Існуючі моделі оцінки часто не встигають за новими загрозами, що робить організації вразливими до кібератак. Розробляючи нові моделі та модернізуючи існуючі, можливо краще узгодити процес оцінки з поточними проблемами кібербезпеки та ефективно вимірювати здатність організації зменшувати ризики. Моделі оцінки дозволяють організаціям прийняти проактивний, а не реактивний підхід до кібербезпеки, що визначає вразливі місця, стимулює впровадження ефективних засобів контролю та швидке реагування на нові загрози. Крім того, будь-яка система управління не є гомогенною, різні елементи спрямовані на різні технічні процеси, які вимагають відповідність різним стандартам, що ускладнює процес комплексної оцінки рівня кібербезпеки організації. Одним із прикладів складної інформаційної системи, яка вимагає відповідності різним нормам, є хмарна система управління охороною здоров’я. Ця система передбачатиме

зберігання, керування та обробку конфіденційних даних пацієнтів із дотриманням різноманітних нормативних рамок, таких як Закон про перенесення та підзвітність медичного страхування (HIPAA) у Сполучених Штатах, Загальний регламент захисту даних (GDPR) у Європейському Союзі та інші регіональні закони про захист даних у сфері охорони здоров’я.

Метою роботи є розробка моделі оцінки рівня кібербезпеки організації, яка відображає ступінь відповідності впровадження контролів різних регулятивних стандартів.

Практичною цінністю роботи є можливість застосування запропонованої моделі оцінки рівня кібербезпеки у існуючоому процесі захисту інформації, як елементу управління кібербезпекою організації.

З одного боку, відповідний рівень кібербезпеки є необхідною складовою ефективної діяльності будь-якої організації, а з іншого - недостатні заходи забезпечення кібербезпеки можуть призвести до значних фінансових втрат і негативного впливу на репутацію компанії.

У зв'язку з цим, розробка моделі оцінки кібербезпеки організації є вкрай актуальною задачею, що дозволить виявляти недоліки в системі захисту інформації та розробляти стратегії щодо їх усунення.

РОЗДІЛ 1
ДОСЛІДЖЕННЯ МОДЕЛЕЙ І МЕТОДІВ ОЦІНКИ КІБЕРБЕЗПЕКИ ОРГАНІЗАЦІЇ

1.1 Класифікація контролей інформаційної безпеки

Впровадження контролів інформаційної безпеки є критично важливим і нетривіальним завданням через низку чинників. Метою введення вимог до безпеки інформаційної системи є застосування заходів безпеки для захисту активів, кінцевих пристроїв, мереж і даних від несанкціонованого доступу та потенційних загроз [2]. До причин появи контролей безпеки належать:

- Використання вразливостей: нові складні методи використання вразливостей у комп'ютерних системах з'являються постійно [3]. Оцінка та впровадження засобів контролю інформаційної та кібер- безпеки має вирішальне значення для запобігання реалізації загроз і захисту критично важливих активів. Аби визначити актуальні вимоги безпеки необхідно всебічне розуміння потенційних вразливостей, методів їх використання та здатністьі розробляти та розгортати ефективні контрзаходи безпеки.
- Управління ризиками безпеки: ефективне впровадження контролів інформаційної безпеки є важливим аспектом управління ризиками. Цей процес залучає науковий аналіз та оцінку потенційних ризиків, включаючи ймовірність їх настання та потенційний вплив. Організації повинні визначати пріоритети заходів безпеки на основі оцінки ризиків, беручи до уваги такі фактори, як конфіденційність даних, бізнес-операції та нормативні вимоги. Це передбачає збалансування розподілу ресурсів при максимальному зниженні ризику.
- Складність систем і мереж. Сучасне IT середовище включає все більше компонентів, що робить інформаційні системи та мережі складними та взаємопов'язаними. Таким чином, впровадження контролів безпеки є складною справою. Розуміння взаємозалежностей і потенційної вразливості різних компонентів інформаційних систем, таких як апаратне забезпечення, програмне забезпечення,

протоколи та людський фактор ускладнює досягнення бажаного рівня захищенності систем.

- Відповідність і юридичні вимоги. Організації повинні дотримуватися різних нормативних актів і законодавчої бази щодо кібербезпеки. Забезпечення відповідності державним і міжнародним стандартам вимагає адаптації в процес управління інформаційною безпекою складних правил, оцінки ефективності контролю та ведення документації для демонстрації дотримання [4].

- Поведінка та освіта користувачів: людський фактор відіграє значну роль у реалізації контролю інформаційної безпеки. Дії та поведінка користувачів можуть ненавмисно наражати системи на вразливі місця [5]. Таким чином, впровадження засобів контролю, спрямованих на людський фактор, а також сприяння обізнаності та освіті користувачів щодо кібербезпеки є критично важливим і передбачає розробку орієнтованих на користувача елементів керування, навчання та постійного вдосконалення «безпечних» практик, що може бути нетривіальним завданням через різноманітну базу користувачів і потребу в ефективній зміні поведінки.

Підсумовуючи, впровадження контролів інформаційної безпеки має першорядне значення через постійну еволюцію загроз, складність систем та IT середовища, управління ризиками, вимоги відповідності, нові технології та вплив людського фактору.

Існують різні способи класифікації засобів контролю кібербезпеки, але більшість з них поділяються на три основні категорії: адміністративні, технічні і фізичні контролі [6]. Засоби адміністративного контролю також відомі як засоби керування безпекою, і вони забезпечують вказівки, правила та процедури для впровадження середовища безпеки. Технічні засоби контролю необхідні для того, щоб гарантувати дотримання політики безпеки та ефективність заходів безпеки для захисту від потенційних загроз. Технічні засоби контролю можуть включати брандмауери, системи виявлення вторгнень, шифрування та інші технології безпеки. Фізичні засоби контролю призначені для захисту фізичних активів організації, таких як будівлі, обладнання та люди. Вони можуть включати спостереження, екологічний

контроль і планування на випадок надзвичайних ситуацій. Ефективна реалізація контролю безпеки базується на його класифікації щодо інциденту безпеки.

Загальні типи класифікації контролів[1]:

- Превентивні - запобіжні засоби контролю, які намагаються запобігти виникненню інциденту;
- Детективні – мають на меті виявлення інцидентів після того, як вони сталися;
- Коригувальні – впроваджуються з метою найшвидшого усунення впливу інциденту;
- Компенсаційні (або стримуючий фактор) - засоби стримування намагаються перешкодити виненкенню інциденту, або використовуються як альтернативні засоби керування, коли неможливе впровадження первинного контролю.

Таким чином, згідно класифікації, контролі безпеки встановлюють вимоги до захищенності інформаційних систем на різних етапах імовірної реалізації загроз.

1.2 Найпоширеніші стандарти управління КБ та методи оцінки на їх відповідність

Стандарти кібербезпеки з’явилися як відповідь на збільшення частоти та складності кібератак. У міру того, як організації та окремі люди стали більше залежати від технологій, ризики, пов’язані з кібератаками, зросли експоненціально. Стандарти кібербезпеки допомагають створити вказівки, описати найкращі практики та вводять спільні поняття для фахівців з кібербезпеки, гарантуючи наявність у них необхідних інструментів і знань для захисту систем і даних.

Стандарти кібербезпеки також сприяють взаємодії та сумісності між різними технологічними системами та продуктами, оскільки вони встановлюють загальний набір вимог і критеріїв, яким мають відповідати всі продукти та послуги. Це може спростити для організацій інтеграцію різних технологій безпеки та рішень, а також гарантувати ефективну співпрацю для забезпечення комплексного захисту від кіберзагроз.

Цілі управління інформаційними та суміжними технологіями (COBIT)— це структура управління організацією, інформацією та технологіями. Цей стандарт містить компоненти та фактори проектування для побудови та підтримки системи управління, яка потрібна організації [7]. З моменту свого заснування COBIT пройшов кілька версій, кожна з яких представляє прогрес і вдосконалення практики управління IT.

Важливо відзначити, що кожна версія COBIT будується на основі попередньої, враховуючи відгуки практиків і адаптуючись до IT та бізнес середовища. Остання версія, COBIT 2019, надає організаціям комплексну та гнучку структуру для вирішення проблем управління IT та досягнення їхніх стратегічних цілей. Деякі з версій стандарту містять значні зміни і широко використовуються у сьогоденні, незважаючи на те, що остання версія стандарту була представлена у 2019 році:

- COBIT 4.0 (2005): має значне оновлення з більш структурованим підходом до управління IT. Включає посібник COBIT Control Practices, що узгоджує IT-процеси з бізнес-цілями. Ця версія також підкреслила важливість управління IT-ризиками та надала вказівки щодо впровадження управління IT.
- COBIT 5 (2012): об’єднує попередні версії та розширює сферу застосування, щоб охопити управління підприємством і управління інформацією та технологіями. В стандарті було представлено цілі управління та менеджменту, концепцію стимулюючих засобів і модель оцінки можливостей процесу.
- COBIT 2019 (2018): значне оновлення COBIT 5, узгоджене з сучасними практиками управління. В цьому стандарті запроваджено спрощену та більш гнучку структуру, зосереджену на цілях управління в масштабах підприємства. COBIT 2019 також підкреслює необхідність цілісного підходу до управління інформацією та технологіями. Надалі в дослідженні буде розглянуто цю версію стандарту.

Структура стандарту управління інформаційними та суміжними технологіями: цілі управління та керівництва, або також відома як базова модель COBIT 2019. COBIT всебічно описує 40 цілей управління. Розробка нових інструкцій, навчання та ресурсів для підтримки COBIT постійно оцінюється на основі ринкового попиту та контролюється за ISACA [8]. У COBIT 5 є сім факторів, які є основними

компонентами для досягнення цілей управління для ефективного використання інформаційних технологій.

COBIT описує 5 доменів IT-процесів організації в межах двох основних областей процесів, а саме [9]:

1) Управління, містить п’ять процесів управління, визначених практикою в кожному процесі оцінки, безпосереднього та моніторингу (EDM).

2) Керування, що містить чотири домени, узгоджені зі сферою відповідальності за планування, створення, виконання та моніторинг, а також охоплюють повну IT-сферу від кінця до кінця, включаючи:

a) Узгодження, планування та організація (APO), включаючи узгодження, планування та налаштування, щоб IT могли сприяти досягненню бізнес-цілей;

b) Побудова, придбання, та впровадження (BAI), включаючи процес створення, придбання та впровадження систем, які підтримують бізнес-процеси;

c) Доставка, обслуговування та підтримка (DSS), включає доставку, обслуговування, підтримку або забезпечення бізнес-процесу;

d) Моніторинг, валідація та оцінка (MEA), включає моніторинг, оцінку управління процесами/ з боку незалежних моніторингових агенцій як всередині організації, так і поза нею.

COBIT має сім активних компонентів управління (рис.1.1). Компоненти управління включають процеси (processes), організаційні структури (organizational structures), принципи, політики та процедури (principles, policies, procedures), інформацію (information), культуру, етику та поведінку (culture, ethics and behavior), людей, їх навички та компетенції (people, skills and competencies), інфраструктуру сервісів та додатків (service infrastructure and competencies).

Рисунок 1.1 Компоненти управління COBIT

Процес захисту інформації або даних компанії для забезпечення та гарантування безпеки інформаційних систем гарантує, що ризики інформаційної системи можуть прийняти відповідні політики безпеки та відповідає за регулювання прав доступу користувачів у компанії таким чином, щоб використання прав доступу користувачів було правилами та повноваженнями, встановленими в компанії. Оцінювання рівня кібербезпеки з фокусом на DSS05 (Manage Security Services) було проведено з використанням розрахунку шкали Лайкерта [10] Це дослідження є якісним, з розрахунком терміну зрілості субдомену DSS05 (керування службами безпеки). В ході дослідження учасники, визначені за класифікацією RACI, повинні були заповнити список питань, за результатами опитування було виділено 0-5 рівні зрілості кібербезпеки організації.

Стандарт NIST 800-53, також відомий як «Спеціальна публікація NIST 800-53», опублікований Національним інститутом стандартів і технологій у Сполучених Штатах, містить вказівки та засоби контролю для забезпечення безпеки федеральних інформаційних систем і захисту конфіденційної інформації [11]. NIST 800-53 пропонує комплексну структуру, яка охоплює різні елементи керування безпекою та конфіденційністю, організовані у 18 різних групах елементів управління. Ці групи охоплюють широкий спектр сфер, включаючи контроль доступу, реагування на інциденти, оцінку ризиків, цілісність системи та інформації та багато інших. Стандарт регулярно оновлюється з урахуванням нових загроз, технологічного прогресу та змін у нормативно-правовій системі. Він служить життєво важливим ресурсом для федеральних агентств, підрядників і організацій, які обробляють федеральну інформацію, забезпечуючи їм структурований підхід до впровадження ефективних заходів безпеки та керування ними. Дотримуючись стандарту NIST 800-53, організації можуть покращити рівень безпеки своїх інформаційних систем, зменшити ризики та забезпечити конфіденційність, цілісність і доступність критично важливих даних.

ISO/IEC 27001 — це міжнародний стандарт систем управління інформаційною безпекою (СУІБ) [12]. Він надає організаціям систематичний підхід до керування безпекою конфіденційних інформаційних активів, включаючи фінансові дані, інформацію про клієнтів, інтелектуальну власність тощо. З часом ISO 27001 зазнав різноманітних переглядів і оновлень, щоб відобразити нові практики безпеки:

ISO/IEC 27001:2005: Перша версія ISO 27001 була опублікована в 2005 році. Вона встановила початкові вимоги для створення, впровадження, підтримки та постійного вдосконалення СУІБ в організації. Ця версія забезпечила систематичну основу для виявлення та управління ризиками інформаційної безпеки.

ISO/IEC 27001:2013: друга версія ISO 27001 була випущена в 2013 році, замінивши попередню редакцію. Ця редакція внесла значні зміни та вдосконалення стандарту. Було запроваджено більш орієнтований на ризик підхід, наголошуючи на оцінці та обробці ризиків як ключових компонентах СУІБ. Версія 2013 року також

узгоджується зі структурою високого рівня, що полегшує інтеграцію з іншими стандартами систем управління.

Остання редакція стандарту відбулася в 2022 році з такими змінами [13]:

- Зміна назви ISO 27002:2022. Ця назва відповідає найновішій редакції серії ISO 27000.
- Зміни в контролях: стандарт тепер включає 93 засоби контролю, порівняно зі 114 у попередній редакціх. Це пов’язано з тим, що деякі елементи керування було об’єднано з іншими для кращого узгодження або видалено через дублювання. Також додано 11 нових елементів керування.
- Деякі терміни видалено або замінено. Терміни «цілі контролю» та «кодекс практики» були видалені..
- Фокус на кіберзагрозах: у стандарті ISO 27002:2022 кіберризикам приділено більше уваги, і організації повинні вживати запобіжні заходи для захисту своїх мереж і систем від кібератак.

Надалі в дипломній роботі розглядається редакція ISO 27002:2022.

СУІБ допомагає організації підтримувати безпечну та безризикову інфраструктуру та бізнес. СУІБ стосується процесів, методів, процедури, політики та інструментів з конкретними організаційними та технічними заходами, які постійно контролюються з поступовим вдосконаленням у контрольованому середовищі. Щоб досягти максимальної безпеки для критично важливих систем, кожен елемент навколишнього середовища, як-от людські ресурси, організація, програмне забезпечення, обладнання тощо, має бути захищеним від ризиків і атак [14].

Серія ISO 27000 містить рекомендації щодо найкращої практики імплементації та використання СУІБ для отримання сертифікату ISO 27001. Окрім, ISO 27001, важливими є також: ISO 27000 (містить вступ і огляд сімейства ISO 27000 із чітким визначенням і словником); ISO 27002 (надає детальний каталог про те, як досягти різних засобів контролю, перелічених у Додатку A ISO 27001); ISO 27005 (містить детальні вказівки щодо управління ризиками, оцінки та обробки, як-от зменшення, уникнення, передача чи прийняття). Окрім наведених стандартів, варто згадати про ISO 27017 - стосується інформаційної безпеки в хмарних середовищах, де захистом

конфіденційності керує ISO 27018. Кілька стандартів онлайн-безпеки, як-от ISO 27032, стосуються кібербезпеки, ISO 27033 керує мережевою безпекою, а ISO 27034 керує безпекою додатків, тоді як ISO 27033 та ISO 27034 складаються з кількох частин.

Засоби контролю за стандартом ISO 27001 — це керівництво з найкращих практик, яке необхідно запровадити, щоб зменшити ризики до прийнятного рівня. У поточній версії стандарту перераховано 93 елементи керування в Додатку А, організованих за чотирма темами, пронумерованими від А.5 до А.8. Перша тема А.5 це 37 організаційних контролів, що стосуються визначення поведінки людей, програмного забезпечення, обладнання та систем. Друга тема А.6 – це люди, 8 засобів контролю, щоб люди могли відповідати стандартам безпеки завдяки належним знанням, освіті, навичкам і досвіду. Третя тема — фізична А.7, яка містить 14 елементів керування для роботи з обладнанням або пристроями, які фізично взаємодіють з людьми та об’єктами. Остання тема — технологічна А.8, що містить 34 контролі, які реалізуються за допомогою програмного та апаратного забезпечення, наприклад антивірусного програмного забезпечення, резервного копіювання даних тощо.

Стандарти ISO 27001 можна впровадити, дотримуючись циклу «Плануй, роби, перевіряй, дій» (PDCA), який виник із забезпечення якості [15]. Якість СУІБ покращуватиметься з плином часу, дотримуючись циклу PDCA. Рис. 1.2 демонструє, як ми можемо запровадити стандарт ISO 27001 із циклом PDCA, щоб підвищити гнучкість, ясність та об’єктивність процесів управління. Обов’язкові пункти 4, 5, 6 (контекст організації, керівництво, планування) увійдуть до фази планування разом із створенням СУІБ. Впровадження СУІБ необхідно буде завершити на етапі виконання, який охоплює обов’язкові пункти 7 і 8. На етапі перевірки здійснюватиметься моніторинг і перегляд СУІБ для виконання обов’язкового пункту 9 (оцінка ефективності). Останньою частиною циклу є акт, який стосується підтримки та вдосконалення СУІБ шляхом дотримання пункту 10 (удосконалення).

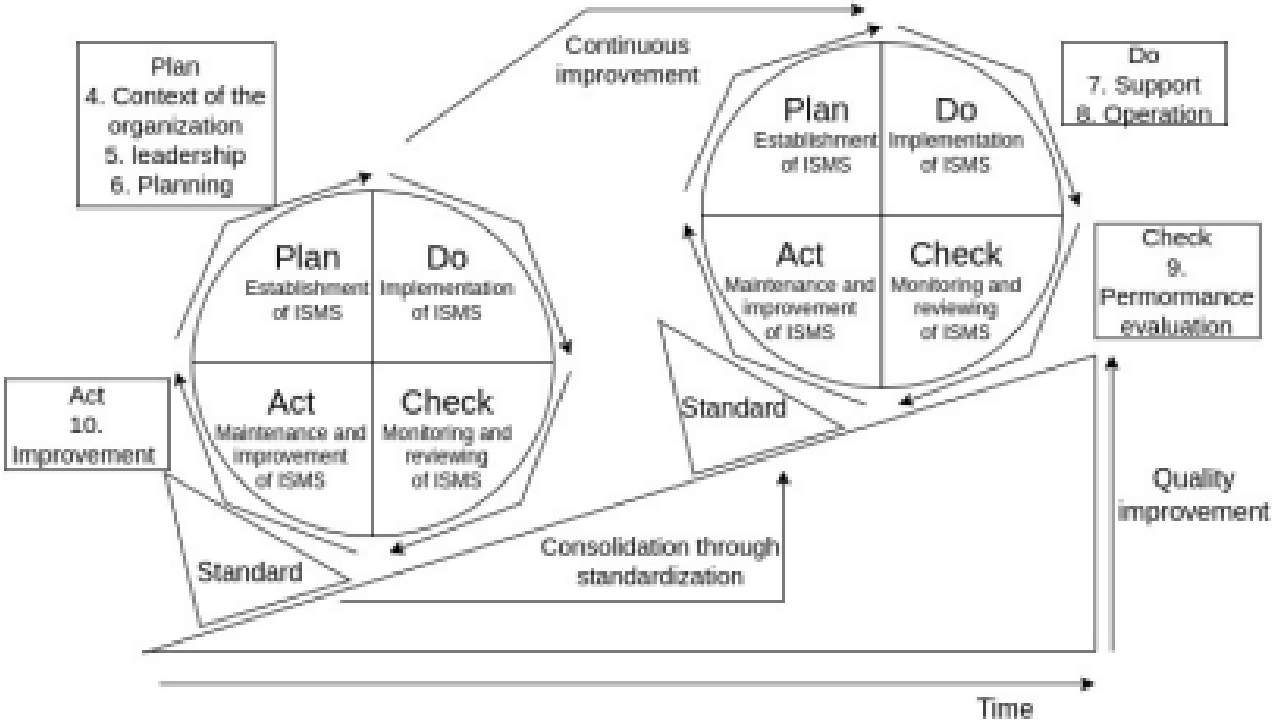

Рисунок 1.2 Цикл "Плануй, роби, перевіряй, дій"

ISO 27001 забезпечує конфіденційність, цілісність і доступність, які є основними цілями безпеки та цілями захисту цього стандарту. Впроваджуючи та приймаючи стандарти ISO 27001, організація може отримати сертифікат, який буде дійсним доказом дотримання рекомендацій ISO. Забезпечуючи відповідність всесвітньо відомому стандарту, сертифікація значно зменшує потребу в повторних аудитах [16].

Для того, щоб спросити, оптимізувати та виміряти впровадження котнролів стандарту, було запропоновано декілька підходів і проведено дослідження.

В дослідженні [17] [18] було запропоновано модель для дослідження впливу організаційних факторів на ефективність впровадження стандарту BS7799 (замінений на ISO/IEC 27001). Висновки визначили чотири фактори, які можуть серйозно вплинути на успішність впровадження управління інформаційною безпекою, серед яких IT-компетентність бізнес-менеджерів, невизначеність середовища, тип галузі та розмір організації. Вплив цих факторів може бути різним для будь-якого типу організацій. Результати показують, що великі організації можуть отримати більше

вигоди від впровадження стандартів безпеки інформаційних систем, оскільки вони більше залежать від формалізації та стандартизації, ніж малі компанії, і мають більшу кількість активів.

Процес розроблення вимог безпеки (SREP) [19] [20] було запропонований для включення вимог безпеки, таких як загальні критерії (ISO/IEC 15480), у модель життєвого циклу програмного забезпечення в структурованому процесі. SREP використовує набір стандартів, процесів і заходів для розробки захищених інформаційних систем відповідно до системного підходу. Структура складалася з дев'яти дій, відомих як мікропроцес, для формування вимог безпеки інженерії, а також зовнішніх і видимих артефактів, пов'язаних із діяльністю. Діяльність включала визначення бачення безпеки, розуміння зацікавлених сторін, ідентифікацію вразливостей та активів, ідентифікацію цілей безпеки та загроз, оцінку ризиків, визначення пріоритетів – перевірку вимог безпеки та вдосконалення рівня кібербезпеки організації.

OntoWorks [21] [22] є онтологічною картою стандарту ISO/IEC 27001, що підтримує процес сертифікації. Автори запропонували структуру для використання онтологічних даних і надання користувачам доступу до онтологічних даних, їх візуалізації та міркувань. Їхній внесок допоміг у підготовці аудиту та перевірці на відповідність правилам щодо засобів контролю ISO/IEC 27001.

З іншого боку, було запропоновано процес розроблення вимог безпеки для лінійок програмних продуктів (SREPPLine) [23], [24]. Це рішення для керування вимогами безпеки на ранній стадії розробки лінійки продуктів на основі стандартів безпеки. Ця структура - структуроване управління вимогами безпеки для сприяння відповідності лінійки програмних продуктів відповідним стандартам безпеки, таким як ISO/IEC 27001 та ISO/IEC 15408.

Насамкінець, великий внесок зробила запропонована системи управління інформацією та подіями безпеки (SIEM) [25]. Структура, яка дозволяє організації оцінювати свою відповідність стандартам ІБ та ефективність їх впровадження шляхом автоматичної генерації ISO 27001 на основі показників безпеки IT [26].

PCI DSS — це глобальний стандарт безпеки даних платіжних карток (кредитних, дебетових, банкоматів) для всіх суб'єктів, які обробляють, зберігають або передають дані власників карток та/або конфіденційні дані автентифікації, що передаються онлайн [27]. PCI DSS складається з технічних і операційних вимог, які обробляють платіжні транзакції, розробників програмного забезпечення, виробників додатків і пристроїв, що використовуються в цих транзакціях, для підвищення безпеки даних платіжних карткових рахунків [28].

PCI-DSS – це стандарт інформаційної безпеки для організацій, які працюють із фірмовими кредитними картками основних карткових схем. PCI DSS надає базові вимоги безпеки, які можуть допомогти підприємствам створювати програми безпеки та визначати, які кроки вжити. Є три кроки для дотримання стандарту PCI DSS:

- Оцінка, ідентифікація даних власника картки, запис інвентаризації активів інформаційних технологій разом із бізнес-процесом обробки платіжних карток та аналіз вразливостей, які потенційно можуть розкрити дані власника картки.
- Усунути, покращити захист від виявлених вразливостей. У цьому випадку, не зберігаючи непотрібні дані власника картки та реалізуючи безпечний бізнес-процес.
- Звіт, двоетапна документація та коригувальні записи. Наступним процесом є надсилання звіту про відповідність відповідному банку для захисту використаної картки.

Вимоги до сертифікації PCI DSS також стосуються сторонніх постачальників центрів обробки даних як засобу зберігання або резервного копіювання даних про власників карток. PCI DSS видає понад 250 (двісті п'ятдесят) підвимог, які згруповані в 6 (шість) цілей і 12 (дванадцять) основних вимог, які повинні бути дотриманими, щоб отримати сертифікат, що організація запровадила стандарт безпеки PCI DSS [29].

Дослідження [30] пропонує незалежний метод вимірювання зрілості інформаційної безпеки, взятий із моделі зрілості інтегрованих можливостей, що ґрунтується на загальній моделі структури кібербезпеки, а також те, що необхідно застосувати до організації (рис. 1.3). Запропонована модель підходу PCI-DSS пропонує механізм вимірювання інформаційної безпеки з використанням запропонованої моделі для досягнення відповідності PCI-DSS. Стандарт PCI-DSS

включає технічні та операційні системні компоненти, які включені в дані про власників карток або пов'язані з ними. Метою PCI DSS є захист даних власників карт, де б вони не оброблялися, не зберігалися та не передавалися. Модель допомагає організаціям легко ідентифікувати ключові фактори успіху та прогалини та може використовувати як інструмент самооцінки та аудиту, допомагаючи організаціям проводити аналіз прогалин та отримувати автоматичні звіти про відповідність та графічне представлення їх стану безпеки. Модель складається із чотирьох етапів, перший етап – вхідні дані, на етапі збираються первинні дані як оцінка зовнішніх факторів та вторинні дані як оцінка внутрішніх факторів. Другий етап - це етап зіставлення шляхом проведення аналізу прогалин для вироблення загальної вимоги кібербезпеки. На третьому етапі виконується зіставлення GCR з PCI-DSS. Останній етап – етап ухвалення рішення. У дослідженні використовується метод анкетування, адаптований до стандартної структури кібербезпеки.

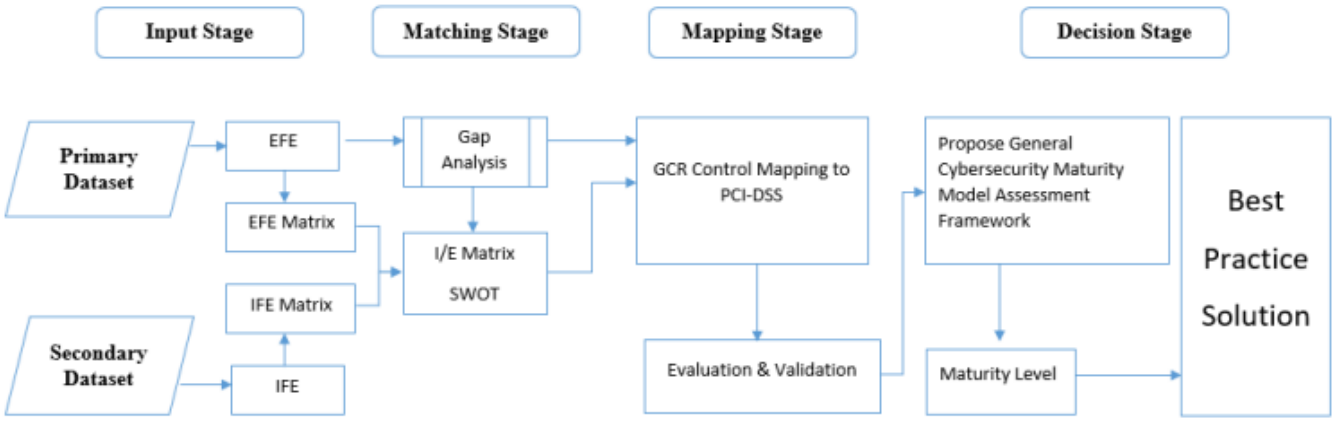

Рисунок 1.3 метод вимірювання зрілості на основі сертифікації на відповідність стандарту PCI DSS

Іншим галузевим стандартом для організацій, які обробляють, передають чи зберігають інформацію в секторі охорони здоров’я є HIPAA. Закон про переносимість та підзвітність медичного страхування 1996 р. (HIPAA) — це федеральний закон, який вимагає створення національних стандартів для захисту конфіденційної медичної інформації пацієнта від розкриття без згоди чи відома пацієнта [31].

GDPR — це закон ЄС, що містить обов'язкові правила про те, як організації та компанії повинні використовувати особисті дані безпечним для цілісності способом. Під персональними даними розуміється будь-яка інформація, яка безпосередньо чи опосередковано може ідентифікувати живе обличчя. Ім'я, номер телефону та адреса є прикладами особистих даних зі шкільних підручників [32].

1.3 Рівні зрілості IT та їх кореляція до рівнів зрілості КБ

У сфері інформаційних технологій (IT) рівні зрілості стосуються прогресивних етапів розвитку та вдосконалення, які проходить організація або система з точки зору її загальних можливостей, процесів та ефективності. Ці рівні зрілості часто оцінюються за допомогою встановлених структур або моделей, таких як інтеграція моделі зрілості можливостей або бібліотека інфраструктури інформаційних технологій (ITIL) [33].

Рівні зрілості IT можна розділити на кілька ступенів, причому кожен ступінь представляє вищий рівень організаційної та операційної зрілості. Конкретні назви та визначення цих рівнів можуть відрізнятися залежно від використовуваної бібліотеки оцінювання, але зазвичай вони поділяється наступним чином (табл. 1.1).

Таблиця 1.1 Рівні зрілості IT

Рівень зрілості IT	Опис
Початковий рівень	організації мають випадкові та хаотичні процеси без формальної структури. Основна увага зосереджена на індивідуальних зусиллях, а видимість і контроль над IT-операціями обмежені.
Рівень повторюваності	Організації на цьому рівні починають впроваджувати базові процеси та процедури, які можна повторити. Однак ці процеси все ще можуть бути не стандартизовані, і існує сильна залежність від індивідуального досвіду.

Продовження табл. 1.2

Визначений рівень	На цьому рівні організації мають чітко визначені та задокументовані процеси, яких дотримуються працівники. Основна увага приділяється стандартизації та інтеграції процесів у різних сферах IT, що веде до підвищення ефективності.
Керований рівень	Цей рівень відповідає організаціям, які впровадили показники та вимірювання для моніторингу та контролю своїх IT-процесів. Частково впроваджено прийняття рішень на основі даних. Пріоритет на постійному вдосконаленні.
Рівень оптимізації	Організації з цим рівнем зрілості повинні мати культуру постійного вдосконалення. Акцент робиться на інноваціях, проактивному управлінні ризиками та оптимізації ресурсів.

Співвідношення між рівнями зрілості IT і рівнями зрілості інформаційної безпеки є критичним етапом, оскільки інформаційна безпека є невід'ємною частиною операцій IT. Зі збільшенням рівня зрілості IT, більш імовірно, що організації матимуть більш надійні методи захисту інформації та засоби контролю. Ця кореляція випливає з того, що вищий рівень зрілості часто означає краще визначені процеси, підвищену обізнаність про ризики, покращене управління та більшу відданість безпеці.

Рівні зрілості інформаційної безпеки відносяться до прогресивних етапів розвитку та вдосконалення, які проходить організація з точки зору її можливостей, практики та ефективності інформаційної безпеки. Ці рівні зрілості часто оцінюються за допомогою встановлених структур або моделей, які досліджуються в наступних

розділах. Рівні зрілості інформаційної і зокрема кібербезпеки, можуть відрізнятися залежно від використовуваного фреймворку, але зазвичай вони охоплюють наступні ступені (табл. 1.2).

Таблиця 1.3 Рівні зрілості ІТ

Рівень зрілості ІБ	Опис
Початковий рівень	На цій стадії організації мають спеціальні або реактивні практики інформаційної безпеки. Існує обмежена обізнаність про ризики інформаційної безпеки, і заходи безпеки зазвичай впроваджуються за потреби. Політики та процедури можуть бути неформальними або взагалі не існувати.
Керований рівень	Організації, що мають керований рівень повинні мати встановлені базові практики та політику інформаційної безпеки, визначені ключові інформаційні активи та пов’язані з ними ризики, запроваджені певні заходи безпеки, такі як контроль доступу та процедури реагування на інциденти. Однак на даному рівні в безпеці все ще можуть бути прогалини та невідповідності.
Визначений рівень	На цьому етапі організації мають чітко визначені та задокументовані політики, стандарти та процедури інформаційної безпеки. Наявна комплексна оцінка ризиків і розроблений формалізований підхід до управління інформаційною безпекою. Засоби контролю безпеки узгоджуються з найкращими галузевими практиками, і існує чітке розуміння ролей і відповідальності за інформаційну безпеку.

Продовження табл. 1.4

Вимірюваний рівень	Цей рівень відповідає організаціям, які впровадили показники та механізми вимірювання для оцінки ефективності засобів контролю та процесів інформаційної безпеки. Інциденти безпеки, вразливості та дотримання вимог регулярно відстежуються, аналізуються. Постійне вдосконалення та періодична переоцінка стану безпеки є пріоритетною.
Рівень оптимізації	На найвищому рівні зрілості організації мають зрілу та проактивну позицію інформаційної безпеки. Підтримується сильна культура безпеки, зосереджена на інноваціях, управлінні ризиками та постійному вдосконаленні. Інформаційна безпека інтегрована в усі аспекти діяльності організації.

Кореляція між рівнями зрілості інформаційної безпеки та рівнями зрілості IT є значною, оскільки інформаційна безпека є критичним аспектом операцій IT. Вищий рівень зрілості інформаційної безпеки часто вказує на вищий рівень зрілості IT, оскільки організації з більш зрілими IT-процесами, як правило, мають краще визначені та ефективніші практики інформаційної безпеки. Таким чином, кожен рівень зрілості IT співвідноситься з відповідним рівнем зрілості інформаційної безпеки і можна припустити, що в міру того, як організації просуваються до вищих рівнів IT-зрілості, вони також впроваджують вищий рівень зрілості інформаційної безпеки.

1.4 Моделі оцінки кібербезпеки організації

На відміну від стандартів, які висувають вимоги до організацій стосовно забезпечення кібербезпеки, моделі оцінки кібербезпеки організації дозволяють визначити, які кроки необхідно виконати для відповідності тому чи іншому рівню зрілості. За допомогою моделй оцінки можна порівняти кібербезпеку конкрентої організації з кращими практиками і таким чином, визначити, що для конкретної компанії буде цільовим станом.

Зрілість Кібербезпеки, типово поділяється на 2 зони: зрілість можливостей, зрілість процесів [34].

Моделі зрілості можливостей кібербезпеки зазвичай структуровані за такими елементами:

- Області чи виміри: область об'єднує загальні концепції організаційних процесів, і кожна область необов'язково незалежна від інших.
- Фактори та індикатори: фактори – це цілі, які мають бути виконані в кожній з областей моделі, а індикатори служать для візуалізації прогресу у досягненні цілей.
- Рівні зрілості: це результат оцінки виконання факторів та показників у галузях чи вимірах організації. Рівні зрілості варіюються від початкового рівня, коли організація, можливо, тільки почала замислюватися про кібербезпеку, до динамічного порівняння, коли організація здатна швидко адаптуватися до змін у ландшафті кібербезпеки, пов'язаних із загрозами, вразливістями, ризиками, економічною стратегією чи зміною потреб організаційної структури [35].

Модель зрілості - це набір характеристик, атрибутів, індикаторів або патернів, які представляють здібності та прогрес у певній дисципліні. Зміст моделі зазвичай є передовим досвідом, що використовується фахівцями в даній дисципліни, і може включати стандарти або інші склепіння правил дисципліни. Таким чином, модель зрілості забезпечує орієнтир, за яким організація може оцінити поточний рівень можливостей своїх практик, процесів та методів, а також встановити цілі та пріоритети для покращення. Крім того, коли модель широко використовується в конкретній галузі (і результати оцінки є загальними), організації можуть порівняти

свою ефективність з іншими організаціями, а галузь може визначити, наскільки добре вона працює загалом, вивчивши можливості своїх організацій-членів. Для вимірювання прогресу моделі зрілості зазвичай мають рівні за шкалою. Кожен рівень визначається набором атрибутів; якщо організація демонструє ці атрибути, кажуть, що вона досягла як цього рівня, і можливостей, які цей рівень представляє. Наявність вимірних перехідних станів між рівнями дозволяє організації використовувати шкалу, щоб: визначити свій поточний стан; визначити своє майбутнє, зріліший стан; визначити цілі, які необхідно досягти, щоб наблизитися цього майбутнього стану.

Модель зрілості потенціалу кібербезпеки (C2M2) призначена для використання організацією для послідовної оцінки своїх можливостей у галузі кібербезпеки, для інформування про рівні своїх можливостей у значущих термінах та для визначення пріоритетів своїх інвестицій у кібербезпеку. Організація виконує оцінку моделі, використовує цю оцінку для виявлення прогалин у можливостях, пріоритизує ці прогалини і розробляє плани їх усунення і, нарешті, реалізує плани усунення прогалин. У міру реалізації планів, зміни бізнес-цілей та розвитку ризикового середовища процес повторюється. Модель організована у 10 доменів. Кожен домен є логічною групою методів кібербезпеки. Практики в домені згруповані за цілями - цільовими досягненнями, що підтримують домен. У межах кожної мети практики впорядковано за рівнем індикатора зрілості (0-3) [36].

Інститут розробки програмного забезпечення (SEI) розробив початкову версію моделі зрілості та анкети зрілості на запит уряду та за сприяння корпорації MITRE. Протягом усього процесу розробки моделі зрілості та анкети SEI приділяла увагу порадам фахівців-практиків, які беруть участь у розробці та покращенні процесу розробки програмного забезпечення. Модель зрілості можливостей CMMI [37] складається з п’яти рівнів зрілості.

Сертифікація моделі зрілості кібербезпеки (CMMC) – це велика програма Міністерства оборони США, створена для захисту оборонно-промислової бази від дедалі більш частих та складних кібератак. Зокрема, вона спрямований на посилення захисту контрольованої несекретної інформації та інформації про федеральні контракти, що передається до оборонно-промислової бази.

CMMC ґрунтується на існуючих правилах, що ґрунтуються на довірі, додаючи компонент перевірки для вимог кібербезпеки.

Наступною важливою бібліотекою впровадження та оцінки на відповідність вимогам інформаційної безпеки є контролі безпеки центру інтернет-безпеки (CIS). CIS – це некомерційна організація, яка займається підвищенням рівня кібербезпеки у всьому світі [38]. CIS було засновано в 2000 році з метою визначення, розробки, перевірки, просування та підтримки найкращих практик у сфері кібербезпеки. Він співпрацює з експертами з уряду, промисловості та наукових кіл, щоб створити практичні та дієві рішення для боротьби з новими загрозами у кіберпросторі.

CIS пропонує низку послуг і ресурсів для підтримки організацій у покращенні стану кібербезпеки. Деякі ключові аспекти Центру безпеки в Інтернеті включають:

- CIS контролі – це набір пріоритетних, всесвітньо визнаних передових практик кібербезпеки. Ці засоби контролю, розроблені спільнотою експертів, надають організаціям основу для створення та підтримки ефективної програми кібербезпеки. Вони охоплюють широкий спектр областей безпеки, включаючи інвентаризацію та контроль апаратних і програмних активів, безперервне керування вразливістю, безпечні конфігурації та реагування на інциденти.
- Еталонні показники CIS - надають рекомендації щодо безпечного налаштування різних технологій, включаючи операційні системи, бази даних, мережеві пристрої та програми. Ці тести, розроблені на основі консенсусного процесу, пропонують організаціям конкретні рекомендації щодо конфігурації для зменшення вразливості та узгодження з прийнятими галузевими стандартами.
- CIS SecureSuite - це комплексне рішення з кібербезпеки, яке поєднує численні ресурси, щоб допомогти організаціям покращити рівень безпеки. Він включає доступ до CIS Controls і CIS Benchmarks, а також до інструментів для оцінки безпеки, розвідки про загрози та вказівок щодо виправлення.
- Центри обміну та аналізу інформації (ISAC): CIS керує кількома центрами ISAC, які є довіреними спільнотами, де організації з певних галузей можуть обмінюватися інформацією про кібербезпеку, найкращими практиками та розвідкою

про загрози. Ці спільні платформи сприяють обміну інформацією про загрози в реальному часі та дозволяють членам бути в курсі нових загроз і вразливостей [39].

Підхід CIS контролів щодо встановлення відповідності між вимогами різних стандартів, та зведення їх до спільної основи робить подальші дослідження на їх основі актуальними та новими. Контролі CIS містять шкалу зрілості 1-3. Так звані, імплементаційні групи контролів CIS поділяються на базовий мінімум виконаних практик безпеки, рекомендований рівень, а також керований/продвинутий рівень, що передбачає впровадження всіх практик безпеки для найбільш можливого захисту. Саме ця шкала і буде використана в подальшому дослідженні.

1.7 Постановка завдання дослідження

Завданнями дипломної роботи є:

- Аналіз наукових досліджень в проблемній області;
- Розробка моделі гетерогенного класифікатора контролів ІБ;
    - підготовка наборів даних;
    - реалізація моделі;
    - перевірка моделі на тестовому наборі даних;
    - валідація моделі на контрольному наборі даних;
- На основі розробленої моделі класифікації - розробка моделі оцінки рівня забезпечення кібербезпеки організації;
- Оцінка адекватності отриманих рішень;
- Автоматизація процесу оцінки рівня забезпечення кібербезпеки;
- Результати та висновки: аналіз результатів дослідження, рекомендації щодо майбутніх досліджень і практичного застосування моделі.

Висновки за розділом 1

У першому розділі було надано комплексний аналіз літератури щодо встановлення вимог захищенності, стандартів управління інформаційною безпекою,

моделей оцінки рівня зрілості кібербезпеки. Проведено аналіз джерел, що включають промислові стандарти, описують моделі оцінки рівня кібербезпеки організації та підходи до їх оцінювання. Академічні дослідження моделей оцінки кібербезпеки були переглянуті та проаналізовані, щоб визначити обмеження та переваги цих моделей. Метою цього аналізу було отримати глибше розуміння поточного стану проблемної області, а також зрозуміти, як можна вдосконалити моделі оцінки рівня кібербезпеки організації.

Аналіз досліджень виявив кілька обмежень моделей оцінки кібербезпеки, включно з їхньою тенденцією зосереджуватися надто вузько на технічних аспектах безпеки, а не розглядати ширші організаційні, соціальні та економічні фактори, які можуть вплинути на результати безпеки. Деякі моделі можуть бути недостатньо актуальними відносно ландшафту загроз, що швидко розвивається, і зростаючою складністю кібератак.

Незважаючи на ці обмеження, аналіз також виявив переваги моделей оцінки кібербезпеки. Наприклад, ці моделі можуть допомогти організаціям ідентифікувати та визначати пріоритети ризиків безпеки, ефективно розподіляти ресурси, а також контролювати та оцінювати ефективність своїх заходів безпеки з часом. Крім того, ці моделі можуть сприяти спілкуванню та співпраці між різними зацікавленими сторонами, залученими до кібербезпеки, такими як IT-фахівці, менеджери та керівники.

Загалом, аналіз академічних досліджень моделей оцінки рівня кібербезпеки підкреслює необхідність постійного розвитку та вдосконалення цих моделей, а також визнає їх потенційну цінність і переваги в допомозі організаціям керувати складним ландшафтом загроз і ризиків кібербезпеці, який швидко розвивається.

Таким чином, за результатами першого розділу було розглянуто класифікацію контролей інформаційної безпеки, визначено завдання дипломної роботи і виконано аналіз літератури.

РОЗДІЛ 2
СТВОРЕННЯ МОДЕЛІ КЛАСИФІКАЦІЇ КОНТРОЛІВ ІНФОРМАЦІЙНОЇ БЕЗПЕКИ ТА ЇЇ ВИКОРИСТАННЯ В ОЦІНЦІ РІВНЯ КІБЕРБЕЗПЕКИ СИСТЕМИ

2.1 Концепція моделі класифікації

Для можливості створення індивідуальної програми оцінки кібербезпеки оргінізації, необхідно визначити контролі, які повинні бути досягнутими та описати конкретні кроки їх виконання, а також рівень оцінки (як саме можна зрозуміти, що дана вимога безпеки виконана?). Для того, аби побудувати таку модель і спростити процес оцінки необхідно побудувати класифікаційну модель, яка могла б розрізняти контролі за доменами (Організаційний, Люди, Технічний, Фізичний), пропонувати відповідні кроки виконання налаштовуваних контролей, а також допомагала визначити, якому рівню зрілості належить виконання чи невиконання цієї контролі. Для досягнення цієї мети було обрано стандарт ISO 27001 як основу для моделі класифікації, оскільки він має повний набір засобів управління інформаційною безпекою, широко прийнятих у галузі.

Пропонована модель класифікації має низку переваг. По-перше, це скорочує час та зусилля, необхідні для впровадження засобів управління ISO 27001. По-друге, це дозволяє організаціям налаштовувати свої елементи управління без шкоди цілісності системи безпеки. Нарешті, модель можна використовувати для виявлення прогалин у поточному стані безпеки та пропозиції відповідних засобів контролю для зниження ризиків.

Таким чином, це дослідження пропонує новий підхід до настроювання реалізації елементів управління ISO 27001, який може значно покращити стан безпеки організацій. Побудована модель забезпечує систематичну основу, яка може

полегшити впровадження стандартних засобів контролю, зберігаючи їх ефективність і дієвість.

Для побудови гнучкої та налаштовуваної моделі оцінки рівня кібербезпеки на основі бази знань необхідно використати алгоритм машинного навчання, що вирішує проблему класифікації.

Різниця між класифікацією з кількома класами та класифікацією з кількома мітками полягає в тому, що в задачах із кількома класами класи є взаємовиключними, тоді як для задач із кількома мітками кожна мітка представляє інше завдання класифікації, але завдання якимось чином пов'язані. Оскільки згідно карти співставлення CIS контролів, деякі рекомендації-контролі можуть бути віднесені одночасно до декількох доменів, було обрано підхід класифікаціх з кількома мітками [40].

Підходи оцінки для однокомпонентних моделей зазвичай відрізняються від багатокомпонентних. В однокомпонентній класифікації, використовуються прості показники, такі як точність, повнота і т. д. [41]. В однокомпонентній класифікації точність дорівнює (1):

$$\frac{1}{N}\sum_{i=1}^{N} I[\check{y}^{(i)} = y^{(i)} \tag{1}$$

Тут N представляє загальну кількість екземплярів у наборі даних, y˅^((i)) представляє передбачену мітку для і-го екземпляра, а y^((i)) представляє справжню мітку для і-го екземпляра.

У класифікації з кількома мітками неправильна класифікація більше не є однозначною помилкою чи правдою. Передбачення, що містить підмножину реальних класів, буде вважатися кращим, ніж передбачення, яке не містить жодного класу, тобто правильне передбачення двох з трьох міток краще, ніж передбачення відсутності міток взагалі.

Мікро- та макро-усереднення зазвичай використовуються на основі міток для оцінки ефективності моделей класифікації з кількома класами [42].

Мікроусереднення – це метод агрегування показників продуктивності для всіх міток або класів у наборі даних, який часто використовується, коли в даних є дисбаланс класів. При мікроусередненні кожному окремому прогнозу надається однакова вага, незалежно від класу, до якого він належить. Це означає, що показники ефективності розраховуються з урахуванням загальної кількості справжніх позитивних, хибних і хибних негативних результатів у всіх класах.

З іншого боку, макроусереднення — це метод обчислення показників ефективності для кожного класу з подальшим усередненням результатів. При макроусередненні кожному класу надається однакова вага, незалежно від кількості екземплярів у кожному класі. Це означає, що показники ефективності обчислюються для кожного класу окремо, а потім усереднюються по всіх класах [43].

Мікроусереднення, як правило, більш доцільне, коли набір даних незбалансований, оскільки воно надає більшої ваги меншим класам. Однак це може бути не ідеальним у випадках, коли в даних домінує клас більшості, оскільки це може призвести до необ'єктивних результатів. З іншого боку, макроусереднення, як правило, більш прийнятне, коли кожен клас однаково важливий або коли набір даних збалансований.

Загалом, як мікро-, так і макро-усереднення є корисними мірками на основі міток для оцінки ефективності моделей класифікації з кількома класами. Вибір між двома методами залежить від конкретних характеристик набору даних і цілей оцінювання моделі.

Щоб виміряти мультикласовий класифікатор, необхідно якимось чином усереднити класи. І тому існує два різних методу, званих мікро-усереднення і макро-усереднення. Під час мікроусереднення всі TP, TN, FP та FN для кожного класу підсумовуються, а потім береться середнє значення (2).

$$Microaveraging\ Precision\ Prc^{micro}(D) = \frac{\sum_{C_i \in C} TPs(C_i)}{\sum_{C_i \in C} TPs(C_i) + FNs(C_i)} \quad (2)$$

$$Microaveraging\ Precision\ Rcl^{micro}(D) = \frac{\sum_{C_i \in C} TPs(C_i)}{\sum_{C_i \in C} TPs(C_i) + FNs(C_i)}$$

У методі мікроусереднення потрібно підсумувати окремі справжні позитивні, хибні позитивні та хибні негативні результати системи для різних наборів і застосовати їх. А мікросередній F1-Score буде просто гармонійним середнім для двох вищевказаних рівнянь [44].

Макроусереднення є прямим. Необхідно використати середнє значення точності та запам'ятовування системи на різних наборах (3).

$$Macroaveraging\ Precision\ Prc^{macro}(D) = \frac{\sum_{C_i \in C} Prc(D, C_i)}{|C|} \quad (3)$$

$$Macroaveraging\ Precision\ Rec^{macro}(D) = \frac{\sum_{C_i \in C} Rcl(D, C_i)}{|C|}$$

Метод макроусереднення можна використовувати, коли необхідно знати, як працює система загалом у наборах даних. Тут немає необхідності приймати якесь конкретне рішення з цим середнім показником. З іншого боку, мікроусереднення може бути корисним заходом, коли набір даних відрізняється за розміром.

Втрати Хеммінга — це міра на основі прикладів, яка зазвичай використовується для оцінки продуктивності моделей класифікації з кількома мітками. Він вимірює частку неправильних міток, передбачених моделлю, відносно загальної кількості міток у наборі даних [45].

У класифікації з кількома мітками кожен екземпляр у наборі даних може бути пов'язаний із кількома мітками або класами, а не лише з однією міткою, як у традиційних задачах класифікації. Втрата Хеммінга враховує той факт, що модель може передбачити більше ніж одну мітку для кожного випадку, і штрафує модель за кожне неправильне передбачення.

Втрати Хеммінга обчислюються шляхом підсумовування кількості неправильно передбачених міток у всіх екземплярах у наборі даних, а потім ділення на загальну кількість міток у наборі даних. Наприклад, якщо набір даних містить 100 екземплярів, кожен з яких пов'язаний з 5 мітками, і модель передбачає 3 неправильні мітки для кожного екземпляра, тоді втрати Хеммінга складатимуть (100 * 3) / (100 * 5) = 0,6.

Втрати Хеммінга є корисним показником для оцінки ефективності моделей класифікації з кількома мітками, оскільки він враховує як кількість міток, так і кількість екземплярів у наборі даних. Однак він може бути чутливим до дисбалансу класів, особливо коли набір даних містить велику кількість екземплярів із дуже малою кількістю міток.

Загалом, втрати Хеммінга є корисним показником на основі прикладів для оцінки продуктивності моделей класифікації з кількома мітками. Він може надати уявлення про точність прогнозів моделі та може використовуватися для порівняння продуктивності різних моделей або алгоритмів.

В цілому, втрата Хеммінга — це частка неправильно передбачених міток, тобто частка неправильних міток до загальної кількості міток (4).

$\frac{1}{|N|*|L|}\sum_{i=1}^{|N|}\sum_{j=1}^{|L|} xor(y_{ij}, z_{ij})$, where y_{ij} is the target and $z_{ij}$is the prediction. (4)

Коефіцієнт точної відповідності, також відомий як точність підмножини, є мірою на основі прикладів, яка зазвичай використовується для оцінки ефективності моделей класифікації з кількома мітками. Він вимірює відсоток екземплярів, які правильно класифіковано з усіма пов’язаними мітками.

У класифікації з кількома мітками кожен екземпляр у наборі даних може бути пов’язаний з кількома мітками або класами, а мета моделі класифікації полягає в тому, щоб передбачити всі правильні мітки для кожного екземпляра. Коефіцієнт точної відповідності вимірює точність моделі в передбаченні всіх правильних міток для кожного екземпляра.

Коефіцієнт точної відповідності обчислюється шляхом підрахунку кількості екземплярів у наборі даних, де передбачені мітки точно збігаються з справжніми мітками, а потім діленням на загальну кількість екземплярів у наборі даних. Наприклад, якщо набір даних містить 100 екземплярів, кожен з яких пов’язаний з 5 мітками, і модель правильно передбачає всі 5 міток для 80 екземплярів, тоді коефіцієнт точної відповідності становитиме 80/100 = 0,8.

Коефіцієнт точної відповідності є корисним показником для оцінки ефективності моделей класифікації з кількома мітками, оскільки він враховує точність прогнозів для всіх міток, пов’язаних з кожним екземпляром. Однак він може бути чутливим до дисбалансу класів, особливо коли набір даних містить велику кількість екземплярів із дуже малою кількістю міток.

Загалом, коефіцієнт точної відповідності є корисним показником на основі прикладів для оцінки продуктивності моделей класифікації з кількома мітками. Він може надати уявлення про точність прогнозів моделі та може використовуватися для порівняння продуктивності різних моделей або алгоритмів.

Коефіцієнт точної відповідності (точність підмножини) - це «найсуворіший» показник, що вказує на відсоток зразків, у яких усі етикетки/класи класифіковані правильно (5) [46].

$$ExactMatchRatio, M\ R = \frac{1}{n}\sum_{i=1}^{n} I(Y_i = Z_i) \quad (5)$$

Недоліком цього виміру є те, що багатокласові проблеми класифікації мають шанс бути частково правильними, але тут ми ігноруємо ці частково правильні збіги. У модулі scikit-learn є функція, яка реалізує точність підмножини і називається accuracy_score. Ця функція буде використана для оцінки моделей у проекті.

2.2 Дослідницький аналіз даних

Дослідницький аналіз даних (EDA) – це критичний процес у галузі науки про дані, який включає аналіз та узагальнення наборів даних для отримання інформації та виявлення закономірностей, які можуть бути приховані у даних. EDA – це науковий підхід, який допомагає фахівцям за даними зрозуміти основну структуру даних та розробити гіпотези для подальшого дослідження [47].

Процес EDA зазвичай включає ряд методів, таких як візуалізація даних, статистичний аналіз та інтелектуальний аналіз даних, які використовуються для

вивчення і розуміння даних. Це включає виявлення відсутніх даних, викидів та інших аномалій, а також дослідження взаємозв'язків між різними змінними в даних.

Виконуючи EDA, фахівці за даними можуть отримати більш глибоке розуміння даних, з якими вони працюють, і можуть розробити обґрунтовані гіпотези та моделі для подальшого аналізу. Це може призвести до більш точного і значного розуміння, а також до більш ефективного прийняття рішень на основі даних. В цілому, EDA – це фундаментальний процес у науці про дані, який допомагає гарантувати, що розуміння та висновки, зроблені на основі даних, ґрунтуються на міцній основі ретельного аналізу та інтерпретації.

Дослідницький аналіз даних є одним із важливих кроків у процесі аналізу даних. Тут основна увага зосереджена на тому, щоб зрозуміти наявні дані — такі речі, як формулювання правильних запитань до набору даних, як маніпулювати джерелами даних для отримання необхідних відповідей тощо.

Для програмної реалізації було використано мову програмування Python. Після імпортування необхідних модулей, було завантажено дані з файлів csv у фрейм даних pandas і і перевірено його атрибути (рис. 2.1).

	ID	Control	Description	Organizational	People	Physical	Technical
0	A5.1	Policies for information security	Information security policy and topic-specific...	1	0	0	0
1	A5.2	Information security roles and responsibilities	Information security roles and responsibilitie...	1	0	0	0
2	A5.3	Segregation of duties	Conflicting duties and conflicting areas of re...	1	0	0	0
3	A5.4	Management responsibilities	Management shall require all personnel to appl...	1	0	0	0
4	A5.5	Contact with authorities	Access rights to information and other associa...	1	0	0	0
...	...	...	...	...	...	...	...
15	C5.5	Establish and Maintain an Inventory of Service...	Establish and maintain an inventory of service...	1	0	0	1
16	C6.2	Establish an Access Revoking Process	Establish and follow a process, preferably aut...	1	1	0	0
17	C6.8	Define and Maintain Role-Based Access Control	Define and maintain role-based access control,...	1	0	0	1
18	C8.1	Establish and Maintain an Audit Log Management...	Establish and maintain an audit log management...	1	0	0	1
19	C8.5	Collect Detailed Audit Logs	Configure detailed audit logging for enterpris...	1	0	0	1

Рисунок 2.1 Фрагмент завантажених даних

Далі було підраховано кількість контролів під кожною міткою (рис. 2.2).

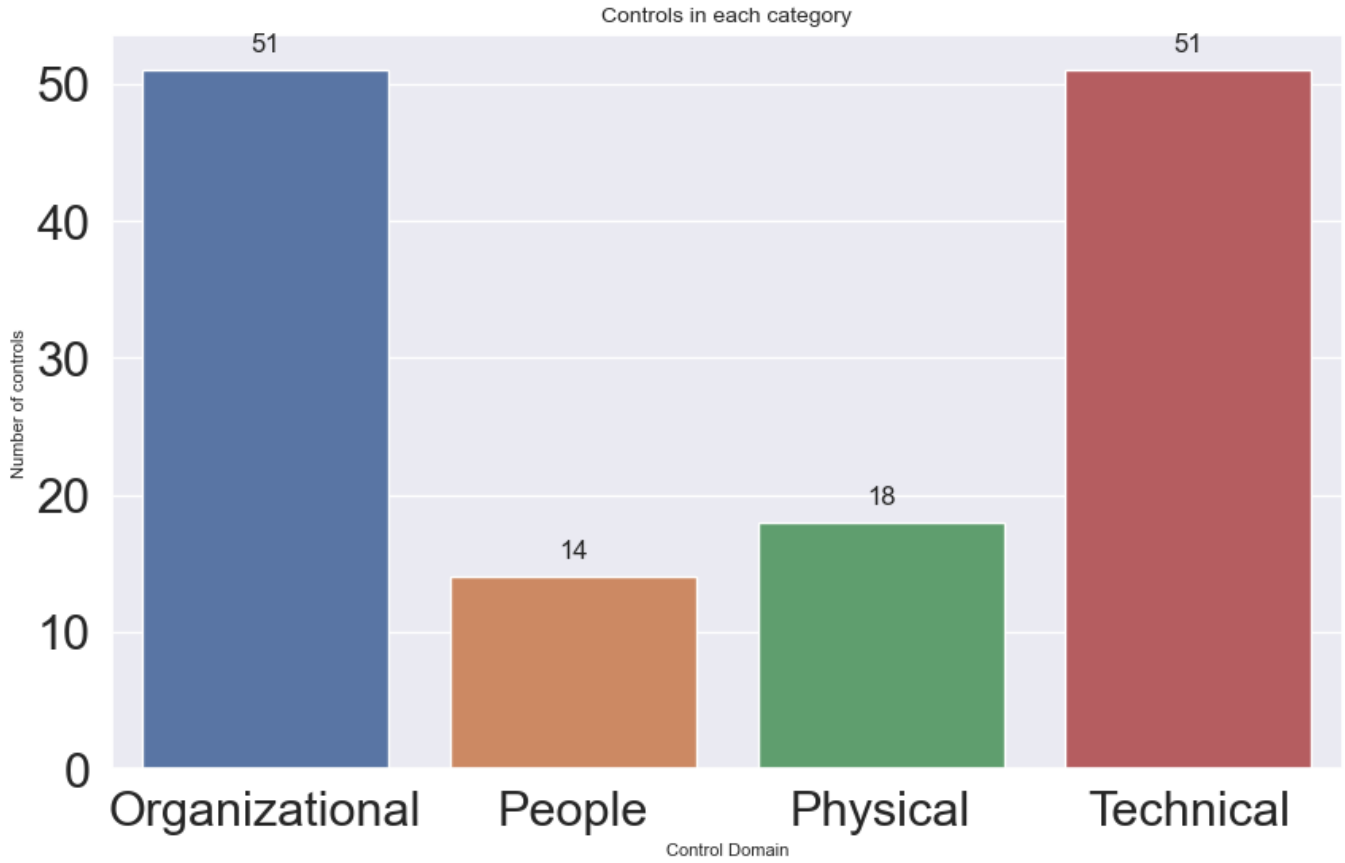

Рисунок 2.2 Кількість контролів під кожною міткою

Генерація хмари слів - це метод візуалізації даних, який використовується для представлення текстових даних у візуально-привабливому вигляді. Цей метод включає створення графічного уявлення слів, які у текстовому боксі, де розмір кожного слова пропорційний частоті його появи у тексті [48].

Найбільш часто зустрічаються слова в корпусі присвоюється найбільший розмір шрифту і вони розташовуються в центрі хмари, а менш часто зустрічаються словами присвоюється менший розмір шрифту і вони розташовуються навколо центру.

Генерація хмари слів може використовуватися для різних цілей, включаючи узагальнення вмісту корпусу текстів, визначення ключових тем або тем у корпусі та візуалізацію тенденцій використання мови з часом. Цей метод широко використовується в науці про дані, маркетингові дослідження та інші галузі, де необхідно аналізувати великі обсяги текстових даних та узагальнювати їх у короткій та візуально привабливій формі. В цілому, генерація хмари слів є потужним

інструментом для аналізу та візуалізації текстових даних та має безліч додатків у дослідному та бізнес-контексті.

WordCloud представлення найбільш вживаних слів у кожній категорії контролів (рис. 2.3-6).

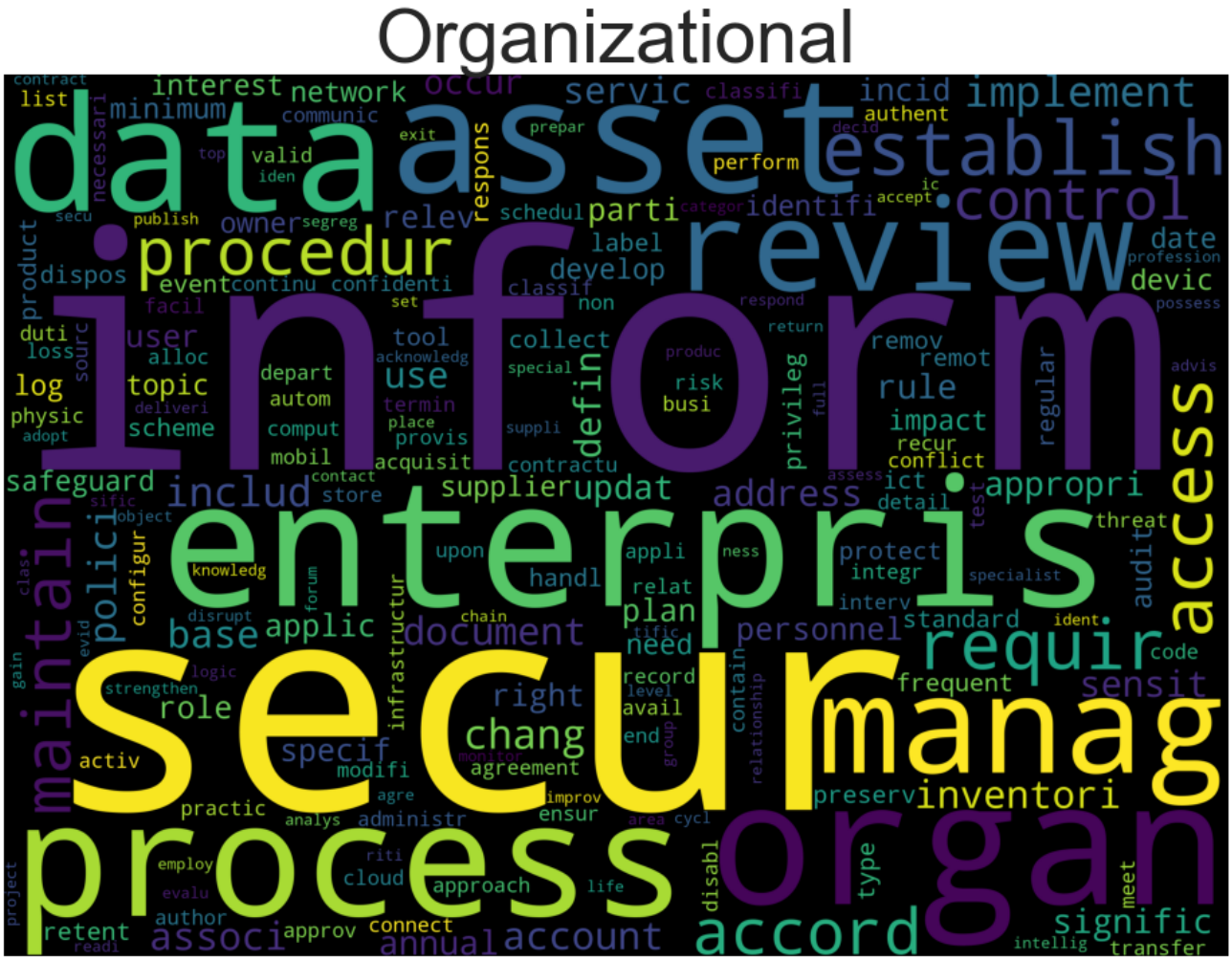

Рисунок 2.3 Слова, які найчастіше зустрічаються в Організаційних контролях

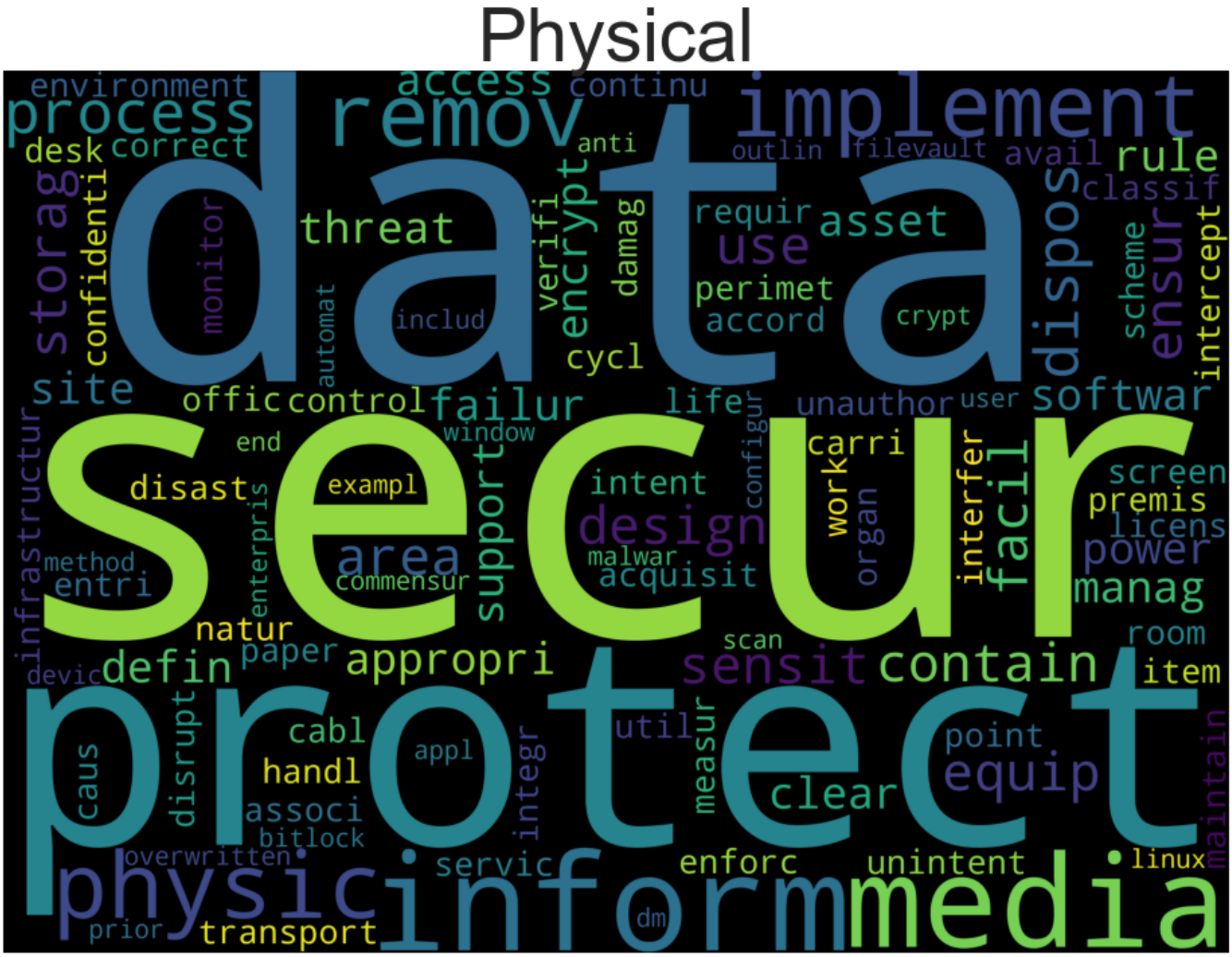

Рисунок 2.4 Слова, які найчастіше зустрічаються у Фізичних контролях

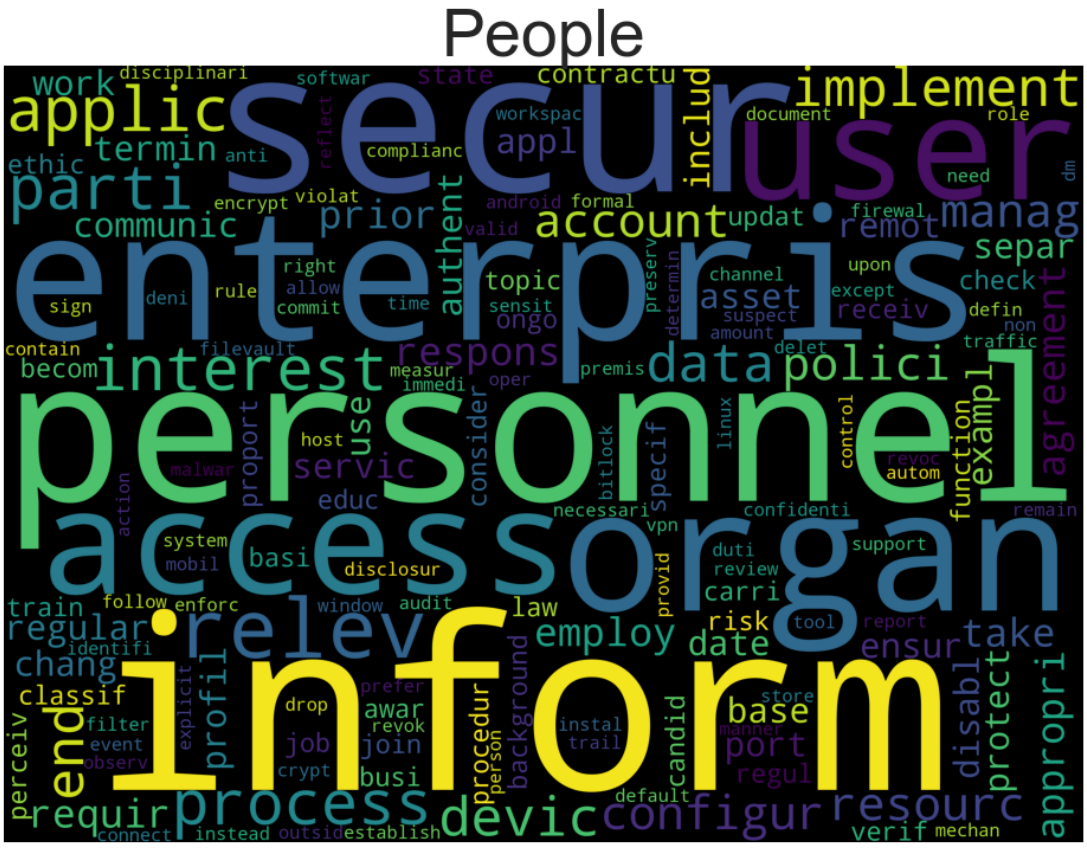

Рисунок 2.5 Слова, які найчастіше зустрічаються у контролях, пов'язаних з людьми

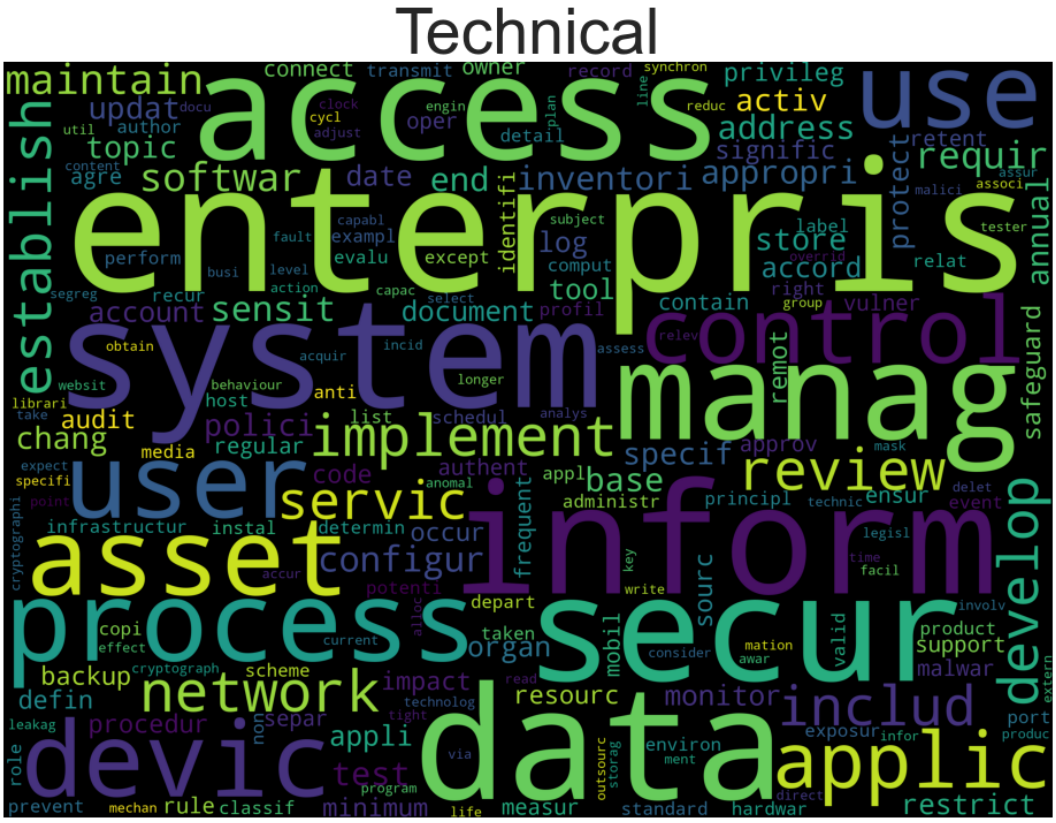

Рисунок 2.6 Слова, які найчастіше зустрічаються у Технічних контролях

2.2 Попередня обробка даних

Попередня обробка даних є важливим етапом у задачах класифікації машинного навчання, що включає перетворення необроблених даних у формат, який може бути легко використаний алгоритмом класифікації. Процес попередньої обробки даних зазвичай включає кілька етапів, включаючи очищення даних, вибір ознак, вилучення ознак та нормалізацію даних [49].

Очищення даних включає виявлення та виправлення помилок, видалення викидів і роботу з відсутніми даними. Це важливо, оскільки алгоритми класифікації зазвичай припускають, що вхідні дані є чистими та безпомилковими, тому будь-які неточності даних можуть призвести до невірних результатів класифікації.

Вибір ознак включає вибір найбільш релевантних ознак або змінних з набору даних, які, ймовірно, вплинуть на ефективність класифікації. Це важливо, тому що включення нерелевантних або надлишкових функцій може призвести до переоснащення, коли модель добре працює на навчальних даних, але погано на нових, невидимих даних.

Витяг ознак включає перетворення необроблених даних на набір ознак або змінних, які можуть використовуватися алгоритмом класифікації. Це може включати такі методи, як зменшення розмірності, при якому кількість функцій зменшується при збереженні якомога більшої кількості релевантної інформації.

Нормалізація даних включає масштабування даних, щоб гарантувати, що функції знаходяться в однаковому масштабі і мають однакові діапазони. Це важливо, тому що алгоритми класифікації можуть бути чутливими до відмінностей у масштабі, що може призвести до неточних результатів.

Загалом, попередня обробка даних є важливим кроком у завданнях класифікації, оскільки вона може значно вплинути на точність та ефективність алгоритму класифікації. Належно попередньо обробляючи дані, фахівці за даними можуть гарантувати, що алгоритм класифікації використовує найбільш актуальну і

точну інформацію для прогнозування, що призводить до більш точним і ефективним результатам.

Спочатку було перетворено контролі в нижній регістр, а потім використано спеціальні функції, щоб видалити з коментарів html-теги, знаки пунктуації та неалфавітні символи, програмний код попередньої обробки даних додається (дод.Б).

Далі було видалено всі стоп-слова, присутні в коментарях, використовуючи стандартний набір стоп-слів, який можна завантажити з бібліотеки NLTK [50]. До стандартного списку також додано кілька стоп-слів.

Стоп-слова – це набір слів, які часто вживаються будь-якою мовою, а не лише англійською. Причина, чому стоп-слова є критичними для багатьох програм, полягає в тому, що якщо ми видалимо слова, які дуже часто вживаються в даній мові, ми зможемо зосередитися на важливих словах замість них (рис. 2.7).

	ID	Control	Description	Organizational	People	Physical	Technical
0	A5.1	Policies for information security	information security policy topic specific po...	1	0	0	0
1	A5.2	Information security roles and responsibilities	information security roles responsibilities s...	1	0	0	0
2	A5.3	Segregation of duties	conflicting duties conflicting areas respons...	1	0	0	0
3	A5.4	Management responsibilities	management shall require personnel apply inf...	1	0	0	0
4	A5.5	Contact with authorities	access rights information associated assets...	1	0	0	0

Рисунок 2.7 Фрагмент даних

Далі було зроблено стемінг. Стеммінг - це метод обробки природної мови, який включає перетворення слів в їх основу або кореневу форму, яка називається основою. Мета визначення основи полягає в тому, щоб привести варіанти слова до загальноприйнятої форми, що може допомогти покращити аналіз тексту та пошук інформації за рахунок зниження складності даних.

Алгоритми стеммінгу використовують лінгвістичні правила для ідентифікації та видалення афіксів зі слів, таких як префікси та суфікси для отримання основи. Наприклад, слово "стрибав" може бути утворено від "стрибати", а слово "біг" може походити від "бігати". Основа може бути недійсним словом сама по собі, але вона все ж таки може надати корисну інформацію про значення вихідного слова.

Стемінг особливо корисний у таких додатках, як пошукові системи та класифікація текстів, де він може допомогти зменшити кількість унікальних слів у

документі чи корпусі та згрупувати пов'язані слова разом. Однак визначення основи може призвести до помилок або двозначності, особливо в мовах зі складною морфологією або неправильними формами.

Було розроблено кілька алгоритмів стеммінгу, у тому числі алгоритм стеммінгу Портера, який широко використовується при обробці англійської мови, та алгоритм стеммінгу Snowball, який є більш загальним алгоритмом, який можна застосовувати до кількох мов.

Загалом, стеммінг – корисний метод обробки природної мови, який може допомогти зменшити складність текстових даних та покращити аналіз тексту та пошук інформації. Тим не менш, важливо ретельно оцінити продуктивність алгоритмів стеммінгу для конкретних додатків та мов, оскільки вони можуть бути відповідними або ефективними не завжди.

Існують різні види стемінгу, які в основному перетворюють слова з приблизно однаковою семантикою в одну стандартну форму (рис. 2.8).

	ID	Control	Description	Organizational	People	Physical	Technical
0	A5.1	Policies for information security	inform secur polici topic specif polici shall ...	1	0	0	0
1	A5.2	Information security roles and responsibilities	inform secur role respons shall defin alloc ac...	1	0	0	0
2	A5.3	Segregation of duties	conflict duti conflict area respons shall segreg	1	0	0	0
3	A5.4	Management responsibilities	manag shall requir personnel appli inform secu...	1	0	0	0
4	A5.5	Contact with authorities	access right inform associ asset shall provis ...	1	0	0	0

Рисунок 2.8 Фрагмент даних з спільнокореневим зведенням

Після поділу набору даних на набори навчання та тестування необхідно узагальнити контролі та перетворити їх у числові вектори. У машинному навчанні поширеною практикою є розділення набору даних на два окремі набори: набір для навчання та набір для тестування. Навчальний набір використовується для навчання моделі або алгоритму, тоді як набір для тестування використовується для оцінки продуктивності моделі на нових, невідомих даних.

Перш ніж навчати модель, важливо попередньо обробити дані, що зазвичай передбачає перетворення даних у числовий формат, який можна легко обробити алгоритмом машинного навчання. Ось тут і з’являється процес підсумовування елементів керування та перетворення їх у числові вектори.

Для перетворення елементів керування в числові вектори зазвичай використовується процес, який називається кодуванням ознак. Це включає в себе перетворення категоріальних змінних у числові значення, наприклад кодування з одним чатом, де кожна категорія представлена двійковим вектором, або кодування міток, де кожній категорії присвоюється унікальне ціле число.

Для безперервних змінних можуть бути використані методи стандартизації або нормалізації, щоб гарантувати, що змінні знаходяться в подібному масштабі та мають подібні діапазони. Це може допомогти покращити продуктивність алгоритму машинного навчання, зменшивши вплив відмінностей у масштабах змінних.

Загалом, узагальнення елементів керування та їх перетворення в числові вектори є важливим кроком у попередній обробці даних для машинного навчання. Перетворюючи дані в числовий формат, стає легше обробляти й аналізувати за допомогою алгоритмів машинного навчання, що може призвести до більш точних і ефективних прогнозів.

Стандартизація використовується в багатьох галузях, зокрема в науці, техніці, бізнесі та освіті, щоб забезпечити узгодженість і порівняльність даних. Під час порівняння різних речей або вимірювання різних кількостей важливо використовувати стандартизований метод, щоб переконатися, що порівняння є точним і значущим. Стандартизація передбачає встановлення узгодженого та єдиного набору вимірювань або критеріїв, які можна застосовувати до всіх речей, що порівнюються [51]. Цей метод гарантує, що відмінності в похідних результатах розглядаються, і порівняння проводиться на рівних умовах.

Метод стандартизації використовує цю формулу (6)

$$z = \frac{x-u}{s} \tag{6}$$

Де z — нове значення, x — вихідне значення, u — середнє значення, а s — стандартне відхилення. Стандартне відхилення є показником того, наскільки рівномірно розподілені числа (7). Більшість даних, ймовірно, близькі до середнього

(середнього) значення, якщо стандартне відхилення низьке. Коли стандартне відхилення велике, значення розподіляються більш рівномірно.

$$\sigma = \sqrt{\frac{\sum(x_i - \mu)^2}{N}} \tag{7}$$

Де N - розмір популяції, хі кожне значення з популяції, а μ - середнє популяції.

Один з методів перетворення у числові вектори полягає у виборі термінів, які найчастіше зустрічаються (слова з високою частотою термінів або TF). Однак слово, яке найчастіше зустрічається, є менш корисним показником, оскільки деякі слова, такі як «це», «а», зустрічаються дуже часто в усіх документах.

Отже, ми також хочемо визначити, наскільки унікальним є слово, тобто наскільки рідко воно зустрічається в усіх документах (інверсна частота документа або IDF).

Таким чином, добуток TF-IDF слова дає добуток того, наскільки часто це слово зустрічається в документі, помножене на те, наскільки це слово є унікальним в контексті всього документу [52].

Слова в документі з високим показником tfidf часто зустрічаються в документі та надають найбільше інформації про цей документ. TF-IDF легко обчислити, але його недоліком є те, що він не фіксує положення в тексті, семантику, одночасне входження в різні документи тощо.

2.3 Методи класифікації за кількома мітками

Більшість традиційних алгоритмів навчання розроблено для задач класифікації з однією міткою. Тому багато підходів у літературі перетворюють проблему з кількома мітками в задачі з кількома одномітками, щоб можна було використовувати існуючі алгоритми з однією міткою.

- Один на противагу іншим (OneVsRest) [53]

Традиційні двокласові та багатокласові задачі можна перетворити на задачі з кількома мітками, обмеживши кожен екземпляр лише однією міткою. З іншого боку,

загальність задач із кількома мітками неминуче ускладнює вивчення. Інтуїтивно зрозумілий підхід до розв'язання задачі з кількома мітками полягає в її розкладанні на кілька незалежних задач бінарної класифікації (по одній на категорію) (рис. 2.9).

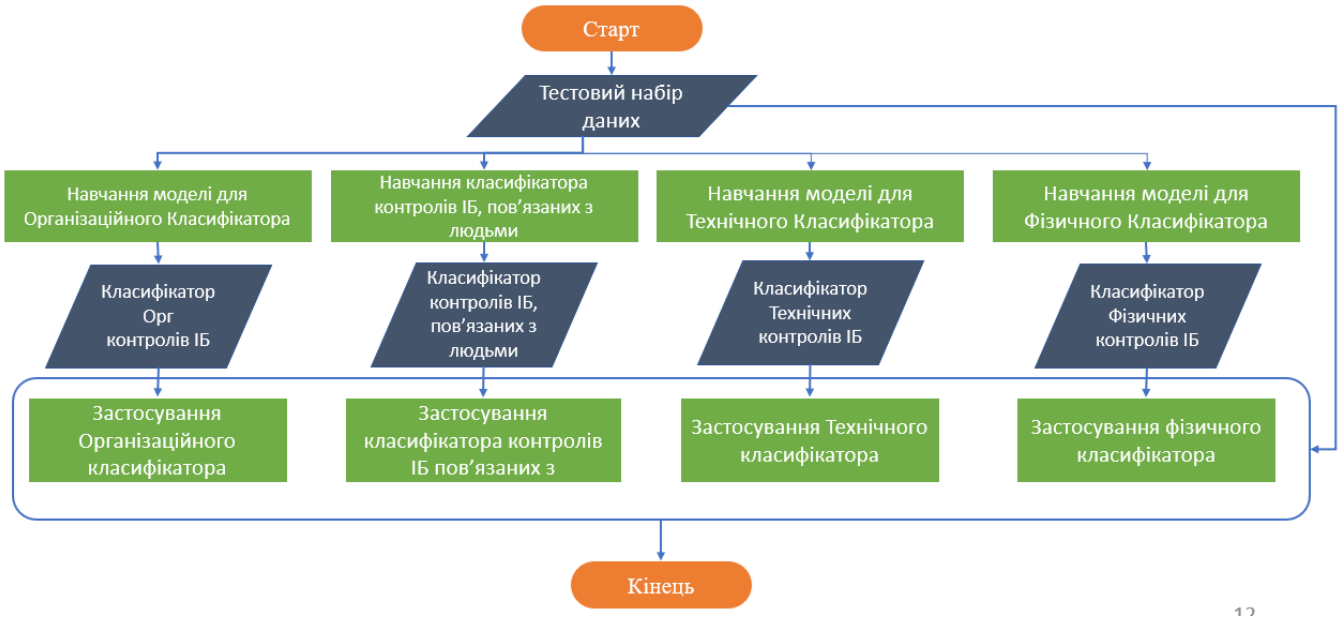

Рисунок 2.9. Один на противагу іншим

У стратегії один на противагу іншим можна створити кілька незалежних класифікаторів і, для невидимого прикладу, вибрати клас, для якого максимізується достовірність.

Головне припущення тут полягає в тому, що мітки є взаємовиключними. Ми не враховуємо жодної основної кореляції між класами в цьому методі. Наприклад, це більше схоже на прості запитання, скажімо, «контроль є організаційною чи ні», «контроль є технічною чи ні?» тощо.

Точність передбачення цієї моделі (рис. 2.10)

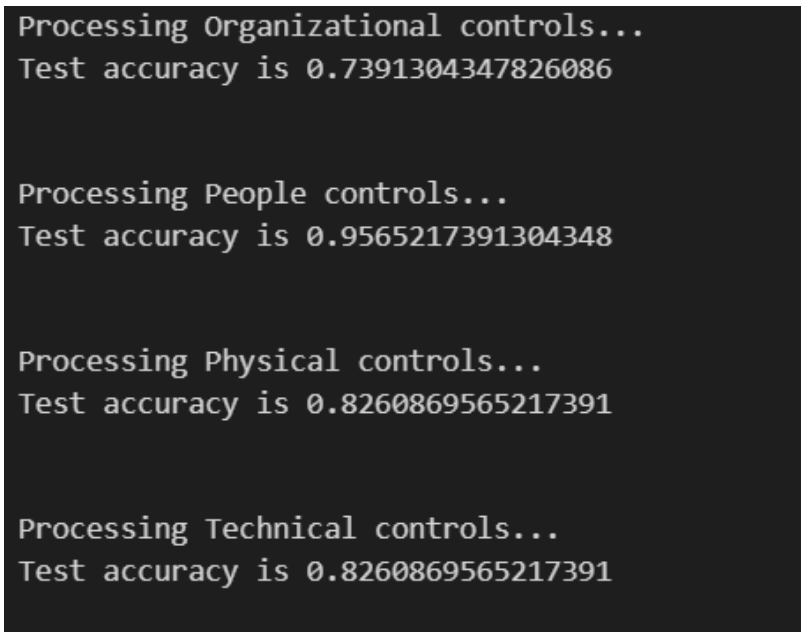

Рисунок 2.10 Точність моделі One vs Rest

- Ланцюги класифікаторів [54]

Ланцюжок бінарних класифікаторів C0, C1, . . . , Cn будується, де класифікатор Ci використовує передбачення всіх класифікаторів Cj , де j < i. Таким чином метод, який також називають ланцюжками класифікаторів (CC), може враховувати кореляції міток.

Загальна кількість класифікаторів, необхідних для цього підходу, дорівнює кількості класів, але навчання класифікаторів є більш залученим. Нижче наведено ілюстрований приклад із проблемою класифікації трьох категорій {K1, K2, K3}, з’єднаних у такому порядку (рис. 2.10).

Точність даної моделі для класифікації контролів процесу управління кібербезпеки оргназіції за доменами = 0.693333333333

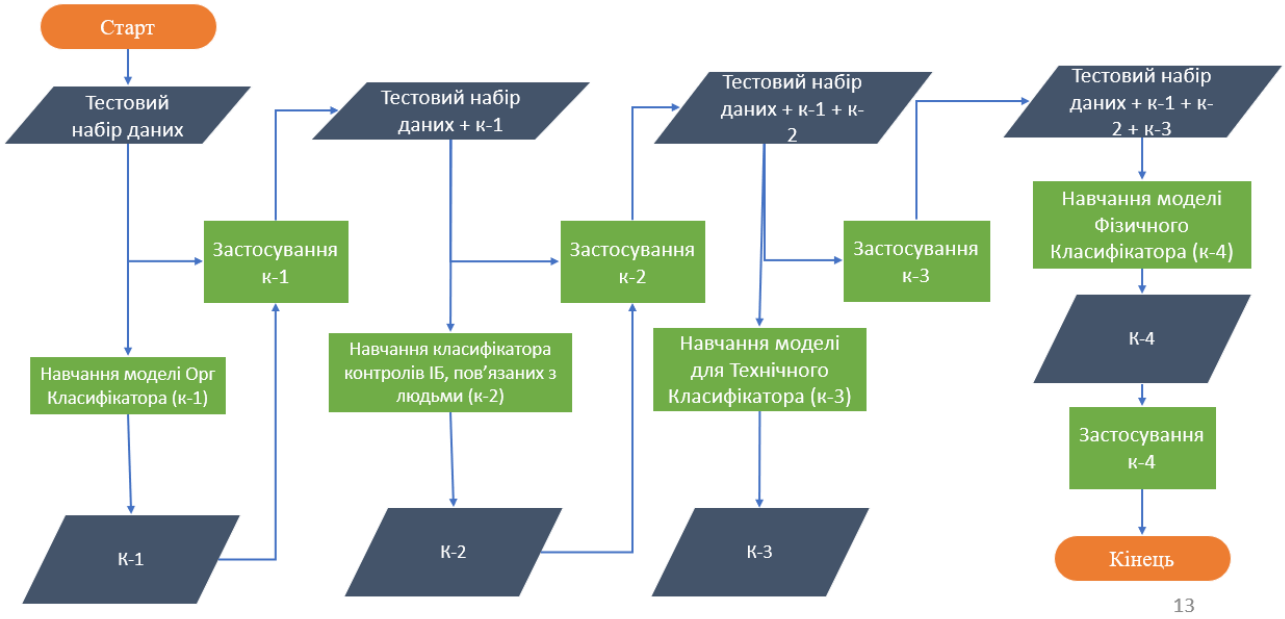

Рисунок 2.11 Ланцюги класифікаторів

- Powerset етикетка [55]

Цей підхід враховує можливі кореляції між мітками класів. Найчастіше цей підхід називають методом набору потужностей міток, оскільки він розглядає кожного члена потужного набору міток у навчальному наборі як одну мітку.

Цей метод потребує класифікаторів найгіршого випадку (2^|C|) і має високу обчислювальну складність (рис. 2.12).

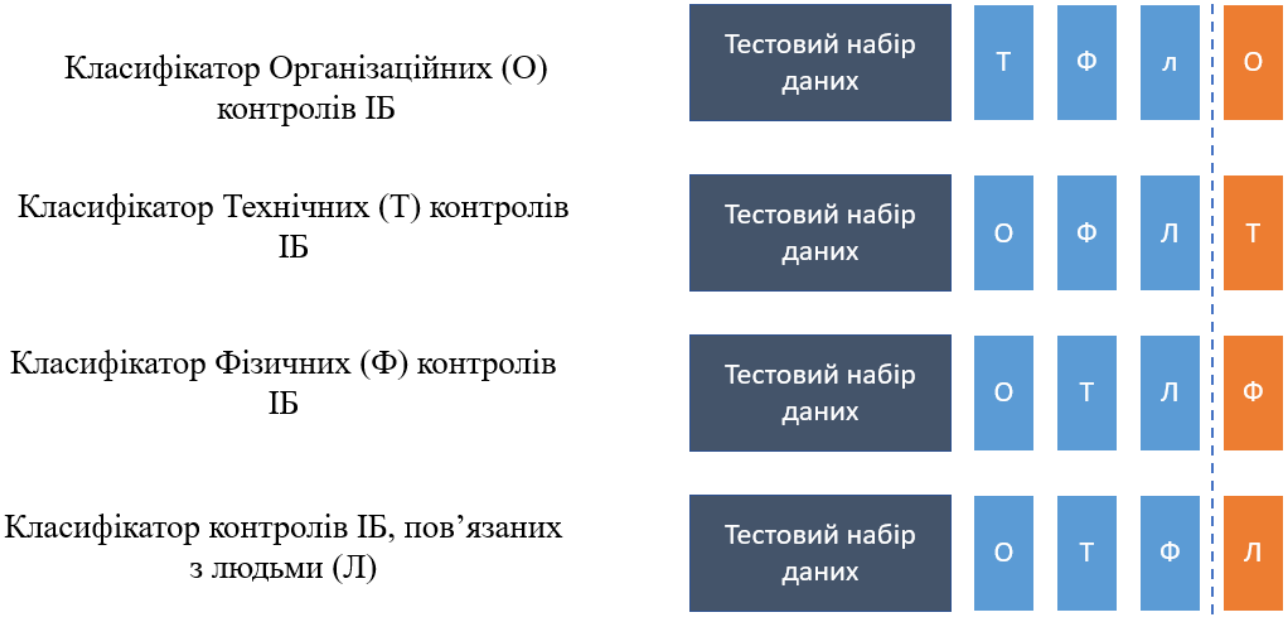

Рисунок 2.12 Метод набору потужностей міток

Однак, коли кількість класів збільшується, кількість різних комбінацій міток може зростати експоненціально. Це легко призводить до комбінаторного вибуху і, отже, до неможливості обчислень. Крім того, деякі комбінації міток матимуть дуже мало позитивних прикладів.

Точність даної моделі для класифікації контролів процесу управління кібербезпеки оргназіції за доменами = 0.7391304347826086

Існує два основні методи вирішення проблеми класифікації з кількома мітками: методи трансформації проблеми та методи адаптації алгоритму.

Методи трансформації проблеми перетворюють проблему з кількома мітками в набір задач бінарної класифікації, які потім можна обробляти за допомогою однокласових класифікаторів. Тоді як методи адаптації алгоритмів адаптують алгоритми для безпосереднього виконання класифікації за кількома мітками. Іншими словами, замість того, щоб намагатися перетворити проблему на більш просту проблему, вони намагаються вирішити проблему в її повній формі.

У масштабному порівнянні з іншими підходами метод «один проти всіх» дає найкращі результати, а за ним іде метод label-powerset.

Що відрізняє модель від традиційних оцінок безпеки, так це її здатність забезпечити адаптивну оцінку ризиків безпеки. Наша модель здатна постійно контролювати систему на наявність нових загроз і вразливостей і відповідно коригувати критерії оцінки. Це гарантує безпеку системи навіть перед лицем нових загроз.

Створена модель має потенціал для трансформації підходу до оцінки кібербезпеки для організацій. Завдяки настроюваним критеріям оцінювання та можливостям динамічного оцінювання наша модель пропонує новий рівень безпеки, який раніше був недоступний. Ми раді бачити, як наша модель вводиться в дію, і очікуємо значного позитивного впливу, який вона матиме на індустрію кібербезпеки.

Висновки за розділом 2

Реалізована модель демонструє великі перспективи для покращення сфери оцінювання кібербезпеки. У даній моделі використовуються розширені алгоритми машинного навчання, щоб забезпечити настроювану та ретельну оцінку безпеки певної системи. За допомогою моделі класифікації, яка проводить аналіз численних факторів, можна покращити комплексну оцінку загального стану кібербезпеки системи та організації.

Отже, в другому розділі було розглянуто методи класифікації тексту з можливістю присвоєння декількох міток, імплементовано підходи: один на противагу іншим, ланцюги класифікаторів, powerset етикетки.

Було перевірено адекватність і точність використаних підходів машинного навчання за допомогою вимірювань ефективності таких як: точність, втрати Хеміннга, мікро-, так і макро-усереднення, коефіцієнт точної відповідності

Таким чином, в другому розділі було розроблено модель машинного навчання, що зможе класифікувати нові контролі за доменами, що робить можливим визначення необхідних рекомендацій/кращих практик для ефективної оцінки рівня кібербезпеки організації та впровадження нових практик безпеки, задля досягнення необхідного рівня захисту.

РОЗДІЛ 3
РОЗРОБКА МОДЕЛІ ОЦІНКИ РІВНЯ ЗАБЕЗПЕЧЕННЯ КБ ОРГАНІЗАЦІЇ

3.1 Модель оцінки рівня забезпечення КБ організації

Процес оцінки кібербезпеки було значно покращено завдяки використанню наших новаторських досліджень. Використовуючи останні досягнення в машинному навчанні та аналітиці даних, ми розробили інноваційні методології та інструменти, які дозволяють організаціям комплексно оцінювати стан цифрової безпеки. Враховуючи результати наших досліджень у процесі оцінювання, компанії тепер можуть ефективніше виявляти та пом'якшувати вразливості, завчасно захищаючи свої системи від складних кіберзагроз. Наше дослідження революціонізувало спосіб проведення оцінювання кібербезпеки, надаючи організаціям знання та інструменти, необхідні для зміцнення їхнього захисту в цифровому ландшафті, що постійно розвивається.

З використанням моделі гетерогенного класифікатора ІБ було сформовано наступний процес оцінки рівня кібербезпеки організації (рис. 3.1)

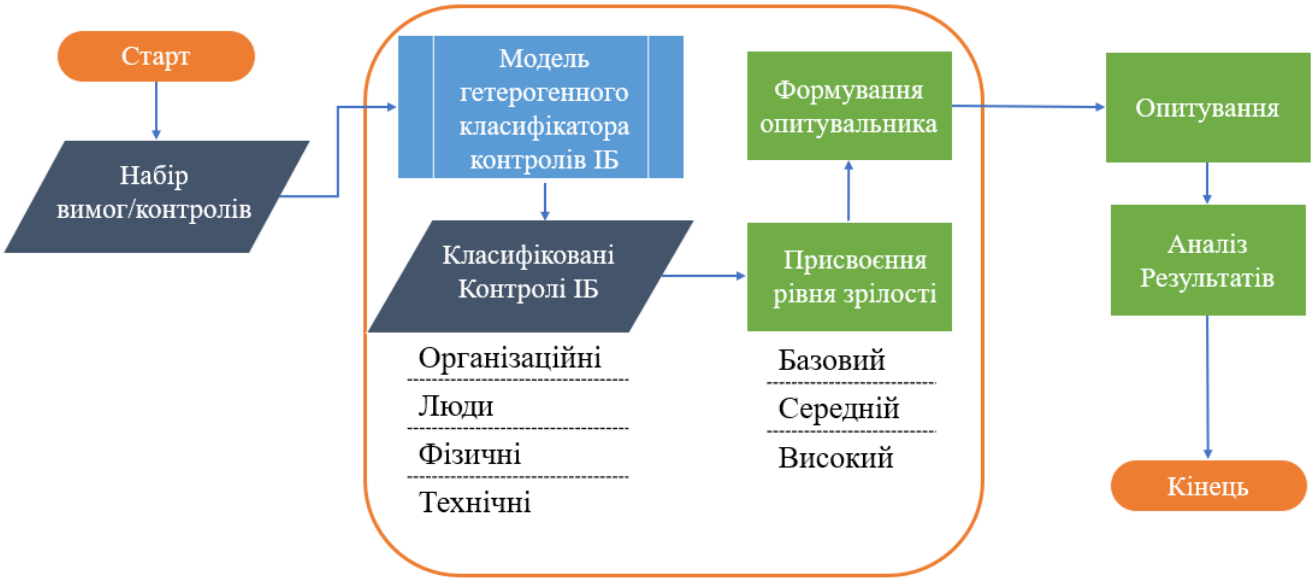

Рисунок 3.1 Модель оцінки рівня забезпечення КБ організації

Розробка матриці зрілості була зумовлена дедалі більшою увагою до правил кібербезпеки. Усвідомлюючи критичність кібербезпеки в сучасному взаємопов'язаному світі, регуляторні органи ввели суворі вказівки та стандарти для забезпечення захисту конфіденційних даних і цифрової інфраструктури. Використовуючи ці правила як основу, ми ретельно розробили матрицю зрілості, яка дозволяє організаціям оцінювати та порівнювати свої можливості кібербезпеки. Ця матриця охоплює різні сфери, такі як управління, управління ризиками, реагування на інциденти та відповідність, забезпечуючи комплексну структуру для оцінки рівня зрілості організації в кожній сфері. Завдяки узгодженню з правилами кібербезпеки матриця зрілості служить цінним інструментом для організацій для вимірювання їхнього прогресу, виявлення прогалин і визначення пріоритетів інвестицій у кібербезпеку, зрештою сприяючи надійній безпеці та дотриманню нормативних вимог.Рівні зрілості моделі сформовані наступним чином (рис. 3.2).

Рівень	Категорія котнролів			
	Організаційні	**Люди**	**Фізичні**	**Технічні**
Базовий	Дотримані основні політики і процедури щодо ІБ та обробки даних.	Обмежена обізнаність і навчання працівників щодо кібербезпеки та потенційних ризиків.	Основні заходи фізичної безпеки застосовуються для захисту фізичних активів.	Зосереджено на кібергігієні та впровадженні основних технічних засобів контролю, таких як антивірусне ПЗ, брандмауери та регулярне встановлення виправлень.
Середній	Ефективно впроваджує політику та процедури, включаючи контроль доступу, реагування на інциденти та навчання співробітників з питань безпеки.	Забезпечена належа обізнаність і навчальні програми для працівників.	Застосовуються передові заходи фізичної безпеки, включаючи біометричний контроль доступу, охорону та системи виявлення вторгнень.	Забезпечені заходи кібербезпеки в складнішому IT-середовищі, включаючи вдосконалений моніторинг мережі, системи виявлення вторгнень і безпечні конфігурації.
Високий	Проактивно керування ризиками та інцидентами безпеки за допомогою надійних політик безпеки, регулярних аудитів і постійного вдосконалення.	Ініціативи та навчання гарантують, що співробітники добре поінформовані про поточні загрози та активно беруть участь у захисті активів організації.	Застосовуються надійні заходи фізичної безпеки, такі як захищені центри обробки даних, резервні системи живлення та охолодження, а також суворий контроль доступу.	Демонструє готовність до розширених постійних загроз шляхом впровадження складних технічних засобів контролю, таких як розвідка загроз, розширена аналітика та шифрування.

16

Рисунок 3.2 Матриця зрілості

Таким чином, визначені рівні зрілості описують наступний стан захищенності організації:

Базовий: Визначає необхідну кібергігієну та представляє мінімальний стандарт інформаційної безпеки для всіх підприємств. Протидія загальним нецільовим атакам

Середній: Допомагає керувати IT-інфраструктурою кількох відділів із різними профілями ризику. Безпека у умовах зростаючої операційної складності

Високий: Допомагає підприємствам із експертами з КБ захищати конфіденційні дані. Спрямований на запобігання та/або зменшення впливу складних атак.

3.2 Автоматизація процесу оцінки рівня забезпечення кібербезпеки

Для практичної побудови моделі оцінки рівня кібербезпеки організації було використано мову програмування Python. Для створення графічного інтерфейсу обрано програмний модуль tkinter [56]. Задля спрощення побудови на даному етапі в ролі бази даних використовується csv файли. Програмний застосунок повинен містити 3 модулі: переглянути результати попередніх оцінок, можливість додати нову контроль безпеки, яка буде класифікована за доменом, а також рекомендації до неї будуть додані автоматично з використанням алгоритмів машинного навчання, а також модуль безпосередньої оцінки, який повинен містити можливість додавання контролів різних доменів, а також видавати результати в розрізі 3 рівнів (імплементаційних груп) відповідно класифікації CIS.

Критичні засоби контролю безпеки CIS були обрані для нашої моделі оцінки кібербезпеки завдяки їхнім унікальним функціям, які забезпечують ефективну навігацію та ефективне картографування. Елементи керування CIS організовані за пріоритетами, що дозволяє легко переміщатися між різними зонами безпеки, що полегшує оцінювачам визначення потенційних вразливостей і визначення пріоритетів заходів із пом’якшення. Крім того, CIS Controls розроблено для надання можливості відображення, що дозволяє експертам відстежувати свій прогрес за допомогою елементів керування, забезпечуючи комплексний і структурований підхід до оцінки безпеки. Ця можливість відображення гарантує, що процес оцінювання є

методичним і ретельним, зменшуючи шанси пропустити важливі питання безпеки. Загалом, CIS Critical Security Controls забезпечує ефективну основу для проведення оцінок кібербезпеки, що робить їх ідеальним вибором для нашої моделі (рис. 3.3).

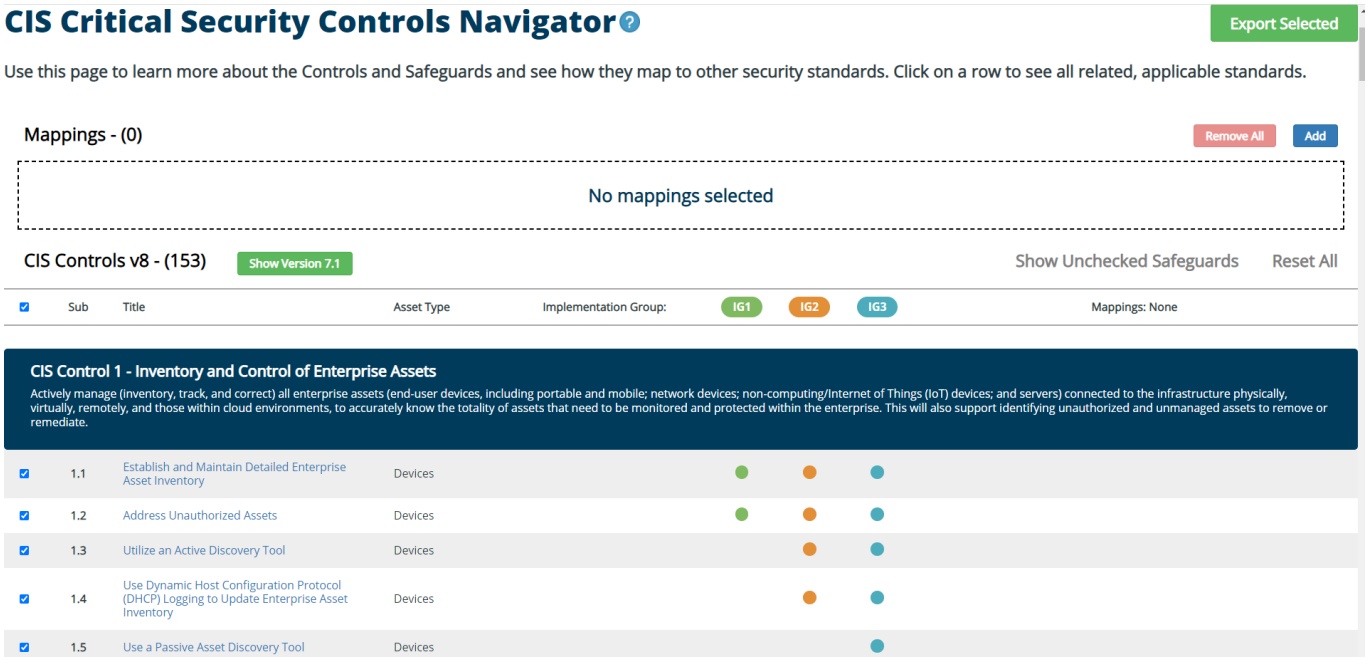

Рисунок 3.3 Веб-інтерфейс CIS

3.3 Розробка програмного застосунку

Pandas — популярна бібліотека обробки даних з відкритим кодом для Python. Він надає потужний набір інструментів для роботи зі структурованими даними, включаючи читання та запис даних із різних форматів файлів (наприклад, CSV, Excel, SQL), очищення та попередню обробку даних, дослідження й аналіз даних, а також візуалізацію даних [57].

Pandas надає дві основні структури даних: Series і DataFrame. Серія — це одновимірний об’єкт, схожий на масив, який може містити будь-який тип даних, тоді як DataFrame — це двовимірна структура даних у вигляді таблиці зі стовпцями потенційно різних типів даних.

Pandas оптимізовано для продуктивності та може ефективно обробляти великі набори даних. Він надає потужні можливості індексування та фільтрації, які дозволяють користувачам легко вибирати та керувати підмножинами даних. Крім

того, Pandas інтегрується з іншими популярними бібліотеками аналізу даних і візуалізації, такими як Matplotlib і Seaborn. Таким чином, Pandas — це універсальна та потужна бібліотека для маніпулювання та аналізу даних у Python, що робить її популярним вибором для широкого спектру додатків у галузі обробки даних, машинного навчання та фінансів, зокрема.

NumPy — це фундаментальна бібліотека для наукових обчислень на Python. Вона надає потужний набір інструментів для роботи з багатовимірними масивами та матрицями разом із великою бібліотекою математичних функцій, які працюють із цими масивами [58].

Основною структурою даних у NumPy є ndarray (n-вимірний масив), який є однорідним масивом елементів фіксованого розміру, який може містити будь-який тип даних. ndarray від NumPy забезпечує ефективне та гнучке зберігання та маніпулювання великими наборами даних, дозволяючи користувачам виконувати математичні операції над цілими масивами даних, а не з окремими елементами. Математичні функції NumPy включають основні операції, такі як додавання, віднімання, множення та ділення, а також більш складні операції, такі як лінійна алгебра, перетворення Фур'є та статистичний аналіз. Оптимізований код на C NumPy дозволяє швидко виконувати ці операції, що робить його популярним вибором для наукових програм і додатків, що інтенсивно обробляють дані.

На додаток до своїх математичних функцій, NumPy також надає інструменти для маніпулювання масивами, включаючи зміну форми, нарізання та індексування. NumPy інтегрується з іншими науковими обчислювальними бібліотеками в Python, такими як SciPy і Matplotlib, щоб забезпечити комплексний набір інструментів для наукових обчислень і аналізу даних.

Загалом, NumPy — це потужна бібліотека для наукових обчислень на Python, яка забезпечує високопродуктивну структуру даних і широкий спектр математичних функцій і функцій обробки даних. Він став стандартним інструментом для аналізу даних, наукових досліджень і машинного навчання в Python.

Matplotlib — це бібліотека Python, яка використовується для візуалізації та графічного представлення даних. Вона надає різноманітні функції для створення

статичних, анімованих та інтерактивних візуалізацій на Python [59]. Matplotlib має широкі можливості налаштування та може створювати широкий діапазон типів графіків, включаючи лінійні діаграми, точкові діаграми, стовпчасті діаграми, гістограми тощо. Він також сумісний з різними форматами даних, включаючи масиви NumPy і кадри даних Pandas.

Ядром Matplotlib є модуль pyplot, який надає низку функцій для створення фігур і графіків і керування ними. Matplotlib надає різноманітні параметри налаштування графіка, включаючи стилі графіка, мітки осей, позначки галочок, легенди, кольори тощо. Він також підтримує складні сюжетні плани та підсюжети.

Matplotlib можна використовувати для широкого спектру програм, включаючи наукову візуалізацію, дослідження даних і презентаційну графіку. Він добре інтегрується з іншими науковими обчислювальними бібліотеками на Python, такими як NumPy і Pandas, а також з інтерактивними інструментами візуалізації, такими як Jupyter notebooks і IPython.

Загалом, Matplotlib — це потужна та гнучка бібліотека для створення високоякісних візуалізацій даних на Python. Його широка функціональність і гнучкість роблять його популярним вибором для вчених, дослідників і аналітиків даних.

Seaborn — це бібліотека візуалізації даних Python на основі Matplotlib. Вона забезпечує інтерфейс вищого рівня для створення статистичної графіки та візуалізації даних з акцентом на естетиці та простоті використання [60].

Seaborn надає різноманітні типи графіків для візуалізації взаємозв’язків у даних, зокрема діаграми розсіювання, лінійні діаграми, стовпчасті діаграми та гістограми. Він також включає спеціалізовані типи графіків, такі як теплові карти, карти кластерів і скрипкові графіки, а також статистичні візуалізації, такі як графіки регресії та графіки розподілу.

Однією з ключових особливостей Seaborn є його здатність швидко та легко створювати складні візуалізації за допомогою мінімального коду. Він надає вбудовані колірні палітри та параметри стилю, що дозволяє легко налаштовувати зовнішній вигляд сюжетів. Seaborn також надає інструменти для візуалізації складних

взаємозв’язків у даних, таких як теплові карти та парні графіки, які можуть допомогти визначити закономірності та кореляції у великих наборах даних.

Seaborn добре працює з фреймами даних Pandas, що дозволяє легко візуалізувати дані безпосередньо з фреймів даних. Він також добре інтегрується з іншими бібліотеками Python для аналізу даних, такими як NumPy і SciPy.

Загалом, Seaborn — це потужна та гнучка бібліотека для створення високоякісних візуалізацій даних на Python. Його акцент на естетиці та простоті використання робить його популярним вибором для вчених, дослідників і аналітиків даних, яким потрібно створювати чіткі та інформативні візуалізації своїх даних.

Використовуючи зазначені модулі було розроблено програмний застосунок, що містить можливість збереження та перегляду попередніх результатів оцінювання, можливість безпосередньої оцінки рівня кібербезпеки організації, та можливість додавання нових контролів безпеки, які за допомогою технологій машинного навчання будуть класифіковані за доменом, а також для нової контролі буде визначений рівень імплементаційної групи і рекомендації щодо її впровадження.

- Функція оцінювання

Ця функція Python розроблена для завантаження опитувальника оцінки в інтерфейсі tkinter, що дозволяє користувачеві проводити самооцінку щодо набору питань - елементів керування. Запитання/елементи керування завантажуються з файлу CSV, який забезпечує гнучкий і настроюваний спосіб визначення запитань і пов’язаних із ними параметрів (так, ні, частково).

Функція спочатку читає файл CSV і витягує питання та варіанти відповідей. Потім створюється графічний інтерфейс користувача (GUI) за допомогою бібліотеки tkinter, який відображає запитання та варіанти відповідей у зручному форматі. Потім користувач може вибрати свої відповіді для кожного запитання/контролю та надіслати свої відповіді.

Згодом функція оброблятиме відповіді користувача, надаючи підсумок їхніх відповідей і загальну оцінку. Ця оцінка ґрунтується на сумі відповідей користувача а також включає більш складні правила оцінки, які враховують необхідність впровадження контролі для досягнення певного рівня імплементаційної групи.

Загалом ця функція забезпечує зручний спосіб проведення самооцінки для широкого спектру додатків, таких як кібербезпека, управління ризиками або контроль якості. Використання файлу CSV для визначення запитань і варіантів відповідей дозволяє легко налаштувати оцінювання для різних контекстів і доменів, тоді як інтерфейс tkinter забезпечує гнучкий та інтуїтивно зрозумілий спосіб взаємодії з оцінюванням.

- Функція зберігання та відображення попередніх результатів

Ця функція Python призначена для зберігання результатів оцінювання у спеціальному файлі з можливістю завантажувати та отримувати ці результати пізніше.

Функція прийматиме як вхідні дані результати оцінювання, які можуть бути у формі списку, словника або іншої структури даних, залежно від специфіки оцінювання. Потім вона створює новий файл, використовуючи вказаний формат, щоб забезпечити послідовність і читабельність даних. Цей формат може бути простим текстовим файлом, файлом із значеннями, розділеними комами (CSV) або більш складним форматом, наприклад JSON або XML.

Ця функція також надає можливість завантажувати та отримувати збережені результати оцінювання, що включає читання файлу та розбір даних у зручний формат, наприклад список або словник. Функція також включає перевірку помилок і перевірку, щоб переконатися, що завантажені дані відповідають очікуваному формату та типам даних.

Загалом ця функція забезпечує надійний і гнучкий спосіб зберігання та отримання результатів оцінювання, що може бути корисно для широкого спектру програм, наприклад для відстеження прогресу з часом або порівняння результатів різних оцінювань. Використання певного формату файлу забезпечує послідовність і сумісність з іншими інструментами та системами, тоді як можливість завантажувати та отримувати дані забезпечує зручний спосіб доступу та аналізу результатів оцінювання.

Результати оцінки подаються у вигляді рахунку (балів) досягнення однієї з імплементаційних груп, а також завдяки поділу на домени, програмний застосунок

надає необхідні рекомендації з метою полегшення пріоритизації впровадження та підтримки заходів інформаційної безпеки.

- Функція додавання нових контролей

Ця функція Python розроблена, щоб приймати як вхідні дані нові контролі, обробляти їх за допомогою вже навченої моделі класифікації та призначати домен, групу реалізації та рекомендації для виконання контролі.

Функція спочатку завантажує навчену модель класифікації, яка є моделлю машинного навчання розробленої в попередньому розділі, навченої на наборі даних контролей та пов’язаних атрибутів, таких як домени, групи впровадження та рекомендовані дії. Потім функція сприймає нові контролі як вхідні дані та попередньо оброблює їх за потреби, щоб зробити їх сумісними з форматом введення моделі.

Потім функція застосує навчену модель класифікації до нових елементів керування, використовуючи вивчені параметри та алгоритми моделі для класифікації елементів керування в один або кілька доменів, груп реалізації та рекомендованих дій. Ці класифікації можуть ґрунтуються на ряді факторів, таких як мета контролю, технічні вимоги, відповідність нормативним вимогам і міркування щодо управління ризиками.

Нарешті, функція виводить результати процесу класифікації, які можуть включати короткий виклад області керування, групи реалізації та рекомендованих дій. Ці результати представлені в зручному для користувача форматі - у вигляді таблиці, для полегшення розуміння та прийняття рішень.

Загалом ця функція надає потужний інструмент для автоматизації класифікації нових елементів керування та призначення їх відповідним доменам, групам реалізації та рекомендованим діям. Використання навченої моделі класифікації забезпечує послідовну та точну класифікацію на основі широкого діапазону факторів, тоді як вихідний формат забезпечує простий у використанні підсумок для подальшого аналізу та дії.

Приклад інтерфейсу програми (рис. 3.4)

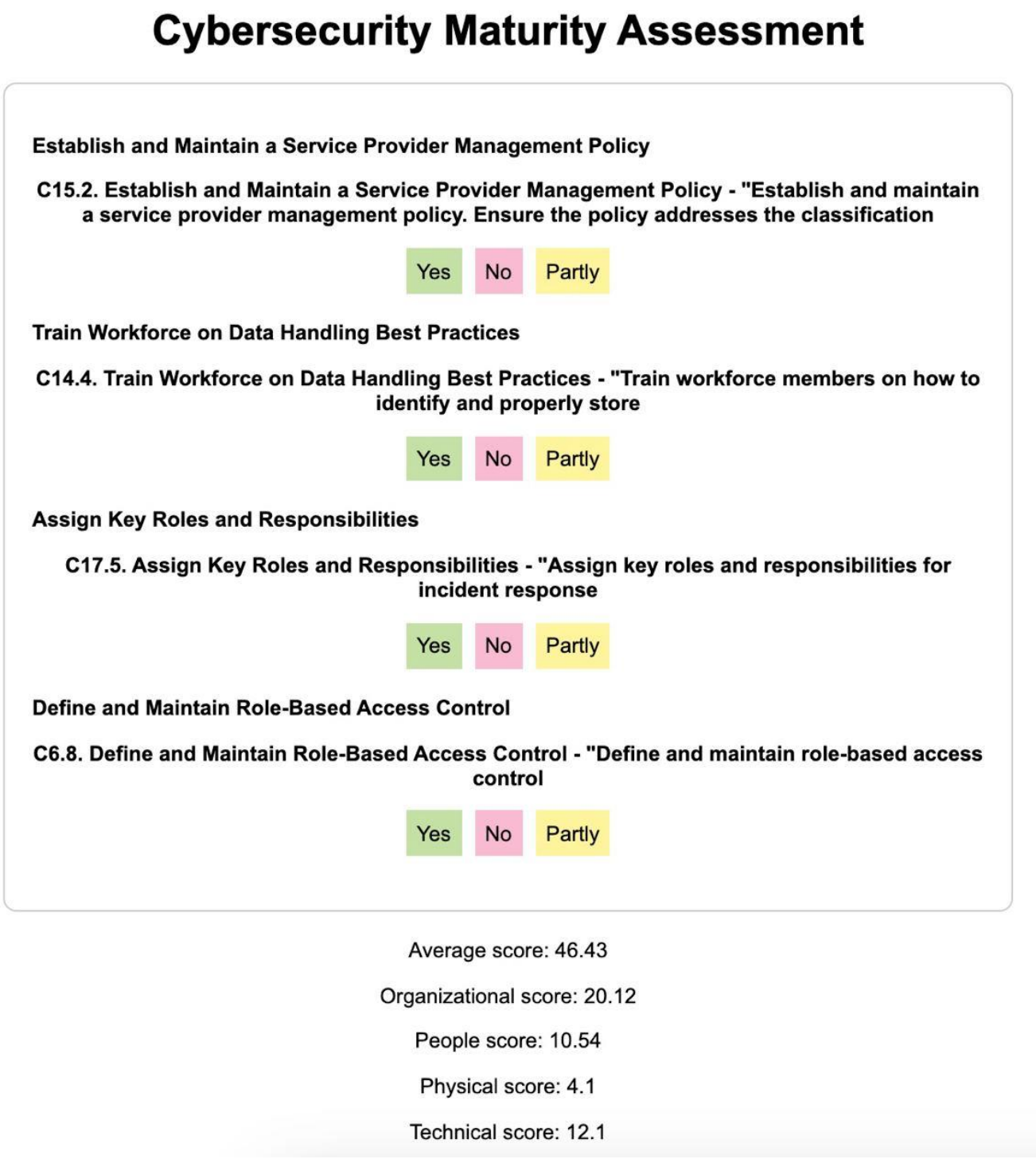

Рисунок 3.4 Приклад роботи програмного застосунку

Завдяки інтеграції вдосконалених додатків процес оцінки кібербезпеки було оптимізовано та автоматизовано, що революціонізувало спосіб, у який організації оцінюють свої заходи безпеки. Використовуючи передові технології, такі як штучний інтелект, машинне навчання та автоматизація, наш інноваційний додаток усунув ручні та трудомісткі завдання, які традиційно пов’язували з оцінками кібербезпеки.

Цей потужний інструмент систематично сканує мережі, системи та програми, проводячи комплексну оцінку вразливості та тестування на проникнення. Використовуючи свої надійні алгоритми, програма швидко виявляє потенційні слабкі місця в безпеці, виявляє аномалії та створює докладні звіти з корисною інформацією. Автоматизація процесу оцінки не тільки економить значний час і ресурси, але й забезпечує більш точну та послідовну оцінку, дозволяючи організаціям оперативно виявляти та усувати потенційні вразливості, зміцнюючи загальний захист кібербезпеки.

Висновки за розділом 4

Як висновок, в цьому розділі подано огляд бібліотек, використаних у розробці програми для оцінки рівня безпеки. Ці бібліотеки зіграли вирішальну роль у підвищенні функціональності, ефективності та надійності програми, завдяки їхній можливості зчитувати, обробляти, аналізувати, та візуалізувати дані.

Таким чином, в третьому розділі було реалізовано модель оцінки рівня кібербезпеки організації з використанням класифікації контролів безпеки на основі алгоритмів машинного навчання.

В результаті розроблена модель може бути адаптована та використана в існуючому процесі захисту інформації, як елемент управління кібербезпекою організації.

ВИСНОВКИ

Зважаючи на швидкий розвиток кіберпростору, організаціям необхідно займати проактивну позицію в області захисту інформації в кібепросторі. У ході дипломної роботи було досліджено – процес оцінки рівня кібербезпеки організації з метою розробки моделі оцінки рівня кібербезпеки організації. Отже, ця кваліфікаційна робота була зосереджена на розробці та впровадженні моделі оцінки для оцінки рівня впровадження контролю безпеки в межах конкретного домену.

Модель оцінки, використана в цьому дослідженні, враховує різні чинники та параметри, які впливають на загальну безпеку. Ці фактори включають технічні аспекти засобів контролю безпеки, такі як їх функціональність, надійність і ефективність, а також організаційні та операційні аспекти, такі як дотримання правил, навчання персоналу та можливості реагування на інциденти.

Оцінки, отримані за допомогою розробленого підходу, слугують кількісним показником, який може бути представлений зацікавленим сторонам щодо оцінювання ефективності впровадження контролю безпеки та відстеження прогресу з часом.

Крім того, модель оцінки є адаптивною та гнучкою, враховуючи зміни та прогрес у технологіях, нові загрози та зміну нормативних вимог. Ця адаптивність гарантує, що модель залишається актуальною та ефективною в умовах динамічного середовища безпеки.

Загалом, розробка та впровадження моделі оцінки, представленої в дипломній роботі, суттєво сприяють посиленню впровадження контролю безпеки в межах домену. Модель забезпечує комплексну структуру оцінки, яка сприяє безперервному вдосконаленню та допомагає організаціям визначити пріоритети своїх зусиль у сфері інформаційної безпеки. Оскільки загрози безпеці продовжують розвиватися, організаціям вкрай необхідно впровадити надійні моделі оцінки, щоб забезпечити постійний захист своїх критично важливих активів і конфіденційної інформації.

В першому розділі дипломної роботи було проаналізовано літературу, визначено основні підходи, переваги та недоліки оцінки рівня зрілості кібербезпеки

організації. Проведено дослідження джерел в області контролів безпеки, складності їх впровадження та підтримки. Було досліджено міжнародні стандартів, що висувають конкретні вимоги щодо інформаційних систем, де обробляється інформація різного характеру. Також були розглянуті моделі оцінки рівня зрілості процесів і технологій/практик інформаційної, зокрема кібербезпеки організацій різного рівня. Наостанок, невід'ємною частиною управління кібербезпекою є взаємозв'язок та залежність рівнів зрілості ІТ та ІБ.

В другому розділі було виконано завдання класифікації нових контролів за доменами, що забезпечує гнучкість моделі оцінки. Було підготовлено набори даних, виконано дослідницький аналіз даних, попередню обробку текстових наборів контролей безпеки з метою синтезу моделі класифікації тексту. Також, було досліджено точність класифікації контролей з кількома мітками.

В третьому розділі на основі розробленої моделі класифікації було розроблено модель оцінки рівня забезпечення кібербезпеки організації, а також було автоматизовано процес оцінки рівня забезпечення кібербезпеки.

Таким чином, були виконані всі завдання дипломної роботи.

Отже, мета роботи, а саме розробка моделі оцінки рівня кібербезпеки організації була досягнута.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Michael Swanagan The 3 Types Of Security Controls [Електронний ресурс]. Режим доступу: https://purplesec.us/security-controls/

2. Krumay, B., Bernroider, E.W.N., Walser, R. (2018). Evaluation of Cybersecurity Management Controls and Metrics of Critical Infrastructures: A Literature Review Considering the NIST Cybersecurity Framework. In: Gruschka, N. (eds) Secure IT Systems. NordSec 2018. Lecture Notes in Computer Science(), vol 11252. Springer, Cham. pp 369–384

3. Reliaquest Threat Research Team 2023 Vulnerabilities: First-Quarter Highlights, 2023, [Електронний ресурс]. Режим доступу: https://www.reliaquest.com/blog/2023-q1-vulnerabilities-cves/

4. Horizons What is Global Compliance and Why Does it Matter?, 2023, [Електронний ресурс]. Режим доступу: https://nhglobalpartners.com/global-compliance-5-reasons-why-it-matters/

5. Sulaiman, N.S.; Fauzi, M.A.; Wider, W.; Rajadurai, J.; Hussain, S.; Harun, S.A. Cyber–Information Security Compliance and Violation Behaviour in Organisations: A Systematic Review. Soc. Sci. 2022, 11, p 386.

6. Brian Willis Three Categories of Security Controls, 2022, [Електронний ресурс]. Режим доступу: https://www.lbmc.com/blog/three-categories-of-security-controls/

7. Rizal A.A., Sarno R., Sungkono K.R. 2020, COBIT 5 for Analysing Information Technology Governance Maturity Level on Masterplan E-Government, International Seminar on Application for Technology of Information and Communication (iSemantic)

8. ISACA, COBIT 2019 Framework: Governance and Management Objectives, ISACA, 2019

9. Motii Malik, Semma Alami. 2017, Towards a new approach to pooling COBIT 5 and ITIL V3 with ISO/IEC 27002 for better use of ITG in the Moroccan parliament, International Journal of Computer Science Issues, Volume 14, Issue 3

10. L. Toyner, L. G., & Sfenrianto, S. (2023), Information system security evaluation using cobit 5 framework, Journal of Information System Management (JOISM), 4(2), 147 – 157

11. NIST Спеціальна публікація 800-53: Контроль безпеки та конфіденційності для інформаційних систем і організацій, [Електронний ресурс]. Режим доступу: https://csrc.nist.gov/publications/detail/sp/800-53/rev-5/final

12. International Organization for Standardization, (2022). ISO/IEC 27001:2022 Information Security, Cybersecurity and Privacy Protection – Information Security Management Systems – Requirements. ISO.

13. P. Magerde Comparative study between ISO 27001:2005, ISO 27001:2013 and ISO 27002:2022, 2022

14. T. Humphreys and A. Plate, Measuring the effectiveness of your ISMS implementations based on ISO/IEC 27001. London: BSi Business Information, 2006

15. C. Carvalho and E. Marques, “Adapting ISO 27001 to a public institution,” 2019 14th Iberian Conference on Information Systems and Technologies (CISTI), 2019

16. Ta-Seen Junaid, ISO 27001: Information Security Management Systems, Faculty of Computer Science and Engineering Frankfurt University of Applied Sciences Frankfurt am Main, Germany, 2023

17. S. E. Chang and C. B. Ho, Organizational factors to the effectiveness of implementing information security management, Industrial Management & Data Systems, vol. 106, no. 3, 2006, pp. 345–361.

18. S. E. Chang and C.-S. Lin, Exploring organizational culture for information security management, Industrial Management & Data Systems, vol. 107, no. 3, 2007, pp. 438–458

19. D. Mellado, E. Fernandez-Medina, and M. Piattini, A common criteria based security requirements engineering process for the development of secure information systems, Computer Standards & Interfaces, vol. 29, 2007, pp. 244–253.

20. A comparison of the common criteria with proposals of information systems security requirements, in 1st International Conference on Availability, Reliability and Security. Vienna, Austria: IEEE, 2006

21. S. Fenz, G. Goluch, A. Ekelhart, and E. Weippl, Information security fortification by ontological mapping of the ISO/IEC 27001 standard, in Pacific Rim International Symposium on Dependable Computing. Melbourne, Victoria, Australia: IEEE Computer Society, 2007, pp. 381– 388.

22. S. Fenz, Ontology-based generation of IT-Security metrics, in ACM Symposium on Applied Computing. Sierre, Switzerland: ACM, 2010, pp. 1833–1839.

23. S. Fenz, G. Goluch, A. Ekelhart, and E. Weippl, Information security fortification by ontological mapping of the ISO/IEC 27001 standard, in Pacific Rim International Symposium on Dependable Computing. Melbourne, Victoria, Australia: IEEE Computer Society, 2007, pp. 381– 388.

24. S. Fenz, Ontology-based generation of IT-Security metrics, in ACM Symposium on Applied Computing. Sierre, Switzerland: ACM, 2010, pp. 1833–1839.

25. R. Montesino, S. Fenz, and W. Baluja, SIEM-based framework for security controls automation, Information Management & Computer Security, vol. 20, no. 4, 2012, pp. 248–263.

26. R. Montesino and S. Fenz, Automation possibilities in information security management, in European Intelligence and Security Informatics Conference. Athens, Greece: IEEE, 2011, pp. 259–262

27. Dupuis M., Bejan C., Bishop M., David S. 2019, Lagesse B, Design Patterns for Compensating Controls for Securing Financial Session, IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation

28. Elluri L., Nagar A., Joshi K.P. 2018 An Integrated Knowledge Graph to Automate GDPR and PCI DSS Compliance, IEEE International Conference on Big Data (Big Data)

29. PCI DSS Quick Reference Guide, Understanding the Payment Card Industry Data Security Standard version 3.2.1, 2018

30. K. Razikin and A. Widodo, General Cybersecurity Maturity Assessment Model: Best Practice to Achieve Payment Card Industry-Data Security Standard (PCI-DSS)

Compliance, CommIT (Communication & Information Technology) Journal 15(2), 91–104, 2021.

31. Health Insurance Portability and Accountability Act of 1996 (HIPAA) [Електронний ресурс]. Режим доступу: https://www.cdc.gov/phlp/publications/topic/hipaa.html#:~:text=The%20Health%20Insurance%20Portability%20and,the%20patient's%20consent%20or%20knowledge.

32. What is GDPR, the EU’s new data protection law? [Електронний ресурс]. Режим доступу: https://gdpr.eu/what-is-gdpr/

33. ITIL. (n.d.). Incident management [Електронний ресурс]. Режим доступу: https://www.itlibrary.org/

34. T. Conkle, Cybersecurity Maturity - Leveraging Standards, 2020. [Електронний ресурс]. Режим доступу: https://www.youtube.com/watch?v=7_b2REug0gg

35. Rea-Guaman, A.M., San Feliu, T., Calvo-Manzano, J.A., Sanchez-Garcia, I.D. (2017). Comparative Study of Cybersecurity Capability Maturity Models. In: Mas, A., Mesquida, A., O'Connor, R., Rout, T., Dorling, A. (eds) Software Process Improvement and Capability Determination. SPICE 2017. Communications in Computer and Information Science, vol 770. Springer, pp 100–113

36. Cybersecurity Capability Maturity Model for Information Technology Services (C2M2 for IT Services), Version 1.0 [Електронний ресурс]. Режим доступу: https://apps.dtic.mil/dtic/tr/fulltext/u2/1026943.pdf

37. Capability Maturity Model Integration Conference: 3rd International Multidisciplinary Research ConferenceAt: Sarhad University of Sceince & Information Technology Peshawar - PakistanVolume: 3rd [Електронний ресурс]. Режим доступу: https://www.researchgate.net/publication/272817266_Capability_Maturity_Model_Integration

38. CIS Critical Security Controls, [Електронний ресурс]. Режим доступу: https://www.cisecurity.org/controls

39. Multi-State Information Sharing and Analysis Center, [Електронний ресурс]. Режим доступу: https://www.cisecurity.org/ms-isac

40. Multi-Label Classification with Deep Learning by Jason Brownlee on August 31, 2020 in Deep Learning [Електронний ресурс]. Режим доступу: https://machinelearningmastery.com/multi-label-classification-with-deep-learning/

41. J. Bogatinovski, L. Todorovski, S. Džeroski, D. Kocev, Comprehensive comparative study of multi-label classification methods, Expert Systems with Applications, Volume 203, 2022, 117215, ISSN 0957-4174

42. Matthew R. Boutell, Jiebo Luo, Xipeng Shen, Christopher M. Brown, Learning multi-label scene classification, Pattern Recognition, Volume 37, Issue 9, 2004, pp1757-1777

43. Guo, Y., Gu, S, 22nd International Joint Conference on Artificial Intelligence, IJCAI 2011; Barcelona, Catalonia; Spain; 16 July 2011 through 22 July 2011; Code 97874, 2011, pp 1300-1305

44. Asch, V.V. Macro-and micro-averaged evaluation measures [[BASIC DRAFT]]. 2013

45. Guoqiang Wu, Jun Zhu Multi-label classification: do Hamming loss and subset accuracy really conflict with each other?, 34th Conference on Neural Information Processing Systems (NeurIPS 2020), Vancouver, Canada

46. Roseberry, et al., Martha Roseberry, Bartosz Krawczyk, and Alberto Cano, ACM Trans. Knowl. Discov. Data., Vol. 1, No. 1, Article 1. Publication date: August 2019.

47. Electronic Design Automation (EDA) - Design And Reuse [Електронний ресурс]. Режим доступу: https://www.design-reuse.com/solutions/electronic-design-automation-eda/

48. WordCloud Generator - Jason Davies [Електронний ресурс]. Режим доступу: https://www.jasondavies.com/wordcloud/

49. Pre-processing of Text Data in Machine Learning [Електронний ресурс]. Режим доступу: https://www.enjoyalgorithms.com/blog/text-data-pre-processing-techniques-in-ml

50. Natural Language Toolkit, [Електронний ресурс]. Режим доступу: https://www.nltk.org/

51. Kateryna Mokliakova, Tetiana Babenko, Dmytro Palko. Cybersecurity level assessment models, International Conference on Next Generation Cybersecurity Systems and Applications, NGSEC.–26-27 April 2023, Kyiv – Conference Proceedings, Section 2

52. Understanding TF-ID: A Simple Introduction, [Електронний ресурс]. Режим доступу: https://monkeylearn.com/blog/what-is-tf-idf/

53. OneVsRestClassifier - Machine Learning Mastery, [Електронний ресурс]. Режим доступу: https://machinelearningmastery.com/one-vs-rest-and-one-vs-one-for-multi-class-classification/

54. Classifier Chain, [Електронний ресурс]. Режим доступу: https://scikit-learn.org/stable/auto_examples/multioutput/plot_classifier_chain_yeast.html

55. Label Powerset, [Електронний ресурс]. Режим доступу: http://scikit.ml/api/skmultilearn.problem_transform.lp.html

56. tkinter — Python interface to Tcl/Tk [Електронний ресурс]. Режим доступу: https://docs.python.org/3/library/tkinter.html

57. Wes Mckinney Pandas: a Foundational Python Library for Data Analysis and Statistics, 2011

58. S. van der Walt, S. C. Colbert and G. Varoquaux, The NumPy Array: A Structure for Efficient Numerical Computation, in Computing in Science & Engineering, vol. 13, no. 2, pp. 22-30, March-April 2011

59. P. L. Shopbell, M. C. Britton, and R. Ebert, eds. Astronomical data analysis software and systems XIVASP Conference Series, Vol. 347, 2005

60. Seaborn: statistical data visualization, [Електронний ресурс]. Режим доступу: https://seaborn.pydata.org/

ДОДАТОК А

СПИСОК ОПУБЛІКОВАНИХ ПРАЦЬ ЗА ТЕМОЮ МАГІСТЕРСЬКОЇ РОБОТИ

Тези наукових доповідей:

1. Kateryna Mokliakova, Tetiana Babenko, Dmytro Palko. Cybersecurity level assessment models, International Conference on Next Generation Cybersecurity Systems and Applications, NGSEC.–26-27 April 2023, Kyiv – Conference Proceedings, Section 2

2. Kateryna Mokliakova, Tetiana Babenko. Cybersecurity maturity assessment models. Збірник матеріалів доповідей та тез; м. Київ, 27 квітня 2023 року; Київський національний університет імені Тараса Шевченка 2023. – 166 с. (с.72-74)

ДОДАТОК Б

КОД ПРОГРАМНОЇ РЕАЛІЗАЦІЇ МОДЕЛІ

```
import pandas as pd
import numpy as np

import seaborn as sns
import matplotlib.pyplot as plt

#for text pre-processing
import re, string
import nltk
from nltk.tokenize import word_tokenize
from nltk.corpus import stopwords
from nltk.tokenize import word_tokenize
from nltk.stem import SnowballStemmer
from nltk.corpus import wordnet
from nltk.corpus import stopwords
from nltk.stem import WordNetLemmatizer

nltk.download('punkt')
nltk.download('averaged_perceptron_tagger')
nltk.download('wordnet')

#for model-building
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from sklearn.linear_model import SGDClassifier
from sklearn.naive_bayes import MultinomialNB
from sklearn.metrics import classification_report, f1_score, accuracy_score, confusion_matrix
from sklearn.metrics import roc_curve, auc, roc_auc_score
from sklearn.multiclass import OneVsRestClassifier
from sklearn.pipeline import Pipeline

# bag of words
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.feature_extraction.text import CountVectorizer

#for word embedding
import gensim
from gensim.models import Word2Vec #Word2Vec is mostly used for huge datasets
```

```
# iso controls: 5 - organizational, 6 - people, 7 - physical , 8 - technical
import os
os.chdir(r"C:/Users/Kateryna_Mokliakova/Desktop/diplom_code")
df_data = pd.read_csv("iso.csv")
print(df_data.shape)
df_data.head()
#print(df_data['Organizational'].to_string(index=False))

#next step is to give a categories in recommends db
df_cis = pd.read_csv("cis.csv")
print(df_cis.shape)
df_cis.head()

df_map = pd.read_csv("map.csv")
print(df_map.shape)
df_map.head()

#iterate through all ID_ISO and add ID_CIS to df_iso marking their domain that are the same as ID_ISO domain

#iterate through all ID_ISO and add ID_CIS to df_iso marking their domain that are the same as ID_ISO domain

buffer_df = pd.DataFrame(columns=['ID','Control','Description','Organizational','People','Physical','Technical']) #buffer data
new_data = [] #we will store the temporary inf here
for ind, row in df_map.iterrows():
    if row[1]: #check if we have value that we will append
        for ind_cis, row_cis in df_cis.iterrows():
            if row_cis[0] == row[1]:
                new_data.extend([row[1], row_cis[1], row_cis[2]]) #the list now have 3 elements ID, Name, Description

        for ind_data, row_data in df_data.iterrows():
            if row_data['ID'] == row[0]:
                new_data.extend([row_data['Organizational'],row_data['People'],row_data['Physical'],row_data['Technical'],])

    buffer_df.loc[len(buffer_df)] = new_data
    new_data.clear()

buffer_df.tail()
```

```
#remove duplicates and merge controls that have more than 1 domain
buffer_df = buffer_df.drop_duplicates()
buffer_df = buffer_df[buffer_df['ID'].duplicated(keep = False) == True]
agg_functions = {'Control': 'first', 'Description': 'first', 'Organizational': 'sum', 'People': 'sum', 'Physical': 'sum', 'Technical':
'sum'}
buffer_df = buffer_df.groupby(buffer_df['ID'], as_index=False).aggregate(agg_functions).reindex(columns=buffer_df.
columns)
buffer_df

#Data with ISO and CIS controls and their domains. Some controls are dedicated to more that 1 domain
df_data = df_data.append(buffer_df)
df_data


rowSums = df_data.iloc[:,2:].sum(axis=1)

print("Total number of controls = ",len(df_data))


categories = list(df_data.columns.values)
categories = categories[3:]
print(categories)

# Calculating number of comments in each category

counts = []
for category in categories:
    counts.append((category, df_data[category].sum()))
df_stats = pd.DataFrame(counts, columns=['category', 'number of controls'])
df_stats


sns.set(font_scale = 2)
plt.figure(figsize=(10,6))
ax= sns.barplot(categories, df_data.iloc[:,3:].sum().values)
plt.title("Controls in each category", fontsize=10)
plt.ylabel('Number of controls', fontsize=8)
plt.xlabel('Control Domain', fontsize=8)
#adding the text labels
rects = ax.patches
labels = df_data.iloc[:,3:].sum().values
for rect, label in zip(rects, labels):
    height = rect.get_height()
    ax.text(rect.get_x() + rect.get_width()/2, height + 1, label, ha='center', va='bottom', fontsize=12)
plt.show()
```