

Міністерство освіти і науки України
Тернопільський національний технічний університет імені Івана Пулюя

Факультет комп'ютерно-інформаційних систем і програмної інженерії
(повна назва факультету)

Кафедра комп'ютерних наук
(повна назва кафедри)

КВАЛІФІКАЦІЙНА РОБОТА

на здобуття освітнього ступеня

бакалавр

(назва освітнього ступеня)

на тему: Аналіз інформаційних ресурсів "розумних міст" на основі методології
великих даних

Виконав: студент IV курсу, групи СН-41

спеціальності 122 Комп'ютерні науки

(шифр і назва спеціальності)

(підпис)

Вітів І.В.

(прізвище та ініціали)

Керівник

(підпис)

Дуда О.М.

(прізвище та ініціали)

Нормоконтроль

(підпис)

Марценко С.В.

(прізвище та ініціали)

Завідувач кафедри

(підпис)

Боднарчук І.О.

(прізвище та ініціали)

Рецензент

(підпис)

Гащин Н.Б.

(прізвище та ініціали)

Тернопіль
2024

Міністерство освіти і науки України
Тернопільський національний технічний університет імені Івана Пулюя

Факультет комп'ютерно-інформаційних систем і програмної інженерії

(повна назва факультету)

Кафедра комп'ютерних наук

(повна назва кафедри)

ЗАТВЕРДЖУЮ

Завідувач кафедри

Боднарчук І.О.

(підпис)

(прізвище та ініціали)

«_24_» _червня_ 2024 р.

**ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУ**

на здобуття освітнього ступеня Бакалавр

(назва освітнього ступеня)

за спеціальністю 122 Комп'ютерні науки

(шифр і назва спеціальності)

Студенту Вітіву Івану Володимировичу

(прізвище, ім'я, по батькові)

1. Тема роботи Аналіз інформаційних ресурсів "розумних міст" на основі методології великих даних

Керівник роботи Дуда Олексій Михайлович, канд. техн.наук, доцент кафедри КН

(прізвище, ім'я, по батькові, науковий ступінь, вчене звання)

Затверджені наказом ректора від «29» квітня 2024 року № 4/7-470

2. Термін подання студентом завершеної роботи 24 червня 2024р.

3. Вихідні дані до роботи Наукові публікації щодо "розумних міст", інформаційних ресурсів та методологічних засад формування та використання великих даних

4. Зміст роботи (перелік питань, які потрібно розробити)

Вступ. 1 Стан та перспективи досліджень в галузі «великих даних» «розумного міста». 1.1

Актуальність проведення аналізу ресурсів «розумного міста». 1.2 Методологія аналізу

ресурсів «розумного міста». 1.3 Опис результатів наукометричного пошуку інформаційних

ресурсів «розумних міст» на основі методології «великих даних». 2 Типи, джерела, аналіз

алгоритмів опрацювання та кластеризація великих даних «розумних міст». 2.1 Типи та

джерела «великих даних» «розумних міст». 2.2 Аналіз алгоритмів опрацювання «великих

даних» «розумних міст». 2.3 Аналіз підходів кластеризації «великих даних». 3 Аналіз

точності обробки інформаційних ресурсів «великих даних». 3.1 Використання каркасів на

основі фреймворків для аналітичного опрацювання. 3.2 Глибинні нейронні мережі при

опрацюванні інформаційних ресурсів «розумних міст». 3.3 Аналіз підходів класифікації

«великих даних». 3.4 Кореляційний аналіз інформаційних ресурсів. 3.5 Аналіз точності

методів обробки інформаційних ресурсів «розумних міст» на основі методології «великих

даних». 4. Безпека життєдіяльності, основи охорони праці. Висновки. Перелік джерел.

5. Перелік графічного матеріалу (з точним зазначенням обов'язкових креслень, слайдів)

1. Титульний слайд. 2. Тема та мета роботи. 3. Завдання роботи. 4. Актуальність роботи.

5. Практичне значення одержаних результатів. 6. Великі дані. 7. Розумне місто. 8. Розумні

міста та великі дані. 9. Етапи систематичного огляду. 10. Пошукові терміни та кількісна

оцінка. 11. Розподіл наукових публікацій. 12. Розподіл наукових публікацій за роками.

13. Методи аналізу. 14. Оцінка точності алгоритмів. 15. Висновки. 16. Завершальний слайд.

6. Консультанти розділів роботи

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		завдання видав	завдання прийняв
Безпека життєдіяльності, основи охорони праці			

7. Дата видачі завдання 29 січня 2024 р.

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів роботи	Термін виконання етапів роботи	Примітка
1.	Ознайомлення з завданням до кваліфікаційної роботи	30.01.2024	Виконано
2.	Підбір джерел про «розумні міста», інформаційні ресурси та методологічні засади опрацювання великих даних	31.01.2024-03.02.2024	Виконано
3.	Опрацювання джерел про «розумні міста», інформаційні ресурси та методологічні засади опрацювання великих даних	04.02.2024-06.02.2024	Виконано
4.	Виконання дослідження щодо аналізу інформаційних ресурсів «розумних міст» на основі методології великих даних	07.02.2024-11.02.2024	Виконано
5.	Оформлення розділу «Стан та перспективи досліджень в галузі «великих даних» «розумного міста»	03.06.2024-05.06.2024	Виконано
6.	Оформлення розділу «Типи, джерела, аналіз алгоритмів опрацювання та Кластеризація великих даних «розумних міст»	06.06.2024-08.06.2024	Виконано
6.	Оформлення розділу «Аналіз точності обробки інформаційних ресурсів «великих даних» для потреб «розумних міст»	09.06.2024-11.06.2024	Виконано
7.	Виконання завдання до підрозділу «Безпека життєдіяльності»	12.06.2024-13.06.2024	Виконано
8.	Виконання завдання до підрозділу «Основи охорони праці»	14.06.2024-15.06.2024	Виконано
9.	Оформлення кваліфікаційної роботи	16.06.2024-17.06.2024	Виконано
10.	Нормоконтроль	18.06.2024-19.06.2024	Виконано
11.	Перевірка на плагіат	20.06.2024	Виконано
12.	Попередній захист кваліфікаційної роботи	21.06.2024	Виконано
13.	Захист кваліфікаційної роботи	25.06.2024	

Студент

(підпис)

Вітів І.В.

(прізвище та ініціали)

Керівник роботи

(підпис)

Дуда О.М.

(прізвище та ініціали)

АНОТАЦІЯ

Аналіз інформаційних ресурсів "розумних міст" на основі методології великих даних // Кваліфікаційна робота освітнього рівня «Бакалавр» // Вітів Іван Володимирович // Тернопільський національний технічний університет імені Івана Пулюя, факультет комп'ютерно-інформаційних систем і програмної інженерії, кафедра комп'ютерних наук, група СН-41 // Тернопіль, 2024 // С. 58, рис. – 4, табл. – 4, кресл. – 16 , додат. – 1, бібліогр. – 37.

Ключові слова: алгоритм, великі дані, інформаційний ресурс, опрацювання даних, розумне місто, споживання ресурсів.

Кваліфікаційна робота присвячена аналізу інформаційних ресурсів «розумних міст» на основі методології великих даних. В першому розділі кваліфікаційної роботи освітнього рівня «бакалавр» висвітлено актуальність проведення аналізу ресурсів розумного міста за допомогою методології великих даних. Подана методологія аналізу ресурсів «розумного міста» на основі «великих даних». Подано опис результатів наукометричного пошуку в царині опрацювання інформаційних ресурсів «розумних міст» на основі методології «великих даних».

В другому розділі кваліфікаційної роботи проаналізовано типи та джерела великих даних «розумних міст». Розглянуто ключові категорії алгоритмів опрацювання «великих даних» «розумних міст». Окремо проаналізовано кластеризацію «великих даних» для потреб «розумних міст».

В третьому розділі кваліфікаційної роботи проаналізовано використання каркасів на основі фреймворків для аналітичного опрацювання інформаційних ресурсів «розумних міст». Також проаналізовано глибинні нейронні мережі при опрацюванні інформаційних ресурсів «розумних міст». Виконано аналіз підходів класифікації «великих даних» для потреб «розумних міст». Розглянуто кореляційний аналіз інформаційних ресурсів для потреб «розумних міст». Проведено аналіз точності методів обробки інформаційних ресурсів «розумних міст» на основі методології «великих даних».

ANNOTATION

Analysis of Information Resources of "Smart Cities" Based on Big Data Methodology // Qualification work of the educational level "Bachelor" // Vitiv Ivan Volodymyrovych // Ternopil Ivan Pulyu National Technical University, Computer and Information Systems and Software Engineering Faculty, Computer Sciences Department, group SN-41 // Ternopil, 2024 // P. 58, fig. – 4, tabl. – 4, chair. – 16, annexes. – 1, references – 37.

Keywords: algorithm, big data, information resource, data processing, smart city, resource consumption.

The qualification work is devoted to the analysis of information resources of "smart cities" based on big data methodology. In the first section of the qualifying work of the "bachelor" educational level, the relevance of analyzing the resources of a smart city using big data methodology is highlighted. The methodology for analyzing the resources of the "smart city" based on "big data" is presented. A description of the results of scientometric research in the field of processing information resources of "smart cities" based on the methodology of "big data" is presented. In the second section of the qualification work, the types and sources of big data of "smart cities" are analyzed. The key categories of algorithms for processing "big data" of "smart cities" are considered. The clustering of "big data" for the needs of "smart cities" is analyzed separately. In the third section of the qualification work, the use of frames based on frameworks for analytical processing of information resources of "smart cities" is analyzed. Deep neural networks for processing information resources of "smart cities" were also analyzed. An analysis of "big data" classification approaches for the needs of "smart cities" was performed. The correlational analysis of information resources for the needs of "smart cities" is considered. An analysis of the accuracy of methods of processing information resources of "smart cities" based on the methodology of "big data" was carried out.

ПЕРЕЛІК СКОРОЧЕНЬ І ТЕРМІНІВ

«Великі дані» (англ. Big Data) – набори структурованої та неструктурованої інформації настільки великих розмірів, що традиційні способи та підходи не можуть бути до них застосовані.

ІКТ – інформаційні та комунікаційні технології.

DNN (англ. Deep neural network) – глибока нейронна мережа.

IoT (англ. Internet of Things) – Інтернет речей.

FIESTA-IoT (англ. Federated Interoperable Semantic IoT/cloud Testbeds and Applications) – об'єднані інтероперабельні хмарні семантичні тестові стенди та IoT-застосунки.

LSTM (англ. long short-term memory) – довга короткочасна пам'ять.

QoL (англ. Quality of Life) – якість життя громадян.

WSN (англ. Wireless Sensor Network) – бездротові сенсорні мережі.

ЗМІСТ

ВСТУП	8
РОЗДІЛ 1. СТАН ТА ПЕРСПЕКТИВИ ДОСЛІДЖЕНЬ В ГАЛУЗІ «ВЕЛИКИХ ДАНИХ» «РОЗУМНОГО МІСТА»	10
1.1 Актуальність проведення аналізу ресурсів «розумного міста» за допомогою методології «великих даних»	10
1.2 Методологія аналізу ресурсів «розумного міста» на основі «великих даних»	12
1.3 Опис результатів наукометричного пошуку інформаційних ресурсів «розумних міст» на основі методології «великих даних» .	15
1.4 Висновок до першого розділу	18
РОЗДІЛ 2. ТИПИ, ДЖЕРЕЛА, АНАЛІЗ АЛГОРИТМІВ ОПРАЦЮВАННЯ ТА КЛАСТЕРИЗАЦІЯ ВЕЛИКИХ ДАНИХ «РОЗУМНИХ МІСТ»	19
2.1 Типи та джерела «великих даних» «розумних міст».....	19
2.2 Аналіз алгоритмів опрацювання «великих даних» «розумних міст»	26
2.3 Аналіз підходів кластеризації «великих даних» для потреб «розумних міст».....	28
2.4 Висновок до другого розділу	33
РОЗДІЛ 3. АНАЛІЗ ТОЧНОСТІ ОБРОБКИ ІНФОРМАЦІЙНИХ РЕСУРСІВ «ВЕЛИКИХ ДАНИХ» ДЛЯ ПОТРЕБ «РОЗУМНИХ МІСТ»	34
3.1 Використання каркасів на основі фреймворків для аналітичного опрацювання інформаційних ресурсів «розумних міст»	34
3.2 Глибинні нейронні мережі при опрацюванні інформаційних ресурсів «розумних міст»	37
3.3 Аналіз підходів класифікації «великих даних» для потреб «розумних міст».....	38
3.4 Кореляційний аналіз інформаційних ресурсів для потреб «розумних міст».....	40

3.5 Аналіз точності методів обробки інформаційних ресурсів «розумних міст» на основі методології «великих даних».....	42
3.6 Висновок до третього розділу	46
РОЗДІЛ 4. БЕЗПЕКА ЖИТТЄДІЯЛЬНОСТІ, ОСНОВИ ОХОРОНИ ПРАЦІ	47
4.1 Діяльність. Її види та розуміння в безпеці праці.....	47
4.2 Електробезпека на будівельному майданчику	50
4.3 Висновок до четвертого розділу	52
ВИСНОВКИ.....	53
ПЕРЕЛІК ДЖЕРЕЛ	54
ДОДАТКИ	

ВСТУП

Актуальність теми. У наукових та виробничих колах впродовж останнього періоду часу популярна концепція «розумного міста», яка характеризує екологічне та здорове навколишнє середовище, що покращує якість життя та добробут громадян [1]. При цьому більшість операцій контролюється за допомогою комплексних інформаційних та комунікаційних технологій [2]. Завдяки різноманітності цифрових послуг, ресурсів та інформаційно-технологічних проектів «розумні міста» керують великими за обсягом наборами та колекціями даних, як правило, у рамках концепції «Великих даних».

За останні десять років кількість датчиків і вимірювальних приладів геометрично зростає. Прагнення контролювати та розуміти все, що нас оточує, стало значним кроком у розвитку інформаційних технологій на основі «розумних» датчиків та «розумних» пристроїв для потреб «розумних будинків», «розумних громад» та «розумних міст», що формуються на основі Інтернету речей (IoT) та кіберфізичних систем. У електронних таблицях або базах даних також знаходиться історична інформація, яка може бути цінною за умови правильного доступу та інтеграції. Громадяни та установи «розумних міст» використовують соціальні мережі для передачі даних чи інформації про ресурси, цифрові послуги чи події. Основні питання полягають у тому, як використовувати ці дані та як отримати корисну, практичну та значущу інформацію з усіх вимірювань. Тому актуальним напрямком сучасних досліджень в галузі комп'ютерних наук є аналіз інформаційних ресурсів «розумних міст» на основі методології «великих даних».

Мета і задачі дослідження. Метою даної кваліфікаційної роботи освітнього рівня «Бакалавр» є виявлення кращих практик впровадження засобів аналітичного опрацювання інформаційних ресурсів «розумних міст» з використанням методології «великих даних». Для досягнення поставленої мети потрібно виконати ряд завдань, зокрема:

- Проаналізувати стан досліджень в галузі опрацювання інформаційних ресурсів «розумних міст» на основі методології «великих даних»
- Проаналізувати типи та джерела «великих даних» «розумних міст».
- Провести аналіз алгоритмів опрацювання «великих даних», що отримані з інформаційних ресурсів «розумних міст».

Практичне значення одержаних результатів. На основі систематичного огляду наукової та популярної літератури щодо опрацювання інформаційних ресурсів «розумних міст» проаналізовано застосування різних алгоритмів для обробки міських даних, що дало можливість розширити розуміння, як дані збиралися, зберігалися, попередньо оброблялися та аналізувалися. Розглянуто обширний перелік алгоритмів та категорій їх застосування. Виявлено, що часто використовуються алгоритми кластеризації, класифікації, кореляції, виявлення аномалій і прогнозування. На основі проведеного аналізу сформовано таблицю з характеристиками точності алгоритмів аналітичного опрацювання інформаційних ресурсів великих даних «розумних міст».

РОЗДІЛ 1. СТАН ТА ПЕРСПЕКТИВИ ДОСЛІДЖЕНЬ В ГАЛУЗІ «ВЕЛИКИХ ДАНИХ» «РОЗУМНОГО МІСТА»

1.1 Актуальність проведення аналізу ресурсів «розумного міста» за допомогою методології «великих даних»

Концепція «розумного міста» виникає у двох основних вимірах. Перший пов'язаний з наданням громадянам якісніших цифрових послуг для покращення комфорту, мобільності, енергоощадності та здоров'я [3]. Другий вимір пов'язаний з покращення процесів постачання та управління ресурсами. У цьому контексті «розумне» та стійке місто використовує ІКТ, щоб зменшити екологічну шкоду навколишньому середовищу міста та покращити якість життя громадян (QoL). Використання ІКТ дає змогу створювати «розумні» програми, а також робити міста стійкішими та приємнішими для життя [4]. Це виклик з проблемами, особливо коли міста поєднують:

- постійно зростаючу кількість населення;
- розширюються;
- стають густонаселенішими.

В основі цього лежить збір великої кількості різноманітних даних з множини міських джерел, локацій та областей [5].

Зі збільшенням обсягів даних розкриття закономірностей і вилучення корисної інформації стає все складнішим завданням. Ручний аналіз та інтерпретація стає неможливим або нездійсненним впродовж адекватного часу через труднощі співвіднесення обширного переліку джерел з інформацією різної природи. Аналіз такої кількості даних повинен спиратися на статистичні підходи або алгоритми, які дають можливість витягувати приховані знання та шаблони в даних. Різноманітність таких технік обробки даних доволі значна [6]:

- описова статистика;

- кореляційний аналіз;
- нейронні мережі;
- алгоритми глибокого навчання.

Ці методи стають основними інструментами для вирішення завдань опрацювання «Великих даних» для потреб «розумних громад», «розумних міст» та «розумних регіонів».

Щоб мати можливість видобувати та розуміти інформацію всередині даних, необхідно:

- розуміти їх формат;
- вибрати найкращі алгоритми для використання.

Описову статистику можна використовувати в структурованих даних, таких як електронні таблиці або бази даних, кореляційний аналіз дає можливість зробити висновок про взаємозалежність між різними стовпцями або даними, нейронні мережі дають змогу прогнозувати, оцінювати та класифікувати неструктуровані дані, такі як часові ряди, текст або звук. Алгоритми глибокого навчання можуть бути особливо корисними для класифікації зображень, інтерпретації відео тощо.

Автори роботи [7] описують дослідницький проєкт проєкт PandIA – для управління соціальною ізоляцією в умовах пандемії на основі міського та соціального інтелекту, який зосереджений на наданні детальної інформації щодо:

- тенденцій споживання ресурсів;
- оцінки людей у кожній локації, місцевості чи домогосподарстві;
- теплові карти підозрілих спалахів захворювань;

Зазначене дослідження [7] проводилось для закладів охорони здоров'я, муніципальних органів та персоналу установ невідкладної допомоги. Для цього було використано інформацію з ряду джерел, включаючи:

- характеристики патогенів;
- статистику інфекцій;
- муніципальну інформацію;

- дані соціальних мереж;
- інформацію та статистику лікарень тощо.

Робота [7] використовує систематичний огляд літератури, щоб зрозуміти характер і призначення даних, створених і зібраних у контексті «розумного міста». Дослідження має на меті зрозуміти:

- які типи даних зазвичай розглядаються;
- як дані збиралися;
- які алгоритми використовуються та для яких цілей.

В процесі аналізу інформаційних ресурсів «розумних міст» на основі методології «великих даних» потрібно розглянути найкращі практики для використання інформації про місто в налаштуваннях, впливі та результатах «великих даних» [8]. У широкому сенсі потрібно систематизувати основні принципи роботи з цифровими даними при формуванні та розвитку «розумних міст». Потрібно підсумувати проведені за останні п'ять років дослідження. Потрібно класифікувати літературу:

- відповідно до алгоритмів, які використовуються;
- підходів до обробки даних;
- характеру даних та результатів їх обробки.

1.2 Методологія аналізу ресурсів «розумного міста» на основі «великих даних»

Основною метою проведеного в кваліфікаційній роботі освітнього рівня «бакалавр» аналізу є спроби зрозуміти:

- структуру та природу даних «розумних міст»;
- джерела даних в міському середовищі;
- процеси збору міських даних;
- процеси зберігання даних;
- алгоритми та завдання, які використовуються для аналізу міських даних;

– цілі або наміри процесів опрацювання «великих даних» для потреб «розумних міст».

При цьому будемо дотримуватися запропонованого авторами [9] та [10] підходу, що включає три фази (див. рисунок 1.1):

- планування;
- функціонування;
- розповсюдження.

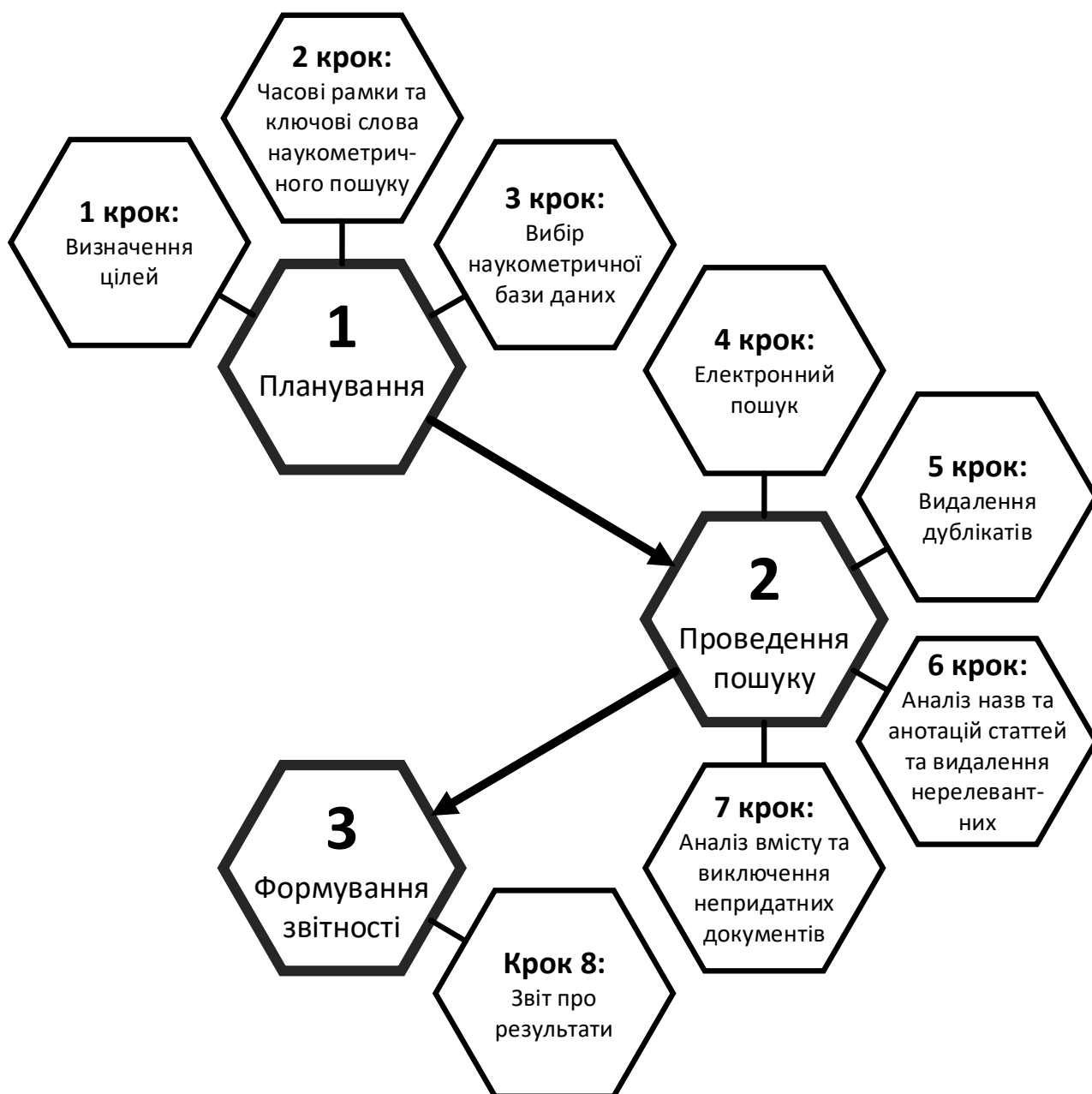


Рисунок 1.1 – Етапи систематичного огляду та аналізу літератури щодо ресурсів «розумного міста» на основі «великих даних»

Для формалізації процесів наукометричного пошуку в [7] було розглянуто ряд дослідницьких питань:

1. Який формат, структура та контекст міських даних в кожній статті?
2. Які алгоритми використовуються для опрацювання міських даних?
3. Яка складність кожного алгоритму аналітичного опрацювання?
4. Які результати отримані після аналізу, що вони означають та як трактуються?

Проводитимемо пошук в наукометричних базах даних Scopus та IEEEExplore. Ці бази даних було обрано, оскільки вони надають широкий набір предметних областей та ключові терміни, забезпечуючи при цьому початкову увагу до основної мети аналізу. На першому кроці було ідентифіковано понад двісті наукових публікацій (див. таблицю 1.1).

Таблиця 1.1 – Пошукові терміни та кількісна оцінка результатів наукометричного пошуку

Наукометрична база даних	Пошукові терміни	Кількість результатів
Scopus	Заголовок («big data» та («urban data» або «smart city» або «geo data» або «social network» або «predictive maintenance» або «algorithms»))	124
IEEEExplore	((«Document Title»: «big data») та («Document Title»:«urban data» або «Document Title»:«smart city» або «Document Title»:«geo data» або «Document Title»:«social network» або «Document Title»:«predictive maintenance» або «Document Title»: «algorithms»))	109
Всього		233

Розглядатимемо лише документи, отримані з наукометричних баз даних Scopus та IEEEExplore, котрі були опубліковані до кінця грудня 2023 року. При цьому текст повинен бути доступний у відкритих інституційних репозиторіях. При проведенні аналізу виключатимемо статті, що не проходили рецензування та написані відмінною від англійської мовою. Після видалення дублікатів загальна кількість документів склала двісті сім.

На наступному кроці визначатимемо рекомендації щодо аналізу назв, анотацій та текстів наукових публікацій. Після первинної оцінки проводитимемо аналіз відповідності обраної публікації темі кваліфікаційної роботи освітнього рівня «бакалавр». При цьому допоможе змістовність назви та пов'язаність з початковим описом. Далі текст буде переглянутий, щоб оцінити, чи можна знайти необхідну інформацію. Таким чином, наукові публікації та документи потрібно буде проаналізувати відповідно до таких факторів:

1. Обмежимо пошук наукових та популярних статей попередньо визначеним періодом часу – десять років.

2. Будемо виключати публікації з нерелевантними назвами, анотаціями та ключовими словами.

3. Виключимо роботи з текстовими матеріалами в яких не згадуються теми, що відповідають галузі «великих даних» для потреб «розумних міст».

Після аналізу назв, анотації та тексту наукових та популярних публікацій будемо виключати та, що не відповідають зазначеним критеріям пошуку.

1.3 Опис результатів наукометричного пошуку інформаційних ресурсів «розумних міст» на основі методології «великих даних»

В [7] описано процес аналізу розпочався з характеристики відібраних джерел. Всього для аналітичного опрацювання залишилося двадцять робіт (див. таблицю 1.2). Потім проводитиметься аналіз змісту наукових та популярних публікацій, щоб оцінити контекст, визначення вмісту та мети предметній

області аналітичного опрацювання «великих даних» для потреб «розумних міст».

Таблиця 1.2 – Результати пошуку наукових публікацій щодо «великих даних» для потреб «розумних міст» [7]

Етап	Scopus	IEEEExplore	Total
Пошук	124	109	233
Видалення дублікатів	115	95	210
Оцінювання назви	43	37	80
Аналіз анотації	17	19	36
Оцінка змісту	11	9	20

Загалом в [7] було знайдено документи з понад п'ятнадцяти різних країн. Розподіл за країнами подано на рисунку 1.2.

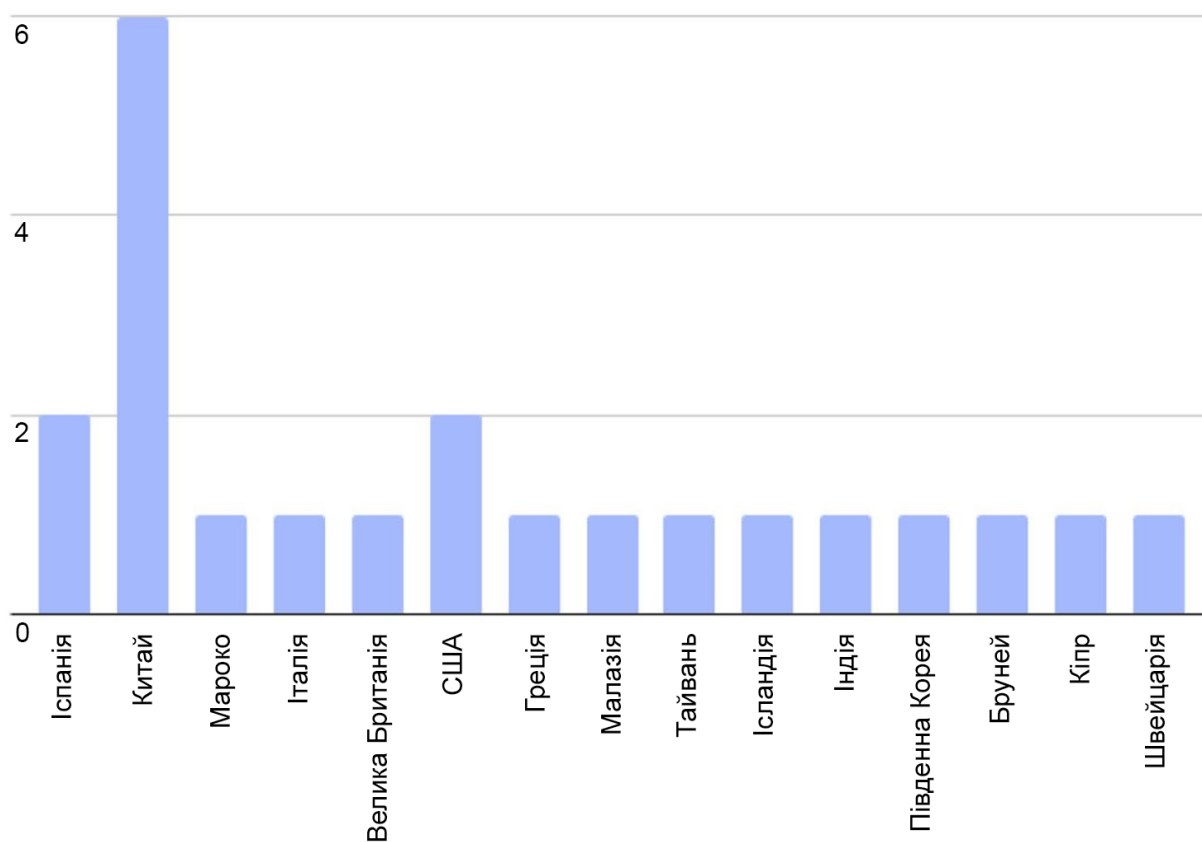


Рисунок 1.2 – Розподіл наукових публікацій щодо «великих даних» для потреб «розумних міст» за країнами [7]

У США та Іспанії опубліковано по дві статті, а також одна стаття у кожній з країн Англія, Бруней, Греція, Індія, Ісландія, Італія, Кіпр, Малайзія, Марокко, Південна Корея, Тайвань і Швейцарія. В Китаї опубліковано найбільше статей щодо інформаційних ресурсів «розумних міст» на основі методології «великих даних».

До розгляду будемо відбирати статті, що опубліковані після 2010 року. Поданий в [7] розподіл показує, що найбільша за роками кількість статей щодо інформаційних ресурсів "розумних міст" на основі методології великих даних була у 2018 році. Тоді було опубліковано вісім статей. Це свідчить про зростання інтересу до цього питання. Хоча в подальшому відбувся спад публікаційної активності, зокрема у 2019 і 2021 роках є знайдено лише по одному документу і жодних документів у 2020 році. Проте за кількістю публікацій впродовж останніх двох років можна спостерігати зростання уваги (див. рисунок 1.3).

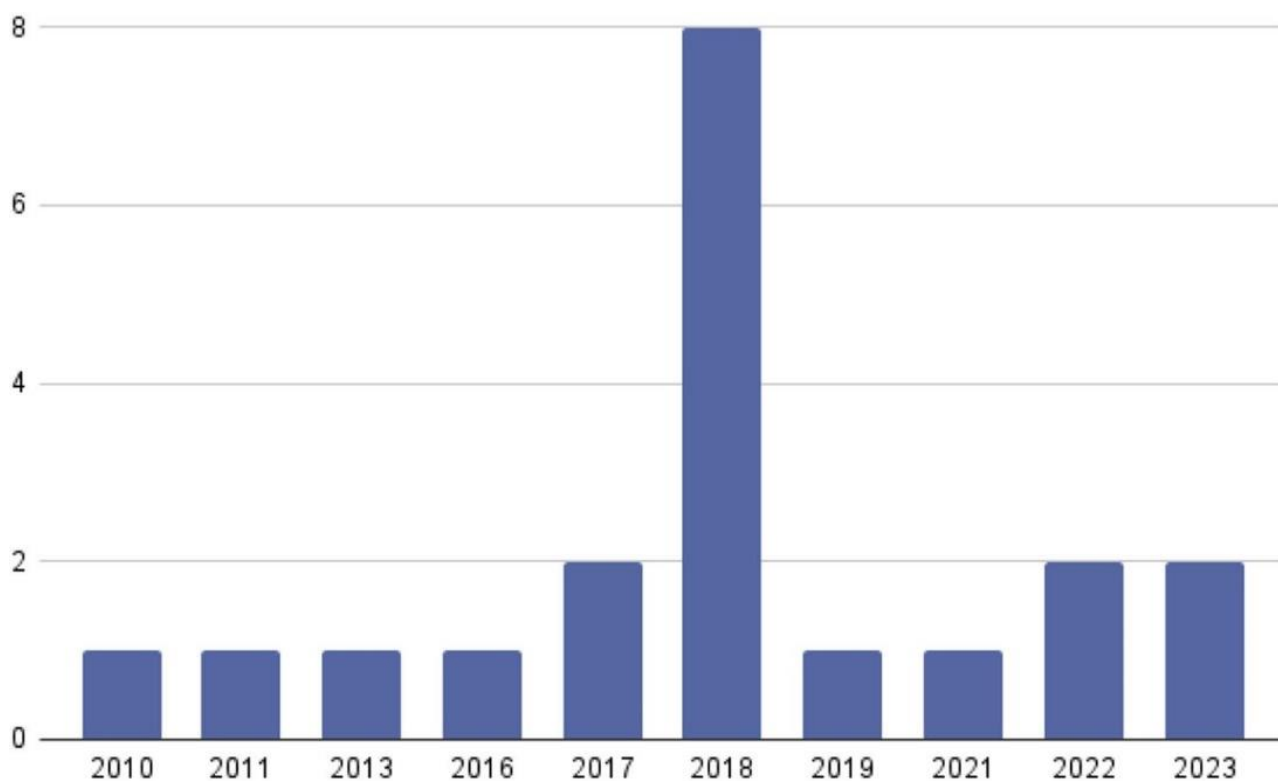


Рисунок 1.3 – Розподіл наукових публікацій щодо інформаційних ресурсів «розумних міст» на основі методології «великих даних» за роками [7]

Незважаючи на значну увагу наукової спільноти до питання впливу цифрових даних на розвиток сучасних складних соціально-економічних систем, включаючи «розумні міста», ця сфера тільки починає розвиватися, а розуміння використання даних як інструменту для розвитку «розумних міст» залишається доволі обмежено висвітленим у науковій та популярній літературі.

Загалом доцільно відзначити, що дослідження все більше зосереджуються на використанні цифрових даних як нового соціально-економічного явища, а також робляться спроби класифікувати, концептуалізувати та оцінити роль різних типів даних у міських соціально-економічних процесах. У більшості випадків такі дослідження пов'язані з використанням «великих даних» у певних сферах міського середовища:

- транспорт;
- громадська безпека;
- захист навколишнього середовища.

Водночас у науковій та популярній літературі відсутні дослідження загального системного характеру щодо використання «великих даних» для «розумних міст» незалежно від сфери застосування.

1.4 Висновок до першого розділу

В першому розділі кваліфікаційної роботи освітнього рівня «бакалавр» висвітлено актуальність проведення аналізу ресурсів розумного міста за допомогою методології великих даних. Подана методологія аналізу ресурсів «розумного міста» на основі «великих даних». Подано опис результатів наукометричного пошуку в царині опрацювання інформаційних ресурсів «розумних міст» на основі методології «великих даних».

РОЗДІЛ 2. ТИПИ, ДЖЕРЕЛА, АНАЛІЗ АЛГОРИТМІВ ОПРАЦЮВАННЯ ТА КЛАСТЕРИЗАЦІЯ ВЕЛИКИХ ДАНИХ «РОЗУМНИХ МІСТ»

2.1 Типи та джерела «великих даних» «розумних міст»

Концепція «розумного міста» передбачає інтеграцію обширного переліку інформаційних та комунікаційних технологій для управління інфраструктурою міст:

- транспорту;
- освіти;
- охорони здоров'я;
- систем житлово-комунального господарства;
- безпеки тощо.

Муніципальні органи влади збирають численну різноманітну інформацію, а термін «міські дані» може означати різні набори даних, зокрема:

- дані з камер відеоспостереження;
- інформація про трафік;
- якість повітря;
- споживання енергії;
- постачання та споживання ресурсів;
- постачання та споживання води;
- зображення для «розумного» розпізнавання тощо.

Тому для проведення аналізу інформаційних ресурсів «розумних міст» на основі методології великих даних важливим завданням є розпізнавання та класифікація різних наборів даних, які використовуються в розглянутих ресурсах.

Автори [11] описують методологію обробки створених дачачами в режимі реального часу «великих даних». Передбачається, що запропонована методологія працює з різними:

- джерелами даних дачачів;

- форматами даних;
- інтерфейсами підключення.

Бездротові сенсорні мережі (WSN) використовуються для:

- моніторингу фізичного стану навколишнього середовища;
- контролю показників забруднення повітря;
- сигналізування про пролісові пожежі;
- інформування про зсуви ґрунту;
- аналізування якості води.

Хоча система, запропонована авторами [11], розроблена для обробки всіх типів даних, WSN в основному продукують числові дані, такі як рівень води та концентрація газу в повітрі, в основному класифіковані як кількісна інформація. Ефективний метод отримання просторово-часового аналізу даних за допомогою кореляцій був запропонований у [12]. Автори використовують дані Bluetooth-давачів, встановлених в опорах міського освітлення. Дані були зібрані з дорожніх давачів у місті Орхус у Данії. Вимірювання проводяться кожні п'ять хвилин. Набір даних містить:

- часові позначки;
- інформацію про місцезнаходження;
- середню швидкість транспортних засобів;
- загальну кількість автомобілів на момент фіксації.

Дані були класифіковані як числові, оскільки в них немає тексту, зображень, аудіо чи відео.

У [13] описано використання великих мобільних соціальних даних, які включають вміст, створений користувачами, із географічними прив'язками та мітками часу. Контент означає текстові дані, які користувачі публікують у сучасних соціальних системах, що розвиваються, зокрема, Twitter, Facebook, Instagram тощо. Таким чином, набір даних можна класифікувати як неоднорідний за вмістом тексту публікацій у соціальних мережах і числових даних про місцезнаходження та час.

Дослідники [14] розглянули інший підхід до аналізу та оцінили ефективність глибоких нейронних мереж. Метою їх роботи був моніторинг і контроль місцевих епідемій ВІЛ. Колекція даних містить статистичні дані про:

- кількісні показники захворюваності;
- смертності;
- смертності за регіонами, віком, статтю чи професією.

Типи даних поділяються на текстові та числові.

В дослідницькій праці [15] запропоновано методологію вилучення моделей споживання електроенергії у часових рядах «великих даних». У дослідженні використовувалися великі часові ряди споживання електроенергії декількома будівлями університету. Дані були отримані за допомогою «розумних» лічильників електроспоживання протягом шести років.

Автори [16] описують нову структуру кластеризації даних для «великих даних», створених IoT-застосунками. Набір даних було зібрано з робочої IoT-інфраструктури «розумного міста» та його множини «розумних будівель», наданої федерацією тестових стендів FIESTA-IoT. Масив неоднорідний і представляє вимірювання різних типів даних:

- температура повітря;
- вологість повітря;
- рівень заряду батарей;
- вологість ґрунту тощо.

Автори [17] представили методику аналізу тривимірних даних із використанням підходу кластеризації дендрограм. Передбачається, що алгоритм можна застосовувати до різнорідних наборів «великих даних», зібраних із:

- датчиків;
- соціальних мереж;
- історичних джерел даних.

В [18] описано перевірку продуктивності двох алгоритмів розділення «K-Means» і «Fuzzy c-Mean» для кластеризації міських наборів «великих даних».

Порівняні методи можна застосовувати до гетерогенних наборів «великих даних» у різних сферах, таких як:

- медицина;
- бізнес;
- біологія тощо.

У статті автори використали міські дані з різних джерел даних:

- Інтернет речей;
- дані LIDAR;
- місцеві метеостанції;
- давачі мобільних телефонів.

У [19] розроблено новий ітераційний алгоритм, який називається «K-sets+» алгоритмом кластеризації точок даних у напівметричному просторі, де міра відстані не обов'язково задовольняє трикутну нерівність. Алгоритм призначений для кластеризації точок даних у напівметричному просторі. Щоб зрозуміти, що таке напівметричний простір, необхідно коротко розглянути поняття метрики в просторі.

Метрика – це відображення деякої множини $d: X \times X \rightarrow \mathbb{R}$, для якої обов'язково мають бути задоволені аксіоми невід'ємності та симетрії, але не обов'язково нерівність трикутника. Якщо відстань між різними точками може дорівнювати нулю, метрика є напівметричною. Метод оцінювали за допомогою двох експериментів:

- спільного виявлення підписаних мереж;
- кластеризації реальних мереж.

Набір даних містив понад двісті серверів у різних локаціях, а затримка вимірювалась часом зворотного проходження між будь-якими двома серверами записується в режимі реального часу.

В [20] порівнюється ефективність у прогнозуванні інфекційних захворювань:

- глибокої нейронної мережі (DNN);
- довгої короткочасної пам'яті (LSTM);

– авторегресійної інтегрованої ковзної середньої (ARIMA).

Дослідження використовує чотири види даних для прогнозування інфекційних захворювань:

- дані пошукових запитів;
- «великі дані» соціальних мереж;
- температуру;
- вологість.

Враховувалися дані щодо малярії, вітряної віспи та скарлатини за понад п'ятсот сімдесят п'ять днів. У результаті дані є частково цифровими, а частково текстовими. Дослідження [21] фокусується на аналізі міських даних із багатьох джерел. Об'єкти інтересу – географічні, вуличні, дорожні мапи та дані про нерухомість. Набір даних містить:

- мережу доріг міста;
- довготу;
- широту;
- назву;
- функціональність споруди в міському середовищі;
- зображення локацій.

Очевидно, що набір даних оцінюється як неоднорідний.

В проаналізованих наукових публікаціях переважають текстові та числові «великі дані», які збираються з обширного переліку джерел (див. таблицю 2.1).

Таблиця 2.1 – Результати аналізу публікацій щодо неоднорідних «великих даних» «розумних міст»

Публікація	Характеристика набору «великих даних» «розумного міста»
1	2
[11]	Дані з різних давачів
[13]	Пости в соціальних мережах, мітка часу, геолокація

Продовження таблиці 2.1

1	2
[14]	Історичні дані про захворюваність на ВІЛ за 10 років: кількість захворюваності, захворюваності, смертності та смертності за регіонами, віком, статтю, професією
[16]	Великі сенсорні дані, вимірювання різних типів: температура, вологість, рівень заряду батареї, вологість ґрунту
[17]	дані розумного міста
[18]	Дані з Інтернету речей, дані LIDAR, місцеві метеостанції, давачі мобільних телефонів
[19]	Розташування та затримка (вимірюється часом проходження туди й назад) між будь-якими двома точками даних
[20]	Дані пошукових запитів, великі дані соціальних мереж, температура та вологість
[22]	Географічні дані, дані про об'єкти інтересу (довгота, широта, назва та функціональність споруди в міському середовищі), дані перегляду вулиць, дані про нерухомість, дані про місцезнаходження мобільного телефону, дані соціальних мереж, дані мікроблогів, траєкторія таксі GPS дані, дані профілю таксі
[23]	Великі геопросторові дані з кількох джерел: визначні місця, транзакції смарт-карт автобусів, траєкторії маршруту таксі
[24]	Соціальні мережі, веб-форуми, зображення, відео, RDBMS, OLAP, сховище даних, дані у форматі XML, CSV, HTML, RDF

Автори [22] описують багатовимірний підхід до видобування даних в послідовний спосіб, знаходячи групи (кластери) спільнот з однаковими мультидинамічними характеристиками. Дані стосуються:

- статистики населення;
- міграції;
- платоспроможності;

- житла;
- зайнятості;
- пасажирів.

В таблиці 2.2 подані проаналізовані статті щодо числових колекцій «великих даних».

Таблиця 2.2 – Результати аналізу публікацій щодо числових «великих даних» «розумних міст»

Публікація	Характеристика набору «великих даних» «розумного міста»
[12]	Дані про дорожній рух, зібрані з дорожніх давачів у місті: географічне розташування, позначка часу, середня швидкість і загальна кількість автомобілів
[15]	Споживання електроенергії за 6 років на декілька будинків
[25]	Вимірювання аналізів крові у вигляді корпускулярного об'єму досліджуваних речовин і кількості півпінтових еквівалентів алкогольних напоїв, випитих за день
[26]	Дані про бронювання готелів: час, розмір вибірки, номер замовлення, номер користувача, старий номер користувача
[27]	Дані про споживання води та електроенергії з міткою часу
[27]	Споживання води
[28]	Валовий внутрішній продукт, споживання вугілля, споживання сирої нафти, споживання природного газу, споживання гідроенергії

Більшість проаналізованих наукових досліджень описує процеси опрацювання неоднорідних наборів «великих даних» різної природи. Одна з них описує дослідження процесів обробки даних зображень.

2.2 Аналіз алгоритмів опрацювання «великих даних» «розумних міст»

Головною метою аналізу інформаційних ресурсів «розумних міст» на основі методології «великих даних» є виявлення застосування сучасних методів опрацювання даних щодо процесів постачання та споживання ресурсів, таких як електроенергія, вода, газ, тепло. При цьому важливо визначитись, яку корисну інформацію можна отримати зі статистики постачання та витрат ресурсів. Враховуючи, що досліджень щодо споживання первинних ресурсів не так багато, доцільно розглянути різні зразки даних і методи їх опрацювання, щоб визначити придатні для використання в майбутньому міські набори «великих даних».

Загалом у [7] було розглянуто понад два десятки різних підходів до обробки міських «великих даних». Методи можна розділити на групи залежно від способу обробки інформації «розумних міст»:

- кластеризація;
- класифікація;
- кореляція;
- глибокі нейронні мережі;
- фреймворки;
- виявлення спільнот.

Рисунок 2.1 ілюструє співвідношення між різними групами підходів до обробки інформації «розумних міст». Найпопулярнішим підходом є кластеризація, яка розглядається в шести рецензованих статтях. Стільки ж разів опубліковано інформацію щодо фреймворків, що мають на увазі архітектури для автоматичної обробки міських наборів «великих даних». всередині фреймворків можуть застосовуватися різні методи в залежності від мети, тому було виокремлено цю категорію. Глибинні нейронні мережі не так часто використовувалися для прогнозування. Відомості щодо цього опубліковані лише двічі.

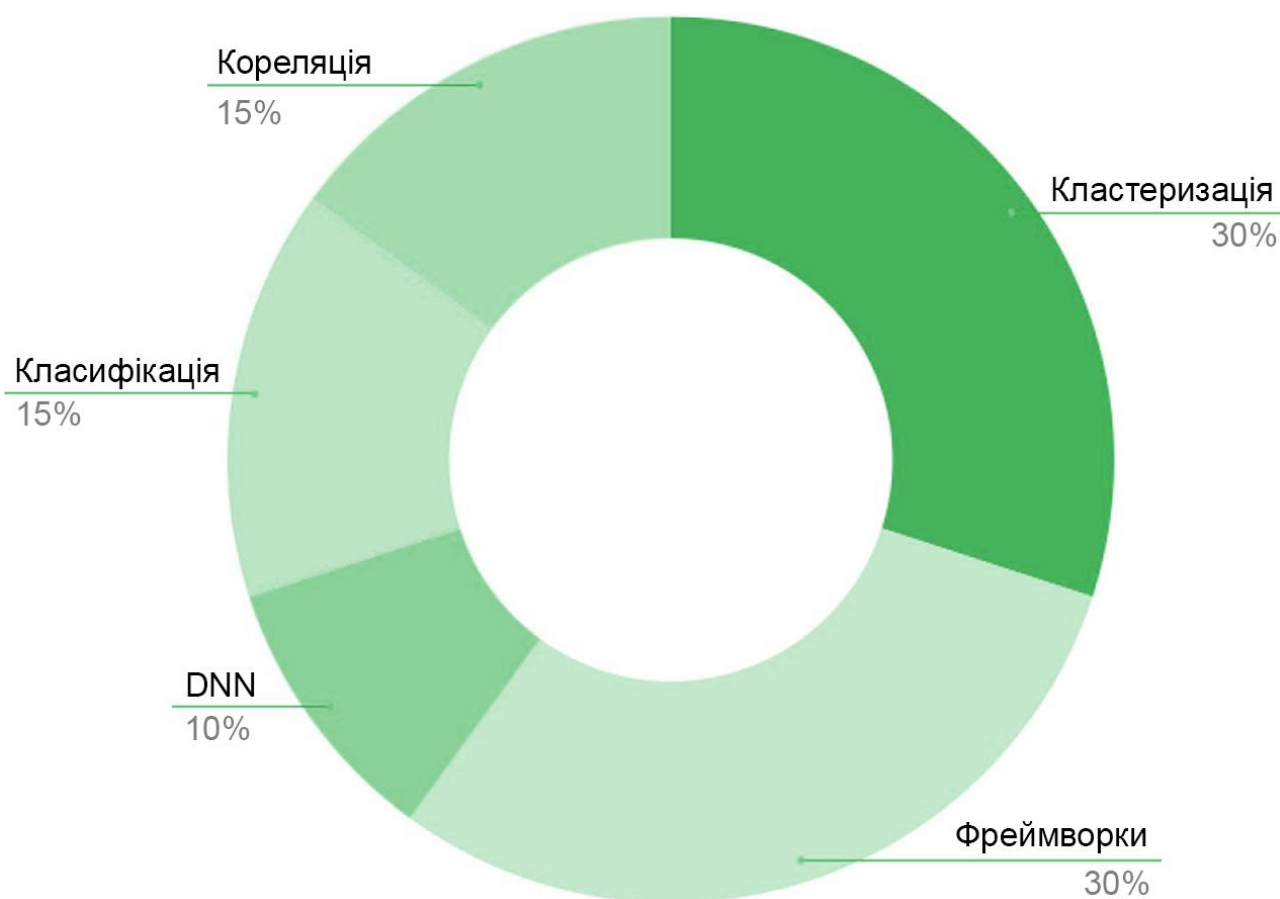


Рисунок 2.1 – Методи аналізу міських наборів «великих даних» [7]

Опубліковані на даний час результати досліджень представляють обширну картину оціночних характеристик ряду методів на основі глибоких нейронних мереж (DNN). У категорії «Класифікація» було розглянуто три наукові праці. Кореляційні методи представлені в 3 статтях. Класифікація та кореляція тут – це типи проблем машинного навчання. Класифікація поєднує в собі цілий набір методів і алгоритмів, призначених для розділення набору інформаційних сутностей, кожен метод описується набором змінних у певному значенні класів. Кореляція є мірою зв'язку сутностей між собою. При цьому мається на увазі як зв'язок між прогностичними та фактичними значеннями змінних. Подана на рис. 2.1 класифікація є умовною. Мета кваліфікаційної роботи освітнього рівня «магістр» полягає в тому, щоб визначити тенденції розвитку засобів аналізу інформаційних ресурсів «розумних міст» на основі методології «великих даних». Тому розглянемо детальніше ключові категорії аналітичних засобів «великих даних» для потреб «розумних міст».

2.3 Аналіз підходів кластеризації «великих даних» для потреб «розумних міст»

Кластеризація – це розділення зібраного в середовищі «розумного міста» набору «великих даних» на групи на основі подібних ознак. Вона використовується в різноманітних задачах обробки «великих даних» для потреб формування «розумних» міських цифрових послуг, зокрема:

- розпізнавання образів;
- машинне навчання;
- автоматична класифікація;
- розробка стратегій керування тощо.

Досі не винайдено ефективного для різних типів зібраних в середовищі «розумного міста» «великих даних» універсального алгоритму. Здебільшого використовуються ітераційні методи кластеризації, які базуються на апріорному встановленні кількості кластерів і певному виборі початкового їх розбиття. При цьому результат застосування істотно залежить від коректності оцінки кількості кластерів «великих даних» «розумного міста».

Кластеризація є одним із найпопулярніших підходів у застосунках для значної кількості міських задач із «великими даними». На даний час її можна вважати базовим кроком для подальшого дослідження, оскільки при цьому формується чіткіша картина закономірностей у будь-якому наборі даних. В процесі наукових розвідок було проаналізовано шість документів з різними алгоритмами кластеризації, які використовуються для різних типів зібраних у середовищі «розумного міста» вхідних наборів «великих даних».

Алгоритм виявлення спільнот «Girvan–Newman GN» [29] був модифікований для кластеризації великих даних IoT-сенсорів «розумного міста» [16]. Зазначений метод організовує складні дані в блоки, які називаються спільнотами або модулями, відповідно до ролей і функцій, організованих у мультиграф. Складність полягає в тому, щоб знайти в мультиграфі розбиття вершин, де цільова функція мінімізована. Щоб досягти цього, ребра графа

видаляються ітераційно, залежно від значення метрики. Найпоширенішим показником є «Edge Betweenness Centrality» (EBC), але для його обчислення потрібно багато часу. Автори запропонували нову міру, що наближає EBC, яка використовує гіперболічне вбудовування мережі та може розглядатися як «гіперболічний» аналог EBC. Ця міра позначається як «Hyperbolic Edge Betweenness Centrality» (HEBC) і обчислюється за допомогою гіперболічних координат вузлів, призначених вбудованим вузлам. Нова метрика підвищує продуктивність без втрати точності опрацювання міських наборів «великих даних».

Інша техніка організації та обробки даних, запропонована авторами [17], використовує тривимірну аналітику «велики даних» «розумного міста» з використанням дендрограмного ієрархічного підходу кластеризації. Тривимірні дані являють собою структуру інформації, яка поєднує одночасно завдання класифікації та кластеризації. Упорядковані дані відображаються в структурі дерева та витягуються за допомогою алгоритмів обходу дерева. Кластеризація дендрограм – це метод об'єднання інформаційних сутностей у групи. У дослідженні використовується висхідний алгоритм кластеризації, що означає, що кожен елемент у класі призначається одному кластеру, а потім комбінуються групи, доки всі об'єкти не об'єднуються. Важливим параметром є відстань між об'єктами в класі. Метрика показує кількісну оцінку співвідношення схожості елементів за різними критеріями. У дослідженні не передбачається вибір конкретного параметра, хоча вибір метрики відбувається на другому кроці методу. Автори [17] вимірювали здатність отримувати інформацію та ефективність структури. В цілому методика демонструє хороші характеристики видобування інформації з міських наборів «великих даних», але не найпривабливіші параметри продуктивності.

Інші алгоритми кластеризації, «Fuzzy c-Mean» (FCM) і «k-Means» були протестовані [18]. Алгоритм k-середніх є одним з найпростіших методів, але в той же час найменш точним. Основна ідея цього методу полягає в тому, що на кожній ітерації центр мас перераховується для кожного кластера, отриманого

на попередньому кроці, потім результати знову розбиваються на кластери з новими центрами. Алгоритм завершується, коли кластер не змінюється під час ітерації. Метод нечіткого «k-Mean» дає змогу отримати «нечітку» кластеризацію числових наборів «великих даних» «розумного міста» для і дає змогу коректно ідентифікувати об'єкти на межах кластерів. Однак виконання цього алгоритму вимагає серйозних обчислювальних ресурсів і початкового налаштування кількості кластерів. Крім того, неоднозначність може виникнути з інформаційними сутностями, віддаленими від центрів усіх кластерів.

Новий підхід для кластеризації точок «великих даних» «розумного міста» був розроблений авторами [19]. По суті, метод є розширенням алгоритму кластеризації K-множин для напівметричного простору. Задача підходу K-множин полягає в тому, що відстань трикутника не є невід'ємною. Таким чином, алгоритм K-множин може взагалі не збігатися, і немає гарантії, що результатом алгоритму K-множин є кластеризація. Для вирішення цієї труднощі було скориговано процедуру визначення відстані трикутника, щоб можна було скасувати вимогу невід'ємності. Експериментальні результати підтверджують ефективність методу для матриці географічної відстані та матриці затримки.

Ще один метод кластеризації використовується в [25]. Реалізовано алгоритми кластеризації «Fuzzy c-Means» (FCM), «k-means» (KM) і «Gustafson-Kessel» (GK). Згідно з опублікованими результатами досліджень [7], найбільш точним і ефективним алгоритмом є «k-means» кластеризація, але інші методи мають свої переваги і показують більшу коректність в окремих випадках вирішення задач опрацювання «великих даних» для потреб «розумних міст».

Великі геопросторові «розумних міст» «великі дані» були кластеризовані за допомогою «алгоритму спектральної кластеризації» [23]. У дослідженні автори подали нещодавно розроблений метод з адаптивними графіками для обмеження кластеризації багатопросторового підпростору з декількома джерелами великих геопросторових даних, якими є GPS-траєкторії таксі, точки інтересу та транзакції смарт-карт автобусів. Дані були зібрані та оброблені

відповідно до зон аналізу руху. Мережі автокодерів використовувалися для відображення великорозмірних і зашумлених оригінальних геопросторових «великих даних» «розумного міста» у приховане представлення. Приховане представлення кожного типу геопросторових «великих даних» «розумного міста» було використано для побудови низьковимірною підпростору. Спільний метод найближчого сусіда був застосований для побудови адаптивних графіків для геопросторових «великих даних» «розумного міста» що характеризуються:

- великою розмірністю;
- неоднорідністю;
- шумністю.

Спектральний алгоритм кластеризації був використаний для отримання результатів кластеризації «великих даних» «розумного міста» за допомогою матриці подібності. Кластери виявляють міські функціональні зони в місті відповідно до діяльності людини. Точність кластеризації склала понад шістьдесят вісім відсотків [23]. Запропонований метод цікавий із соціально-економічної точки зору, його можна екстраполювати на більші «розумні регіони», а статистику споживання ресурсів «розумного міста» можна використовувати замість геоданих. Тим не менш, точність все ще досить низька, і метод потребує подальшого вдосконалення. В опублікованих результатах досліджень не подана оцінка складності.

Алгоритм «K-sets+» забезпечує найвищу продуктивність серед усіх кластерних алгоритмів опрацювання «великих даних» для потреб «розумних міст». Часова складність є лінійною:

$$O((Kn+m)I),$$

де I – кількість ітерацій.

Інший метод із лінійним часом – нечітке с-середнє з $O(nCI)$, де C – кількість кластерів, I – кількість ітерацій. Якщо порівняти експоненти для цих двох підходів, очевидний факт полягає в тому, що «K-sets+» дає невелику

перевагу. Зважаючи на точність, «K-sets+» демонструє точність дев'яносто п'ять відсотків як найгірший результат аналітичного порачування «великих даних» «розумного міста». Алгоритм «Fuzzy k-Means» демонструє складність в середньому понад вісімдесят один процент. Слід зазначити, що модифікація Гірвана-Ньюмана забезпечує стовідсоткову точність для більшості наборів «великих даних» «розумних міст» і лише п'ятдесят відсотків у випадку викидів. Його можна використовувати для набору «великих даних» «розумного міста» із низькою розрідженістю, якщо потрібна висока точність. Метод кластеризації дендрограм повільніший за інші при опрацюванні «великих даних» «розумних міст», але може створювати ієрархічну структуру дерева для даних. Кластеризація «K-Means» є найпростішим методом, але має квадратичну складність і точність не більше восьмидесяти восьми відсотків для різних вхідних наборів «великих даних», що зібрані в середовищі «розумного міста».

Один із внесків, який полягає у використанні підходу кластеризації для групування кривих навантаження споживання води, описаний в [27]. Кластеризація спрямована на класифікацію моделей щоденного споживання води відповідно до інтенсивності, наприклад:

- нормальний чи ненормальний режим споживання;
- вихідні чи робочі дні;
- відпустка та канікули.

В дослідженні [27] використовувався метод ієрархічної агломеративної кластеризації, складність якого є квадратичною. Було розглянуто три різні підходи до кластеризації:

- евклідова відстань;
- динамічне викривлення часу;
- використання самоорганізуючої карти.

Метод кластеризації з евклідовою відстанню виявився найбільш ефективним при опрацюванні наборів «великих даних» для потреб «розумних міст» з точки зору точності та складності. Даний метод не є найефективнішим

серед усіх методів кластеризації, але заслуговує на увагу при категоризації даних про споживання ресурсів «розумного міста».

В цілому методи кластеризації ефективні і дають можливість виявити обширний перелік закономірностей, корисних для аналізу процесів, що протікають в середовищі «розумних міст». Однак результати аналітичного опрацювання «великих даних» суттєво залежить від типів міських даних, використаних процедур попередньої обробки при середній точності та високій ресурсомісткості. В результаті проведеного аналізу можна зробити висновок про позитивну динаміку розвитку засобів кластеризації для опрацювання «великих даних» «розумних міст» та можуть бути позитивно вдосконалені за допомогою додаткових аналітичних методів.

2.4 Висновок до другого розділу

В другому розділі кваліфікаційної роботи проаналізовано типи та джерела великих даних «розумних міст». Розглянуто ключові категорії алгоритмів опрацювання «великих даних» «розумних міст». Окремо проаналізовано кластеризацію «великих даних» для потреб «розумних міст».

РОЗДІЛ 3. АНАЛІЗ ТОЧНОСТІ ОБРОБКИ ІНФОРМАЦІЙНИХ РЕСУРСІВ «ВЕЛИКИХ ДАНИХ» ДЛЯ ПОТРЕБ «РОЗУМНИХ МІСТ»

3.1 Використання каркасів на основі фреймворків для аналітичного опрацювання інформаційних ресурсів «розумних міст»

Обробка зібраних в середовищі «розумних міст» наборів та колекцій «великих даних» є складним нетривіальним завданням. При цьому доволі зручно створити автоматизований ланцюжок або конвеєр для автоматизації процедур аналізу. Зазвичай для цієї мети формується структура, що складається з конвеєра для обробки «великих даних». В процесі аналізу було опрацьовано шість статей, які описують різні реалізації фреймворків.

Структура візуального аналізу для дослідження та розуміння неоднорідних наборів та колекцій міських «великих даних» була описана в [22]. Візуальна модель запитів – це підхід на основі візуального аналізу для дослідження просторово-часових наборів міських «великих даних» (VAUD) є основою для інтерактивного дослідження в поєднанні з простими, але потужними структурними абстракціями та функціями міркування. VAUD використовує візуалізацію різноманітних міських «великих даних», що складаються з інформації про:

- мобільність громадян;
- мобільних телефонних дзвінків;
- дорожнього руху тощо.

Підхід, який обговорюють автори [22], базується на запитах до бази даних. При цьому не може бути оцінена часова складність. Таким чином, дані витягуються, пов'язуються та візуалізуються за допомогою спеціально розробленої мови запитів. У середньому точність за запитами становить сімдесят шість процентів.

Описаний у [11] підхід містить три рівні:

- рівень вмісту;

- рівень послуг;
- рівень застосунків.

Рівень вмісту узагальнює джерела даних сенсорних мереж «розумного міста», а рівень служб забезпечує:

- підключення до баз та сховищ «великих даних»;
- процедури перетворення даних;
- протоколи зв'язку для обробки даних у режимі реального часу.

Прикладний рівень містить програмно алгоритмічні засоби для обслуговування потреб користувачів. Сервісний рівень реалізує алгоритм виявлення аномалії «Cumulative SUM» (CUSUM). Даний метод розглядає набір спостережень за нормальним розподілом. Для кожної колекції вимірювань в середовищі «розумного міста» розраховується кумулятивна сума. Коли оцінка долає порогове значення, алгоритм виявляє аномалії «великих даних» «розумного міста». Якщо параметр перевищує порогове значення, аномалія буде пов'язана зі збільшенням «up-event», а якщо сума перевищує порогове значення, це буде пов'язано зі зменшенням «down-event». Різні типи даних з багатьох джерел «розумного міста» обробляються спеціалізованою оболонкою та перетворюються в стандартну форму. Трансформовані дані спостереження середовища «розумного міста» кодуються в рядки відповідно до стандарту Відкритого геопросторового консорціуму (OGC) для спостережень і вимірювань.

Моделі споживання електроенергії в «розумних містах» були досліджені авторами [15]. Методологія описує всі етапи обробки даних:

- збір даних;
- очищення;
- трансформація;
- індексний аналіз;
- кластеризація;
- результати.

Перший етап спрямований на попередню обробку «великих даних», щоб їх можна було кластеризувати. Другий етап полягає в отриманні оптимальної кількості кластерів «великих даних» «розумного міста» для набору даних шляхом аналізу та інтерпретації різних індексів перевірки кластерів. Далі k-середнє використовується для кластеризації та, нарешті, отримує центроїди для кожного кластера «великих даних» «розумного міста». Конвеєрна обробка виконується засобами Apache Spark. Алгоритми включають індекси дійсності кластеризації «великих даних» (BD-CVI) і k-середні.

В [24] описано багатоагентний підхід був представлений. Запропонована авторами інформаційно-технологічна архітектура складається з трьох рівнів:

- збору та зберігання «великих даних» «розумного міста»;
- керування та обробки даних;
- сервісного рівня, заснованого на багатоагентній системі для автоматизації аналізу «великих даних».

Система складається з множини програмно-алгоритмічних агентів із різними ролями:

- Агент приймача, що обробляє великі за обсягом дані різних типів і з різних джерел в середовищі «розумного міста», таких як давачі, зовнішні бази даних, веб-сервіси цифрових «розумних» міських послуг тощо.
- Агент сховища даних відповідає за всі операції, пов'язані з керуванням сховищем «великих даних» і читанням даних із кластера Hadoop.
- Сервісний агент взаємодіє з застосунками і службами та надає результати аналітичного опрацювання наборів та колекцій даних «розумного міста».
- Агент офлайн аналітики обробляє запити на аналіз даних «розумного міста», використовуючи великі за обсягом історичні дані.
- Агент потокової аналітики обробляє потоки даних у режимі реального часу. Як було зазначено, вхідні дані неоднорідні з різних джерел в середовищі «розумного міста».

Для обробки даних автори [24] використовують алгоритми класифікації, зокрема порівнювали:

- випадкові ліси;
- наївний Байєс;
- багат шарові перцептрони.

Найвищу точність при опрацюванні «великих даних» «розумного міста» показав метод випадкового лісу – вісімдесят п'ять відсотків. Найменшу точність опрацювання «великих даних» «розумного міста» процемонстрував наївний Байєс – це шістдесят відсотків. Точність багат шарового перцептрона сімдесят три відсотки.

Після видобування групи «великих даних» «розумного міста» її було використано для прогнозування в режимі реального часу. Запропонований підхід використовується в міському управлінні та плануванні.

3.2 Глибинні нейронні мережі при опрацюванні інформаційних ресурсів «розумних міст»

Нейронні мережі є математичними моделями, а їхня програмна реалізація заснована на структурі нервової системи людини. Головною особливістю нейронних мереж є можливість вчитися на потенційно великому наборі прикладів. В продовж останнього періоду часу нейронні мережі використовуються практично у всіх наукових сферах, в тому числі в галузі аналітичного опрацювання інформаційних ресурсів «розумних міст» на основі методології великих даних. В процесі аналізу було опрацьовно дві статті, в яких розглядалися нейронні мережі, і досліджено та оцінено шість методів:

- моделі нейронної мережі довгої короткострокової пам'яті (LSTM);
- моделі авторегресійної інтегрованої ковзної середньої (ARIMA);
- моделі нейронної мережі узагальненої регресії (GRNN)
- моделі експоненціального згладжування (ES) для оцінки захворюваності громадян.

Автори [14] дослідити, яка модель є найкращою та найточнішою для місцевого прогнозування захворюваності поширення інфекційних захворювань в середовищі «розумного міста». Було проаналізовано:

– ARIMA – це модель, яка використовується для прогнозування часових рядів.

– LSTM – це рекурентна нейронна мережа, яка характеризується здатністю вивчати довготривалі залежності.

У [7] було побудовано ряд моделей. Модель з найменшою середньоквадратичною помилкою (MSE) вважалася оптимальною при опрацюванні «великих даних» для потреб «розумних міст». GRNN – це нейронна мережа прямого зв'язку, яка оцінює значення безперервних залежних змінних. Основними перевагами GRNN є швидке навчання та збіжність до оптимальної поверхні регресії, оскільки кількість вибірок міських наборів «великих даних» стає дуже великою.

GRNN має особливу перевагу з розрідженими даними в середовищі реального часу, оскільки поверхня регресії миттєво визначається скрізь, навіть із лише одним зразком. Метод зазвичай використовується для апроксимації функцій, тому він може забезпечити дуже високу точність, але для великих вибірок вимагає високопродуктивних обчислень.

Модель ES є однією з найпростіших і найпоширеніших практик вирівнювання рядів. Метод можна представити як фільтр, який отримує на вхід зібрані в середовищі «розумного міста» вихідні члени ряду «великих даних», а на виході формує поточні значення експоненціального середнього.

3.3 Аналіз підходів класифікації «великих даних» для потреб «розумних міст»

Завдання класифікації «великих даних» для потреб «розумних міст» характеризується кінцевою множиною допустимих відповідей, що містить мітки класів. Клас – це набір усіх об'єктів «великих даних», які мають однакову

мітку. Класифікація використовується для визначення категорії «великих даних», до якої належить окрема вибірка даних «розумних міст». В процесі виконання кваліфікаційної роботи освітнього рівня «бакалавр» було проаналізовано три статті, що присв'ячені дослідженню методів класифікації.

Загальний підхід для трансформації вихідної задачі в традиційну точкову класифікаційну задачу за допомогою методу на основі вибірки був запропонований в [26]. «Загальний підхід» означає, що цей метод можна використовувати для будь-якого набору даних «розумних міст» на основі розподілу. Для тестування були використані реальні набори «великих даних» про бронювання готелів «розумних міст». Основне призначення класифікації – рекомендаційні системи для потреб цифрових послуг «розумного міста». Автори [26] запропонували власну систему оцінки точності на основі рейтингу попадань «Тор-К». Це означає, що система рекомендує K цифрових продуктів «розумних міст» користувачеві один раз, і якщо користувач остаточно вибере серед K рекомендованих цифрових продуктів, то відбувається точне попадання. Показник стабільно зростає зі збільшенням кількості рекомендацій [26]. Часова складність дорівнює $O(TR)$, де T – кількість навчальних кортежів, кожен з яких містить R атрибутів. З грубою оцінкою можна сказати, що складність є квадратичною ($O(n^2)$).

Класифікація моделей споживання води та електричної енергії в середовищі «розумного міста» в режимі реального часу була описана в [27]. Це передбачає навчання алгоритмів машинного навчання для визначення нормальних і ненормальних моделей споживання води та розрізнення різних типів електричних пристроїв в середовищі «розумного міста». Сигнатури електричних пристроїв класифікуються за допомогою трьох алгоритмів машинного навчання:

- багат шарового перцептрона (MLP),
- k -найближчих сусідів (KNN)
- дерева рішень (DT).

Зразок вхідного набору «великих даних» – це одиниця споживання, наприклад один літр води або одна ват-година електроенергії, і кожна подія має позначку часу. З проведених експериментів автори дійшли висновку, що «One Against All» (ОАА) – MLP і вихідні коди з виправленням помилок (ЕСОС) – MLP Р працюють краще, ніж MLP, DT і KNN. Крім того, можна підкреслити, що ОАА-MLP перевершив усі інші підходи з точки зору продуктивності та точності, але він складніший, ніж ЕСОС-MLP при опрацюванні «великих даних» для потреб «розумних міст».

Класифікаційний підхід розглядає споживання води в середовищі «розумного міста». Платформа, запропонована в [27] використовує вдосконалені методології обробки сигналів у поєднанні з контрольованими класифікаторами машинного навчання для класифікації потоків води «розумних міст», таким чином ідентифікуючи побутові водопровідні прилади з високою точністю. Моделі досягають точності дев'яносто один відсоток у класифікації чотирьох найбільш використовуваних побутових приладів для води. Вхідним набором «великих даних» є потік води з часовими мітками, упорядкованими за площею та приладами. Наприклад, розглядалися офіси, будинки та фабрики, а також різна техніка: кухонні змішувачі, душові кабінки тощо. Зазначена інформаційна система спрямована на розширення моніторингу споживання води та підвищення ресурсозбереження «розумних міст».

3.4 Кореляційний аналіз інформаційних ресурсів для потреб «розумних міст»

Ще один широко розповсюджений статистичний метод, застосований до «великих даних» – кореляція. У проаналізованих статтях один алгоритм розглядає кореляцію, застосовану до «великих даних» «розумного міста». У дослідженні [7] порівнювали два типи методів:

- кореляцію Пірсона;
- взаємну інформацію.

Часова складність обох підходів є кубом, але кореляція Пірсона може виявити лінійний розподіл «великих даних», а взаємна інформація може виявити залежності в більш загальних випадках розподілу даних «великого міста». Однак, якщо програма надає перевагу відповіді в реальному часі над точністю, кореляція Пірсона буде придатною, оскільки вона дасть лише декілька хибно негативних результатів. В інших сценаріях «розумних міст» з різними типами потоків «великих даних», температура, забруднення тощо, краще використовувати взаємну інформацію без попереднього знання потенційних кореляцій, оскільки наперед невідомо відсоток випадків, коли кореляція Пірсона не зможе виявити кореляції [12].

Автори [7] спробували застосувати кореляційні методи до аналізу міських «великих даних». Вони запропонували ефективний метод для отримання просторово-часового аналізу «великих даних», використовуючи:

- кореляції, за допомогою методів Пірсона;
- кореляції, за допомогою ентропії;
- порівняння результатів обох алгоритмів.

Коефіцієнт Пірсона характеризує наявність лінійної залежності між двома величинами. Слабкою стороною кореляції Пірсона є низька точність, коли змінні не розподіляються нормально. Взаємна інформація – це статистична функція двох випадкових величин, яка описує кількість інформації однієї випадкової величини в іншій. Обмеження взаємної інформації полягає в тому, що вона має вищу складність обробки, ніж кореляція Пірсона. При цьому безперервно обчислюється середня кореляція для сенсорних даних «розумного міста» про дорожній рух, розділених на два сектори, доки набір «великих даних» не закінчиться. В результаті було проаналізовано різні типи кореляції.

Дослідження про взаємозв'язок між споживанням енергії та економічним зростанням «розумних міст» було опубліковано в [28]. Для поглибленого аналізу кореляції між економічним розвитком і структурою споживання енергії «розумних міст» автори вибирають основний ресурс «великих даних», щоб

отримати сіру кореляцію релевантності. Параметри вхідних наборів «великих даних»:

- ВВП, мільярдів одиниць;
- витрата вугілля, од. т умовного вугілля;
- споживання сирової нафти, одиниця тонн умовного вугілля;
- споживання природного газу, одиниця тонн умовного вугілля;
- споживання гідроенергії, одиниця тонн умовного вугілля.

В дослідженні використано модель аналізу релевантності Грея. Основою обстеження «великих даних» «розумного міста» є відстані між опорними точками та точками порівняння. Аналізуючи відмінності та близькість факторів, можна кількісно визначити невизначеності. Кореляція результату між споживанням енергії та економічною ефективністю «розумного міста» становить сімдесят два відсотки, що можна вважати достовірним. Автори не забезпечили перевірки точності та складності.

3.5 Аналіз точності методів обробки інформаційних ресурсів «розумних міст» на основі методології «великих даних»

За допомогою методів дистанційного спостереження важко вловити соціально-економічні характеристики та соціальну динаміку «розумних міст», які тісно пов'язані з міським землекористуванням [23].

Оцінка часу та точності всіх проаналізованих в кваліфікаційній роботі освітнього рівня «бакалавр» алгоритмів показує, що якщо метою є прогнозування, то найкращим варіантом є глибокі нейронні мережі на кшталт LSTM.

Для ефективною кластеризації найпотужнішими є алгоритми «K-sets+» або «Fuzzy c-Means». Якщо необхідно здійснити розширений аналіз «великих даних» «розумного міста», то можна задіяти кореляційні алгоритми.

В контексті аналізу муніципальних даних «розумних міст», фреймворки є вигідними, оскільки вони реалізують всі етапи обробки «великих даних» від зберігання до візуалізації.

Алгоритм «CUSUM» має лінійну складність у часі при опрацюванні «великих даних» для потреб «розумного міста». Програмно-алгоритмічне рішення є простим і швидким, але має обмеження, які необхідно враховувати, наприклад те, що всі ряди «великих даних» повинні відповідати нормальному розподілу, а ряди спостережень не можуть мати трендів [11].

Проаналізовані в кваліфікаційній роботі освітнього рівня «бакалавр» алгоритми мають характеристики продуктивності та масштабованості, які слід розуміти. В додатку А подано сформовану на основі [7] порівняльну характеристику складності та точності розглянутих алгоритмів.

Для значної частини проаналізованих алгоритмів було непросто оцінити складність, оскільки час залежить від характеристик обчислювальних платформ. Тому в додатку А надано лише приблизні оцінки. При цьому всі подані оцінки стосуються найгірших значень.

Виявлення аномалії за допомогою алгоритму CUSUM [7] створює попереджувальне повідомлення для сторони клієнта у випадку рідкісних подій в середовищі «розумного міста». Кожна подія містить ідентифікатор дачача – поле відправника та ідентифікатор конкретного спостереження, яке спричинило подію – поле ідентифікатора. Панель подій візуалізує ці дані. Панель показує всі вузли зондування мережі на карті за допомогою маркерів. Усередині кожного маркера відображається кількість подій, які були виявлені для цього конкретного вузла зондування в середовищі «розумного міста».

Аналіз, заснований на кореляції та взаємній інформації, використовувався для моніторингу дорожнього руху «розумного міста». Автори [12] описують три серії експериментів. У першому випадку порівнювали продуктивність кореляції Пірсона та взаємної інформації «великих даних» «розумного міста». Результати були візуалізовані на Google Maps. Можна зробити висновок, що кореляція Пірсона ефективна для лінійного розподілу «великих даних»

«розумного міста», а взаємна інформація життєво важлива для нелінійних залежностей, але вимагає більше часу для опрацювання.

Результати, отримані в [14] є прогнози інфекційних захворювань. Кожен порівнюваний алгоритм має свої показники. Наприклад, ARIMA включає процес ковзного середнього, процес автоматичної регресії ковзного середнього, процес автоматичної регресії ковзного середнього та процес ARIMA відповідно до різних частин регресії та того щоб оцінити, чи вихідні дані стабільні. Щоб оцінити точність даних, вони порівняли з вихідною інформацією. Такий самий тип результатів демонструє [20]. Автори порівняли однакові параметри для LSTM, DNN і ARIMA, щоб оцінити правильність прогнозування інфекційних захворювань. Усі кластерні алгоритми дають однаковий результат щодо кількості кластерів та їх точності.

Дані про споживання електроенергії в «розумному місті» були згруповані в чотири групи, а потім у вісім [15]. Результати подані у вигляді діаграм. Кластери поділяються на категорії залежно від:

- характеристик «розумних» будівель;
- пори року;
- днів тижня.

Модифікація методу Гірвана–Ньюмана з новою метрикою, надана в [16], була застосована до багатовимірних «великих даних» «розумного міста», отриманих з робочої інфраструктури «розумного міста» на основі IoT-пристроїв в «розумних» будівлях. Автори подають оцінку точності, модульності та порівняння часу HGN та GN, порівнюючи час виконання алгоритмів GN та HGN для графів із відомими спільнотами та порівняння модульності вибірок за періоди часу різної тривалості. З огляду на те, що статистика демонструє ефективність обчислень, а алгоритм може давати точні результати.

Кластерна візуалізація в дендрограмах, протестована на інформації про сотні тисяч «розумних» будівель, була опублікована в [17]. Також було надано

аналіз часу відгуку, який демонструє, що час відгуку для запропонованого методу на 50% швидший, ніж для несуккупних даних.

Метою кваліфікаційної роботи освітнього рівня «бакалавр» було з'ясувати тенденції в обробці «великих даних» інфраструктури «розумних міст» [30]. В процесі аналізу нас цікавили дані про постачання та споживання ресурсів, зокрема електроенергії, води, тепла, дані про міський трафік, а також методи створення прогнозних моделей, кластеризації та класифікації. «Великі дані» все частіше розглядаються як ключовий ресурс розвитку «розумного» міського середовища, що відкриває можливості для:

- оптимізації економічних процесів;
- створення інновацій у соціальній сфері;
- формування нових моделей управління.

Проведений аналіз є основою для подальшої роботи з аналізу «великих даних» про витрати ресурсів та створення системи управління містом.

Вхідні та вихідні дані різняться залежно від методу та цілей дослідження. Переважно розглядаються різноманітні джерела даних в середовищі «розумного міста» [31]. В одному випадку використовуються зображення, а в двох випадках – числова інформація. Різноманітні дані означають, що використовується інформація різної бази: зображення, текст, числова. Відео чи звукові дані здебільшого не використовуються в рецензованих статтях. Основна частина опублікованих результатів досліджень пропонує моделі кластеризації в якості вихідних результатів для потреб «розумних міст». Друге місце за поширеністю займають каркаси. Решта результатів можна розділити на глибокі нейронні мережі, класифікацію та моделі кореляції.

Найцікавішим підходом є кредитне плече LSTM. Виходячи з досліджених статей, LSTM дає найвищу точність прогнозів при опрацюванні «великих даних» для потреб «розумних міст та є найшвидшим рішенням у порівнянні з аналогічними рішеннями. Прогнозування соціальних явищ на основі міських даних є найбільш бажаним результатом. Хоча в контексті сучасної ситуації розвитку «розумних міст» це все ще складне завдання, оскільки важливий весь

конвеєр зі складання, обробки та аналізу «великих даних». Як показує проведений аналіз, для різних задач «розумних міст» необхідні різні дані. При цьому, різні алгоритми дають різні результати. Дилема практичних переваг і стандартизації «великих даних» «розумного міста» все ще залишається відкритою.

3.6 Висновок до третього розділу

В третьому розділі кваліфікаційної роботи проаналізовано використання каркасів на основі фреймворків для аналітичного опрацювання інформаційних ресурсів «розумних міст». Також проаналізовано глибинні нейронні мережі при опрацюванні інформаційних ресурсів «розумних міст». Виконано аналіз підходів класифікації «великих даних» для потреб «розумних міст». Розглянуто кореляційний аналіз інформаційних ресурсів для потреб «розумних міст». Проведено аналіз точності методів обробки інформаційних ресурсів «розумних міст» на основі методології «великих даних».

РОЗДІЛ 4. БЕЗПЕКА ЖИТТЄДІЯЛЬНОСТІ, ОСНОВИ ОХОРОНИ ПРАЦІ

4.1 Діяльність. Її види та розуміння в безпеці праці

Кваліфікаційна робота освітнього рівня «Бакалавр» прив'ячена аналізу аналізу інформаційних ресурсів «розумних міст» на основі методології «великих даних». Розвиток сучасних міст тісно пов'язаний з різнотиповими видами діяльності. Тому в параграфі «Безпека життєдіяльності» доцільно розглянути поняття «діяльність» та охарактеризувати її види та розуміння в безпеці праці.

Слово «життєдіяльність» поєднує в собі два поняття: «життя» та «діяльність». Життя – це унікальна форма існування матерії, яка відрізняється від інших здатністю до розмноження, росту, розвитку, активного регулювання своїх функцій та різноманітних форм руху.

Отже, поняття «життя» вже певною мірою включає в себе «діяльність». Іншими словами, життя – це не просто статична форма існування, а динамічний процес, що постійно розвивається. Всі живі істоти постійно взаємодіють з навколишнім середовищем, шукають їжу, розмножуються, захищаються від хижаків та пристосовуються до мінливих умов. Ця активність, або «діяльність», є невід'ємною частиною життя.

Людська діяльність – це не просто реакція на навколишнє середовище, а цілеспрямована трансформація світу на користь людини. Вона включає в себе постановку мети, вибір способів її досягнення, сам процес діяльності та її результат [32]. На відміну від інших живих істот, людина не лише пристосовується до довкілля, але й активно змінює його під свої потреби. Вона усвідомлено ставить перед собою цілі, обирає способи їх досягнення та діє, щоб досягти бажаного результату.

Під життєдіяльністю людини слід розуміти не просто дії в навколишньому середовищі, а збалансований процес існування та самореалізації. Це стосується як окремих людей, так і груп, суспільства та

людства загалом. Життєдіяльність людини ґрунтується на єдності її життєвих потреб та можливостей.

Іншими словами:

- Людина не просто існує в світі, вона творить світ навколо себе.
- Людська діяльність спрямована на задоволення потреб та реалізацію цілей.
- Життєдіяльність людини – це процес саморозвитку та самовдосконалення.

Важливо зазначити:

- Людська діяльність має як позитивні, так і негативні наслідки.
- Відповідальність за наслідки своєї діяльності лежить на самій людині.
- Людська діяльність повинна бути спрямована на гармонійне співіснування з природою та на благо всього людства.

Види діяльності:

- виробнича;
- побутова;
- наукова;
- освітня і т.п.

Життя і діяльність взаємопов'язані та обумовлюють одне одного. Неможливо уявити життя без діяльності і навпаки. Людина постійно взаємодіє з навколишнім середовищем, змінюючи його, а середовище, у свою чергу, впливає на її життєдіяльність. Ця взаємодія здійснюється через наявність прямих і зворотних зв'язків.

Будь-яка господарська діяльність має приносити користь суспільству, а в бізнесі – бути ще й рентабельною. Водночас така діяльність може мати негативні наслідки для життя і здоров'я її учасників або викликати інші види шкоди, такі як моральна, що призводить до травм, захворювань, повної втрати працездатності чи смерті [33]. Шкода працівникам може бути завдана як під час роботи на виробництві, так і під час відпочинку, розваг або навіть у процесі навчання. Історія та практика показують, що будь-яка діяльність у сучасних

умовах є потенційно небезпечною. Абсолютна безпека трудової діяльності, коли виключені всі ризики для працівників, не існує. Тому роботодавці мають усвідомити важливість забезпечення безпеки праці. Без належного методичного інструментарію, який починається з ідентифікації поняття «безпека праці», практичне забезпечення захисту працівників від різних небезпек і загроз неможливе.

Життєдіяльність (ЖД) – це форма організації життя і цілеспрямованої діяльності, яка повністю або частково задовольняє всі потреби і запити людини. Життєдіяльність (ЖД) – це не просто сукупність дій, які виконує людина, а цілеспрямована система, що організовує життя та діяльність для задоволення потреб людини [34]. Вона має ряд ключових характеристик. Цілеспрямованість. ЖД завжди має певну мету, яку прагне досягти людина. Це може бути задоволення базових потреб, таких як їжа, вода та житло, або більш складні цілі, пов'язані з саморозвитком, творчістю чи служінням суспільству. Організованість. ЖД не є хаотичним набором дій. Вона має чітку структуру, яка включає в себе планування, виконання та оцінку результатів. Задоволення потреб. ЖД спрямована на задоволення як біологічних, так і соціальних потреб людини. Це включає в себе потребу в їжі, воді, безпеці, спілкуванні, самореалізації та інших важливих аспектах життя.

– Взаємодія з середовищем. ЖД неможлива без взаємодії з навколишнім середовищем. Людина використовує ресурси середовища для задоволення своїх потреб і водночас впливає на нього своїми діями.

Вивчення ЖД неможливе без урахування взаємозв'язків між людиною та її оточенням. Тому важливо досліджувати людину і її середовище в рамках системи «людина – життєве середовище». Цей системний підхід дозволяє нам: зрозуміти, як фактори навколишнього середовища впливають на людину; оцінити вплив людини на навколишнє середовище; розробити стратегії стійкого розвитку, які враховують потреби як людей, так і довкілля. Вивчення ЖД – це багатогранна та складна область, яка має велике значення для розуміння людини та її місця в світі.

4.2 Електробезпека на будівельному майданчику

Кваліфікаційна робота присвячена аналізу інформаційних ресурсів «розумних міст» на основі методології «великих даних». Протягом усієї історії людства міста були центрами інновацій та технологій. Сучасні інформаційно-технологічні концепції «Розумних міст» розвиваються шляхом вдосконалення та розширення існуючих міст і міських агломерацій. Це включає широкий спектр різноманітних будівельних робіт. Тому у розділі «охорона праці» варто приділити увагу питанням електробезпеки на будівельному майданчику.

Електрифіковані, пневматичні та інші інструменти, які видаються працівникам, мають бути справними та відповідати вимогам, зазначеним в інструкції або паспорті виробника, і застосовуватися відповідно до виду виконуваної роботи. Заборонено працювати механізованими інструментами з приставних, розсувних і навісних драбин. До роботи з електрифікованими та пневматичними інструментами допускаються особи, які досягли 18 років, пройшли медичний огляд і ознайомилися з правилами користування інструментами, технікою безпеки та методами штучного дихання [35].

Не рідше одного разу на квартал ізоляцію струмопровідних частин електрифікованих інструментів слід перевіряти мегометром, а результати заносити в спеціальний журнал. Забороняється користуватися електрифікованими інструментами під час дощу або снігопаду, якщо робоче місце не має накриття. Працювати з такими інструментами потрібно у спецодязі (комбінезон і головний убір), користуючись діелектричними гумовими рукавичками, килимками і захисними окулярами. Під час роботи в обмежених умовах, лежачи або на колінах, необхідно використовувати налокітники, наколінники та діелектричні шоломи. У всіх випадках не можна працювати в тканинних рукавичках. Необхідно уникати пошкодження струмопровідного проводу робочим інструментом.

Електропроводку в приміщеннях, де проводяться оздоблювальні роботи з водними розчинами, вимикають на час фарбування. Тимчасова переносна

електропроводка для зовнішніх і внутрішніх робіт на відкритому повітрі та в особливо небезпечних приміщеннях повинна мати напругу не вище 12 В, а в приміщеннях з підвищеною безпекою ураження електричним струмом – не вище 36 В.

При організації будівельного майданчика, визначенні розташування дільниць, робочих місць, проїздів для будівельних машин і транспортних засобів, а також проходів для людей, встановлюють небезпечні зони для людей. У цих зонах постійно діють або можуть потенційно діяти небезпечні виробничі фактори, що робить їх небезпечними для перебування людей. Небезпечні зони позначають знаками безпеки та відповідними написами встановленої форми; ці зони можуть бути постійними або тимчасовими [36]. До зон постійно діючих небезпечних виробничих факторів належать:

- поблизу неізольованих струмопровідних частин електроустановок;
- поблизу необгороджених перепадів заввишки 1,3 м і більше;
- у місцях переміщення машин і обладнання або їхніх частин і робочих органів;
- у місцях, де є шкідливі речовини, що перевищують граничнодопустимі концентрації або діє шум, що перевищує гранично допустиму інтенсивність;
- у місцях, над якими переміщують вантажі вантажопідіймальними кранами.

До зон потенціально діючих небезпечних виробничих факторів належать:

- дільниці території поблизу споруджуваних будинків;
- поверхи, яруси та частини будівель, над якими проходить монтаж (демонтаж) конструкцій чи обладнання.

Тимчасовими вважаються небезпечні зони, що утворюються під час виконання робіт протягом однієї робочої зміни, наприклад, під час вибухових робіт або монтажу крана. Небезпечні зони та їхні межі визначаються проектами організації будівництва і виконання робіт, і можуть бути уточнені безпосередньо на об'єктах у процесі будівництва.

Визначення меж небезпечних зон для різних видів і умов робіт має велике значення, оскільки, за аналізом виробничого травматизму в будівництві, майже 20% нещасних випадків трапляються з працівниками, які перебувають у таких зонах. Інженерно-технічним працівникам будівельних організацій потрібно звертати особливу увагу на умови роботи в цих зонах.

Якщо окремі робочі місця розміщені на висоті (монтажники, покрівельники, оздоблювальники), то небезпечною зоною вважається ділянка, розташована внизу під робочим майданчиком, межі якого визначають горизонтальною проекцією майданчика, збільшеною по своєму контуру на безпечну величину $0,3 H$, де H — висота, на якій ведуть роботи .

Межі небезпечних зон, де бувають ураження електричним струмом, залежать від напруги в електроустановках [37]. При земляних роботах небезпечною вважається зона в межах призми обвалення ґрунту плюс 1 м.

Під час підривних робіт найбільший радіус небезпечної зони визначають дією повітряної хвилі. Межі небезпечних зон поблизу частин і робочих органів машин, що рухаються, становлять до 5 м, якщо інших підвищених вимог немає в паспорті чи інструкції заводу-виготовлювача. Усі входи в будівлю мають бути захищені зверху щільним навісом завширшки не менше ширини входу з відстанню не ближче ніж 2 м від стіни будівлі. Захисні огорожі розраховують на міцність, вони повинні мати висоту не менш як 1,1 м, а відстань між горизонтальними елементами – не більш як 0,45 м.

4.3 Висновок до четвертого розділу

В четвертому розділі кваліфікаційної роботи описано діяльність, її види та розуміння в безпеці праці. Окремо висвітлено питання електробезпеки на будівельному майданчику, оскільки сучасні інформаційно-технологічні концепції «Розумних міст» розвиваються шляхом вдосконалення та розширення існуючих міст і міських агломерацій.

ВИСНОВКИ

В першому розділі кваліфікаційної роботи освітнього рівня «Бакалавр»:

- Висвітлено актуальність проведення аналізу ресурсів «розумного міста» за допомогою методології «великих даних».
- Подано методологію аналізу ресурсів «розумного міста» на основі «великих даних».
- Описано результати наукометричного пошуку в царині опрацювання інформаційних ресурсів «розумних міст» на основі методології «великих даних».

В другому розділі кваліфікаційної роботи:

- Проаналізовано типи та джерела великих даних «розумних міст»
- Розглянуто ключові категорії алгоритмів опрацювання «великих даних» «розумних міст».
- Проаналізовано кластеризацію «великих даних» для потреб «розумних міст».

В третьому розділі кваліфікаційної роботи:

- Проаналізовано використання каркасів на основі фреймворків для аналітичного опрацювання «великих даних» «розумних міст».
- Проаналізовано глибинні нейронні мережі при опрацюванні «великих даних» для потреб «розумних міст».
- Виконано аналіз підходів класифікації «великих даних» для потреб «розумних міст».
- Описано кореляційний аналіз інформаційних ресурсів для потреб «розумних міст».
- Проведено аналіз точності методів обробки інформаційних ресурсів «розумних міст» на основі методології «великих даних».

У розділі «Безпека життєдіяльності, основи охорони праці» описано діяльність, її види та розуміння в безпеці праці. Окремо висвітлено питання електробезпеки на будівельному майданчику.

ПЕРЕЛІК ДЖЕРЕЛ

- 1 Kwon O, Kim YS, Lee N, Jung Y. When collective knowledge meets crowd knowledge in a smart city: a prediction method combining open data keyword analysis and case-based reasoning. *J Healthc Eng* 2018;2018. <https://doi.org/10.1155/2018/7391793>.
- 2 Wieclaw L.; Pasichnyk V.; Kunanets N.; Duda O.; Matsiuk O.; Falat P. Cloud computing technologies in “smart city” projects. In *Proceedings of the 2017 9th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS)*. Bucharest: Romania, 21–23 September 2017. pp. 339–342.
- 3 Chehri A, Fofana I, Yang X. Security risk modeling in smart grid critical infrastructures in the era of big data and artificial intelligence. *Sustainability*. 2021;13(6):3196. <https://doi.org/10.3390/su13063196>.
- 4 Duda O., Kunanets N., Matsiuk O., Pasichnyk V., Cloud-based IT Infrastructure for “Smart City” Projects, in *Dependable IoT for Human and Industry: Modeling, Architecting, Implementation*. River Publishers, 2018. P. 389-410.
- 5 Wang X, Zou Z. Open data based urban for-profit music venues spatial layout pattern discovery. *Sustainability*. 2021;13(11):6226. <https://doi.org/10.3390/su13116226>.
- 6 Gutman AJ, Goldmeier J. *Becoming a data head: how to think, speak, and understand data science and machine learning*, indianapolis: statistics. 2021.
- 7 Gubareva, Regina, and Rui Pedro Lopes. "Literature Review on the Smart City Resources Analysis with Big Data Methodologies." *SN Computer Science* 5.1 (2024): 152.
- 8 Naeem M, Jamal T, Diaz-Martinez J, Butt SA, Montesano N, Tariq MI, De-la-Hoz-Franco E, De-La-Hoz-Valdiris E. Trends and future perspective challenges in big data. In: Pan J-S, Balas VE, Chen C-M, editors. *Advances in intelligent data analysis and applications*. Singapore: Springer; 2022. p. 309–25.

- 9 Materla T, Cudney EA, Antony J. The application of Kano model in the healthcare industry: a systematic literature review. *Total Qual Manag Bus Excell.* 2017;1–22. <https://doi.org/10.1080/14783363.2017.1328980>.
- 10 Subhash S, Cudney EA. Gamified learning in higher education: a systematic review of the literature. *Comput Hum Behav.* 2018;87:192–206. <https://doi.org/10.1016/j.chb.2018.05.028>.
- 11 Trilles S, Belmonte O, Schade S, Huerta J. A domain-independent methodology to analyze IoT data streams in real-time. A proof of concept implementation for anomaly detection from environmental data. *Int J Digit Earth.* 2017;10(1):103–20. <https://doi.org/10.1080/17538947.2016.1209583>.
- 12 Bermudez-Edo M, Barnaghi P, Moessner K. Analysing real world data streams with spatio-temporal correlations: entropy vs. pearson correlation. *Autom Construct.* 2018;88:87–100. <https://doi.org/10.1016/j.autcon.2017.12.036>.
- 13 Bordogna G, Cuzzocrea A, Frigerio L, Psaila G. An effective and efficient similarity-matrix-based algorithm for clustering big mobile social data. 2017;514–521. <https://doi.org/10.1109/ICMLA.2016.188>.
- 14 Wang G, Wei W, Jiang J, Ning C, Chen H, Huang J, Liang B, Zang N, Liao Y, Chen R, Lai J, Zhou O, Han J, Liang H, Ye L. Application of a long short-term memory neural network: a burgeoning method of deep learning in forecasting HIV incidence in Guangxi, China. *Epidemiol Infect.* 2019;147:194. <https://doi.org/10.1017/S095026881900075X>.
- 15 Pérez-Chacón R, Luna-Romera JM, Troncoso A, Martínez-Alvarez F, Riquelme JC. Big data analytics for discovering electricity consumption patterns in smart cities. *Energies* 2018;11 (3). <https://doi.org/10.3390/en11030683>.
- 16 Karyotis V, Tsitseklis K, Sotiropoulos K, Papavassiliou S. Big data clustering via community detection and hyperbolic network embedding in IoT applications. *Sens (Switzerl)* 2018;18(4). <https://doi.org/10.3390/s18041205>.
- 17 Azri S, Ujang U, Abdul Rahman A. Dendrogram clustering for 3D data analytics in smart city. *Issue: 4/W9.* 2018;42:247–53. <https://doi.org/10.5194/isprs-archives-XLII-4-W9-247-2018>.

18 AlShami A, Guo W, Pogrebna G. Fuzzy partition technique for clustering Big Urban dataset. In: 2016 SAI Computing Conference (SAI), 2016; 212–216. <https://doi.org/10.1109/SAI.2016.7555984>.

19 Chang C-S, Chang C-T, Lee D-S, Liou L-H. K-sets+: a lineartime clustering algorithm for data points with a sparse similarity measure. 2018;1–8. <https://doi.org/10.1109/USIC-ATC.2017.8397636>.

20 Chae S, Kwon S, Lee D. Predicting infectious disease using deep learning and big data. *Int J Environ Res Public Health*. 2018;15(8):1596. <https://doi.org/10.3390/ijerph15081596>.

21 Chen W, Huang Z, Wu F, Zhu M, Guan H, Maciejewski R. VAUD: a visual analysis approach for exploring spatio-temporal urban data. *IEEE Trans Visual Comput Graph*. 2018;24(9):2636–48. <https://doi.org/10.1109/TVCG.2017.27583>.

22 Simhachalam B, Ganesan G. Performance comparison of fuzzy and non-fuzzy classification methods. *Egyp Inform J*. 2016;17(2):183–8. <https://doi.org/10.1016/j.eij.2015.10.004>.

23 Sassite F, Addou M, Barramou F. A machine learning and multiagent model to automate big data analytics in smart cities. *Int J Adv Comput Sci Appl*. 2022;13(7):441–51. <https://doi.org/10.14569/IJACSA.2022.0130754>.

24 Huang J, Zhu L, Liang Q, Fan B, Li S. Efficient classification of distribution-based data for internet of things. *IEEE Access*. 2018;6:69279–87. <https://doi.org/10.1109/ACCESS.2018.2879652>.

25 Liu Q, Huan W, Deng M. A method with adaptive graphs to constrain multi-view subspace clustering of geospatial big data from multiple sources. *Remote Sens*. 2022;14(17):4394. <https://doi.org/10.3390/rs14174394>.

26 Bashir S. Real-time water and electricity consumption monitoring using machine learning techniques. *IEEE Access*. 2023;11:11511–28. <https://doi.org/10.1109/ACCESS.2023.3241489>.

27 Charalampous A, Papadopoulos A, Hadjiyiannis S, Philimis P. Towards hydro-informatics modernization with real-time water consumption classification. In:

2021 IEEE AFRICON, 1– 6 2021. ISSN: 2153-0033. <https://doi.org/10.1109/AFRICON51333.2021.9570909>.

28 Cao W, Zhang H, Li J. A Grey relevancy analysis on the relationship between energy consumption and economic growth in Henan Province. In: 2011 2nd International Conference on Artificial Intelligence, Management Science and Electronic Commerce (AIMSEC). 27– 30. 2011. <https://doi.org/10.1109/AIMSEC.2011.6010196>.

29 Girvan M, Newman MEJ. Community structure in social and biological networks. Proc Natl Acad Sci. 2002;99(12):7821–6. <https://doi.org/10.1073/pnas.122653799>.

30 Duda O., Kunanets N., Matsiuk O., Pasichnyk V. Information-Communication Technologies of IoT in the "Smart Cities" Projects", CEUR Workshop Proceedings. 2018. Vol. 2105. P. 317- 330.

31 Duda, O., Pasichnyk, V., Kunanets, N., Antonii, R., Matsiuk, O. Multidimensional Representation of COVID-19 Data Using OLAP Information Technology. International Scientific and Technical Conference on Computer Sciences and Information Technologies, 2020, 2, pp. 277–280, 9321889.

32 Категорійно-понятійний апарат з безпеки життєдіяльності, таксономія небезпек. <http://opcb.kpi.ua/wp-content/uploads/2014/09/%D0%9A%D0%BE%D0%BD%D1%81%D0%BF%D0%B5%D0%BA%D1%82%D0%9E%D0%9F%D0%A6%D0%91-%D0%A0%D0%BE%D0%B7%D0%B4%D1%96%D0%BB1.pdf>.

33 Северин, Олександр Петрович, et al. "Охорона праці та безпека військової діяльності." (2018).

34 Левченко, Олег Григорович, et al. "Безпека життєдіяльності та цивільний захист." (2019).

35 Безпека в надзвичайних ситуаціях. Методичний посібник для здобувачів освітнього ступеня «магістр» всіх спеціальностей денної та заочної (дистанційної) форм навчання / укл.: Стручок В. С. Тернопіль: ФОП Паляниця В. А., 2022. 156 с.

- 36 Курепін, Вячеслав Миколайович, and Валерій Миколайович Курепін. "Актуальні питання охорони праці в енергетичної галузі України." (2019).
- 37 Левченко, Олег Григорович, et al. "Охорона праці та цивільний захист." Київ: КПІ ім. Ігоря Сікорського (2019).

ДОДАТКИ

Оцінка точності проаналізованих алгоритмів опрацювання «великих даних» для потреб «розумних міст» [7]

Алгоритм	Призначення	Точність
1	2	3
CUSUM	Виявлення аномалій	–
Взаємна інформація та кореляція Пірсона	Знайти кореляцію між сенсорними даними	+
LSTM	Передбачення	85%+
ARIMA	Передбачення	80%
ГРНН	Передбачення	76%
ES	Передбачення	74%
К-означає кластеризацію	Виділити закономірності споживання електроенергії	78%
модифікація алгоритму Гірвана-Ньюмена	Виявлення спільноти	65–100%
Кластеризація дендрограми	Створення ієрархічної деревовидної структури	–
К-Means	Кластеризація	87,94%
Нечітке с-середнє	Кластеризація	81,91%
К-набори+	Кластеризація в метричному просторі	95%
DNN	Прогнозування інфекційних захворювань	77%
VAUD	Просторово-часова візуалізація даних	78.6%+
К-Means	Кластеризація	60.41–87.94%
Нечіткі с-середні (FCM)	Кластеризація	56.25–81,91%
Густафсон-Кессель (GK)	Кластеризація	66.19–95.83%
Кластеризація на основі матриці подібності	Кластеризація траєкторій	–
Кластеризація підпростору з кількома видами	Кластеризація	68.11%
Наївний Байєс	Класифікація	61%
Багатошаровий перцептрон	Класифікація	73%
Випадковий ліс	Класифікація	85%

1	2	3
Класифікація даних на основі розподілу	Класифікація	90%,44%
Класифікація електричних пристроїв	Класифікація	99,93%
Класифікація водоспоживання	Класифікація	91%
Ієрархічна агломеративна кластеризація з евклідовою відстанню	Кластеризація	72%