МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Західноукраїнський національний університет
Факультет комп'ютерних інформаційних технологій
Кафедра інформаційно-обчислювальних систем і управління

КНИШ Тетяна Олегівна

Модуль оцінювання аварій на промислових об'єктах на основі опису ситуації присутніми / Module for assessing accidents at industrial facilities based on the description of the situation by those present

Спеціальність 122 – Комп'ютерні науки
Освітньо-професійна програма – Комп'ютерні науки

Дипломний проект

Виконала студентка групи КН-41
Т. О. Книш

Науковий керівник:
к.т.н., доцент Х.В. Ліп'яніна-Гончаренко

Дипломний проект допущено до захисту
«___»_____ 2023 р.

Завідувач кафедри
_____ М.П. Комар

Тернопіль – 2023

Факультет комп'ютерних інформаційних технологій
Кафедра інформаційно-обчислювальних систем і управління
Освітній ступінь «бакалавр»
Спеціальність 122 – Комп'ютерні науки
Освітньо-професійна програма – Комп'ютерні науки

ЗАТВЕРДЖУЮ
Завідувач кафедри
_____ М.П. Комар
«____» _____2022р.

З А В Д А Н Н Я

НА ДИПЛОМНИЙ ПРОЕКТ СТУДЕНТЦІ

Книш Тетяні Олегівні
(прізвище, ім'я, по батькові)

1. Тема проекту: Модуль оцінювання аварій на промислових об'єктах на основі опису ситуації присутніми / Module for assessing accidents at industrial facilities based on the description of the situation by those present
керівник проекту к.т.н., доцент Х.В. Ліп'яніна-Гончаренко

затверджені наказом по університету від 08 грудня 2022 р. № 491.

2. Строк подання студентом закінченого проекту 01 червня 2023 р.

3. Вихідні дані до проекту: технічне завдання.

4. Зміст розрахунково-пояснювальної записки (перелік питань, які потрібно розробити)
– провести дослідження особливостей проведення оцінювання аварій на промислових об'єктах та аналіз існуючих підходів.
– провести аналіз існуючих підходів до оцінювання аварій на промислових об'єктах з метою визначення їх переваг та недоліків.
– розробити концептуальну модель оцінки наслідків техногенних катастроф та створення модуля оцінювання аварій на промислових об'єктах на основі опису ситуації присутніми.
– розробити архітектуру мобільного додатку та інформаційного забезпечення.
– провести дослідницький аналіз даних з використанням методів аналізу тексту та класифікації на основі текстових даних ансамблевим навчанням.
– розробити модуль оцінювання аварій на промислових об'єктах на основі опису ситуації присутніми.

5. Перелік графічного матеріалу (з точним зазначенням обов'язкових креслень)
   - концептуальну модель оцінки наслідків техногенних катастроф
   - архітектура мобільного додатку та інформаційного забезпечення

6. Консультанти розділів проекту

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		Завдання видав	Завдання прийняв
Н. контроль	к.т.н., доцент Х.В. Ліп'яніна-Гончаренко		

7. Дата видачі завдання 08 грудня 2022 р.

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів дипломного проекту	Строк виконання етапів проекту	Примітка
1	Аналіз предметної області і постановка задачі дослідження	30.12.2022	
2	Алгоритмічне та інформаційне забезпечення	24.03.2023	
3	Програмно-технологічне забезпечення	12.05.2023	
4	Повне завершення та оформлення дипломного проекту	01.06.2023	

Студентка_____ Т. О. Книш
(підпис)

Керівник проекту_____ Х.В. Ліп'яніна-Гончаренко
(підпис)

РЕФЕРАТ

Пояснювальна записка до дипломного проекту: 70 с., 22 рис., 4 табл., 2 додатки, 60 джерел.

Метою дипломного проекту є розробка модуля оцінювання аварій на промислових об'єктах на основі опису ситуації присутніми.

Об'єктом дослідження є процес аварій на промислових об'єктах, а предметом дослідження є розробка модулю оцінювання, що дозволяє оцінити ступінь небезпеки аварії та рекомендувати заходи щодо її недопущення в майбутньому.

Предметом дослідження є розробка модулю оцінювання аварій на промислових об'єктах на основі опису ситуації присутніми.

Розроблено та досліджено програмне забезпечення для модуля модулю в промислових підприємствах з метою запобігання аварій та забезпечення безпеки працівників та населення, що проживає в районах розташування промислових об'єктів.

Розроблений модуль може знайти застосування в системах автоматизованого керування технологічними процесами та забезпечення безпеки промислових об'єктів.

МАШИННЕ НАВЧАННЯ, АНАЛІЗ ДАНИХ, ОЦІНЮВАННЯ АВАРІЙ, ТЕХНОГЕННІ КАТАСТРОФИ.

ABSTRACT

The bachelor's thesis report: 70 pages, 22 figures, 4 tables, 2 appendices, 60 references.

The aim of the diploma project is to develop a module for assessing accidents at industrial facilities based on the description of the situation provided.

The object of the research is the process of accidents at industrial facilities, and the subject of the research is the development of an assessment module that allows evaluating the degree of danger of an accident and recommending measures to prevent it in the future.

Software for the module has been developed and studied for implementation in industrial enterprises to prevent accidents and ensure the safety of workers and the population residing in the areas near industrial facilities.

The developed module can be applied in automated control systems for technological processes and ensuring the safety of industrial facilities.

MACHINE LEARNING, DATA ANALYSIS, ACCIDENT ASSESSMENT, TECHNOGENIC CATASTROPHES.

ТЕХНІЧНЕ ЗАВДАННЯ

1. НАЙМЕНУВАННЯ ТА ОБЛАСТЬ ЗАСТОСУВАННЯ

1.1 Модуль оцінювання аварій на промислових об'єктах на основі опису ситуації присутніми.

1.2 Область застосування – розроблений модуль може бути використаний у промислових підприємствах для оцінювання аварій та запобігання їм.

2. ОСНОВА ДЛЯ РОЗРОБЛЕННЯ

Основою для розроблення є завдання на дипломний проект, затверджене кафедрою інформаційно-обчислювальних систем і управління факультету комп’ютерних інформаційних технологій Західноукраїнського національного університету.

3. ПРИЗНАЧЕННЯ РОЗРОБЛЕНОГО КОМПЛЕКСУ

Метою дипломного проекту є розробка модуля оцінювання аварій на промислових об'єктах на основі опису ситуації присутніми.

4. ДЖЕРЕЛА РОЗРОБЛЕННЯ

Джерелами даної розробки є матеріали навчальної і реферативної літератури, технічна документація, науково-дослідні статті, журнали, Інтернет.

5. ТЕХНІЧНІ ВИМОГИ

5.1 Основні функціональні вимоги до програмної системи:

– вибір даних (пакетних даних / даних у реальному часі) з різних архітектур великих даних;

– попередня обробка даних для перетворення у необхідний формат;

– провести дослідницький аналіз даних з використанням методів аналізу тексту та класифікації на основі текстових даних ансамблевим навчанням.

– розробка модуля оцінювання аварій на промислових об'єктах на основі опису ситуації присутніми.

5.2 Вимоги до апаратних засобів:

– кластер запускається на фізичній машині Macbook Pro з 16 гігабайт оперативної пам’яті та процесором 2.3 GHz Intel Core i5.

5.3 Вимоги до програмних засобів:

– для розробки програмне забезпечення - Python 3.7;

– для створення графічного інтерфейсу користувача використано – tkinter, Adobe XD;

– для реалізації моделей навчання – фреймворк sklearn.

6. ПОРЯДОК КОНТРОЛЮ

6.1 Представлення дипломного проекту на попередній захист.

6.2 Представлення дипломного проекту на захист.

Завдання прийняла до виконання _____ Т. О. Книш

(підпис) (прізвище та ініціали)

Керівник дипломного проекту_____ Х.В. Ліп’яніна-Гончаренко

(підпис) (прізвище та ініціали)

ЗМІСТ

Вступ 9
1 Аналіз предметної області і постановка задачі дослідження 11
1.1 Особливості проведення оцінювання аварій на промислових об'єктах 11
1.2 Аналіз існуючих підходів 15
1.3 Підходи дослідницького аналізу даних 17
1.4 Вибір перспективного шляху і постановка задачі дослідження 22
2 Алгоритмічне та інформаційне забезпечення 24
2.1 Концептуальна модель оцінки наслідків техногенних катастроф 24
2.2 Архітектура мобільного додатку 25
2.3 Методи аналізу тексту 26
2.4 Методи класифікації текстових даних ансамблевим навчанням 28
3 Програмно-технологічне забезпечення 38
3.1 Інформаційне забезпечення 38
3.2 Дослідницький аналіз даних 39
3.3 Класифікація на основі текстових даних 43
3.4 Модуль оцінювання аварій на промислових об'єктах на основі опису ситуації присутніми 47
Висновки 50
Список використаних джерел 52
Додаток А 57
Додаток Б 58

ВСТУП

Актуальність теми. Потенційна небезпека на промислових об'єктах може бути викликана різноманітними факторами, такими як технічні несправності, людський фактор, природні катастрофи тощо. У випадку аварії на промисловому об'єкті, наслідки можуть бути катастрофічними для людей, навколишнього середовища та економіки. Тому, розробка модуля оцінювання аварій на основі опису ситуації є важливим кроком у покращенні безпеки на промислових об'єктах.

Такий модуль може допомогти швидко та ефективно оцінити потенційний ризик аварії на основі опису ситуації. Це може допомогти зменшити час реакції на небезпечну ситуацію та запобігти можливим наслідкам, що можуть виникнути в результаті аварії. Крім того, модуль може допомогти в удосконаленні технічних засобів безпеки та покращенні процесів, пов'язаних з аварійною ситуацією на промислових об'єктах.

Отже, розробка модуля оцінювання аварій на промислових об'єктах на основі опису ситуації присутніми є важливим кроком у покращенні безпеки на промислових об'єктах та запобіганні можливим наслідкам, що можуть виникнути в результаті аварії.

Метою дипломного проекту є розробка модуля оцінювання аварій на промислових об'єктах на основі опису ситуації присутніми.

Досягнення цієї мети зумовило потребу теоретичних розробок, визначення та послідовного вирішення таких завдань:

1. провести дослідження особливостей проведення оцінювання аварій на промислових об'єктах та аналіз існуючих підходів.
2. провести аналіз існуючих підходів до оцінювання аварій на промислових об'єктах з метою визначення їх переваг та недоліків.
3. розробити концептуальну модель оцінки наслідків техногенних катастроф та створення модуля оцінювання аварій на промислових об'єктах на основі опису ситуації присутніми.

4. розробити архітектуру мобільного додатку та інформаційного забезпечення.
5. провести дослідницький аналіз даних з використанням методів аналізу тексту та класифікації на основі текстових даних ансамблевим навчанням.
6. розробити модуль оцінювання аварій на промислових об'єктах на основі опису ситуації присутніми.

Об'єктом дослідження є процес аварій на промислових об'єктах, а предметом дослідження є розробка модулю оцінювання, що дозволяє оцінити ступінь небезпеки аварії та рекомендувати заходи щодо її недопущення в майбутньому.

Предметом дослідження є розробка модулю оцінювання аварій на промислових об'єктах на основі опису ситуації присутніми.

Методи дослідження. У роботі використані методи математичного програмування, статистичного опису, візуалізації даних та машинного навчання.

Практичне значення полягає в можливості застосування розробленого модулю в промислових підприємствах з метою запобігання аварій та забезпечення безпеки працівників та населення, що проживає в районах розташування промислових об'єктів. Крім того, розроблений модуль може знайти застосування в системах автоматизованого керування технологічними процесами та забезпечення безпеки промислових об'єктів.

Структура та обсяг роботи. Дипломна робота складається із вступу, трьох розділів, висновків, списку використаних джерел. Повний обсяг роботи становить 53 ст. комп’ютерного тексту, який включає 14 рисунки та 3 таблиця. Список використаних джерел із 35 найменувань викладено на 5 сторінках.

1 АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ І ПОСТАНОВКА ЗАДАЧІ ДОСЛІДЖЕННЯ

1.1 Особливості проведення оцінювання аварій на промислових об'єктах

Оцінювання аварій на промислових об'єктах є важливим етапом у забезпеченні безпеки на робочому місці та попередженні аварій. Оцінювання передбачає вивчення ситуації, визначення потенційної небезпеки та розробку заходів для запобігання аварії в майбутньому.

Типи аварій на промислових об'єктах можуть бути класифіковані за різними ознаками, такими як масштаб, ступінь небезпеки, вид виробництва, наслідки та інші. Основні типи аварій на промислових об'єктах, з якими можуть стикатися робітники і власники підприємств, включають наступні:

1. Техногенні аварії виробничого обладнання - це аварії, що виникають на обладнанні і механізмах підприємства, таких як вибух, пожежа, замикання, витік рідини чи газу і т. д.
2. Надзвичайні ситуації на транспорті - аварії, що виникають на дорогах, залізницях, в морському, річковому і повітряному транспорті.
3. Надзвичайні ситуації в енергетиці - аварії, що пов'язані з генерацією, трансформацією та передачею електроенергії, які можуть призвести до перебоїв в електропостачанні, замикання, пожежі та інших наслідків.
4. Надзвичайні ситуації в хімічній і нафтогазовій промисловості - аварії, що пов'язані з виробництвом і зберіганням хімічних речовин, нефти та газу, які можуть призвести до викидів шкідливих речовин, пожеж, вибухів та інших небезпек.
5. Надзвичайні ситуації природного характеру - аварії, що виникають в результаті стихійних лих, таких як повені, землетруси, цунамі, урагани, лісові пожежі та інші.

Аварії на промислових об'єктах також можна класифікувати за рівнем їх наслідків. Зазвичай використовуються чотири рівні класифікації (Рис.1.1.):

- Перший рівень - місцеві аварії, які мають незначні наслідки та обмежені в межах промислового об'єкта.
- Другий рівень - регіональні аварії, які можуть мати значні наслідки на довкілля та здоров'я людей у межах регіону.
- Третій рівень - національні аварії, які мають значні наслідки для здоров'я людей та довкілля в цілому.
- Четвертий рівень - глобальні аварії, які мають катастрофічні наслідки на міжнародному рівні та можуть вплинути на стан довкілля та здоров'я на весь світ.

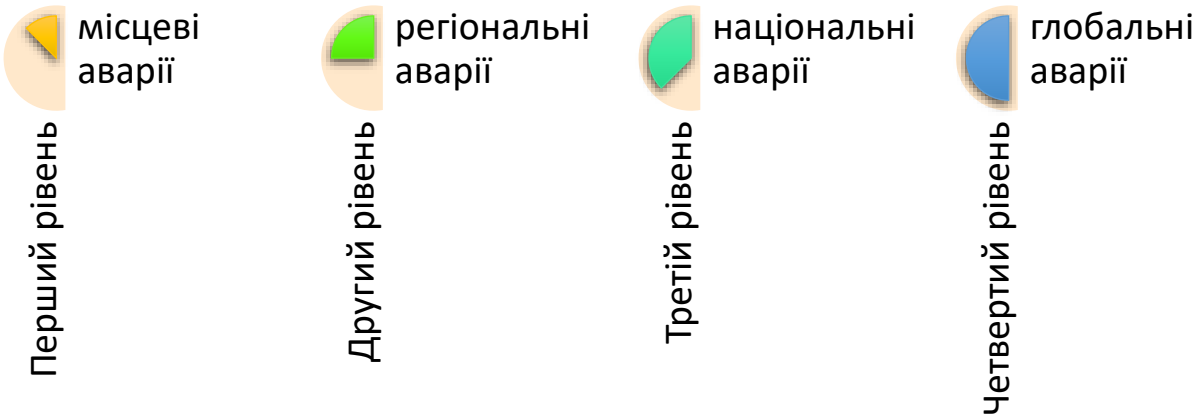


Рисунок 1.1 - Класифікація аварій за рівнем їх наслідків

Кожен рівень аварії вимагає відповідного підходу до її оцінки та управління наслідками.

Особливості проведення оцінювання аварій на промислових об'єктах включають (Рис.1.2) різні підходи.

Рисунок 1.2 - Особливості проведення оцінювання аварій на промислових об'єктах

Комплексність підходу. Оцінювання аварій повинно проводитись з урахуванням всіх аспектів, що впливають на небезпеку аварії, включаючи технічні, організаційні та соціально-економічні.

Використання експертних методів. Оцінювання аварій часто базується на думках та досвіді експертів з різних галузей, що дозволяє враховувати багато нюансів та особливостей конкретної ситуації.

Використання математичних моделей. Для проведення оцінювання можуть бути використані математичні моделі, що дозволяють враховувати різні фактори та прогнозувати наслідки аварій.

Врахування інформації від присутніх на місці події. При проведенні оцінювання аварій важливо враховувати інформацію, що надається

присутніми на місці події, а також використовувати різні джерела інформації, такі як відеозаписи, свідчення свідків тощо.

Системність підходу. Оцінювання аварій повинно бути вбудоване у систему управління безпекою на промислових об'єктах, яка передб ачувати регулярні аудити та перевірки, щоб переконатися у відповідності планування та реалізації запобіжних заходів.

Розгляд можливих наслідків. При оцінюванні аварій важливо розглянути можливі наслідки аварії та визначити способи їх запобігання. Це дозволяє забезпечити не тільки безпеку робочого місця, але й зменшити витрати на відновлення обладнання та інфраструктури після аварії.

Систематизація даних та їх аналіз. Для проведення оцінювання аварій важливо мати належну систему збору та аналізу даних, що дозволяє визначати тенденції та виявляти проблеми вчасно.

Використання сучасних технологій та інструментів. Сучасні технології та інструменти можуть значно полегшити та прискорити процес оцінювання аварій, а також забезпечити більш точні результати. До таких технологій можна віднести системи моніторингу та діагностики, системи відслідковування стану обладнання та його ремонтів, системи резервування енергії та інші.

Залучення кваліфікованих фахівців. Проведення оцінювання аварій на промислових об'єктах вимагає наявності висококваліфікованих спеціалістів з досвідом роботи в цій галузі. Їх залучення дозволяє не тільки проводити оцінювання комплексно та системно, але й ефективно розробляти плани запобігання аваріям та зменшення їх наслідків.

Моніторинг та оновлення оцінювання. Після проведення оцінювання аварій важливо забезпечувати моніторинг та оновлення результатів з метою виявлення нових ризиків та вдосконалення системи запобігання аварій. Також важливо регулярно оцінювати ефективність запроваджених заходів та коригувати їх у разі потреби.

Проведення оцінювання аварій на промислових об'єктах є важливим етапом у забезпеченні безпеки працівників та запобіганні аварій. Оцінювання повинно бути комплексним, системним та базуватися на використанні різних методів та джерел інформації. Його результати дозволяють розробити заходи для запобігання аварій в майбутньому та зменшити їх наслідки.

1.2 Аналіз існуючих підходів

Існує ряд методів для визначення техногенних катастроф, включаючи візуальне спостереження, використання сенсорів та моніторів, аналіз даних зі сторонніх джерел тощо. Однак, ці методи мають свої обмеження у точності та швидкості реакції, тому з'являється потреба у розробці нових та більш ефективних способів визначення рівня техногенної катастрофи. Серед нових напрямків досліджень з прогнозування та управління ризиками є оцінка катастроф на основі даних соціальних мереж. Використання інформації з соціальних мереж може значно сприяти виявленню та розпізнаванню небезпек, що пов'язані з катастрофами. Наприклад, аналіз публікацій та повідомлень користувачів у соціальних мережах може допомогти визначити рівень попередження, ступінь поширення та масштаб катастроф.

Застосування даних соціальних мереж може забезпечити більш ефективну систему управління катастрофами та скоротити час між виявленням катастрофи та прийняттям необхідних заходів. Однак, такі методи мають свої обмеження та ризики, такі як недостатня точність та надійність даних соціальних мереж, поширення неправдивої інформації тощо. Тому, перед прийняттям рішень щодо управління ризиками та катастрофами, необхідно ретельно аналізувати та перевіряти отриману інформацію.

Існує декілька робіт, що присвячені використанню даних соціальних мереж для керування надзвичайними ситуаціями. У [1] відображено один з підходів, який використовує інформацію з соціальних мереж для управління

надзвичайними ситуаціями. Робота [2] містить огляд передових методів керування даними та аналізу для створення кращих систем управління інформацією про катастрофи. У [3] були виявлені стадії катастроф, які були відображені в соціальних мережах, та встановлено зв'язок між методами, що базувалися на даних із соціальних мережах та зусиллями служб реагування. У [4] було запропоновано прототип системи, що використовує твіти для підтримки завдання мобілізації ресурсів під час надзвичайних ситуацій. Конвеєр рішень запропонованої системи включає класифікацію твітів з наступним ранжуванням твітів для визначення пріоритетності передачі ресурсу.

У статтях [5], [6], [7] та [8] досліджено різні підходи до аналізу соціальних мереж для виявлення та аналізу надзвичайних ситуацій. Наприклад, у [5] запропоновано мультиагентну систему для управління та аналізу даних мікроблогів у реальному часі, у [6] порівняно різні алгоритми аналізу даних для класифікації твітів про стихійні лиха, у [7] розроблено модель для оптимізованого розподілу ресурсів у сценарії після катастрофи на основі інформації з мікроблогів, а у [8] запропоновано нейронний підхід для ідентифікації ситуаційних твітів під час катастрофи. Проте, продуктивність цих підходів залежить від кількості тегів в соціальних мережах, які стосуються надзвичайних ситуацій, що може впливати на оперативність реагування на аварії.

Крім цього, у статтях [9] та [10] представлено дві інші системи екстреного виявлення подій в соціальних мережах. EAIMS використовує хешування, чутливе до місцевості, для ідентифікації, відстеження, узагальнення та керування останніми надзвичайними подіями з соціальних мереж, а Civique зосереджена на твітах з геотегами та використовує опорні векторні машини (SVM) і Naive Bayes для навчання.

У статтях [11], [12], [13], [14] та [15] представлені різноманітні методи та моделі для виявлення техногенних катастроф та аварійних ситуацій. У [11] описано систему ChangeOS, яка базується на глибокій об’єктно-орієнтованій

структурі та використовується для виявлення семантичних змін. У [12] розроблено модель для визначення часу реагування екстрених служб. У [13] запропоновано концептуальну та архітектурну основу системи виявлення аварійних ситуацій на основі парадигми «людина як датчик». У [14] описано гібридний підхід, який поєднує методи аналізу тексту та визначення місця розташування для ефективного виявлення постраждалих людей. Експериментальні результати показали, що комбінація аналізу тексту та визначення місця розташування дала більш високу точність в порівнянні з використанням лише аналізу тексту. У [15] представлено гібридну структуру, яка включає в себе попередньо підготовлену модель DistilBERT та алгоритм вибору функцій для виявлення надзвичайних ситуацій. Ця система демонструє перевагу щодо точності ідентифікації подій і зменшення ознак порівняно з іншими. Незважаючи на наявність цих інструментів, прийняття ефективних рішень щодо техногенних катастроф є проблемою.

1.3 Підходи дослідницького аналізу даних

Дослідницький аналіз даних (EDA) — це процес візуалізації та аналізу даних для отримання з них розуміння. Іншими словами, EDA — це процес узагальнення важливих характеристик даних для кращого розуміння набору даних.

Дуже важливо глибоко зрозуміти це, перш ніж виконувати аналіз даних і запускати дані за допомогою алгоритму. Вам потрібно знати закономірності у своїх даних і визначити, які змінні є важливими, а які не відіграють суттєвої ролі в результатах. Крім того, деякі змінні можуть мати кореляції з іншими змінними. Ви також повинні розпізнавати помилки у своїх даних.

Усе це можна зробити за допомогою дослідницького аналізу даних. Це допомагає вам збирати статистичні дані та краще розуміти дані, а також усуває нерівності та непотрібні значення з даних.

- Допомагає підготувати набір даних для аналізу.
- Дозволяє моделі машинного навчання краще прогнозувати наш набір даних.
- Дає більш точні результати.
- Це також допомагає нам вибрати кращу модель машинного навчання.

Етапи дослідницького аналізу даних можуть відрізнятися в залежності від конкретної задачі дослідження та методів, які використовуються для аналізу. Однак, загальні етапи дослідницького аналізу даних можуть включати наступне (Рис.1.3):

Рисунок 1.3 - Етапи дослідницького аналізу даних

Збір даних - це перший етап дослідження, на якому збираються необхідні дані для подальшого аналізу. Цей етап включає в себе збір

інформації з різних джерел, таких як літературні джерела, статистичні дані, інтерв'ю з експертами, опитування та спостереження.

В залежності від типу дослідження, збір даних може вимагати різних методів та інструментів. Наприклад, у соціологічних дослідженнях для збору даних можуть використовуватися опитування, анкети, фокус-групи, а в медичних дослідженнях можуть проводитися клінічні випробування.

Для збору даних можна також використовувати різні інструменти, такі як апарати вимірювання, датчики, програмне забезпечення для аналізу даних, тощо.

Важливо, щоб збір даних був проведений з урахуванням етичних та правових вимог, таких як дотримання конфіденційності, згода учасників дослідження на участь та забезпечення безпеки під час збору даних.

Збір даних є важливим етапом дослідницького процесу, оскільки від якості та достовірності даних залежить подальший аналіз та отримання висновків.

Після збору даних наступним етапом є обробка даних. Обробка даних - це процес перетворення сирого матеріалу даних в структуровану форму, що дозволяє їх аналізу та отриманню корисної інформації.

Першим кроком в обробці даних є їх очистка. Це включає в себе видалення аномальних значень, помилок, дублювання, відсутніх значень та інших неточностей. Очищені дані потім можуть бути структуровані із використанням різних методів, таких як фільтрація, сортування, групування, агрегування та індексування.

Далі настає етап інтерпретації та перетворення даних. Це включає в себе створення нових змінних на основі наявних даних, їх перетворення з одного формату в інший, а також створення зведених таблиць та графіків для подальшого аналізу. Наприклад, можна використовувати математичні та статистичні методи для створення нових змінних, таких як середнє значення, медіана, дисперсія та інші.

Останній етап - аналіз даних. На цьому етапі дослідники використовують методи статистики, машинного навчання та інші методи для отримання корисної інформації з даних. Аналіз може бути проведений на основі створених зведених таблиць та графіків, або на основі математичних моделей. Результатом аналізу даних може бути отримання нових знань, виявлення закономірностей, залежностей та трендів, що можуть бути використані для прийняття рішень.

Візуалізація даних може допомогти зрозуміти взаємозв'язки та тенденції між даними, а також показати структуру і розподіл даних. Вона може допомогти знайти залежності, які можуть бути невидимими при простому аналізі даних. Крім того, візуалізація даних може допомогти зробити дані більш доступними та зрозумілими для широкої аудиторії.

При візуалізації даних, використовуються різні типи діаграм, графіків та карт, залежно від характеристик даних та завдання, яке необхідно вирішити. Наприклад, лінійні графіки можуть використовуватися для відображення залежностей між двома змінними, діаграми розсіювання можуть показати взаємозв'язки між декількома змінними, а гістограми можуть відображати розподіл даних.

Після візуалізації даних, можна зробити висновки та прийняти рішення на основі цих даних. Візуалізація даних є важливою складовою в аналітиці даних та допомагає зробити дані зрозумілі та доступні для використання.

Статистичний аналіз є одним з ключових етапів дослідницького аналізу даних, що включає в себе розрахунок числових параметрів та побудову графіків та діаграм для аналізу структури даних та виявлення закономірностей.

Основні задачі статистичного аналізу полягають у:

1. Описі структури даних: розрахунок середніх значень, дисперсій, медіан, коефіцієнтів асиметрії та ексцесу, квартилів, інтерквартильного діапазону тощо.

2. Побудові графічних та діаграматичних зображень: діаграм розсіювання, гістограм, полігонів частот, діаграм ящиків з вусами тощо.
3. Виявленні взаємозв'язків та залежностей між різними змінними: кореляційний аналіз, регресійний аналіз, аналіз дисперсії та інші методи.
4. Перевірці статистичних гіпотез: оцінка значимості різниць між групами, аналіз ANOVA, t-тесту та інших методів.

Після проведення статистичного аналізу отримані результати можуть бути використані для формулювання висновків та розробки подальших рекомендацій.

Інтерпретація результатів - це останній етап дослідження, який полягає у зрозумінні та поясненні знайдених під час аналізу даних висновків та результатів.

На цьому етапі дослідник розуміє та аналізує значення знайдених результатів, визначає, наскільки вони відповідають поставленим гіпотезам та цілям дослідження. Для цього важливо не тільки зрозуміти самі результати, але й з’ясувати, що вони означають для вивчуваної проблеми, як вони можуть бути використані та які можуть бути наслідки їх використання.

Інтерпретація результатів може бути складною, оскільки деякі результати можуть бути багатозначними або незрозумілими. Тому важливо використовувати додаткові дослідницькі методи та джерела інформації для більш точної та повної інтерпретації результатів.

Крім того, важливим аспектом інтерпретації результатів є їх візуалізація та подання відповідної інформації за допомогою графіків, таблиць, діаграм та інших візуальних засобів. Це допоможе дослідникові та іншим зацікавленим особам краще зрозуміти та оцінити результати дослідження.

Підготовка звіту є важливим етапом дослідницького аналізу даних, оскільки вона дозволяє оформити та передати отримані результати в зручному для сприйняття форматі.

1.4 Вибір перспективного шляху і постановка задачі дослідження

Тема автоматизованого виявлення техногенних катастроф та швидкого реагування на них є дуже актуальною в наш час, оскільки з розвитком промисловості і технологій зростає ймовірність виникнення таких подій. Техногенні катастрофи можуть мати серйозні наслідки для життя і здоров'я людей, довкілля і економіки. Швидке та ефективне виявлення та реагування на такі події може зменшити їх наслідки та зберегти життя.

Застосування штучного інтелекту, машинного навчання та обробки природних мов у розробці систем виявлення техногенних катастроф та вибору ефективних стратегій реагування може значно поліпшити швидкість та точність реагування на такі події. Такі системи можуть бути особливо корисними для екстрених служб, що займаються ліквідацією наслідків техногенних катастроф, а також для організацій, що відповідають за безпеку на підприємствах. Отже, тема є дуже актуальною і має великий потенціал у забезпеченні безпеки людей та ефективного вирішення наслідків техногенних катастроф.

У зв’язку з цим можна вважати, що можна вважати, що розробка такого модуля може мати значний вплив на підвищення безпеки на промислових об'єктах та запобігання можливим техногенним катастрофам. Оцінка ситуації з використанням зібраних даних та інформації допоможе швидко та ефективно приймати рішення в разі надзвичайних ситуацій на промислових об'єктах та попереджати можливі аварії, зберігаючи при цьому життя людей та уникнення екологічної катастрофи. Такий модуль може бути корисним не тільки для промислових підприємств, але і для інших організацій, де можливі надзвичайні ситуації.

Метою дипломної роботи є розробка модуля оцінювання аварій на промислових об'єктах на основі опису ситуації присутніми. Для досягнення мети потрібно вирішення такі завдань:

1. провести дослідження особливостей проведення оцінювання аварій на промислових об'єктах та аналіз існуючих підходів.
2. провести аналіз існуючих підходів до оцінювання аварій на промислових об'єктах з метою визначення їх переваг та недоліків.
3. розробити концептуальну модель оцінки наслідків техногенних катастроф та створення модуля оцінювання аварій на промислових об'єктах на основі опису ситуації присутніми.
4. розробити архітектуру мобільного додатку та інформаційного забезпечення.
5. провести дослідницький аналіз даних з використанням методів аналізу тексту та класифікації на основі текстових даних ансамблевим навчанням.
6. розробити модуль оцінювання аварій на промислових об'єктах на основі опису ситуації присутніми.

Завдання 1-2 розглянуто в параграфах 1.1 - 1.4 відповідно, а виконання завдань 3-6 представлені у параграфах 2.1-3.4.

2 АЛГОРИТМІЧНЕ ТА ІНФОРМАЦІЙНЕ ЗАБЕЗПЕЧЕННЯ

2.1 Концептуальна модель оцінки наслідків техногенних катастроф

Концептуальна модель оцінки наслідків техногенних катастроф передбачає використання комп'ютерних технологій та аналізу даних для забезпечення більш точної та оперативної оцінки наслідків катастроф.

Концептуальна модель оцінки наслідків техногенних катастроф може бути такою (див.дод.А):

1. Збір даних: збирання різноманітних даних про техногенну катастрофу, такі як місце, час, причина, тип техніки, що викликала катастрофу, та інші фактори, які можуть вплинути на наслідки катастрофи.
2. Аналіз даних: аналіз зібраних даних, щоб визначити масштаб техногенної катастрофи та наслідки, які вона може мати на довкілля, людей та майно. Для цього можуть використовуватись різні методи аналізу даних, включаючи статистичний аналіз, машинне навчання та інші методи.
3. Моделювання наслідків: розроблення моделі, яка допоможе прогнозувати наслідки техногенної катастрофи, використовуючи зібрані дані та аналіз. Ця модель може включати різні складові, такі як модель поширення забруднення в повітрі, воді та грунті, модель оцінки впливу на здоров'я людей та тварин, модель оцінки екологічних наслідків та інші.
4. Оцінка наслідків: використання моделі для оцінки наслідків техногенної катастрофи на довкілля, людей та майно. Оцінка може включати різні показники, такі як ризик для здоров'я людей, вартість збитків, екологічну шкоду та інші.
5. Визначення параметрів: далі визначаються параметри, які впливають на наслідки катастрофи. Це можуть бути такі параметри, як тип

виробництва, кількість людей та обладнання на місці події, метеорологічні умови, відстань до населених пункті.

6. Розроблення рекомендацій: розроблення рекомендацій для зменшення наслідків техногенної катастрофи в майбутньому, таких як рекомендації щодо поліпшення безпеки на робочому місці, встановлення систем попередження про можливі аварії та швидкий реагування на виявлення проблем, аналіз попередніх аварій та проведення відповідних заходів для попередження подібних ситуацій в майбутньому та інше.

2.2 Архітектура мобільного додатку

Для реалізації запропонованого інтелектуального методу розроблено мобільний додаток (див.дод.Б).

Додаток для мобільних пристроїв розроблено для свідків аварій та може входити до складу системи "розумного міста". Під час першого етапу свідок проходить коротке опитування, в якому необхідно вказати сферу праці (промисловість, працівник або третя сторона), опис критичних ризиків (Блок 1). Крім того, додаток автоматично визначає дані про країну, дату та місце (Блок 2).

Потім усі зібрані дані надсилаються до бази даних, що використовується для навчання системи. Після цього дані розділяються (Блок 3) для класифікації потенційних аварійних ситуацій. Класифікація проводиться як на основі текстових даних, так і на основі кількісних даних. Опрацювання відбувається паралельно.

Процес класифікації даних на основі текстової інформації складається з трьох етапів. Спочатку здійснюється очищення даних (Блок 4), яке включає в себе вилучення зайвих символів, заміну різноманітних абревіатур та інших операцій, що сприяють зменшенню шуму в даних. Наступним етапом є перетворення тексту (Блок 5), під час якого з тексту створюється числовий

вектор, що відображає внутрішній зміст даних. На останньому етапі проводиться класифікація (Блок 6) і визначається категорія, до якої належить текст. Результати класифікації виводяться на екран (Блок 9).

При класифікації даних на основі кількісних показників (Блок 7), спочатку значення індексуються, що дозволяє відобразити структуру даних та відносини між ними. Після індексації проводиться класифікація (Блок 8) і визначається рівень потенційної небезпеки. Результати також виводяться на екран (Блок 9).

Додаток може також надавати користувачу інформацію про застережні заходи та рекомендації щодо поведінки відповідно до рівня потенційної небезпеки, який визначається (Блок 9). Це може допомогти зменшити ризики та підвищити безпеку відповідно до виявлених потенційних небезпек.

2.3 Методи аналізу тексту

Методи аналізу тексту - це різноманітні методи та техніки, які використовуються для виявлення, опису та розуміння текстів. Вони дозволяють отримувати інформацію з текстових даних, зокрема з документів, електронних листів, повідомлень в соціальних мережах, новин тощо.

Деякі з найбільш поширених методів аналізу тексту:

1. Тематичний аналіз: цей метод дозволяє визначити теми, які обговорюються в тексті, та виділити ключові слова, які пов'язані з цими темами.
2. Сентимент-аналіз: цей метод використовується для визначення настрою тексту, чи то позитивного, чи негативного.
3. Кластерний аналіз: цей метод використовується для групування текстів за їхніми спільними ознаками, наприклад, за темою чи жанром.
4. Аналіз ключових слів: цей метод дозволяє виявляти найбільш важливі слова в тексті та їх зв'язки між собою.

5. Лексичний аналіз: цей метод використовується для вивчення лексичної структури тексту, тобто для виявлення найчастіших слів, фраз та їх співвідношення.

Ці методи дозволяють аналізувати велику кількість текстових даних та використовувати їх для прийняття рішень, вивчення трендів та поведінки користувачів, а також для багатьох інших цілей.

Завдяки розвитку комп'ютерних технологій та машинного навчання, методи аналізу тексту стали все більш ефективними та широко використовуються у різних галузях, таких як природні мови, маркетинг, фінанси, медицина, суспільствознавство тощо.

До методів аналізу тексту відносяться наступні:

- Токенизація - розділення тексту на окремі слова (токени).
- Лематизація - перетворення слів на їх базову форму (лему).
- Стемінг - видалення закінчень слів з метою зменшення кількості унікальних слів у тексті.
- Векторизація - перетворення тексту на числову матрицю, що дозволяє використовувати машинне навчання для аналізу тексту.
- Тематичне моделювання - виявлення тем та їх зв'язків у тексті.
- Кластерний аналіз - групування текстів за схожістю за певними ознаками.
- Аналіз емоційного відтінку - визначення позитивного, негативного або нейтрального відтінку тексту.
- Розпізнавання іменованих сутностей - виявлення та класифікація різних типів іменованих сутностей, таких як імена людей, місця, організації та інше.
- Розуміння природної мови - відтворення відносин між словами, розуміння контексту та виведення логічних висновків з тексту.
- Машинне перекладання - автоматичний переклад тексту з однієї мови на іншу.

Для поставленої задачі проведено класифікацію текстових даних ансамблевим навчанням складається з наступних кроків:

1. Підготовка даних: необхідно зібрати та підготувати текстові дані для подальшої обробки. Це може включати в себе очистку даних від зайвої інформації (наприклад, знаків пунктуації та стоп-слів), токенізацію (розділення тексту на окремі слова), та векторизацію (перетворення тексту у числовий вектор).
2. Вибір алгоритмів: для ансамблевої класифікації необхідно вибрати кілька алгоритмів класифікації. Можуть бути використані різні алгоритми машинного навчання, такі як наївний Баєс, метод опорних векторів, рішучі дерева та інші.
3. Навчання моделей: для кожного вибраного алгоритму необхідно навчити модель на підготовлених даних. Це може зайняти деякий час, залежно від розміру та складності даних та використаного алгоритму.
4. Об'єднання моделей: після того, як всі моделі навчені, необхідно їх об'єднати в ансамбль. Це може бути зроблено шляхом голосування, де кожна модель має право на один голос, або шляхом усереднення ваг моделей.
5. Оцінка результатів: останній крок - оцінка точності та ефективності класифікації ансамблю. Для цього можна використовувати метрики, такі як точність (precision), відгук (recall) та F1-оцінка.

Ці методи застосовуються в різних областях, таких як машинне навчання, обробка природної мови, аналіз соціальних мереж, маркетинг, лінгвістика та інші. Використання методів аналізу тексту дозволяє зробити текстові дані більш доступними для подальшої обробки та зробити з них корисну інформацію.

2.4 Методи класифікації текстових даних ансамблевим навчанням

Методи класифікації текстових даних ансамблевим навчанням використовуються для покращення точності класифікації шляхом поєднання декількох моделей. Ансамбль моделей може бути створений за допомогою різних методів, таких як бустинг, багатокласова класифікація та стекінг.

У бустингу створюється послідовність моделей, кожна з яких намагається виправити помилки попередньої моделі. Це дозволяє покращити точність класифікації з кожною наступною моделлю.

У багатокласовій класифікації створюється набір моделей, кожна з яких відповідає за класифікацію окремого класу. Кожен зразок тексту потім може бути призначений до кількох класів, якщо він підходить для декількох категорій.

У стекінгу кілька моделей різних типів об'єднуються в єдиний ансамбль. Кожна модель може бути використана для створення нових функцій, які потім використовуються в іншій моделі.

Ансамблеві методи класифікації текстових даних дозволяють зменшити ризик перенавчання моделі та покращити точність класифікації. Вони є корисним інструментом для аналізу великих обсягів текстових даних, таких як соціальні медіа або новинні статті.

Далі проведемо опис основних методів класифікації текстових даних ансамблевим навчанням.

AdaBoost. AdaBoost - це скорочення від Adaptive Boosting, що означає адаптивний бустинг. Цей метод дозволяє перетворити слабкі навчальні алгоритми на сильні рішення для задач класифікації. Остаточна модель для класифікації може бути представлена у наступному вигляді:

$$F(x) = \sum_{t=1}^{T} f_t(x),$$

де кожен f_t є слабким учнем, який бере предмет x як вхідні дані і повертає значення, що вказує клас об'єкта. Наприклад, у двокласовій задачі знак слабкого результату для учнів визначає передбачуваний клас об’єктів, а абсолютне значення дає впевненість у цій класифікації. Подібним чином із T , якщо цей класифікатор є позитивним, то вибірка знаходиться в позитивному класі, а в іншому випадку - негативна.

Кожен слабкий учень формує вихідну гіпотезу $\left(h(x_i)\right)$ для кожного зразка в навчальному наборі. На кожній ітерації t, слабкий учень вибирається і йому присвоюється коефіцієнт (α_t), що дозволяє зменшити суму похибки навчання (E_t) на Т-му класифікаторі. Після цього ваги навчальних зразків оновлюються з урахуванням нової інформації, яку надає доданий на цій ітерації слабкий учень. Кожна наступна ітерація зосереджується на навчанні нового слабкого учня на підмножині даних, для яких попередній класифікатор здійснив найбільші помилки. Нарешті, остаточний класифікатор формується шляхом комбінування результатів всіх слабких учнів з урахуванням їх ваг і коефіцієнтів (α_t), отриманих на кожній ітерації.

$$E_t = \sum_i E\left[F_{(t-1)}(x_i) + \alpha_t h(x_i)\right],$$

де $F_{(t-1)}(x_i)$ - представляє попередній посилений класифікатор, що був побудований до поточної ітерації навчання. Функція помилки E(F) визначає, наскільки точним є класифікатор F на даних. Слабкий учень $f_t(x)$ представлений у вигляді $\alpha_t h(x)$, де h(x) є базовим класифікатором, а α_t - це коефіцієнт, який визначає вагу учня $f_t(x)$ у фінальному класифікаторі.

Кожну ітерацію тренувального процесу в методі AdaBoost відображається вага $\left(\omega_{(i,t)}\right)$, яка призначається кожному зразку в навчальному наборі. Ця вага дорівнює поточній помилці $\left(E\left(F_{(t-1)}(x_i)\right)\right)$ на цьому зразку, де $F_{(t-1)}$ - це класифікатор, що був отриманий на попередній ітерації. Таким

чином, зразки, на яких $F_{(t-1)}$ робив більші помилки, отримують більшу вагу. Ці ваги можуть бути використані для навчання слабкого учня, наприклад, можуть бути використані для побудови дерев рішень, які сприяють розбиттю навчального набору на підмножини зразків з більшою вагою.

Після того, як на кожній ітерації був навчений новий слабкий учень, ваги кожного зразка в навчальному наборі оновлюються. Вага кожного зразка збільшується, якщо цей зразок був неправильно класифікований, та зменшується, якщо класифікація була правильною. При цьому, зразки, які були правильно класифіковані на попередніх ітераціях, мають меншу вагу на наступних ітераціях, тоді як вага зразків, які були неправильно класифіковані, зростає, що забезпечує зосередження уваги на зразках, які були класифіковані неправильно. Остаточний класифікатор формується шляхом комбінування результатів всіх слабких учнів з урахуванням їх ваг і коефіцієнтів, отриманих на кожній ітерації.

Gradient Boost. Метод Gradient Boost (Gradient Boosting Machines, GBM) навчає моделі послідовно з метою отримання більш точної оцінки змінної відповіді. Основна ідея цього алгоритму полягає у побудові нових базових моделей навчання, які максимально корелюють з негативним градієнтом функції втрат, що пов'язано з усім ансамблем. Використовувані функції втрат можуть бути будь-якими, але якщо функція помилки є класичною квадратичною функцією помилки, процес навчання буде призводити до послідовної підгонки помилок. Вибір функції втрат залежить від дослідника та конкретної задачі. Ця гнучкість дозволяє легко настроювати GBM для будь-якої конкретної задачі, пов'язаної з даними.

Однак, контрольоване навчання має свої обмеження, оскільки дані повинні мати достатній набір відповідних цільових міток, а вилучення цих міток може бути дуже дорогим, наприклад, через дорогі експерименти. У наявності даних $(x, y)Ni = 1, \text{де}\, x = (x1, \ldots, xd)$ є пояснювальними вхідними змінними, а у є відповідними мітками змінної відповіді, метою є

реконструкція невідомої функціональної залежності $\left(x \xrightarrow{f} y\right)$ з нашою оцінкою $f\widehat{\left((x)\right)}$, так що деяка визначена функція втрат Ψ(y, f) мінімізується:

$$f(\widehat{x}) = y,$$
$$f(\widehat{x}) = \underset{f}{argmin}\,(x)\,\psi\left(y, f(x)\right)$$

Таким чином, спочатку ініціалізується початкова модель $f\,\hat{0}$ з фіксованою константою, після чого на кожній ітерації вираховується негативний градієнт $g_t(x)$, який використовується для побудови нової базової моделі навчання $h(x, \theta_t)$. Далі знаходиться оптимальний крок градієнтного спуску, щоб зменшити функцію втрат. Цей процес повторюється на кожній ітерації до досягнення заданої кількості ітерацій M або до досягнення заданої точності.

$$\rho_t = argmin_\rho \sum_{i=1}^{N} \Psi\left[y_i, \hat{f}_{t-1}(x_i) + \rho h(x_i, \theta_t)\right]$$

І на останньому кроці потрібно провести оновлення оцінки функції:

$$\hat{f}_t \leftarrow \hat{f}_{t-1} + \rho_t h(x, \theta_t)$$

XGBoost. XGBoost - це алгоритм машинного навчання, який базується на дереві пошуку рішень та використовує градієнтний бустинг. Одним з головних переваг XGBoost є його здатність до паралелізації, що досягається шляхом зміни порядку циклів побудови дерев. В результаті, обчислення розподіляються між потоками, що прискорює роботу алгоритму.

Для навчання XGBoost необхідно мати навчальний набір (x, y)Ni = 1, де y - цільова змінна, а також вибрати диференційовану функцію втрат, ряд навчання та параметр швидкості навчання α. Модель ініціалізується з постійним значенням, після чого виконується сортування, що використовує

паралельні потоки, та проводиться навчання з використанням градієнтного бустингу.

$$f(\widehat{x}) = argmin_{\theta} \sum_{i=1}^{N} L(y_i, \theta)$$

Визначаємо «градієнти» та «гессиани» для $m = 1$ до M:

$$\hat{g}_m(x_i) = \left[\frac{\partial^2 L(y_i, f(x_i))}{\partial f(x_i)}\right]_{f(x)=\hat{f}_{(m-1)}(x)}$$

$$\hat{h}_m(x_i) = \left[\frac{\partial^2 L(y_i, f(x_i))}{\partial f(x_i)^2}\right]_{f(x)=\hat{f}_{(m-1)}(x)}$$

Потім здійснюється вибір слабкого учня з використанням навчального набору.

$$\left\{x_i - \frac{\hat{g}_m(x_i)}{\hat{h}_m(x_i)}\right\}_{i=1}^{N}$$

розв'язавши задачу оптимізації:

$$\hat{\phi}_m = argmin_{\phi \in \Phi} \sum_{i=1}^{N} \frac{1}{2} \hat{h}_m(x_i) \left[-\frac{\hat{g}_m(x_i)}{\hat{h}_m(x_i)} - \phi(x_i)\right]^2$$

$$\hat{f}_m(x) = \alpha \hat{\phi}_m(x)$$

На заключному етапі необхідно оновити оцінку функції.

$$\hat{f}_t \leftarrow \hat{f}_{t-1} + \hat{f}_m(x)$$

CatBoost. CatBoost [19] - це бібліотека, яка забезпечує ефективну реалізацію градієнтного бустингу над деревами рішень. Однією з особливостей CatBoost є можливість використовувати бінарні дерева рішень як базові предиктори. Нехай, є дані зі зразками $D = \{(X_j, y_j)\}_{j=1,\dots,m}$, де $X_j = (x^1_j, x^2_j, \dots, x^n_j)$ - вектор ознак n, які можуть бути двійковими або числовими, і відповіді y_j, які також можуть бути двійковими або числовими. Кожен зразок (X_j, y_j) є незалежним та ідентично розподіленим за деяким

розподілом $p(\cdot,\cdot)$. Метою навчального завдання є тренування функції $H: R^n \rightarrow R$, яка мінімізує очікувані втрати, наведені в

$$L(H) := EL(y, H(X))$$

де $L(\cdot,\cdot)$ — функція плавних втрат і (X, y) — тестові дані, відібрані на основі навчальних даних D. Алгоритм підвищення градієнта є ітеративним процесом, який використовує функцію гладких втрат L(·, ·) та тестові дані (X, y), що були відібрані з навчальних даних D. На кожній ітерації будується послідовність наближень $H: R^m \rightarrow R$, де $t = 0,1,\ldots,$ і кожне наступне наближення H^t отримується в адитивному процесі, який має вигляд $H^t = H^{t-1} + \alpha g^t$. Тут α - розмір кроку, а $g^t: R^n \rightarrow R$ - базовий предиктор, який обирається з набору функцій G з метою мінімізації очікуваних втрат, що визначаються як:

$$g^t = arg\,min_{g \in G} L(H^{t-1} + g) = arg\,min_{g \in G} EL(y, H^{t-1}(X) + g(X))$$

Часто для вирішення задачі мінімізації використовують метод Ньютона, який полягає в апроксимації другого порядку функції втрат L(H^(t-1) + g^(t)) вздовж H^(t-1), або виконують негативний крок градієнта. Один з варіантів для виконання цих функцій - градієнтний спуск.

LightGBM Classifier. LightGBM [20, 21], - це відкрите програмне забезпечення для здійснення градієнтного бустингу. Градієнтний бустинг дерев рішень (GBDT) є дуже популярним алгоритмом класифікації. Навчальний набір містить зразки даних $\{(x_1, y_1), (x_2, y_2), L, [\![(x]\!]_n, y_n)\}$, де x - це вхі д н і дані, а y - мітки класів. Оцінювальна функція представлена як F(x), а оптимізація GBDT полягає у мінімізації втрат $L(y, F(x))$:

$$\hat{F} = \arg\min_F E_{x,y}[L(y, F(x))]$$

Ітераційний критерій GBDT - це процес послідовної побудови дерев рішень, в якому кожне дерево намагається скоригувати помилки, які були зроблені попереднім деревом. Після будь-якої ітерації ми можемо використовувати лінійний пошук, щоб знайти найкращий шлях у вихідному просторі параметрів, який мінімізує функцію втрат. Цей процес відбувається ітеративно до тих пір, поки не буде досягнуто зазначену кількість дерев або досягнуто певного рівня точності. Кожне нове дерево робить внесок у функцію втрат, що покращує точність класифікації. Лінійний пошук на кожній ітерації забезпечує знаходження найкращого шляху для покращення класифікації.

$$F_m(x) = F_{m-1}(x) + \gamma_m h_m(x)$$

де

$$\gamma_m = arg\min_{\gamma} \sum_{i=1}^{n} L(y_i, F_{m-1}(x_i) + \gamma h_m(x_i))$$

, m - число ітерації, $h_m(x)$ представляє базове дерево рішень.

Однак, коли кількість зразків велика або розмір характеристик високий, GBDT все ще має недостатню ефективність та точність. Оскільки GBDT є ансамблевим алгоритмом, базовим класифікатором якого є дерево рішень, основна вартість полягає у пошуку кращих точок розщеплення при побудові дерев. Щоб покращити ефективність, Ke et al [27] запропонував використовувати високоефективне дерево рішень під назвою LightGBM, яке використовує GOSS та EFB для підвищення градієнту.

Замість використання інформаційного приросту для розділення кожного вузла, LightGBM використовує GOSS для вибору точки розщеплення на основі розрахунку коефіцієнта посилення дисперсії. Спочатку, сортується абсолютне значення градієнтів навчальних зразків у порядку спадання, та вибирається

верхня a×100% вибірка даних градієнтів, яка позначається як A. З нерозподілених зразків A^c випадковим чином вибирається підмножина B розміром $b \times |A^c|$. Нарешті, зразки розбиваються за допомогою розрахункової дисперсії на A ∪ B.°

$$V_j\,(d) = \frac{1}{n}\left(\frac{\left(\sum_{x_i \in A_i} g_i + \frac{1-a}{b}\sum_{x_i \in B_i} g_i\right)^2}{n_i^j(d)} + \frac{\left(\sum_{x_i \in A_i} g_i + \frac{1-a}{b}\sum_{x_i \in B_i} g_i\right)^2}{n_r^j(d)}\right)$$

де

$$A_i = \{x_i \in A : x_{ij} \le d\}, A_r = \{x_i \in A : x_{ij} > d\}, B_i = \{x_i \in B : x_{ij} \le d\}, B_r = \{x_i \in B : x_{ij} > d\}$$

g_i представляє негативний градієнт функції втрат, $\frac{1-a}{b}$ використовується для нормалізації суми градієнтів.

Так як об'ємні функції зазвичай є розрідженими та ексклюзивними, то можна використовувати EFB для прискорення процесу навчання GBDT. LightGBM може об'єднувати ексклюзивні функції в одну, а алгоритм сканування функцій може бути застосований для створення гістограм з пакетів функцій. Таким чином, обчислювальна складність LightGBM становить O(data * bundle) замість O(data * feature), де bundle * feature.

LightGBM є ефективним виконанням GBDT з GOSS та EFB, що дозволяє підвищити обчислювальну ефективність без зниження точності. GOSS використовується для розбиття вузлів дерева за допомогою розрахунку коефіцієнта дисперсії, тоді як EFB прискорює процес навчання GBDT, об'єднавши багато унікальних функцій з меншою кількістю густо розташованих функцій. LightGBM - це ансамблева модель, яка базується на дереві рішень, яке може бути навчено послідовно, підганяючи негативні градієнти функції втрат. Модель LightGBM $F_M(x)$ може бути отримана за допомогою взваженої комбінації схем.

3 ПРОГРАМНО-ТЕХНОЛОГІЧНЕ ЗАБЕЗПЕЧЕННЯ

3.1 Інформаційне забезпечення

Для проведення розробленого інтелектуального методу оцінки наслідків техногенних катастроф обрано мову Python. У якості вхідних даних використано записи про аварії на 12 різних підприємствах у 3 різних країнах, кожен рядок у яких містить інформацію про аварію [23] (табл.3.1). Кожен рядок у записах містить інформацію про аварію, включаючи дату та час аварії, місце виникнення, тип промисловості, причини та наслідки аварії, відомості про кількість потерпілих, рівень катастрофи та потенційно можливий рівень катастрофи.

Далі дані про аварії були оброблені та підготовлені підходом Exploratory Data Analysis [30, 31] для використання в розробленому методі. Для цього були виконані наступні дії:

— Зчитування даних з файлу та перетворення у формат, придатний для подальшої обробки.

— Очищення та нормалізація даних, включаючи видалення непотрібної інформації та стандартизацію формату записів.

— Створення бази даних для зберігання оброблених даних та побудова відповідного інтерфейсу для доступу до цих даних.

Для проведення інтелектуального методу було використано бібліотеки машинного навчання, такі як scikit-learn, pandas та numpy. Після попередньої обробки даних, які включали в себе зведення та очищення даних, а також інженерію ознак, було побудовано модель машинного навчання, яка використовувала ансамблеві методи класифікації.

Для оцінки точності класифікації було використано крос-валідацію, що дозволило отримати точність моделі на методах. Після навчання та перевірки

моделі на валідаційній вибірці, вона була застосована для оцінки наслідків техногенних катастроф на інших підприємствах та в інших регіонах.

Таблиця 3.1 – Опис даних

№	Параметр	Опис	Non-Null, Count, D type
1	Data	timestamp or time/date information	411, non-null, object
2	Countries	which country the accident occurred (anonymized)	411, non-null, object
3	Local	the city where the manufacturing plant is located (anonymized)	411, non-null, object
4	Industry sector	which sector the plant belongs to	411, non-null, object
5	Accident level	from I to VI, it registers how severe was the accident (I means not severe but VI means very severe)	411, non-null, object
6	Potential Accident Level	Depending on the Accident Level, the database also registers how severe the accident could have been (due to other factors involved in the accident)	411, non-null, object
7	Genre	if the person is male of female	411, non-null, object
8	Employee or Third Party	if the injured person is an employee or a third party	411, non-null, object
9	Critical Risk	some description of the risk involved in the accident	411, non-null, object
10	Description	Detailed description of how the accident happened	411, non-null, object

3.2 Дослідницький аналіз даних

Для більш глибокого розуміння даних проведемо аналіз даних на етапі дослідження [30, 31], на початковій стадії проведемо візуальну оцінку показників. За результатами аналізу даних було встановлено, що гірничодобувна промисловість має значно більшу кількість аварій порівняно з металургійною промисловістю (див. рис. 3.2а). Також відзначено, що в обох

галузях промисловості переважає чоловіча стать серед працівників (див. рис. 3.2b). Додатково, було встановлено, що набір даних має певне упередження, оскільки працівники чоловічої статі були представлені значно більшою кількістю порівняно з працівницями. Населення прямих та сторонніх співробітників було приблизно однаковим (див. рис. 3.2c), проте кількість віддалених співробітників третіх сторін була меншою.

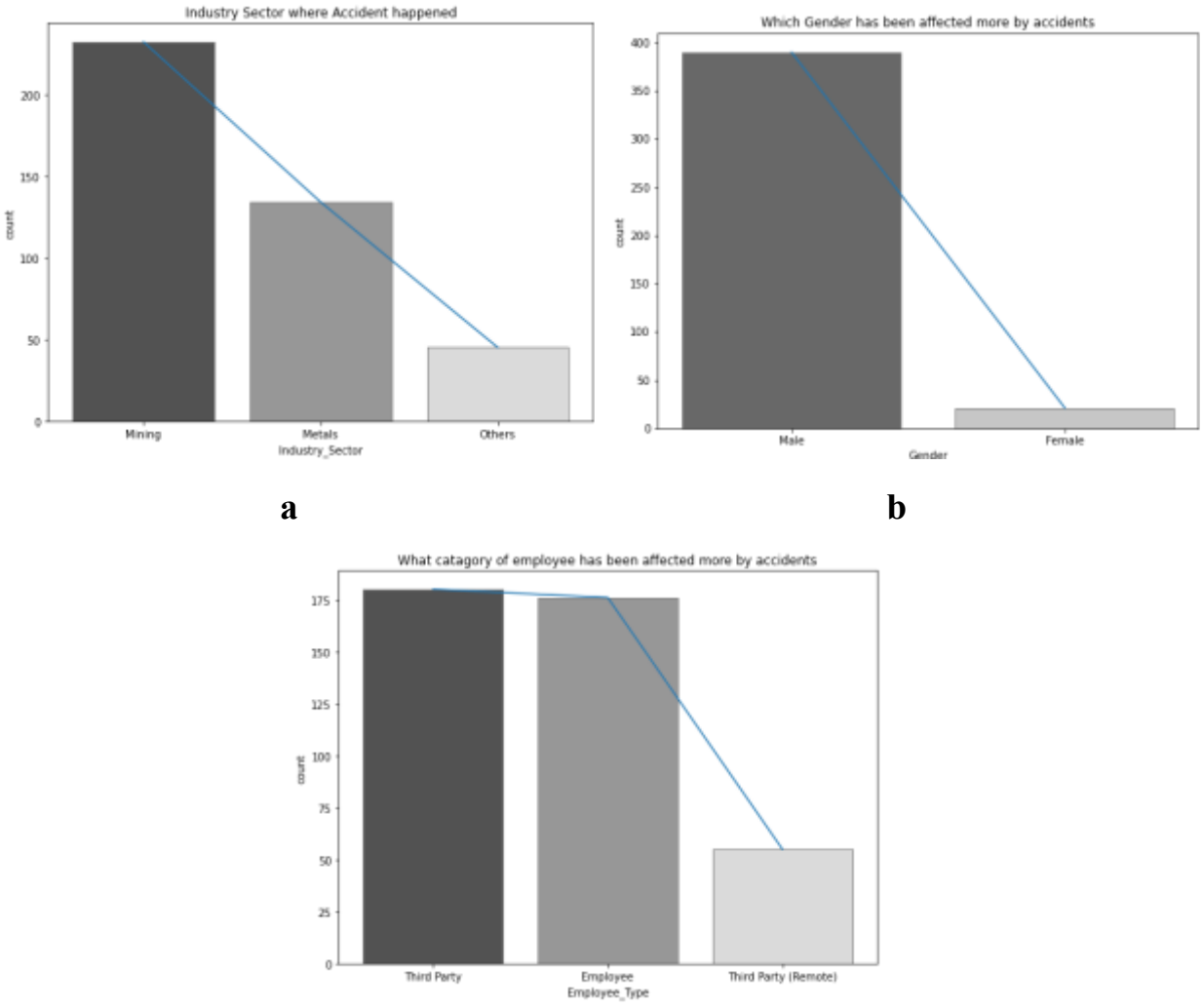


Рисунок 3.2 - Кількісна оцінка показників

На підприємстві Local_03 відбулася найбільша кількість аварій порівняно з іншими заводами країни, що становить приблизно 21% від усіх аварій. Найбільш поширеним типом критичних ризиків є класифікація "Інші", яка охоплює майже 50% від усього набору даних. Це вказує на важливість

детальної класифікації численних ризиків для кращого розуміння та управління їхніми наслідками.

За рівнем ризику наступні в списку є пресовані, ручні інструменти, хімічні речовини, різання та інші види дій. Щодо рівнів небезпеки, I рівень означає легку небезпеку, а V рівень відображає високу серйозність. Найчастіше відбуваються аварії з рівнем небезпеки I, які зазвичай пов'язані з дрібними помилками, такими як забування використовувати засоби індивідуального захисту або випадкове випущення інструменту з рук.

Потенційний рівень аварії вказує на можливу серйозність аварії, яка може виникнути через додаткові фактори, пов'язані з конкретною ситуацією. Рівень потенційної аварії IV є найбільш поширеним та відповідає помірній важкості аварій. Врахування рівнів небезпеки та потенційних аварій допомагає зрозуміти та краще керувати ризиками, а також розробляти стратегії з метою зниження аварійності та підвищення рівня безпеки на виробництві.

Для детальнішої оцінки важливо дослідити взаємозв'язок між потенційним рівнем аварійності та загальним рівнем аварійності (рис. 3.3), а також взаємозв'язок з промисловим сектором (рис. 3.4). Рисунок 3.3 демонструє значну відмінність між серйозністю інцидентів та їх потенційною серйозністю.

З рисунку 3.4 можна спостерігати, що зі зростанням кількості аварій спостерігається зменшення потенційного рівня аварійності. Водночас, із збільшенням рівня аварійності відбувається зростання потенційного рівня аварійності. Найбільш поширеними є аварії 1-го рівня. Вивчення цих взаємозв'язків може допомогти у розробці стратегій забезпечення безпеки в різних секторах промисловості та зниження аварійності в цілому.

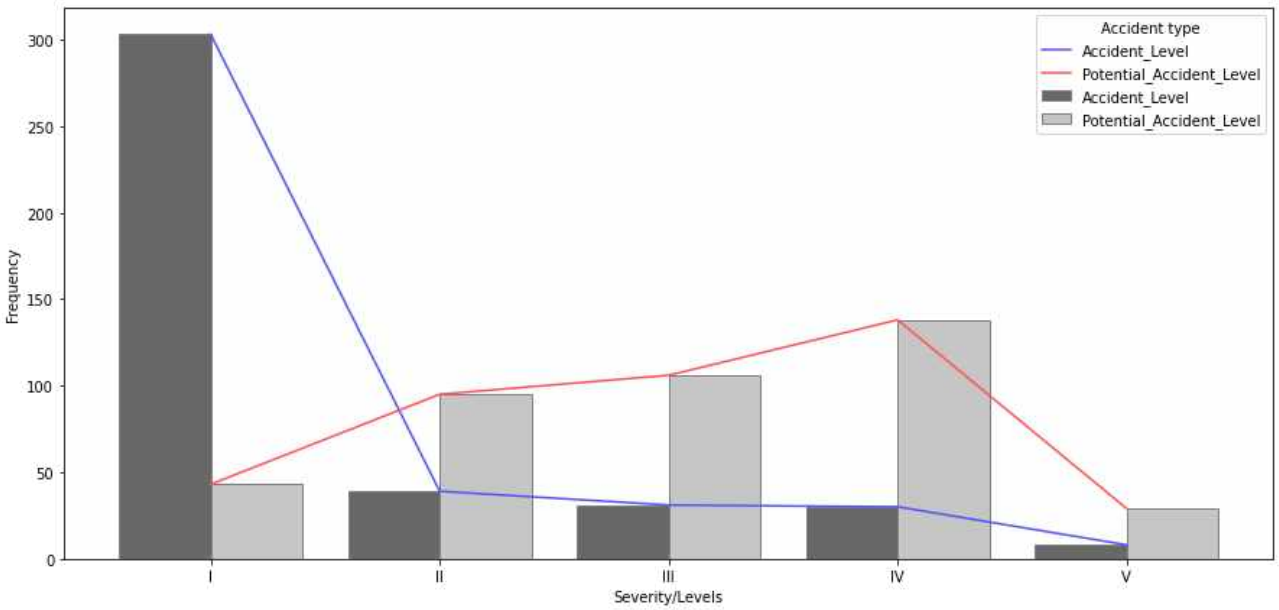
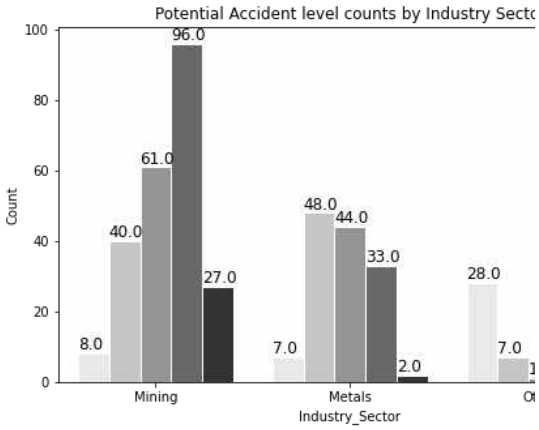
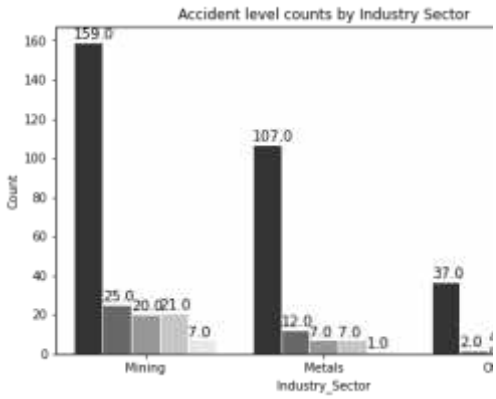


Рисунок 3.3 - Співвідношення між рівнем потенційної аварії та явним рівнем аварії

У гірничодобувній промисловості відбулися аварії з найсерйознішим рівнем (див. рис.3.4a,b) серед всіх галузей, а також мають найвищий потенційний рівень аварійності. Наступними за серйозністю йдуть металургія та інші галузі. В гірничодобувному секторі рівні серйозності інцидентів (рис.3.4b) виявляються вищими, зокрема рівень 4 є трохи вищим, ніж рівні 2 і 3.



a



b

Рисунок 3.4 - Співвідношення між рівнем потенційної, явним рівнем аварії та сектором промисловості

З аналізу кореляційної матриці (див. рис.3.5) можна зрозуміти, що між значеннями двох змінних спостерігається помірна кореляція. Найсильніша зв'язок спостерігається між II та III рівнями потенційної небезпеки та I рівнем очевидної небезпеки. Відповідно, 88 та 89 випадків відбулися при такій комбінації рівнів небезпеки.

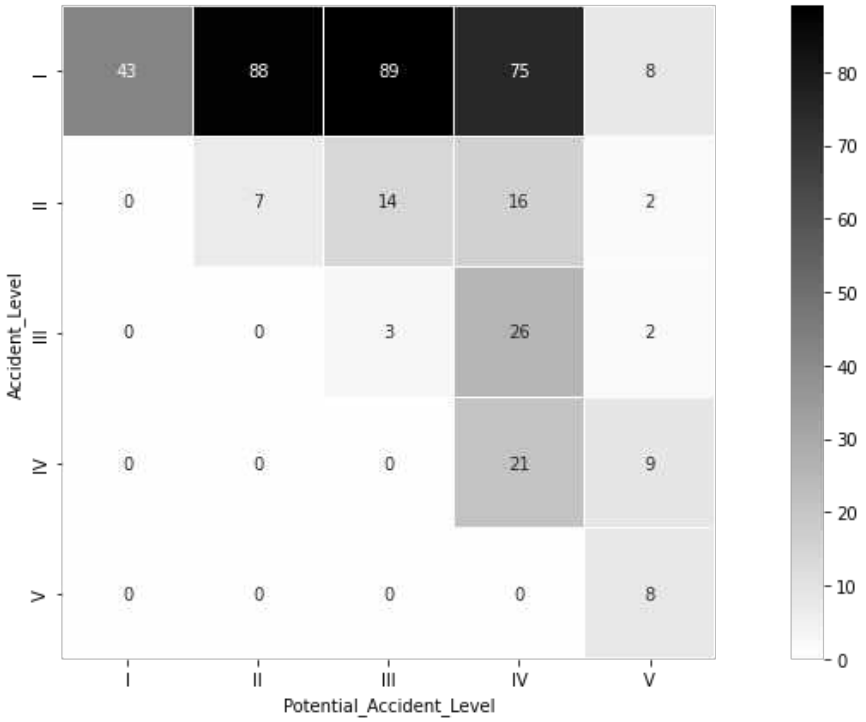


Рисунок 3.5 - Кореляція між рівнем потенційної аварії та явним рівнем аварії

Вивчення взаємозв'язків між рівнями аварійності та потенційної аварійності допомагає розробляти стратегії безпеки в різних секторах промисловості. Найчастіше відбуваються аварії з небезпекою I, пов'язані з дрібними помилками. Потенційний рівень аварії IV є найбільш поширеним, відповідаючи помірній важкості аварій.

3.3 Класифікація на основі текстових даних

Після попередньої обробки даних переходимо до аналізу текстових записів про аварійні ситуації, що дозволить класифікувати рівень техногенної катастрофи за описами очевидців.

На першому етапі здійснимо очистку тексту: видалимо стоп слова, проведемо нормалізацію, лематизацію та токенізацію тексту. Результати такої обробки ілюструються на рисунку 3.6. Використання цих методів полегшує розуміння контексту та підвищує точність класифікації рівня техногенних катастроф.

```
'activation sodium sulphide pump, piping uncoupled sulfide solution designed area reach maid.
immediately made use emergency shower directed ambulatory doctor later hospital. note: sulphid
e solution gram liter.'
```

Рисунок 3.6 - Очищений текст

У наступному етапі було визначено N-грами стосовно категорій промисловості (див. табл. 3.2). Виявлено численні слова, асоційовані з руками, рухами, працівниками або їхніми посадами, що підкреслює їх важливість у контексті промислових аварій.

Таблиця 3.2. N-grams відносно категорій промисловості

Index	Words_ Metals	Count_ Metals	Words_ Mining	Count_ Mining	Index	Words_ Metals	Count_ Metals	Words_ Mining	Count_ Mining
0	employee	50	causing	103	**15**	worker	14	accident	47
1	left	46	right	100	**16**	performed	13	collaborator	44
2	causing	43	operator	96	**17**	mr.	13	safety	44
3	right	37	time	96	**18**	center	13	mesh	44
4	hit	27	left	92	**19**	hose	12	work	43
5	hand	25	hand	88	**20**	area	12	hit	42
6	operator	25	moment	62	**21**	face	12	employee	38
7	activity	25	level	60	**22**	remove	12	one	37
8	medical	24	assistant	59	**23**	sheet	12	fall	36
9	report	23	worker	53	**24**	cut	12	circumstance	35
10	finger	19	support	51	**25**	reaching	12	height	35
11	moment	18	rock	49	**26**	pipe	11	floor	34
12	one	16	pipe	48	**27**	fall	11	injured	32
13	collaborator	15	equipment	47	**28**	acid	11	used	32
14	cleaning	15	finger	47	**29**	contact	11	mr.	31

Далі ми перетворюємо текст на частини мови за допомогою тегів: NN - іменник; NNP - власний іменник; NNS - іменник множини; CD - кількісний

числівник; DT - детермінант; VB - дієслово; JJ - прикметник; RB - прислівник; VBD - дієслово минулого часу. У більшості пояснень найбільш поширеними є іменники, їх кількість становить 35.

Наступним кроком є застосування методів класифікації на основі машинного навчання з використанням техніки SMOTE [24]. Спочатку розділимо дані на навчальну та тестову вибірки, потім перетворимо вибірки у масиви та проведемо передискретизацію навчального набору даних за допомогою SMOTE (див. рис. 3.7). Це допомагає збільшити точність моделі класифікації.

```
array([[   0,    0,    0, ...,   83,  337,  298],
       [   0,    0,    0, ...,  207, 1492, 1493],
       [   0,    0,    0, ...,   15,    2,    8],
       ...,
       [   0,    0,    0, ...,   81,   60,   12],
       [   0,    0,    0, ...,  275, 1151,  909],
       [   0,    0,    0, ...,  260,  383,  412]], dtype=int32)
```

Рисунок 3.7 - Результат передискретизації набору навчальних даних за допомогою SMOTE

При використанні ансамблевих методів машинного навчання з бустинговим підходом та параметрами n_estimators = 50, learning_rate = 0.05 для швидкого моделювання, були отримані такі результати (див. табл. 3): AdaBoost - 0.59; Gradient Boost - 0.92; XGBoost - 0.90; CatBoost – 0.79; LGBoost – 0.98. LGBoost показав найкращий результат у швидкісному оцінюванні, тому на наступному етапі буде застосовано цей підхід для моделювання на тестових даних.

Таблиця 3.3. - Результати моделюванні ансамблевими методами машинного навчання

	model	accuracy
0	AdaBoost	0.595376
1	Gradient Boost	0.924855
2	XGBoost	0.901734
3	CatBoost	0.791908
4	LGBoost	0.988439

Після моделювання на тестовій вибірці було сформовано звіт у вигляді матриці невідповідностей (Confusion Matrix) [25] (див. рис. 3.8). Оглядаючи цю матрицю, можна зробити висновок, що модель успішно класифікує більшість тестових точок даних відповідно до реальних значень. З 173 правильно визначених значень, 34 значення відносяться до рівня загрози I, тоді як рівень небезпеки V було правильно класифіковано 33 рази.

Точність класифікації виявилася кращою для потенційних рівнів небезпеки V та I, з показниками 0,97 та 0,94 відповідно. Середня точність отриманих результатів становить 0,68, що свідчить про достатню якість моделі.

**********Classification Report**********

	precision	recall	f1-score	support
I	0.97	0.81	0.88	42
II	0.59	0.59	0.59	34
III	0.62	0.55	0.58	38
IV	0.29	0.45	0.35	22
V	0.94	0.89	0.92	37
accuracy			0.68	173
macro avg	0.68	0.66	0.66	173
weighted avg	0.73	0.68	0.70	173

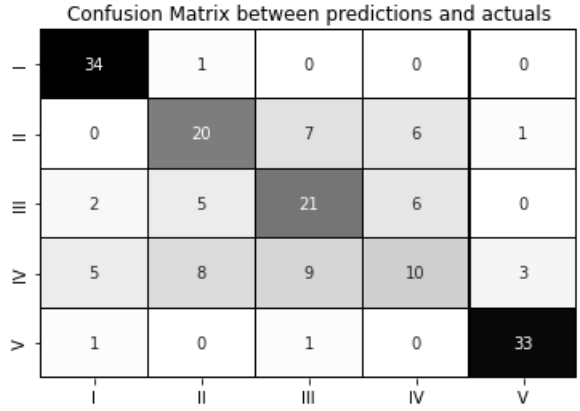


Рисунок 3.8 - Confusion Matrix between predictions and actuals

Отже, зараз спробуємо прогнозувати рівень небезпеки, враховуючи введений текст з тестової вибірки і значеннями: Country_02, Local_02, Mining, Accident level I, Potential Accident Level V, Male, Employee, Others, а також текст:

«When performing cleaning with LHD in block F 9970 at level 420, the operator was surprised by a rock block displacement of the side of the gallery, reaching his right leg causing him superficial injury»

На рисунку 3.9 можна спостерігати, що система коректно передбачила рівень потенційної небезпеки. Це свідчить про надійність та ефективність моделі в роботі з тестовими даними.

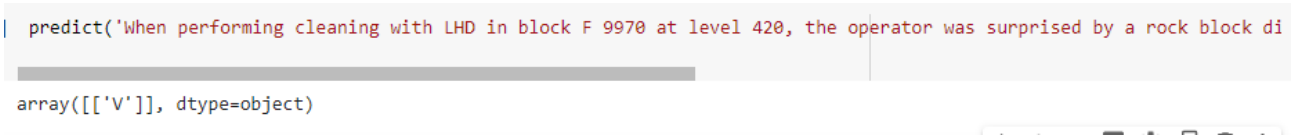
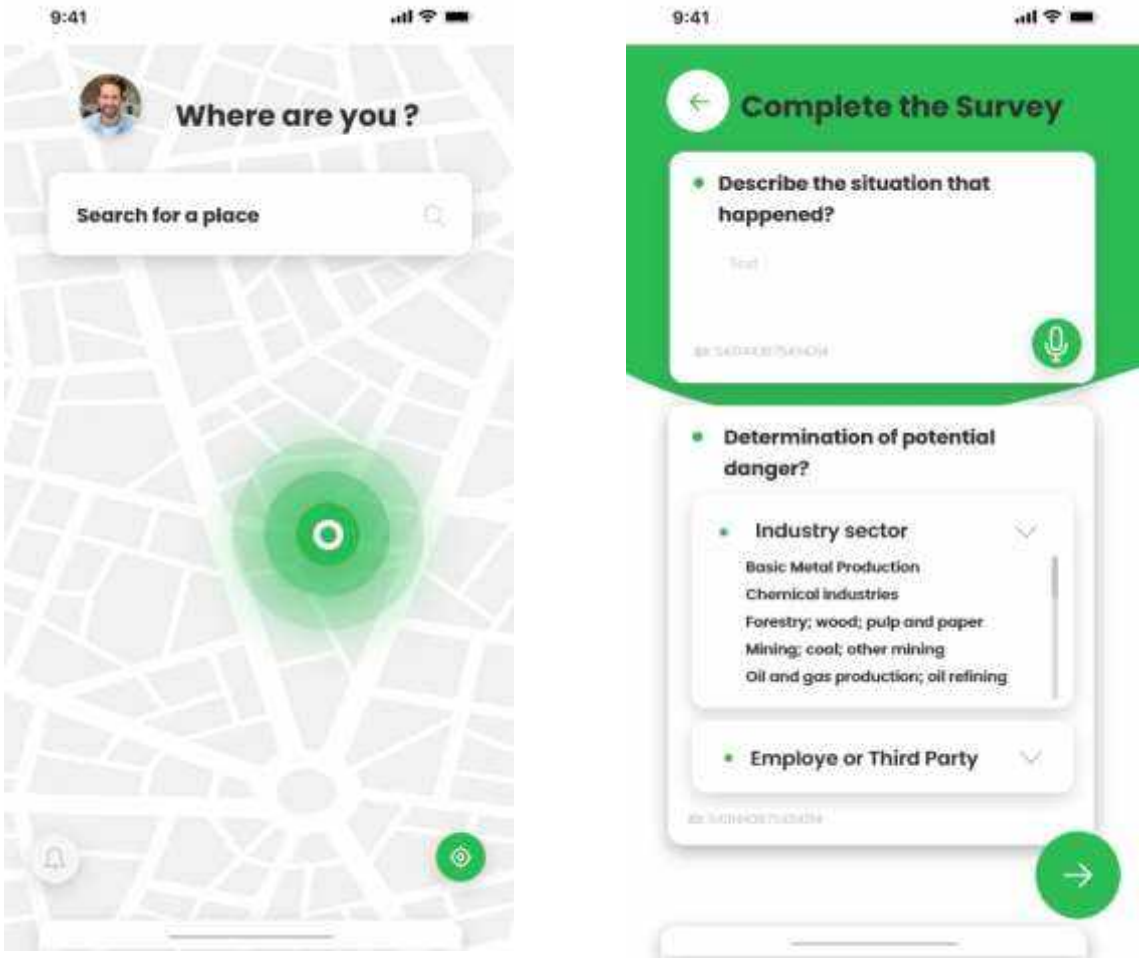


Рисунок 3.9 - Результат передбачення

3.4 Модуль оцінювання аварій на промислових об'єктах на основі опису ситуації присутніми

На рисунку 3.10 зображено процес введення вхідних даних, а на рисунку 3.11 - відображення отриманих результатів. Спочатку система визначає геолокацію користувача (див. рис. 3.10 a), що допомагає встановити місце техногенної катастрофи. Наступним кроком користувач надає голосовий опис ситуації, вибирає відповідний сектор промисловості, де мала місце аварія, та вказує своє ставлення до ситуації (як співробітник або стороння особа) (див. рис. 3.10 b).



a b

Рисунок 3.10 - Введення вхідних даних

В цілому, додаток дозволяє автоматизувати процес визначення потенційного рівня небезпеки техногенних катастроф на основі аналізу текстових даних та кількісних показників. Зокрема, за допомогою застосування методів машинного навчання та класифікації тексту, можна ефективно визначити категорію небезпеки на основі опису ситуації. Однак, важливо розуміти, що додаток не може замінити професійних експертів та спеціалістів, які мають досвід та знання для визначення рівня небезпеки в конкретній ситуації. Також, можуть існувати інші фактори, які не враховуються у додатку, і які можуть впливати на рівень небезпеки в конкретних обставинах.

На підставі отриманого результату (рис. 3.11) формується відомості про рівень потенційної небезпеки. Це допомагає службам надзвичайних ситуацій

швидше орієнтуватися на місці події та підготуватися відповідно ще під час дороги до місця аварії. Такий підхід забезпечує зменшення часових витрат та покращення оперативності реагування на надзвичайні ситуації.

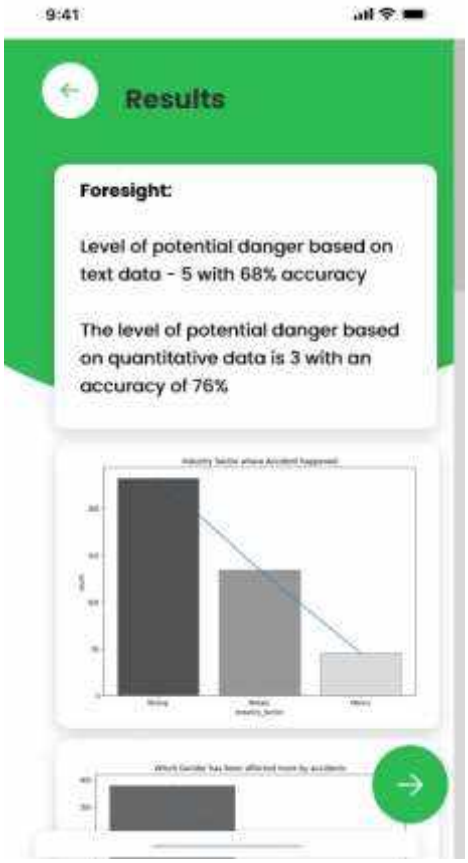


Рисунок 3.11 - Вивід отриманих результатів

ВИСНОВКИ

У ході даної теми роботи було досліджено особливості проведення оцінювання аварій на промислових об'єктах та розглянуто різноманітні типи аварій, що можуть виникнути на таких об'єктах. Було проаналізовано існуючі підходи та підходи дослідницького аналізу даних, а також розроблено концептуальну модель оцінки наслідків техногенних катастроф.

Розглянуті типи аварій демонструють різноманітність ситуацій, які можуть виникнути на промислових об'єктах, тому важливо розуміти їхні наслідки та вміти приймати ефективні рішення в разі аварійної ситуації.

Дослідження існуючих підходів та підходів дослідницького аналізу даних дозволяє зрозуміти, які методи можуть бути використані для покращення оцінювання наслідків техногенних катастроф на промислових об'єктах.

Концептуальна модель оцінки наслідків техногенних катастроф та архітектура мобільного додатку можуть бути використані для створення програмних продуктів, які допоможуть вирішувати проблеми безпеки на промислових підприємствах.

Методи аналізу тексту та класифікації текстових даних ансамблевим навчанням можуть бути корисними для обробки інформації про аварійні ситуації на промислових об'єктах, що дозволить швидко та ефективно реагувати на них.

Для покращення процесу оцінювання аварій на промислових об'єктах, було розроблено архітектуру мобільного додатку та модуль оцінювання аварій на промислових об'єктах на основі опису ситуації присутніми. Для класифікації текстових даних використовувались методи аналізу тексту та методи класифікації текстових даних ансамблевим навчанням.

Під час тестування системи на тестовій вибірці було виявлено, що модель може точно класифікувати більшість тестових даних відповідно до

дійсних значень. Найточніше класифікуються значення I та V рівнів небезпеки при аварії. Загальна середня точність отриманих результатів складає 0,68.

Таким чином, система класифікації аварій на промислових об'єктах на основі текстових даних є ефективним інструментом для попередження техногенних катастроф та покращення безпеки на робочому місці. Використання такої системи дозволяє швидко та точно оцінювати рівень небезпеки при аварії, що зменшує ризики для промислових об'єктів та працівників.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Luna, S., & Pennock, M. J. (2018). Social media applications and emergency management: A literature review and research agenda. International Journal of Disaster Risk Reduction, 28, 565–577. https://doi.org/10.1016/j.ijdrr.2018.01.006

2. Li, T., Xie, N., Zeng, C., Zhou, W., Zheng, L., Jiang, Y., Yang, Y., Ha, H.-Y., Xue, W., Huang, Y., Chen, S.-C., Navlakha, J., & Iyengar, S. S. (2017). Data-Driven Techniques in Disaster Information Management. ACM Computing Surveys, 50(1), 1–45. https://doi.org/10.1145/3017678

3. Nazer, T. H., Xue, G., Ji, Y., & Liu, H. (2017). Intelligent Disaster Response via Social Media Analysis: A Survey. SIGKDD Explor. Newsl., 19(1), 46–59. https://doi.org/10.1145/3137597.3137602

4. Bhoi, A., Pujari, S.P., & Balabantaray, R.C. (2020). A deep learning-based social media text analysis framework for disaster resource management. Social Network Analysis and Mining, 10(78), 1-14. https://doi.org/10.1007/s13278-020-00692-1

5. Şahin, C., Rokne, J., & Alhajj, R. (2019). Emergency Detection and Evacuation Planning Using Social Media. In Özyer, T., Bakshi, S., & Alhajj, R. (Eds.), Social Networks and Surveillance for Society (Lecture Notes in Social Networks). Springer, Cham. https://doi.org/10.1007/978-3-319-78256-0_9

6. Gopnarayan, A., & Deshpande, S. (2020). Tweets Analysis for Disaster Management: Preparedness, Emergency Response, Impact, and Recovery. In Raj, J., Bashar, A., & Ramson, S. (Eds.), Innovative Data Communication Technologies and Application. ICIDCA 2019 (Lecture Notes on Data Engineering and Communications Technologies, vol 46). Springer, Cham, pp. 760–764. https://doi.org/10.1007/978-3-030-38040-3_87

7. Basu, M., Bit, S.D., & Ghosh, S. (2022). Utilizing microblogs for optimized real-time resource allocation in post-disaster scenarios. Social Network Analysis and Mining, 12, 15. https://doi.org/10.1007/s13278-021-00841-0

8. Madichetty, S., & M, S. (2021). A Neural-Based Approach for Detecting the Situational Information From Twitter During Disaster. IEEE Transactions on Computational Social Systems, 8(4), 870-880. https://doi.org/10.1109/TCSS.2021.3064299

9. McCreadie, R., Macdonald, C., & Ounis, I. (2016). Eaims: emergency analysis identification and management system. Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '16, pp. 1101–1104.

10. Kanojia, D., Kumar, V., & Ramamritham, K. (2016). Civique: using social media to detect urban emergencies. CoRR, abs/1610.04377.

11. Zheng, Z., Zhong, Y., Wang, J., Ma, A., & Zhang, L. (2021). Building damage assessment for rapid disaster response with a deep object-based semantic change detection framework: From natural disasters to man-made disasters. Remote Sensing of Environment, 265, 112636. Elsevier Inc.

12. Bandyopadhyay, M., & Singh, V. (2016). Development of agent based model for predicting emergency response time. Perspectives in Science, 8, 138-141. Elsevier GmbH.

13. Avvenuti, M., Cimino, M.G.C.A., Cresci, S. et al. (2016) A framework for detecting unfolding emergencies using humans as sensors. SpringerPlus, 5, 43. https://doi.org/10.1186/s40064-016-1674-y

14. Karam, E., Hussein, W., & Gharib, T.F. (2021). Integrating location and textual information for detecting affected people in a crisis. Social Network Analysis and Mining, 11, 5. Springer Nature. https://doi.org/10.1007/s13278-020-00715-x

15. Adel, H., Dahou, A., Mabrouk, A., Abd Elaziz, M., Kayed, M., El-Henawy, I. M., Alshathri, S., & Amin Ali, A. (2022). Improving Crisis Events Detection Using DistilBERT with Hunger Games Search Algorithm. Mathematics, 10(3), 447. https://doi.org/10.3390/math10030447

16. Gramyak, R., Lipyanina-Goncharenko, H., Sachenko, A., Lendyuk, T., & Zahorodnia, D. (2021). Intelligent Method of a Competitive Product Choosing based on the Emotional Feedbacks Coloring. In IntelITSIS, pp. 246-257.

17. Jin, S., Chen, S., & Xie, X. (2021, November). Property-based Test for Part-of-Speech Tagging Tool. In 2021 36th IEEE/ACM International Conference on Automated Software Engineering (ASE), pp. 1306-1311. IEEE.

18. Guo, S., Liu, Y., Chen, R. et al. Improved SMOTE Algorithm to Deal with Imbalanced Activity Classes in Smart Homes. Neural Process Lett, 50, 1503–1526 (2019). https://doi.org/10.1007/s11063-018-9940-3

19. Abdullahi, A., Raheem, L., Muhammed, M., Rabiat, O., & Ganiyu, A. (2020). Comparison of the CatBoost Classifier with other Machine Learning Methods. International Journal of Advanced Computer Science and Applications, 11(11). https://doi.org/10.14569/ijacsa.2020.0111190

20. Chen, C., Zhang, Q., Ma, Q., Yu, B. (2019). LightGBM-PPI: Predicting protein-protein interactions through LightGBM with multi-information fusion. Chemometrics and Intelligent Laboratory Systems. https://doi.org/10.1016/j.chemolab.2019.06.003

21. Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.-Y. (2017). LightGBM: a highly efficient gradient boosting decision tree. In Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17), Curran Associates Inc., Red Hook, NY, USA, 3149–3157.

22. Chicco, D., & Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. BMC Genomics, 21, 6. https://doi.org/10.1186/s12864-019-6413-7

23. Industrial Safety and Health Analytics Database. Kaggle: Your Machine Learning and Data Science Community. https://www.kaggle.com/datasets/ihmstefanini/industrial-safety-and-health-analytics-database

24. Lipianina-Honcharenko, K., Lukasevych-Krutnyk, I., Butryn-Boka, N., Sachenko, A., & Grodskyi, S. (2021). Intelligent Method for Identifying the Fraudulent Online Stores. In 2021 IEEE 8th International Conference on Problems of Infocommunications, Science and Technology (PIC S&T), pp. 218-222. doi: 10.1109/PICST54195.2021.9772195.

25. Krysovatyy, A., Lipianina-Honcharenko, H., Sachenko, S., Desyatnyuk, O., Banasik, A., & Lukasevych-Krutnyk, I. (2022). Recognizing the fictitious business entity on logistic regression base. Paper presented at the CEUR Workshop Proceedings, 3156, 218-227.

26. Classification Report — Yellowbrick v1.5 documentation. Yellowbrick: Machine Learning Visualization — Yellowbrick v1.5 documentation. https://www.scikit-yb.org/en/latest/api/classifier/classification_report.html

27. Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.Y. (2017). LightGBM: a highly efficient gradient boosting decision tree. 31st Conference Neural Information Processing Systems NIPS, pp. 3146-3154.

28. Turchenko, I., Osolinsky, O., Kochan, V., Svyatnyy, V., & Komar, M. (2009). Approach to neural-based identification of multisensor conversion characteristic. In Proceedings of the 5th IEEE International Workshop on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS'2009), pp. 27–31.

29. Wang, C., Shakhovska, N., Sachenko, A., & Komar, M. (2020). A new approach for missing data imputation in big data interface. Information Technology and Control, 49(4), 541–555.

30. Komorowski, M., Marshall, D.C., Salciccioli, J.D., & Crutain, Y. (2016). Exploratory Data Analysis. In Secondary Analysis of Electronic Health Records. Springer, Cham. https://doi.org/10.1007/978-3-319-43742-2_15

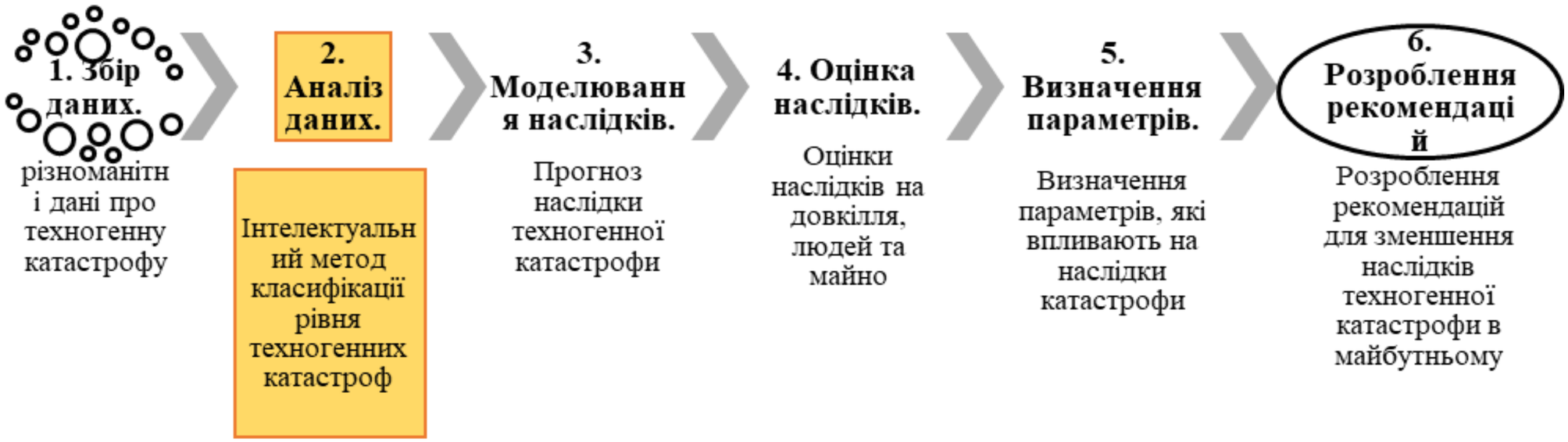
31. Vigni, M. L., Durante, C., & Cocchi, M. (2013). Exploratory data analysis. In Data handling in science and technology (Vol. 28, pp. 55-126). Elsevier.

32. Gramyak, R., Lipyanina-Goncharenko, H., Sachenko, A., Lendyuk, T., & Zahorodnia, D. (2021). Intelligent Method of a Competitive Product Choosing based on the Emotional Feedbacks Coloring. In *IntelITSIS* (pp. 246-257).

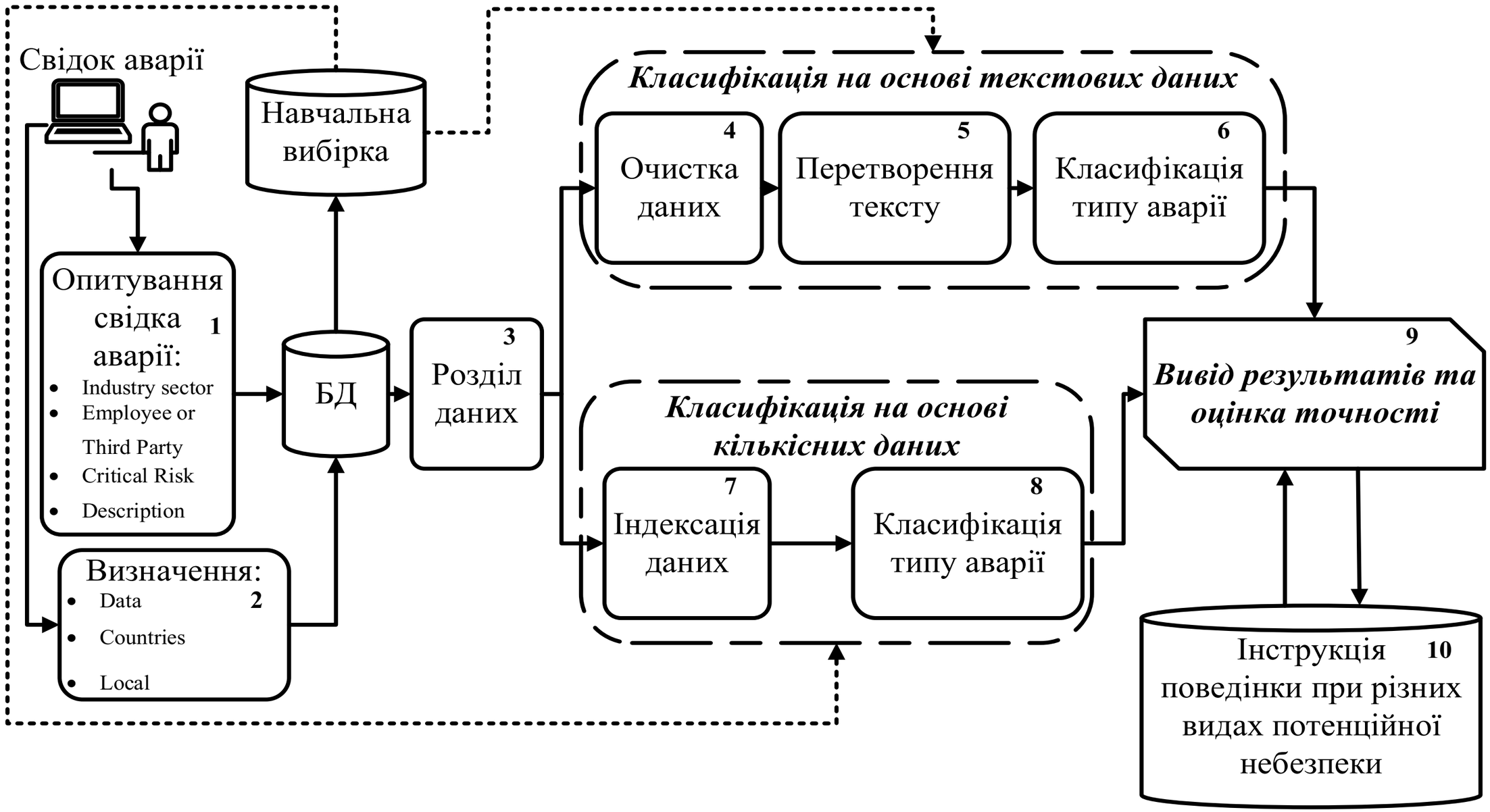
33. Lipyanina-Goncharenko, H., Brych, V., Sachenko, S., Lendyuk, T., Bykovyy, P., & Zahorodnia, D. (2021). Method of Forming a Training Sample for Segmentation of Tender Organizers on Machine Learning Basis. In *COLINS* (pp. 1843-1852).

34. Lipianina-Honcharenko, K., Wolff, C., Chyzhovska, Z., Sachenko, A., Lendiuk, T., & Grodskyi, S. (2022, October). Intelligent Method for Forming the Consumer Basket. In *Information and Software Technologies: 28th International Conference, ICIST 2022, Kaunas, Lithuania, October 13–15, 2022, Proceedings* (pp. 221-231). Cham: Springer International Publishing.

35. Комар М.П., Саченко А.О., Васильків Н.М., Гладій Г.М., Коваль В.С. Методичні рекомендації до виконання кваліфікаційної роботи з освітньопрофесійної програми «Комп'ютерні науки» спеціальності 122 «Комп'ютерні науки» за першим (бакалаврським) рівнем вищої освіти. – Тернопіль: ЗУНУ, 2021. – 56 с.



					ДП.КН.8351780.001 А1		
Змн.	Арк.	№ докум.	Підпис	Дата	Концептуальна модель оцінки наслідків техногенних катастроф	Літера	Маса / Масштаб
Розроб.		Книш Т..					
Перевір.		Ліп'яніна-Гончаренко Х.В.				Аркуш 1	Аркушів 1
Консультант							
Т. Контр.						ЗУНУ.ФКІТ.КН-41	
Н. Контр.		Ліп'яніна-Гончаренко Х.В.					
Затверд.		Комар М.П.					



					ДП.КН.8351780.001 А1			
Змн.	Арк.	№ докум.	Підпис	Дата	Архітектура мобільного додатку	Літера	Маса	Масштаб
Розроб.		Книш Т..						
Перевір.		Ліп'яніна-Гончаренко Х.В.				Аркуш 1	Аркушів 1	
Консультант								
Т. Контр.								
Н. Контр.		Ліп'яніна-Гончаренко Х.В.				ЗУНУ.ФКІТ.КН-41		
Затверд.		Комар М.П.						