

НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ БІОРЕСУРСІВ
І ПРИРОДОКОРИСТУВАННЯ УКРАЇНИ

НУБІП України

Факультет інформаційних технологій

УДК 004.9:502.3
«ПОГОДЖЕНО» «ДОПУСКАЄТЬСЯ ДО
ЗАХИСТУ»

Декан факультету

інформаційних технологій

Завідувач кафедри комп'ютерних наук

Глазунова О.Г., д.п.н., професор

Голуб Б.Л., к.т.н., доцент

НУБІП України 2022 р. 2022 р.

МАГІСТЕРСЬКА КВАЛІФІКАЦІЙНА РОБОТА

на тему «Система дослідження атмосферного повітря на основі сучасних
інформаційних технологій»

Спеціальність 121 «Інженерія програмного забезпечення»

Освітня програма «Програмне забезпечення інформаційних систем»

Орієнтація освітньої програми освітньо-професійна

Гарант освітньої програми

к.т.н., доцент

Голуб Б.Л.

Керівник магістерської кваліфікаційної роботи

к.т.н., доцент

Голуб Б.Л.

Виконав

Кондратенко В.С.

НУБІП України

КИЇВ-2022

НУБІП України

ЗАТВЕРДЖУЮ

Завідувач кафедри комп'ютерних наук

к.т.н., доцент, Голуб Б.Л.

(науковий ступінь, вчене звання) (підпис) (ІПБ)

“01” листопада 2021 року

ЗАВДАННЯ

ДО ВИКОНАННЯ МАГІСТЕРСЬКОЇ КВАЛІФІКАЦІЙНОЇ РОБОТИ СТУДЕНТУ

Кондратенку Владиславу Сергійовичу

(прізвище, ім'я, по батькові)

Спеціальність 121 «Інженерія програмного забезпечення»

Освітня програма «Програмне забезпечення інформаційних систем»

Орієнтація освітньої програми: освітньо-професійна

Тема магістерської кваліфікаційної роботи «Система аналізу атмосферного повітря на основі сучасних технологій»

затверджена наказом ректора НУБіП України від “01” листопада 2021 р. №1861 «С»

Термін подання завершеної роботи на кафедру 27 жовтня 2022 р.

Вихідні дані до магістерської кваліфікаційної роботи:

- дані про станції;
- дані про екологічні показники;
- дані про стан атмосферного повітря.

Перелік питань, що підлягають дослідженню:

- Аналіз предметної області
- Дослідження інструментів OLAP
- Проектування системи
- Дослідження інструментів Data Mining
- Розробка алгоритмів аналізу даних
- Дослідження отриманих результатів

Дата видачі завдання “01” листопада 2021 р.

Керівник магістерської кваліфікаційної роботи

_____ (підпис)

Голуб Б.Л.
(прізвище та ініціали)

Завдання прийняв до виконання

_____ (підпис)

Кондратенко В.С.
(прізвище та ініціали)

ЗМІСТ

НУБІП України	ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ	5
	ВСТУП	6
	1 СИСТЕМНИЙ АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ	10
НУБІП України	1.1 Процеси аналізу стану атмосферного повітря	10
	1.2 Технічне завдання	15
	1.3 Аналіз наявних рішень	16
	2 МОДЕЛЮВАННЯ ПРОЦЕСІВ ПРЕДМЕТНОЇ ОБЛАСТІ	20
НУБІП України	2.1 Загальні відомості про моделювання бізнес-процесів засобами UML	20
	2.2 Діаграма прецедентів	20
	2.3 Діаграма послідовності	23
НУБІП України	2.4 Діаграма активності	25
	3 РОЗРОБКА СИСТЕМИ	28
	3.1 Архітектура системи	28
НУБІП України	3.2 Оперативні джерела інформації	29
	3.3 Сховище даних	30
	3.4 Побудова розгорнутого куба підсистеми аналізу, SSAS	33
	3.5 Реалізація міграції даних за допомогою Data Flow, SSIS	37
НУБІП України	3.6 Побудова звітності засобами SSRS	40
	3.7 Розрахунок KPI	42
	3.8 Проектування моделей інтелектуального аналізу засобами Data Mining	43
НУБІП України	3.9 Застосовані алгоритми аналізу даних засобами Data Mining	45
	4 РЕЗУЛЬТАТИ ДОСЛІДЖЕННЯ	51

4.1 Результати дослідження на основі оперативного аналізу 51

4.2 Результати дослідження на основі інтелектуального аналізу 55

ВИСНОВКИ..... 59

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ..... 62

НУБІП України

НУБІП України

НУБІП України

НУБІП України

НУБІП України

НУБІП України

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ

БД – база даних.
СД – сховище даних.

СППР – система підтримки прийняття рішень.

СУБД – система управління базами даних.

DSV – Data Source View.

DF – Data Flow.

KPI – ключовий показник ефективності.

OLAP – On-Line Analytical Processing

SSAS – SQL Server Integration Services.

SSIS – SQL Server Integration Services

SSRS – SQL Server Reporting Services.

НУБІП України

НУБІП України

НУБІП України

НУБІП України

ВСТУП

Актуальність

Для покращення рівня життя людство постійно розширює сфери виробничої діяльності. Але результатом цієї діяльності є не тільки блага, що задовольняють потреби сучасної людини, але й збільшення антропогенного навантаження на навколишнє середовище. На сьогоднішній день забруднення атмосферного повітря, являє собою одним із основних факторів ризику для здоров'я людини.

Крихітні, невидимі частинки у повітрі глибоко проникають у наші легені та кров. Ці забруднювачі є причиною приблизно третини смертей від інсульту, хронічних респіраторних захворювань та раку легень. Приземний озон, що утворюється при взаємодії багатьох різних забруднювачів при сонячному світлі, веде до астми та хронічних респіраторних захворювань.

Від наслідків забруднення повітря щороку вмирають сім мільйонів людей. Лікування захворювань, викликаних загазованістю, коштує приблизно 1 трильйон доларів на рік [1].

Забруднення повітря – серйозна загроза не тільки здоров'ю населення, а й довкілля. Воно знижує вміст кисню в наших океанах, веде до скорочення біорізноманіття та сприяє зміні клімату.

Через викиди парникових газів виникає дисбаланс у газовому складі повітря. Це призводить до підвищення температури Землі – глобального потепління. В результаті спостерігається танення льодовиків і підвищення рівня моря.

Спалювання викопного палива виділяє в повітря шкідливі гази, такі як оксиди азоту та оксиди сірки. Краплі води поєднуються з цими забруднювачами та випадають у вигляді кислотних дощів, що завдає шкоди людям, тваринам і рослинам.

Тому в сучасних реаліях дуже важливо досліджувати атмосферне повітря, аналізувати поточний його стан, визначати чинники, що впливають на його стан

та прогнозувати його зміну у майбутньому для прийняття відповідних заходів для його поліпшення.

Об'єкт та предмет дослідження

Об'єктом дослідження є стан атмосферного повітря, чинники, що впливають на його зміну та заходи прийняття відповідних рішень для його покращення.

Предметом дослідження є система дослідження атмосферного повітря, що дозволить аналізувати поточний стан атмосферного повітря, прогнозувати його зміну для прийняття відповідних заходів для його покращення.

Мета дослідження

Метою дослідження є визначення доцільності використання сучасних інформаційних технологій OLAP та Data Mining для аналізу стану атмосферного повітря та його змін.

Перелік поставлених завдань

Процеси дослідження та розробки інтелектуальної системи, що є об'єктом дослідження даної наукової роботи були розбиті на етапи:

1. Провести аналіз предметної області. Визначити її проблему, актуальність та методи вирішення цієї проблеми.
2. Виходячи з аналізу предметної області визначити набір інструментів, що будуть використані для створення СУПР.
3. Обрати та побудувати архітектуру програмної системи, створити сховище даних для збереження інформації для аналізу.
4. Обрати алгоритми для аналізу стану атмосферного повітря.
5. Проаналізувати результати дослідження.
6. Сформулювати висновки щодо доцільності використання OLAP та Data Mining технологій для аналізу стану атмосферного повітря.

Методи дослідження

За допомогою інструментів OLAP було створено структуру даних у вигляді багатомірного кубу для подальшого аналізу, побудовано ряд звітів та графіків стосовно тенденцій зміни стану атмосферного повітря.

Інструменти Data Mining були використані для прогнозування зміни стану атмосферного повітря в майбутньому на основі існуючих даних та для знаходження чинників та залежностей, що впливають на те чи інше значення екологічних показників.

Наукова новизна

1. Було вдосконалено архітектуру системи дослідження атмосферного повітря.
2. Запропоновано використання технологій OLAP та Data Mining для інтелектуального аналізу стану атмосферного повітря.

Апробація результатів дослідження

Кондратенко В.С., Голуб Б.Л. Система аналізу атмосферного повітря на основі сучасних інформаційних технологій. Збірник матеріалів XIII Міжнародної науково-практичної конференції молодих вчених “Інформаційні технології: економіка, техніка, освіта”. 26-27 жовтня 2022 року, НУБіП України, Київ. – С. 22. Режим доступу: 22

Структура магістерської роботи

Структура магістерської роботи складається з чотирьох розділів.

У першому розділі описано результати аналізу предметної області, формування технічного завдання та наведено інформацію про виявні рішення.

Другий розділ присвячено функціональному моделюванню предметної області засобами UML, побудовано діаграму прецедентів, активності та послідовності.

Третій розділ описує технічну складову магістерської роботи, а саме: побудова архітектури системи, створення сховища даних, застосування інструментів OLAP та алгоритмів Data Mining для проведення аналізу даних.

У четвертому розділі описано аналіз даних, що отримано під час застосування алгоритмів Data Mining та сформовано відповідні висновки.

Магістерська робота складається з текстового матеріалу, що доповнюється діаграмами, таблицями, графіками.

- Кількість ілюстрацій – 60.

1 СИСТЕМНИЙ АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ

1.1 Процеси аналізу стану атмосферного повітря

Одним із основних життєво важливих компонентів навколишнього середовища є атмосферне повітря.

Забруднення атмосферного повітря відбувається різними елементами: твердими частинками, хімічними речовинами і біологічними матеріалами, що здатні заподіяти шкоду як людині, так й іншим живим організмам.

Всесвітня організація охорони здоров'я констатує, що забруднення повітря призводить до збільшення захворюваності та смертності в світі. За її даними, погіршення стану атмосферного повітря є основним чинником ризику для здоров'я населення. Понад 80% захворювань тією чи іншою мірою залежать від якості повітря [1].

Від забруднення повітря щороку в світі передчасно помирають мільйони людей. Мільярди щодня змушені дихати насиченим пилом та отруйними сполуками повітрям. На сьогоднішній день загалом близько 90% дітей проживають у містах, у яких повітря забруднене різними шкідливими речовинами [2].

Забруднення повітря відбувається за рахунок природних та антропогенних джерел. Природними джерелами вважаються пилові бурі, пожежі в лісах, виверження вулканів, космічний пил, руйнування гір тощо. Разом з тим, значно більше забруднення відбувається безпосередньо через діяльність людини.

Антропогенне забруднення – це розвиток промисловості: робота теплових електростанцій, металургійних заводів, видобуток і використання радіації, хімічна промисловість, виробничі і побутові відходи, всі види транспорту, використання хімічних засобів у сільському господарстві і т.д. [3].

Відповідно для того, щоб проводити оцінку та аналіз забруднення атмосферного повітря існують спеціальні концентрації тих чи інших шкідливих речовин, які потрапляють в атмосферне повітря чинниками, що були описані вище.

Це так звані граничні допустимі концентрації (ГДК), які бувають середньодобовими, середньомісячними та максимально-разовими (за 1 годину).

Середньодобові ГДК стосуються тривалості дні забруднюючих речовин з ними порівнюються середньомісячні, середньодобові і максимально разові концентрації згідно з гігієнічними регламентами допустимого вмісту хімічних і біологічних речовин в атмосферному повітрі населених місць [5].

Середньомісячні, середньодобові і максимальні разові значення концентрації забруднювальних речовин, вираховуються на основі вимірювань, що фіксуються на автоматичних станціях мережі громадських та державних

системах моніторингу. Для діоксиду сірки максимальна одногодинна концентрація в 350 мкг/м^3 не повинна перевищувати більше ніж 24 раз на календарний рік, максимальна добова концентрація в 125 мкг/м^3 – більше ніж три рази на календарний рік.

Згідно Постанови КМУ № 827 при оцінюванні стану атмосферного повітря необхідно враховувати також верхній та нижній пороги оцінювання. Дані постановою особливо жорсткі вимоги встановлені для твердих частинок (ТЧ_{10} – до 10 мкм і $\text{ТЧ}_{2,5}$). У таблиці 1 наведено пороги оцінювання для твердих частинок (ТЧ_{10} та $\text{ТЧ}_{2,5}$).

Пороги оцінювання для твердих часток (ТЧ_{10} , $\text{ТЧ}_{2,5}$)

Таблиця 1

Пороги оцінювання	Середнє значення ТЧ_{10} за 24 години	Середнє значення ТЧ_{10} на рік	Середнє значення $\text{ТЧ}_{2,5}$ на рік
1	2	3	4
Верхній поріг оцінювання	70 % ГВ (35 мкг/м^3) не має бути перевищено більше 35 разів у будь-який календарний рік)	70 % ГВ (28 мкг/м^3)	70 % ГВ ГВ (17 мкг/м^3)

Продовження таблиці 1

1	2	3	4
Нижній поріг оцінювання	50 % ГВ (25 мкг/ м ³) не має бути перевищено більше 35 разів у будь-який календарний рік	50 % ГВ (20 мкг/м ³)	50 % ГВ (12 мкг/м ³)

Також для охорони здоров'я людини та захисту рослин максимальні концентрації діоксиду сірки SO₂ і діоксиду азоту NO₂ необхідно порівнювати граничні величини (ГВ) поллютантів з нормованими значеннями верхнього і нижнього порогів, що наведені у таблицях 2,3.

Для діоксиду і оксидів азоту максимальне значення одномодинної концентрації становить 200 мкг/м³, для оксиду вуглецю (CO) і озону (O₃) встановлені середні значення концентрацій за восьми годинний період, що описано у таблиці 3. Верхній поріг оцінювання, наприклад, для оксиду вуглецю (CO) на рівні 70% від граничної величини (7 мкг/м³) і нижній поріг оцінювання на рівні 50% від граничної величини (5 мкг/м³).

Пороги оцінювання для CO та O₃

Таблиця 2

Пороги оцінювання	CO	Цільовий показник	O ₃
1	2	3	4
Верхній поріг	70% від ГВ (7 мкг/м ³)	Охорона здоров'я людини	120 мкг/м ³ не повинен бути перевищений більше ніж 25 днів на календарний рік; середнє значення вимірюється протягом 3-х років

Продовження таблиці 2

1	2	3	4
Нижній поріг	50 % ГВ (5 мг/м ³)	Охорона рослинності	АОТ 40 (розраховується із значень за одну годину) 80 мг/м ³ за травень-липень із використанням значень, вимірених щогодини кожний день у часовому проміжку між 8-ю і 20-ю годинами 18 000 мкг/м ³ x середнє щогодинне значення за 5 років

Пороги оцінювання для SO₂ та NO₂

Таблиця 3

Пороги оцінювання	Охорона здоров'я		Захист рослинності		
	SO ₂	NO ₂	SO ₂	NO ₂	
		Щогодинна ГВ	Щорічна ГВ		
1	2	3	4	5	6
Нижній поріг	40 % ГВ 24-годинної ГДК (50 мкг/м ³) не має бути перевищено більш як три рази	50% ГВ (100 мкг/ м ³) не має бути перевищено більше 18 разів у будь-який календарний рік	65 % ГВ граничної величини (26 мкг/м ³)	40 % зимового критичного рівня (8 мкг/м ³)	65 % критичного рівня (19,5 мкг/м ³)

Продовження таблиці 3

1	2	3	4	5	6
Верхній поріг	60% 24-годинної ГДК (75 мкг/м ³) не має бути перевищено більш як три рази у будь-який календарний рік)	50% ГВ (140 мкг/м ³) не має бути перевищено більше 18 разів у будь-який календарний рік)	80% ГВ (32 мкг/м ³)	60 % (зимового критичного рівня) (12 мкг/м ³)	80 % критичного рівня (24 мкг/м ³)

Згадана вище Постанова регламентує щорічний критичний рівень для захисту рослинної та природної екосистем від впливу NO₂ (в розмірі 80% від ГДК, або 24 мкг/м³, для верхнього рівня і 65%, або 19,5 мкг/м³ - для нижнього), а для SO₂ - 60% зимового критичного рівня (12 мкг/м³) [6].

Для озону перевищення інформаційного порогу 180 мкг/м³ повинні вимірюватися протягом 3-х послідовних годин (з обов'язковим інформуванням населення в разі перевищення), середньогодинний поріг безпеки встановлено на рівні 240 мкг/м³, а максимальне середньодобове восьмигодинне значення 120 мкг/м³ для озону не може бути перевищене більше ніж 25 днів на календарний рік.

Виходячи з описаного вище можна зробити висновок, що стан атмосферного повітря – це сукупність показників елементів, що входять в склад атмосферного повітря. Наявність шкідливих елементів в атмосферному повітрі обумовлена природними та антропогенними факторами. Відповідні вимірювання та фіксація вмісту тих чи інших шкідливих елементів здійснюється

за допомогою спеціальних станцій моніторингу [4]. Також КМУ та ЄС було прийнято допустимі значення шкідливих елементів, які можуть бути використані під час проведення аналізу стану атмосферного повітря.

1.2 Технічне завдання

У результаті аналізу предметної області було сформовано технічне завдання, що описує вимоги та набір функціональності яким система аналізу атмосферного повітря повинна відповідати.

Для розробки системи необхідно створити сховище даних, що містить всю необхідну інформацію для аналізу стану атмосферного повітря на базі існуючої оперативної бази даних, до якої надходять результати вимірювання екологічних показників від станцій.

Також необхідно реалізувати два варіанти занесення даних до СД:

- за допомогою міграції даних із оперативної бази даних;
- за допомогою спеціального алгоритму зчитування CSV-файлів із зафіксованими показниками та занесення їх до СД.

Сховище даних повинно зберігати таку інформації стосовно вимірювань:

- інформація про одиниці вимірювань;
- інформація про оптимальні значення показників;
- інформація про категорії значень показників;
- інформація про станції, на яких було зафіксовано вимірювання;
- інформація про вимірювання (дата, станція, одиниця вимірювання).

Міграцію даних з оперативної БД до СД необхідно реалізувати за допомогою OLAP інструментів. Реалізацію мануального зчитування CSV-файлів та запис зафіксованих показників необхідно реалізувати утилітарну програму з можливістю завантаження файлу та запуску процесу обчислення середньодобового значення показників та формування SQL-запитів для занесення цих даних до СД. Фрагмент CSV-файлу наведено на рис. 1.

Також за допомогою інструментів OLAP та Data Mining необхідно знайти відповіді на наступні питання:

1. В якій області було зафіксовано найбільший середньодобовий вміст твердих частинок $PM_{2.5}$ в розрізі вказаного часового періоду?
2. В якій області було зафіксовано найбільший середньодобовий вміст твердих частинок PM_{10} в розрізі вказаного часового періоду?
3. Скільки разів було перевищено середньодобове граничне значення вмісту твердих частинок $PM_{2.5}$ на певній станції?
4. Скільки разів було перевищено середньодобове граничне значення вмісту твердих частинок PM_{10} на певній станції?
5. Яка кількість вимірювань озону, що були проведені в розрізі певного часового періоду на певній станції відповідають кожній категорії допустимих значень?

```

1 device_id,phenomenon,value,logged_at,value_text
2 21041,pm1,7.3960,2022-08-15 19:20:00,NULL
3 21041,pm1,7.3960,2022-08-15 19:20:00,NULL
4 21041,pm25,10.5500,2022-08-15 19:20:00,NULL
5 21041,pm25,10.5500,2022-08-15 19:20:00,NULL
6 21041,pm10,19.4780,2022-08-15 19:20:00,NULL
7 21041,pm10,19.4780,2022-08-15 19:20:00,NULL
8 21041,temperature,26.3140,2022-08-15 19:20:00,NULL
9 21041,humidity,68.0640,2022-08-15 19:20:00,NULL
10 21041,temperature,26.3140,2022-08-15 19:20:00,NULL
11 21041,humidity,68.0640,2022-08-15 19:20:00,NULL
12 21041,pm1,7.3270,2022-08-15 19:21:00,NULL
13 21041,pm25,10.1170,2022-08-15 19:21:00,NULL
14 21041,pm10,18.1280,2022-08-15 19:21:00,NULL
15 21041,temperature,26.2980,2022-08-15 19:21:00,NULL
16 21041,humidity,67.6680,2022-08-15 19:21:00,NULL
17 21041,pm1,7.3010,2022-08-15 19:22:00,NULL
18 21041,pm25,10.7650,2022-08-15 19:22:00,NULL
19 21041,pm10,20.7900,2022-08-15 19:22:00,NULL
20 21041,temperature,26.2850,2022-08-15 19:22:00,NULL
21 21041,humidity,67.7190,2022-08-15 19:22:00,NULL

```

Рис. 1 Фрагмент CSV-файлу із зафіксованими показниками

1.3 Аналіз наявних рішень

В результаті аналізу наявних систем аналізу та моніторингу якості атмосферного повітря було вкореєкшено українську розробку під назвою SaveEcoBot. SaveEcoBot – це єдиний в Україні екологічний чат-бот, що поєднує в собі дані про забруднювачів, забруднення та інструменти замісту довідки [7].

SaveEcoBot включає в себе веб-додаток із мапою забруднення атмосферного повітря та чат-боти у Telegram, Messenger та Viber.

Розглянемо більш детально інформацію, що SaveEcoBot агрегує в собі:

- інформацію про радіоакційний фон на території країни;
- інформацію про пожежі;
- інформація про моніторинг якості повітря (стан атмосферного повітря в режимі 24/7, можливість пошуку найближчої станції та отримання автоматичних повідомлень в разі погіршення якості повітря);
- аналітика стану повітря за певний період;
- реєстр ОВД (моніторингу планової господарської діяльності, що має вплив на довкілля, шаблони заяв зауважень тощо);
- дозвільні документи (ліцензії на поводження з небезпечними відходами і т.д.);
- екологічні податки.

Показники якості повітря в Україні збираються зі всіх наявних станцій моніторингу – громадських, комерційних, державних. SaveEcoBot поєднує всю екологічну інформацію про забруднювачів, яку можна отримати онлайн в режимі реального часу [7]. Для прикладу на рис. 2 відображено мапу якості повітря.

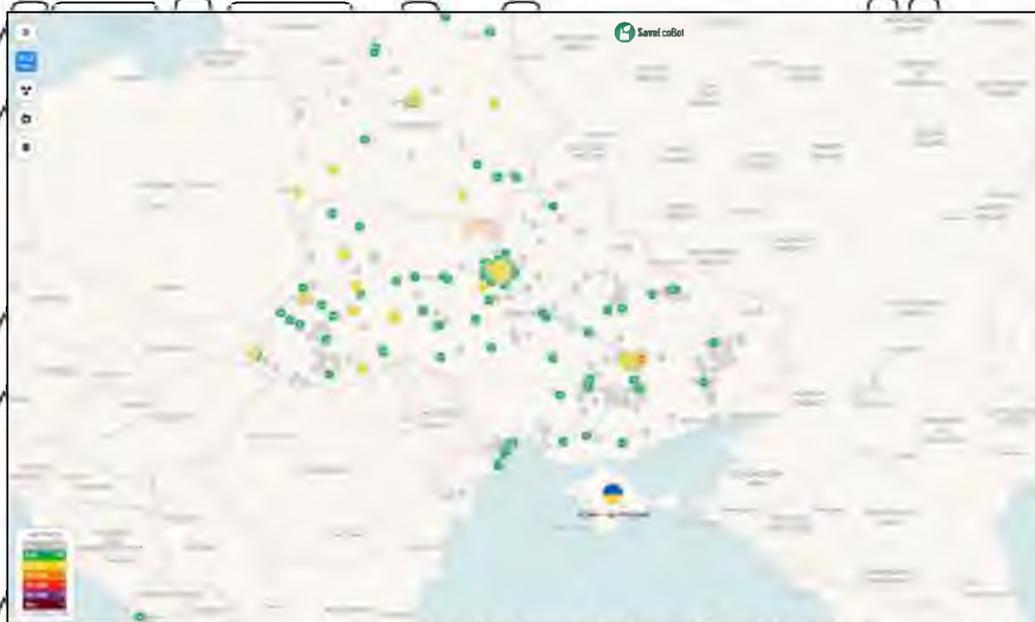


Рис. 2 Мапа якості повітря моніторингової системи SaveEcoBot

SaveEcoBot надає не тільки візуалізацію поточного стану атмосферного повітря, він також має свій API для його інтеграції в сторонні сервіси, а також надає можливість завантажити CSV-файли із зафіксованими екологічними показниками повітря, які саме були використані для проведення дослідження в рамках магістерської роботи.

Також було виділено ще одну систему моніторингу стану атмосферного повітря, що була розроблена силами викладачів та студентів Національного університету біоресурсів та природокористування України в рамках бакалаврських робіт. Дана система підтримує інтеграцію з вище описаною

системою моніторингу SaveEcoBot. Також студентами та викладачами було встановлено на території університету та інтегровано в розроблену систему моніторингу власну станцію вимірювання AirFreshMax.

Дана система має такий функціонал:

- додавання нових станцій моніторингу;
- агрегування екологічних показників вимірювання у власній оперативній базі даних;
- візуалізувати результати вимірювань по станціях;
- надавати користувачам системи звітної інформації стосовно

вимірювань.

Приклад інтерфейсу розробленої моніторингової системи відображено на рис. 3 та рис. 4.

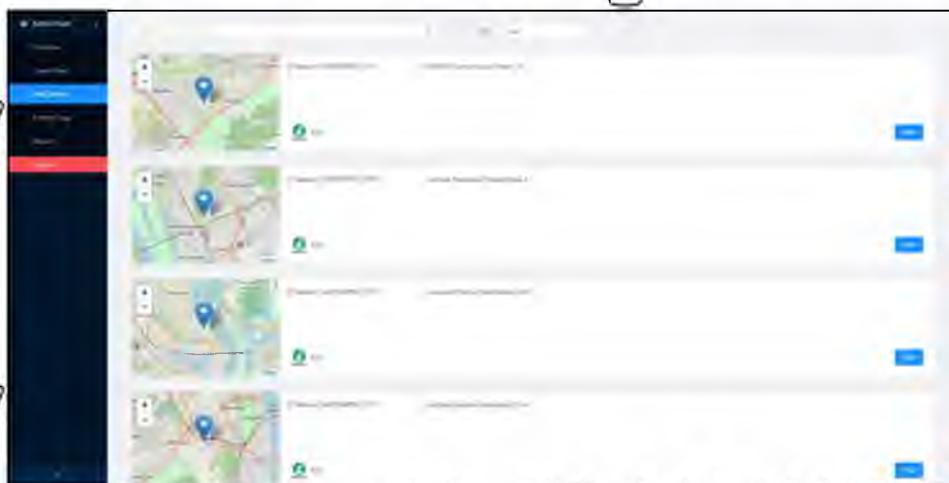


Рис. 3 Інтерфейс додавання нової станції до системи



Рис. 4 Звіт з кількості вимірювань

Також серед подібних моніторингових систем можна виокремити такі, як Eco City, Air Pollution.

НУБІП України

НУБІП України

НУБІП України

НУБІП України

НУБІП України

2 МОДЕЛЮВАННЯ ПРОЦЕСІВ ПРЕДМЕТНОЇ ОБЛАСТІ

2.1 Загальні відомості про моделювання бізнес-процесів засобами

UML

Мова UML (Unified Modeling Language) – це стандартизована мова моделювання, що складається з набору діаграм. Вона була розроблена для того, щоб моделювати та візуалізувати як бізнес-процеси, так і аналізувати, проектувати та реалізовувати програмне забезпечення. UML являє собою колекцію найкращих інженерних практик, які довели свою успішність у моделюванні великих та складних систем. UML є дуже важливою частиною розробки об'єктно-орієнтованого програмного забезпечення та процесу розробки програмного забезпечення. UML використовує переважно графічні позначення для відображення бізнес-процесів та дизайну, архітектур програмних продуктів.

Використання UML допомагає командам розробки ПЗ спілкуватися більш ефективно, досліджувати різні підходи до моделювання та перевіряти архітектурні дизайни програмного забезпечення [8].

Існує кілька типів діаграм UML, і кожна з них має різну мету, незалежно від того, розробляється вона до впровадження чи після (як частина документації). Дві найбільш широкі категорії, які охоплюють усі інші типи, – це поведінкові діаграми та структурні діаграми UML. Із назви категорій можна зробити висновок, що структурні діаграми намагаються проаналізувати та зобразити структуру системи або процесу, тоді як поведінкові описують поведінку системи, її акторів та структурних компонентів. До поведінкових діаграм відносять такі, як: діаграма активності, прецедентів, взаємодії, станів, послідовності та ін. До структурних – діаграму класів, об'єктів, компонентів, розгортання, пакетів [9].

2.2 Діаграма прецедентів

Діаграма варіантів використання – одна із базових діаграм UML, що використовується для узагальнення процесів предметної області або деталей

системи та користувачів у цій системі або акторів, що приймають участь у процесах предметної області. Діаграми варіантів використання відображають події в системі та те, як ці події протікають, однак діаграма варіантів використання не описує, як ці події реалізуються [10].

Основними цілями, для яких використовується діаграма прецедентів використання є наступні.

- відобразити основні цілі системи або процесів;
- відобразити контекст, у рамках якого слід розглядати систему;
- вказати системні вимоги;
- відобразити взаємодію акторів в системі, процесі;
- відобразити загальний вигляд системи;
- представити зовнішній та внутрішній вплив на систему або процес.

Виділяють три основні компоненти діаграми використання UML:

- функціональні вимоги – представлені у вигляді варіантів використання;
- актори – вони взаємодіють із системою; діючою особою може бути людина, організація, внутрішній або зовнішній додаток [10].

В ДОДАТКУ А відображено діаграму прецедентів використання для предметної області аналізу стану атмосферного повітря.

Розглянемо дану діаграму більш детально. В результаті аналізу предметної області було виокремлено таких акторів, як: станція, адміністратор, аналітик, Міністерство екології та основні прецеденти, що більш детально описані у таблиці 4.

Перелік акторів та опис основних прецедентів використання

Таблиця 4

Актор	Опис
1	2
Станція	Фіксує показники стану атмосферного повітря та надсилає їх до MQTT серверу

Адміністратор	Управляє оперативною базою даних, управляє станціями, управляє сховищем даних.
Аналітик Міністерство екології	Формує аналітичні запити, аналізує екологічний стан повітря, формує аналітичну звітність та переглядає її. Приймає рішення щодо поліпшення стану атмосферного повітря на основі отриманої інформації про його стан.

Також розглянемо прецедент використання «Аналіз екологічного стану атмосферного повітря» більш детально.

Короткий опис: на основі зафіксованих станціями показників вимірювання аналітик шляхом формування аналітичних запитів та аналітичних звітів проводить аналіз стану атмосферного повітря.

Передумови: до сховища даних передано всю необхідну інформацію стосовно зафіксованих середньодобових показників екологічного стану повітря.

Головний потік.

1. Аналітик проводить аналітичні запити.
2. Застосовує алгоритми прогнозування зміни стану повітря.
3. Формує припущення щодо чинників, які впливають на зміну стану повітря.

// E1 – зроблено висновок про неможливість проведення прогнозування через недостатній об'єм даних

4. Робить відповідні висновки та передає результати аналізу Міністерству екології.

Альтернативний потік.

// E1 – зроблено висновок про неможливість проведення прогнозування через недостатній об'єм даних.

1. Аналітик робить припущення стосовно того, яких даних недостатньо для проведення дослідження.

2. Аналітик передає інформацію стосовно того, яких даних додатково необхідно зібрати до адміністратора системи моніторингу.

3. Адміністратор надає додаткову інформацію.

4. Аналітик повторно проводить аналіз та надає результати Міністерству екології.

2.3 Діаграма послідовності

Діаграма послідовності — це тип діаграми взаємодії, оскільки вона описує, як і в якому порядку група об'єктів працює разом. Ці діаграми використовуються розробниками програмного забезпечення щоб зрозуміти вимоги до нової системи або задокументувати існуючий процес.

Діаграми послідовності можуть бути корисними для документування того, як має поводитися майбутня система. На етапі проектування архітектори та розробники можуть використовувати діаграму, щоб виключити взаємодію об'єктів системи, таким чином конкретизуючи загальний дизайн системи [11].

Одним із основних застосувань діаграм послідовності є перехід від вимог, виражених як варіанти використання, до наступного та більш формального рівня моделювання. Варіанти використання часто уточнюються в одну або кілька діаграм послідовності. На додаток до їх використання при проектуванні нових систем, діаграми послідовності можна використовувати для документування того, як об'єкти в існуючій системі в даний момент взаємодіють. Ця документація дуже корисна під час передачі системи іншій особі чи організації.

Основна мета діаграми послідовності — визначити послідовності подій, які призводять до певного бажаного результату. Основна увага приділяється не самим повідомленням, а більше порядку їх появи; незважаючи на це, більшість діаграм послідовності повідомлятимуть, які повідомлення надсилаються між об'єктами системи, а також порядок, у якому вони відбуваються. Діаграма передає цю інформацію вздовж горизонтального та вертикального вимірів: вертикальний вимір показує зверху вниз часову послідовність

повідомлень/викликів, а горизонтальний вимір показує зліва направо екземпляри об'єктів, яким надсилаються повідомлення [12].

На рис. 5 представлена діаграма послідовності програмної системи

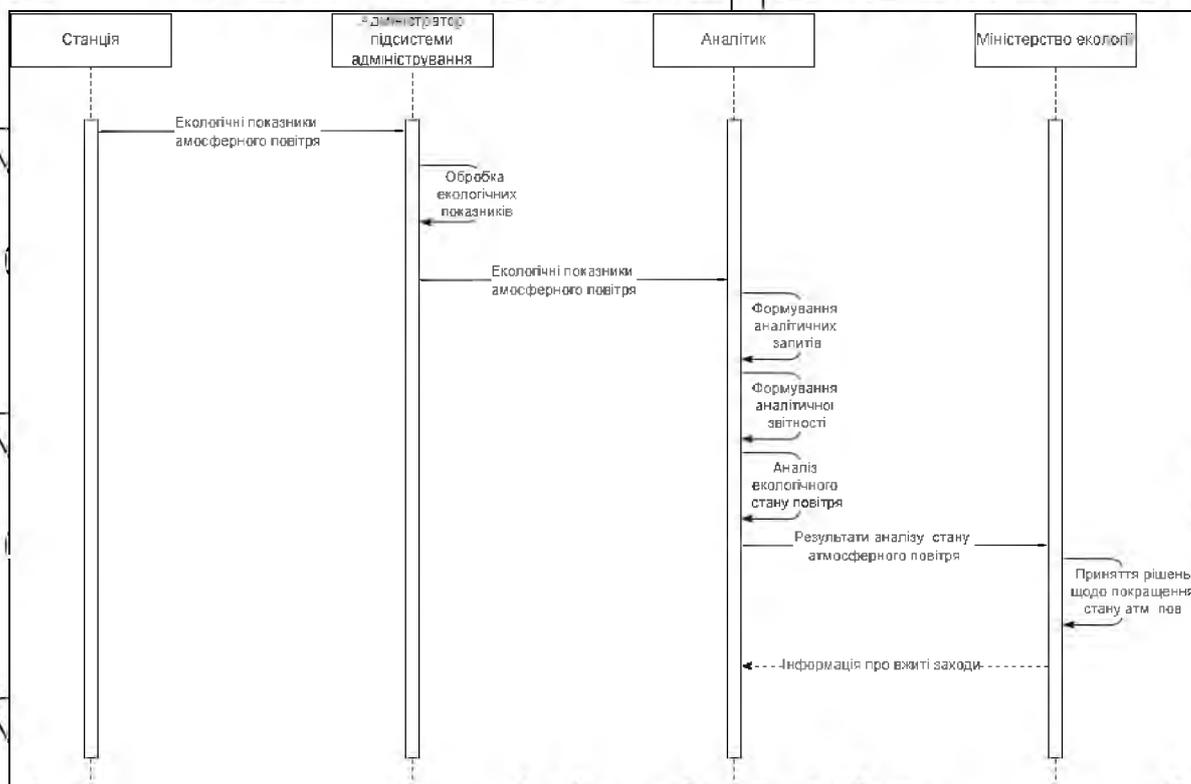


Рис. 5 Діаграма послідовності

Розглянемо діаграму послідовності більш детально. На ній зображено чотири об'єкти, які взаємодіють між собою шляхом обміну повідомленнями (станція, адміністратор підсистеми адміністрування, аналітик та міністерство екології).

Основним завданням станції є передача зафіксованих екологічних показників до підсистеми адміністрування, що відображено у вигляді передачі повідомлення про зафіксовані показники.

Наступним кроком об'єкт «Підсистема адміністрування» оброблює отримані від станції вимірювання показників та агрегує їх в оперативній базі даних. Далі ці показники передаються до аналітика та записуються безпосередньо в СД.

Аналітик на основі даних із СД проводить формування аналітичних запитів, формування аналітичної звітності та аналіз екологічного стану повітря.

Результати аналізу відповідно у вигляді повідомлення передаються безпосередньо до Міністерства екології.

Міністерство екології в свою чергу приймає відповідні рішення на базі отриманих даних від аналітика для покращення екологічної ситуації. Також Міністерство екології передає інформацію про вжиті заходи до аналітика для проведення подальших досліджень в цій області.

2.4 Діаграма активності

Діаграма активності – це одна з найбільш використовуваних діаграм під час моделювання бізнес-процесів. При розробці ПЗ вона зазвичай використовується для опису потоку різних видів діяльності. Вони можуть бути як послідовними, так і паралельними. Вона описує об'єкти, що використовуються, використовуються чи виробляються якоюсь діяльністю, та взаємозв'язок між різними видами діяльності. Все вищезазначене є суттєвим для моделювання бізнес-процесів [13].

Найпоширеннішими компонентами діаграми діяльності є:

Дія – етап діяльності, на якому користувачі або програмне забезпечення виконують певне завдання. Представляє прямокутник із заокругленими кроями.

Вузол рішення – умовна гілка в потоці, представлена ромбом. Він включає один вхід і два або більше виходів.

Потоки керування відображені у вигляді стрілочок

Початковий вузол символізує початок діяльності. Початковий вузол представлений білим колом.

Кінцевий вузол представляє останній крок у дії. Кінцевий вузол представлений окресленим чорним колом [14].

На рис. 6 зображено діаграму активності, яка ілюструє процеси збору екологічних показників атмосферного повітря, їх аналіз та прийняття відповідних заходів для покращення стану атмосферного повітря.

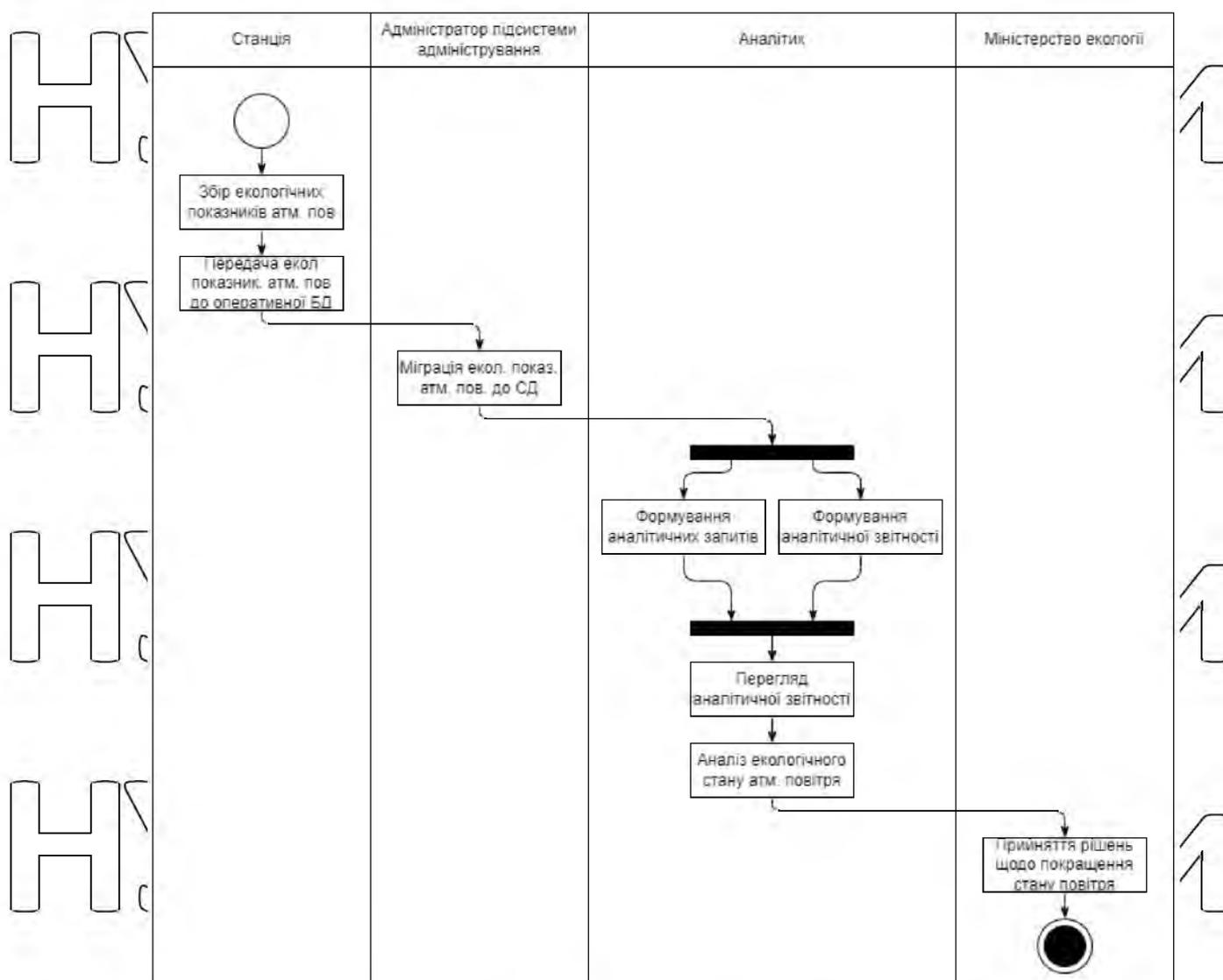


Рис. 6 Діаграма активності

Розглянемо діаграму активності для предметної області аналізу стану атмосферного повітря. На діаграмі представлено основні чотири актори предметної області (станція, підсистема адміністрування, аналітик та міністерство екології).

У станції представлено дві активності – збір екологічних показників атмосферного повітря та передача цих показників до оперативної БД підсистеми адміністрування.

Основними активностями підсистеми адміністрування є агрегування даних отриманих від станцій вимірювання.

В свою чергу аналітик формує аналітичні запити до СД та формує аналітичні звітності для подальшого проведення аналізу. Також він має

можливість переглядати сформовану аналітичну звітність та проводити аналіз екологічного стану атмосферного повітря на основі отриманих звітностей. Результати аналізу передаються до Міністерства екології для подальшого прийняття рішень.

Відповідно Міністерство екології на базі отриманих даних аналізу приймає рішення для поліпшення стану атмосферного повітря в країні.

НУБІП України

НУБІП України

НУБІП України

НУБІП України

НУБІП України

3 ПРОЕКТУВАННЯ СИСТЕМИ

3.1 Архітектура системи

Коли мова йде про архітектуру програмного забезпечення, ми говоримо про певний план, що описує сукупність аспектів та рішень, що є важливими під час розробки програмного забезпечення. Це передбачає врахування всіх видів вимог (швидкодія, безпека тощо), організацію системи, способів взаємодії різних частин системи між собою, чи існують якісь зовнішні залежності, які основні принципи та технології впровадження та які ризики необхідно брати до уваги.

Архітектура програмного забезпечення, як вже було зазначено, є своєрідним планом системи і є первісною для розуміння, ведення переговорів та спілкування між усіма «стейкхолдерами» (або зацікавленими у проекті сторонами - на стороні користувача, замовника, управління тощо). Це полегшує розуміння всієї системи та робить процес прийняття рішень більш ефективним.

Архітектура програмного забезпечення визначає його модель та те, як воно буде функціонувати. Наявність цього дозволяє повторно використовувати цю модель для іншого програмного забезпечення. Завдяки цьому вимоги та код можуть бути з легкістю використані повторно в інших проектах. Весь досвід, який ми отримуємо, під час розробки архітектури, також передається. В результаті, правильний вибір архітектури забезпечить ефективну комунікацію між всіма зацікавленими сторонами проекту, а розмовника, в свою чергу, бажаними функціональними можливостями та атрибутами якості [15].

Для відображення архітектури системи зачату використовують діаграми розгортання або розміщення. Діаграми розгортання використовуються для візуалізації зв'язку між програмним та апаратним забезпеченням. Якщо бути більш точним, за допомогою діаграм розгортання ми можемо побудувати фізичну модель того, як програмні компоненти (артефакти) розгортаються на апаратних компонентах, відомих як вузли. Типова спрощена схема розгортання веб-програми включає:

- Вузли (сервер додатків та сервер баз даних).
- Артефакти (клієнт програми та схема бази даних).

Вузли містять артефакти. Схема бази даних працює на сервері бази даних, а клієнт програми працює на сервері додатків. Як впливає з назви, схема розгортання показує, де саме розгортається кожен компонент програмного забезпечення [16]. Діаграма розгортання представлена в ДОДАТКУ Б.

Розглянемо її більш детально.

Основними фізичними вузлами діаграми розгортання для розробленої системи є:

- станції вимірювання показників;
- MQTT сервер, що відправляє зафіксовані станціями показники до оперативної БД;
- сервер підсистеми адміністрування;
- сервер оперативної БД, до якої постійно надсилаються зафіксовані показники;
- сервер сховища даних, у якому зберігаються дані для аналізу;
- сервер аналітики, за допомогою якого аналітик отримує доступ до даних із сховища даних.

3.2 Оперативні джерела інформації

Як вже було зазначено в попередніх розділах розроблена система була основана на базі розробленої підсистеми моніторингу стану атмосферного повітря. На рис. 7 відображена логічна схема оперативної бази даних, яка є основним джерелом даних для проведення дослідження.

Розглянемо її більш детально. На етапі розробки логічної моделі даних для системи моніторингу було виділено 9 сутностей:

- Station – таблиця для опису станцій, наявних в системі;
- Coordinates – таблиця в якій зберігаються координати розташування станції;
- MQTT Server – таблиця для опису серверів, з яких здійснюється опитування;
- Measurment – таблиця, що описує вимір станції;

- **Measurement_Unit** – таблиця для опису величин, в яких здійснюються вимірювання;
- **Optimal_Value** – таблиця для зберігання оптимальних значень вимірювань;
- **Category** – таблиця, що зберігає словесне визначення для оптимального значення.
- **Favorite_Station** – таблиця для зберігання відмітки про додавання користувачем станції до списку обраних;
- **MQTT_Message_Unit** – таблиця, що зберігає відповідності між повідомленнями MQTT сервера до вимірюваних величин.

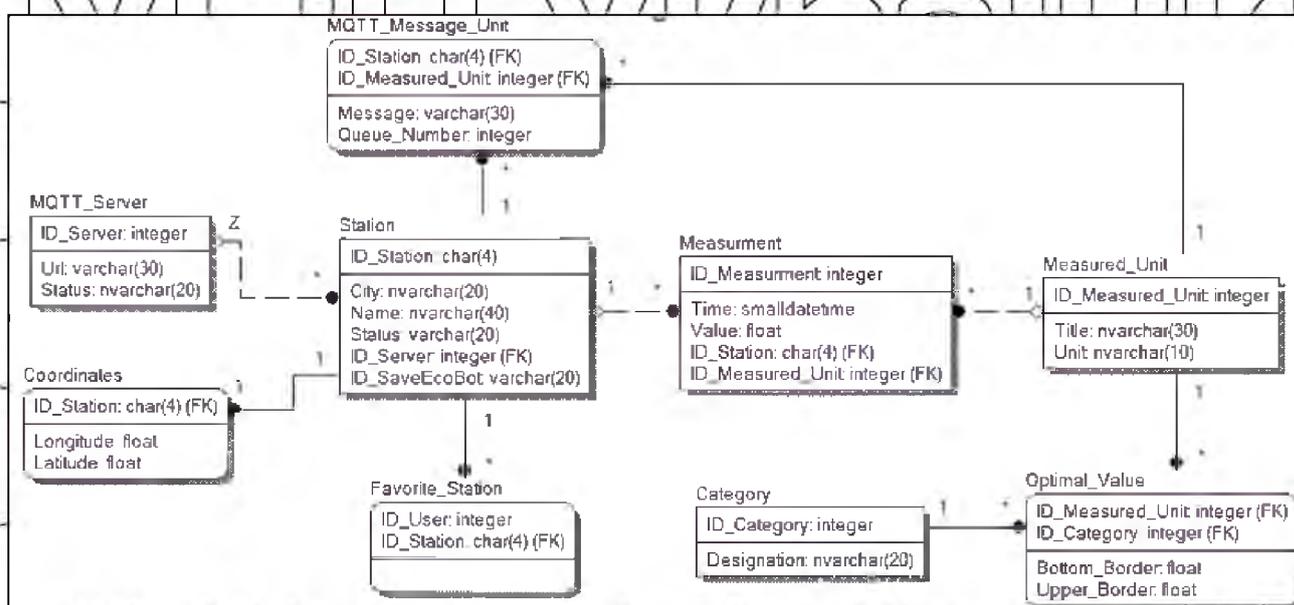


Рис. 7. Логічна модель БД підсистеми моніторингу атмосферного повітря

Реляційність бази даних дозволяє не лише накопичувати дані, а і збільшувати кількість показників, що вимірюються, а також доповнювати в відповідних пороговими і цільовими показниками.

3.3 Сховище даних

Сховище даних (data warehouse) — це тип системи керування даними, яка призначена для забезпечення та підтримки діяльності бізнес-аналітики (BI).

Сховища даних призначені для виконання запитів та аналізу і часто містять великі обсяги історичних даних. Дані в сховищі даних зазвичай отримують з різних джерел, таких як файли журналів програм і програми транзакцій і т.д.

Сховище даних централізує та агрегує великі обсяги даних із кількох джерел. Його аналітичні можливості дозволяють організаціям отримувати цінну бізнес-ідею зі зворотк даних для покращення процесу прийняття рішень.

Розроблене сховище даних представляє собою багатовимірну структуру, в якій дані діляться на дві групи: виміри та факти.

Вимір – це множина об'єктів одного або декількох типів, що організовані у вигляді ієрархічної структури і які забезпечують інформаційний контекст числового показника. Вимірювання прийнято візуалізувати у вигляді ребра багатовимірного куба.

Факт – це числова величина, яка розташовується в рамках гіперкубу. Один куб може характеризуватися одним або декількома показниками.

На рис. 8 представлено модель сховища даних для розробленої системи.

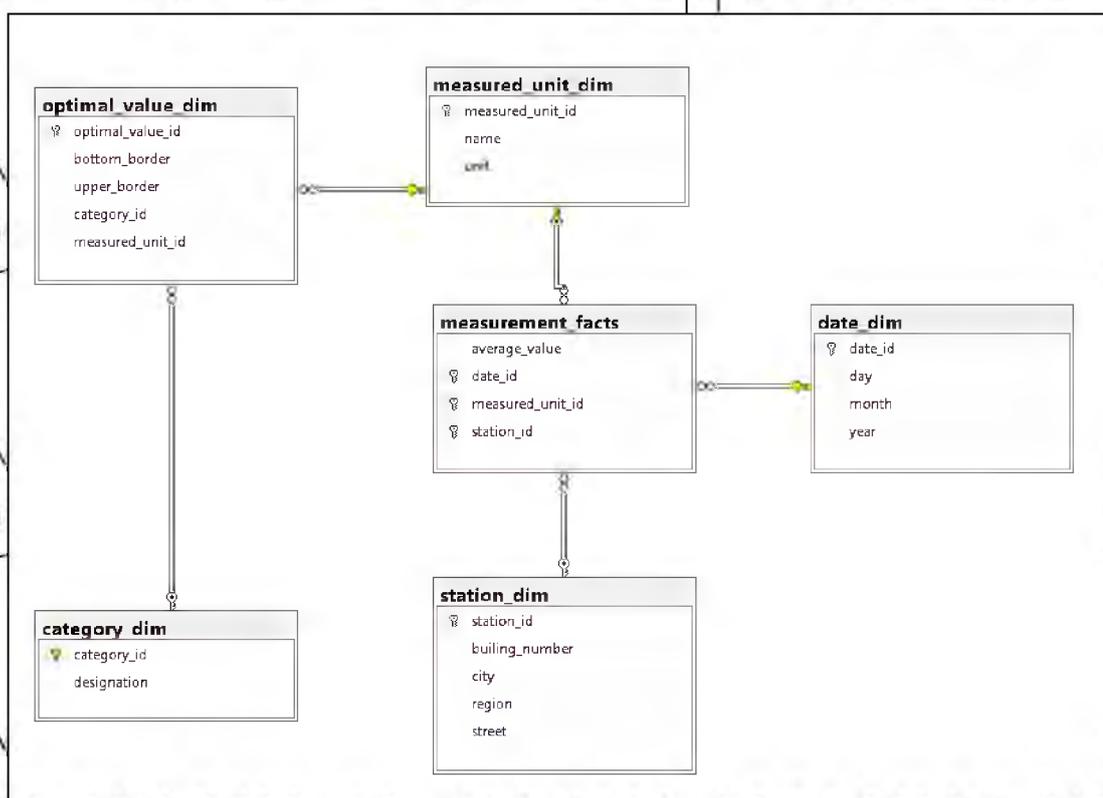


Рис. 8 Модель сховища даних

Сховище даних було описано та створено за допомогою специфікації JPA та ORM бібліотеки Hibernate мови програмування Java (повний опис наведено у ДОДАТОКУ В). На рис. 9 наведено приклад опису вимірів наведеном вище інструментів.

```

@Getter
@Setter
@Entity
@AllArgsConstructor
@NoArgsConstructor
@Table(name = "STATION_DIM")
public class StationDim {

    @Id
    private String stationId;

    private String address;

}

```

Рис. 9 Опис вимірів станцій мовою Java

Розглянемо модель СД більш детально. На моделі відображено 5 вимірів (category_dim, optimal_value_dim, measured_unit_dim, station_dim, date_dim) та таблиця фактів (measurement_facts). Нижче наведено загальний опис створених сутностей.

- Category Dim – вимір, що містить інформацію про категорії оптимальних значень із полями:
 - category_id – ідентифікатор категорії;
 - designation – визначення категорії;
- Measured Unit Dim – вимір, що зберігає показники та їх одиниці вимірювання
 - measured_unit_id – ідентифікатор показника;
 - name – назва показника;
 - unit – одиниця вимірювання;
- Optimal Value Dim – вимір, що містить інформацію про оптимальні значення показників вимірювання.
 - optimal_value_id – ідентифікатор оптимального значення;
 - bottom_border – нижня границя;
 - upper_border – верхня границя;
 - category_id – ідентифікатор категорії;
 - measured_unit_id – ідентифікатор показника

- **Station_Dim** – вимір, що зберігає станції вимірювання:
 - **station_id** – ідентифікатор станції;
 - **address** – адреса станції;

- **Date_Dim** – вимір, що зберігає часовий проміжок:
 - **date_id** – ідентифікатор дати;
 - **day** – день;
 - **month** – місяць;
 - **year** – рік;

- **Measurement_Facts** – таблиця фактів, із зафіксованими показниками:
 - **date_id** – ідентифікатор дати;
 - **station_id** – ідентифікатор станції;
 - **measured_unit_id** – ідентифікатор показника;
 - **value** – середньодобове значення.

3.4 Побудова розгорнутого куба підсистеми аналізу, SSAS

Служби аналізу SQL Server (SSAS) – це багатовимірний OLAP-сервер, а також механізм аналітики, який дозволяє розділяти великі обсяги даних. Він є частиною Microsoft SQL Server і допомагає виконувати аналіз за допомогою різних вимірів. Він має 2 варіанти багатовимірний і табличний [17].

Для розробки кубу було використано середовище Visual Studio з розширенням SSAS. На першому етапі необхідно визначити джерело даних – база даних OLAP або сховище даних. На основі визначеного джерела даних імпортуються дані для аналізу. На рис. 10 зображено формування підключення джерела даних за допомогою модуля Data Source Wizard, де ми обираємо попередньо створене сховище даних.

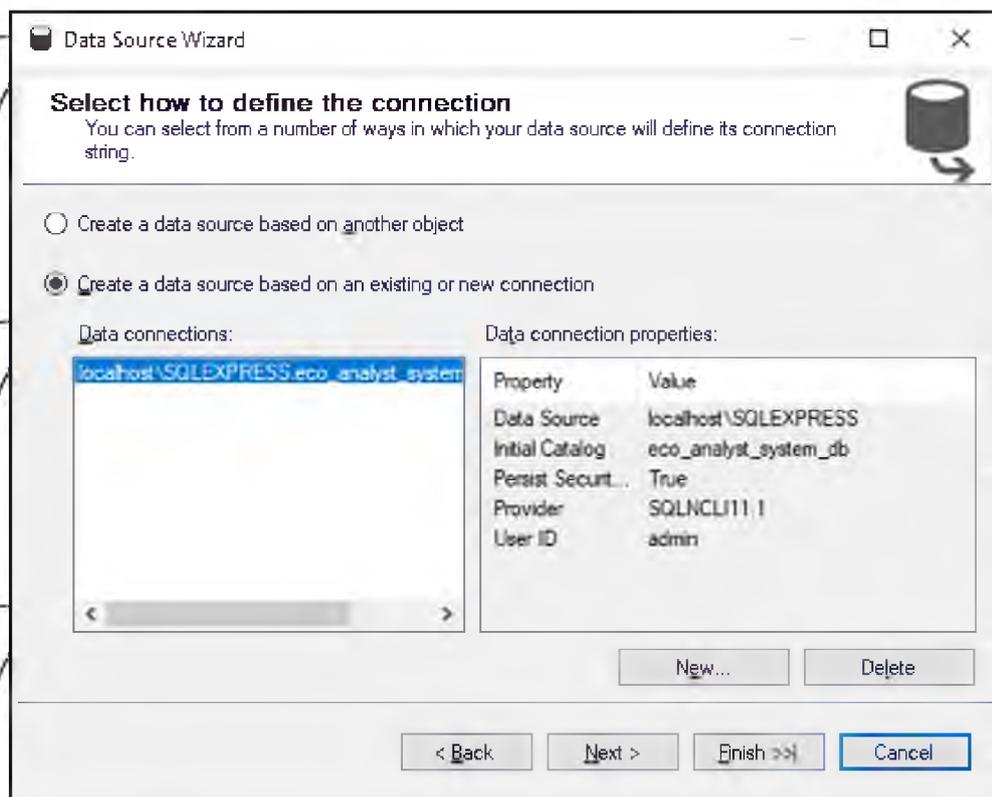


Рис. 10 Додавання джерела даних

Наступним етапом у розробці куба є створення уявлення джерела даних, яке являє собою абстракцію реляційного джерела даних, яка стає основою кубів і вимірів, які ви створюєте в багатовимірному проєкті. Метою DSV є надання контролю над структурами даних, які використовуються у проєкті, і працювати незалежно від базових джерел даних (наприклад, можливість перейменувати або об'єднати стовпці без безпосередньої зміни вихідного джерела даних). У процесі розробки можна створити кілька представлень джерел даних у проєкті або бази даних Analysis Services на одному або кількох джерелах даних і створити кожне з них, щоб задовольнити вимоги до іншого рішення [17].

На рис. 11 зображено додавання уявлення за допомогою Data Source View Wizard. У розроблюваному проєкті уявленням виступає створене СД.

На основі створеного джерела та уявлення джерела даних необхідно створити виміри для OLAP куба. Виміри створюються на базі вимірів СД: category_dim, optimal_value_dim, measured_unit_dim, station_dim, date_dim.

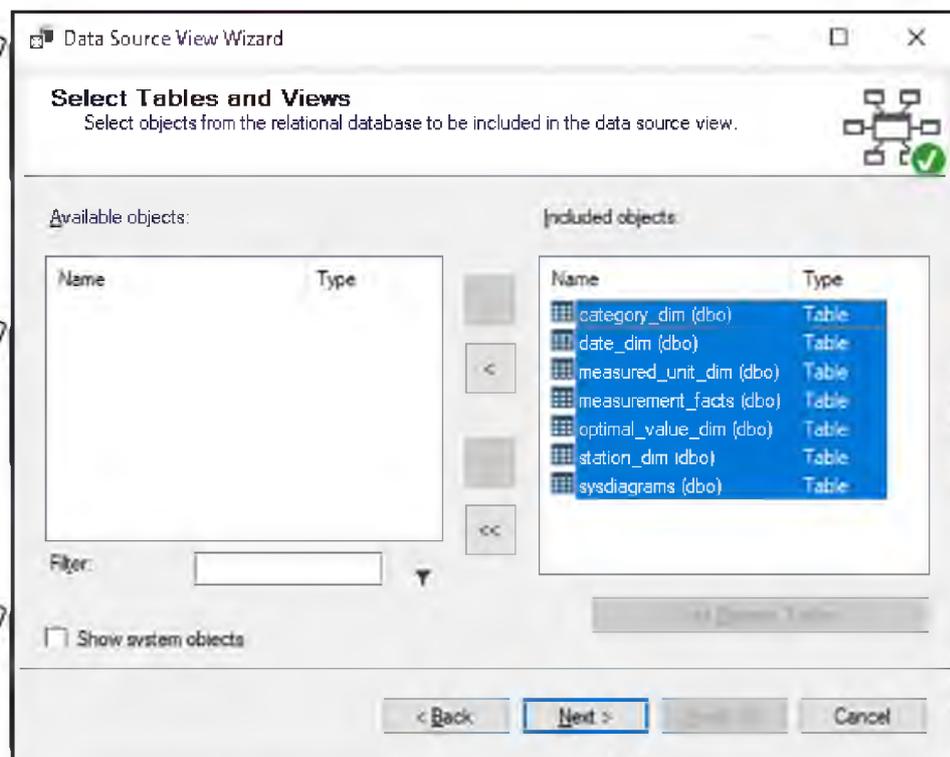


Рис. 11 Створення уявлення на основі таблиць-вимірів та фактів СД

На рис. 12 зображено процес створення виміру категорій оптимальних значень. Таким чином створюємо інші виміри.

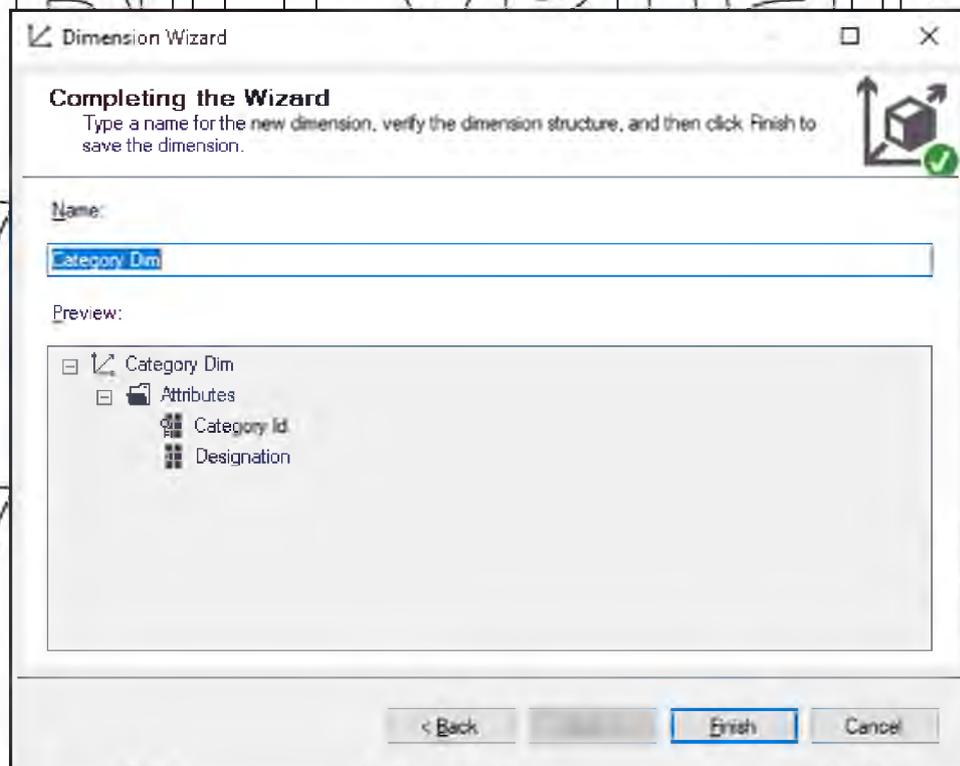


Рис. 12 Створення виміру категорій оптимальних значень

Після створення всіх необхідних вимірів наступним кроком є створення кубу OLAP за допомогою майстра кубів (Cube Wizard). На рис. 13 зображено

створення кубу. На першому етапі необхідно обрати таблицю фактів, яка виступає основою для створення кубу.

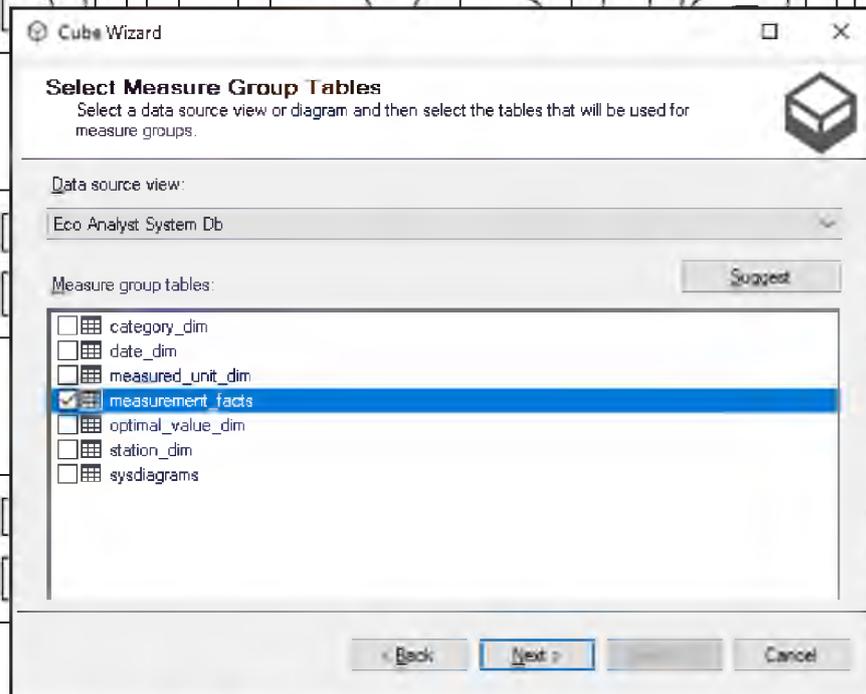


Рис. 13 Вибір таблиці для створення кубу

Наступним кроком визначаються виміри та їх поля для формування кубу

рис. 14.

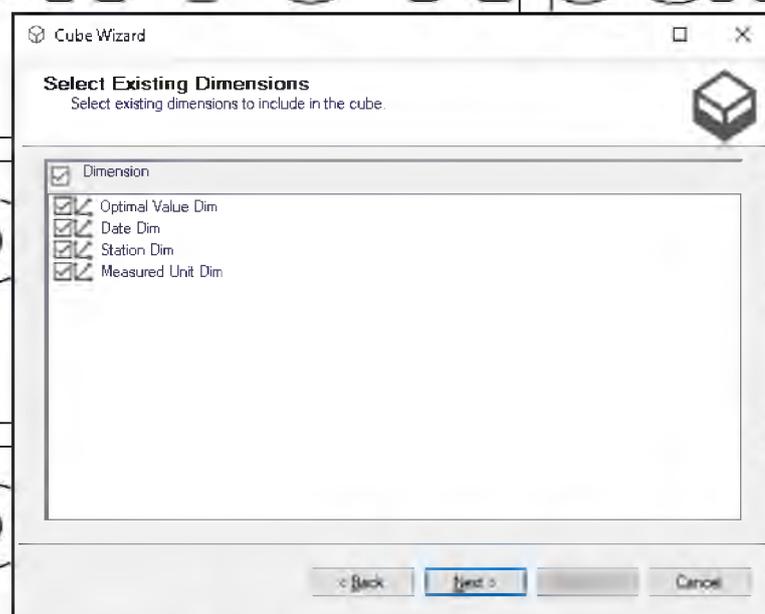


Рис. 14 Формування кубу

Результатом є створений куб, що зображений на рис. 15.

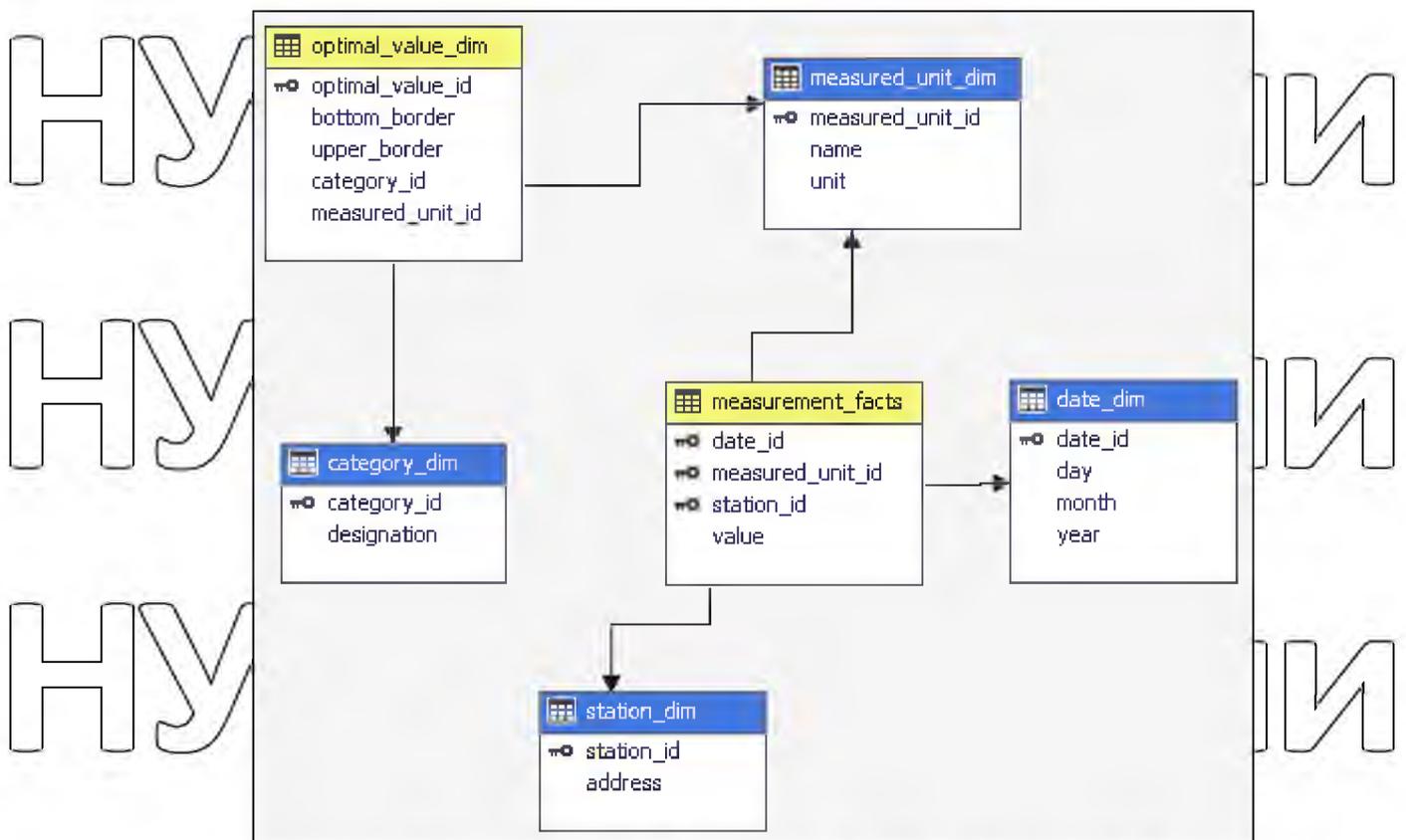


Рис. 15 Створений OLAP куб

3.5 Реалізація міграції даних за допомогою Data Flow, SSIS

Процес передачі даних було реалізовано за допомогою служби SQL Server Integration Services. SSIS – це інструмент, що дозволяє у зручному вигляді реалізувати інтеграцію, тобто. реалізувати процес перенесення даних з одного джерела до іншого. Цей процес іноді називають ETL (від англ. Extract, Transform, Load – дослівно «вищукання, перетворення, завантаження») [18].

В інструменті SSIS є служба Data Flow, за допомогою якої було проведено заповнення таблиць вимірів та фактів.

Data Flow – це технологія підготовки даних. Потoki даних дозволяють клієнтам отримувати, перетворювати та завантажувати дані в середовища Microsoft Dataverse, робочі області Power BI або обліковий запис Azure Data Lake Storage. Потoki даних створюються за допомогою Power Query – уніфікованого засобу підключення даних, який уже представлений у багатьох продуктах Microsoft, включаючи Excel і Power BI [25].

В результаті роботи наповнення сховища даних було розділено на два етапи наповнення вимірів та таблиці фактів (рис. 16).

На рис. 17 зображено потоки, які реалізують передачу даних з БД у СД.

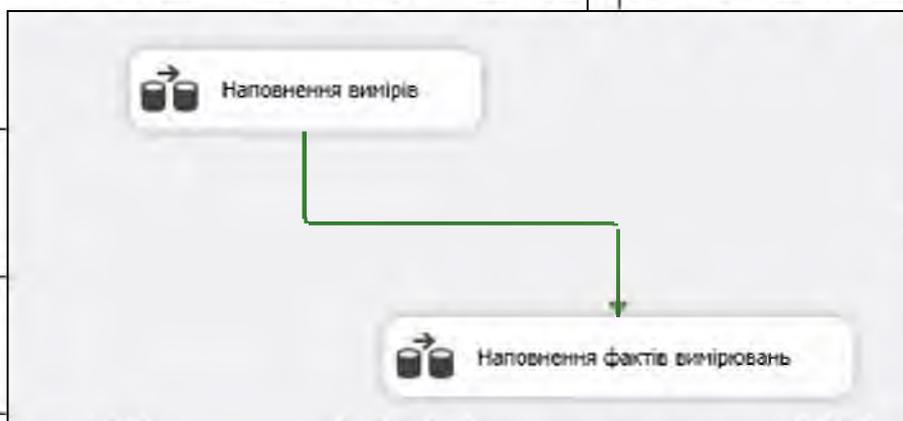


Рис. 16 Потоки даних для наповнення СД



Рис. 17 Наповнення таблиць-вимірів та фактів

На рис. 18 відображено приклад заповнення виміру Date Dim за допомогою SQL-запиту.

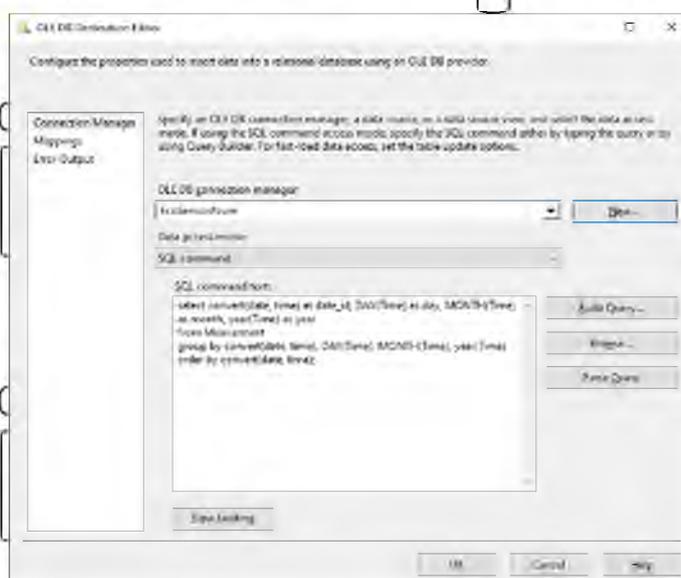


Рис. 18 Формування запиту для заповнення виміру Date_Dim

Наступним кроком проводиться заповнення таблиць фактів. Запит, що наведений на рис. 19 агрегує дані з таблиці вимірювань оперативної БД та переносить їх у таблицю фактів СД.

```
select avg(m.Value)          as value,
       convert(date, m.Time) as date_id,
       m.ID_Measured_Unit    as measured_unit_id,
       m.ID_Station          as station_id
from Measurement m
group by ID_Station, ID_Measured_Unit, convert(date, m.Time);
```

Рис. 19 Запит для заповнення таблиці фактів вимірювань

У результаті виконання потоків у таблиці СД було додано дані з оперативної БД. На рис. 20-21 зображено заповнені таблиці вимірів та фактів. У результаті отримано підготовлене сховище даних для аналізу.

category_id	designation
1	Excellent
2	Fine
3	Moderate
4	Poor
5	Very Poor
6	Severe

Рис. 20 Результат заповнення таблиці category_dim

optimal_value_id	bottom_border	upper_border	category_id	measured_unit_id
1	0	12	1	2
2	12	25	2	2
3	25	50	3	2
4	50	90	4	2
5	90	180	5	2
6	180	NULL	6	2
7	0	7	1	3
8	7	15	2	3
9	15	30	3	3
10	30	55	4	3
11	55	110	5	3
12	110	NULL	6	3
13	0	50	1	9
14	50	100	2	9
15	100	150	3	9
16	150	200	4	9
17	200	300	5	9
18	300	NULL	6	9

Рис. 21 Результат заповнення таблиці optimal_value_dim

3.6 Побудова звітності засобами SSRS

SQL Server Reporting Services – це засіб створення звітів, розроблений компанією Microsoft, що безкоштовно постачається разом із SQL Server. За допомогою SSRS створюються форматовані звіти з таблицями даних, графіки та звіти різних форматів. Звіти зберігаються на сервері та можуть бути налаштовані для запуску за допомогою параметрів, наданих користувачами. Дані для створення звітності можуть бути під'єднані з поточної БД на сервері та інших джерел даних (файлів XML або іншого джерела даних) [19].

Розглянемо створення звітності на основі створеного сховища даних. Для цього необхідно налаштувати джерело даних. На рис. 22 відображено обраний спосіб підключення до СД.

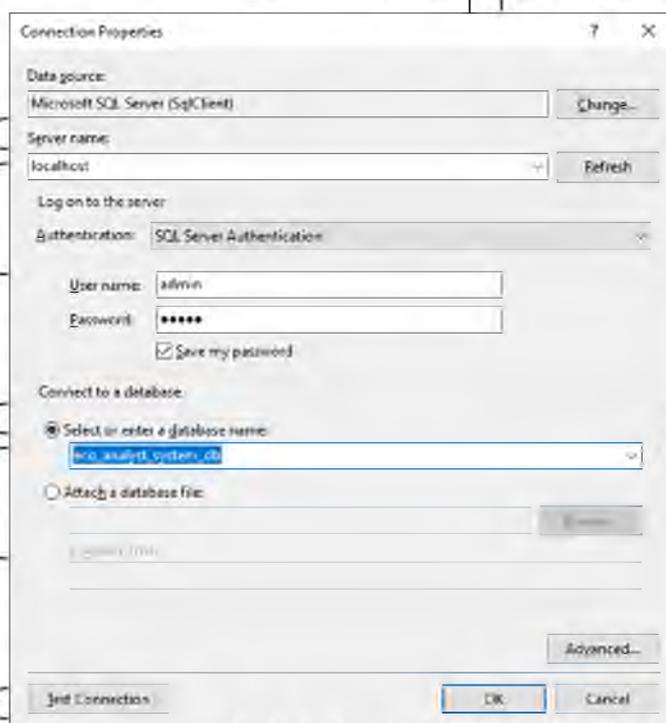


Рис. 22 Підключення до СУБД

За допомогою команди SQL SELECT здійснюється вибірка даних для побудови звітності. На рис. 23 відображено приклад запиту для вибірки динаміки зміни вмісту середньодобового значення важких частинок PM_{10} за певною станцією.

```

select mud.name, mf.value, dd.date_id, sd.address, sd.station_id
from measurement_facts as mf
    inner join date_dim dd on dd.date_id = mf.date_id
    inner join station_dim sd on mf.station_id = sd.station_id
    inner join measured_unit_dim mud on mud.measured_unit_id = mf.measured_unit_id
where sd.station_id in ('0020', '0004', '0001')
and mud.measured_unit_id = '2'
and dd.month = 6
and dd.year = 2021
order by sd.address

```

Рис. 23 Приклад запиту для вибірки даних

Також необхідно обрати тип звіту (матричний або табличний) та необхідні поля як показано на рис. 24.

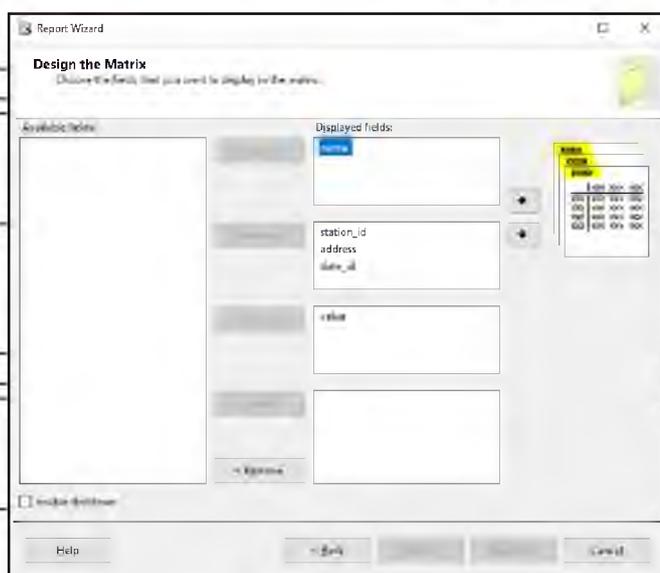


Рис. 24 Налаштування полів звіту

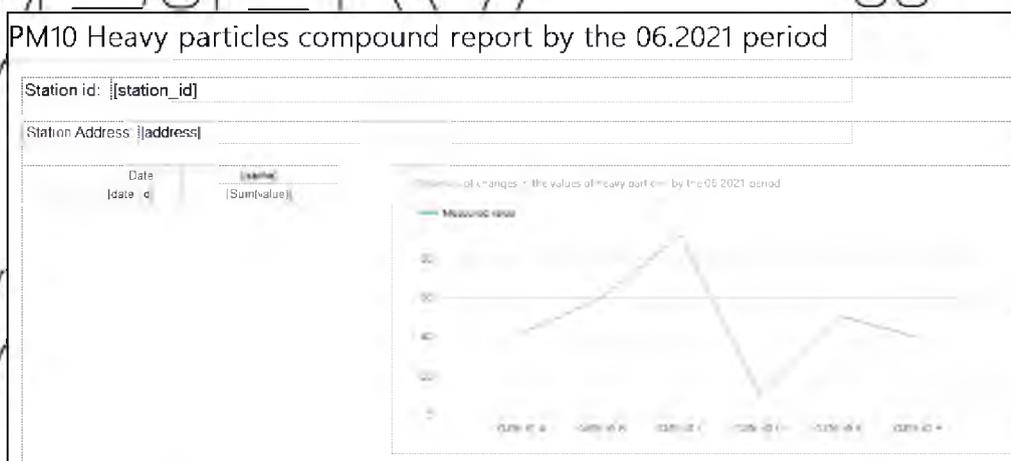


Рис. 25 Конструктор звітів SSRS

Також SSRS надає можливість налаштування звітності за допомогою вбудованого дизайнера звітів, як показано на рис. 25. Він надає можливість

обрати різні формати звітів, додати графіки та в цілому більш гнучко налаштувати звіт.

3.7 Розрахунок KPI

Щоб проаналізувати досягнення цілей підприємства за певними критеріями обчислюють ключовий показник ефективності.

KPI – числовий показник, який пов'язаний з групою вимірів у кубі, які використовуються для оцінки успіху бізнесу. Зазвичай ці обчислення є комбінацією виразів багатовимірних виразів (MDX) або обчислюваних елементів. KPI також мають додаткові метадані, які надають інформацію про те, як клієнтські програми повинні відображати результати обчислень KPI.

KPI обробляє інформацію про поставлену мету, фактичне значення продуктивності, записану в кубі і вимірювання, щоб показати тенденцію та статус ефективності.

Компоненти для розрахунку KPI:

- Ціль (Goal expression) – числовий вираз MDX або обчислення, яке повертає цільове значення KPI.
- Значення (Value expression) – числовий вираз MDX, який повертає фактичне значення KPI.
- Статус (Status expression) – вираз MDX, який представляє стан KPI на певний момент часу. Вираз MDX стану має повертати нормалізоване значення від -1 до 1. Значення, рівні або менші за -1, будуть інтерпретуватися як "погані" або "низькі". Значення нуль (0) інтерпретуються як «прийнятне» або «середнє». Значення, що дорівнюють або перевищують 1, будуть інтерпретуватися як «добре» або «високе» [20].

Для аналізу фактів, занесених у куб, було визначено такі KPI:

1. KPI_PM2_5_AVG_VALUE – визначення ефективності середньодобового значення важких частинок PM_{2.5} на станції
2. KPI_PM10_AVG_VALUE – визначення ефективності середньодобового значення важких частинок PM₁₀ на станції.

Для розрахунку KPI необхідно розраховане фактичне значення, цільове значення та зазначені умови для визначення статусу (рис. 26–28).

```
Value Expression
([Measures].[Value], [Station Dim].[Station_Id].&[0001], [Measured Unit Dim].[Measured_Unit_Id].&[3])/([Measures].[
Measurement Facts Count], [Station Dim].[Station_Id].&[0001], [Measured Unit Dim].[Measured_Unit_Id].&[3])
No issues found Ln: 1 Ch: 222 Col: 108 SPC CRLF
```

Рис. 26 Розрахунок фактичного значення

```
Goal Expression
([Measures].[Bottom Border],[Measured Unit Dim].[Measured_Unit_Id].&[3],[Category].[Designation].&[Excellent])
No issues found Ln: 1 Ch: 1 SPC CRLF
```

Рис. 27 Розрахунок цільового значення

```
Status
Status indicator: Gauge
Status expression:
CASE
WHEN KPIVALUE( "KPI_MEASURED_UNIT_VALUE" ) = KPIGOAL( "KPI_MEASURED_UNIT_VALUE" ) THEN 1
WHEN KPIVALUE( "KPI_MEASURED_UNIT_VALUE" ) >= ([Measures].[Bottom Border],[Measured Unit Dim].[Measured_Unit_Id].&[3],
[Category].[Designation].&[Excellent]) AND
KPIVALUE( "KPI_MEASURED_UNIT_VALUE" ) <= ([Measures].[Upper Border],[Measured Unit Dim].[Measured_Unit_Id].&[3],
1 0 ← → Ln: 4 Ch: 145 Col: 33 SPC CRLF
```

Рис. 28 Визначення статусу за зазначеними умовами

Результат розрахунку KPI зображено на рис. 29.

KPI_PM10_AVG_VALUE	11.068	0	
KPI_PM2_5_AVG_VALUE	3.476	0	

Рис. 29 Результати розрахунку KPI

На основі отриманих даних KPI можна зробити висновок, що середньодобове значення показників важких частинок знаходяться у межах норми.

3.8 Проектування моделей інтелектуального аналізу засобами Data Mining

SQL Server Business Intelligence (BI) — це набір інструментів для перетворення необроблених даних в інформацію, що може бути використана для прийняття рішень. SQL Server BI надає безліч алгоритмів для обробки даних та подальшого прийняття рішень на основі цих даних [21]. Прикладом таких алгоритмів є:

- дерева рішень;
- лінійна регресія;
- часові ряди;

- метод наївного Байєса;
- кластеризація
- асоціативні правила.

Розглянемо застосування інструментів SQL Server BI на прикладі алгоритму наївного Байєса [22].

На основі розгорнутого кубу створюємо нову структуру інтелектуального аналізу даних. Обираємо алгоритм Байєса (рис. 30).

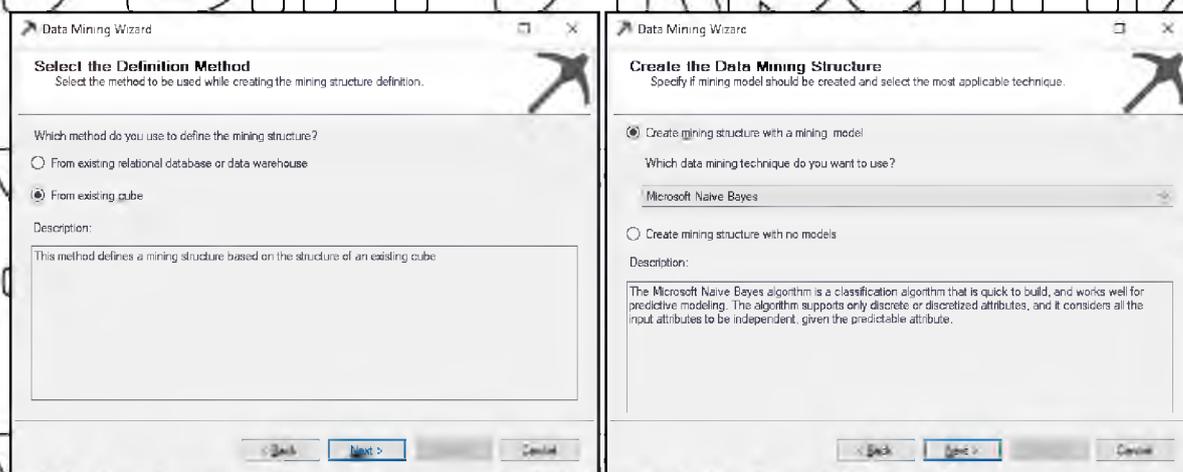


Рис. 30 Створення структури аналізу на основі алгоритму Байєса

Далі необхідно обрати вимір, на базі якого буде проводитися аналіз (рис.

31). Також обираємо стовпці для аналізу та прогнозування (рис. 32)

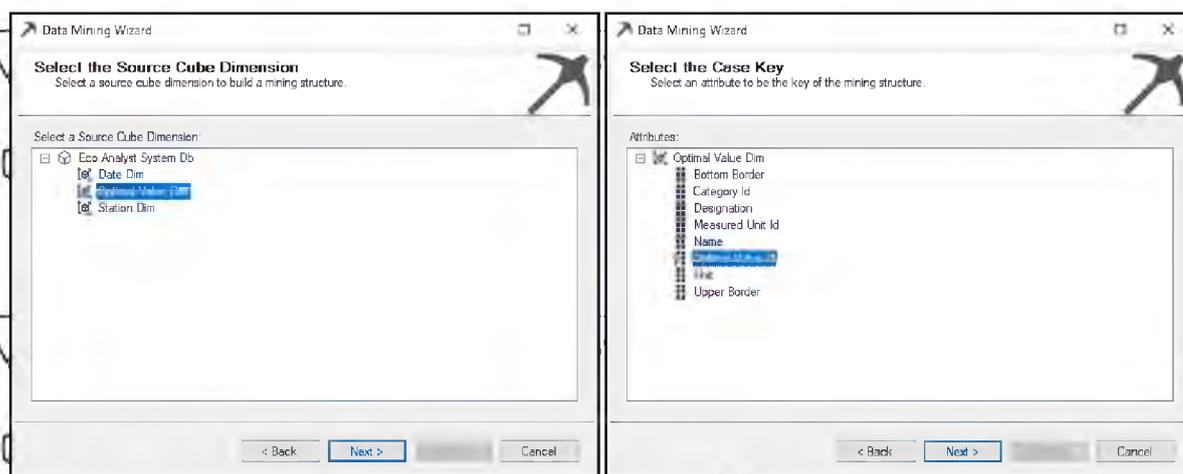


Рис. 31 Вибір виміру для аналізу

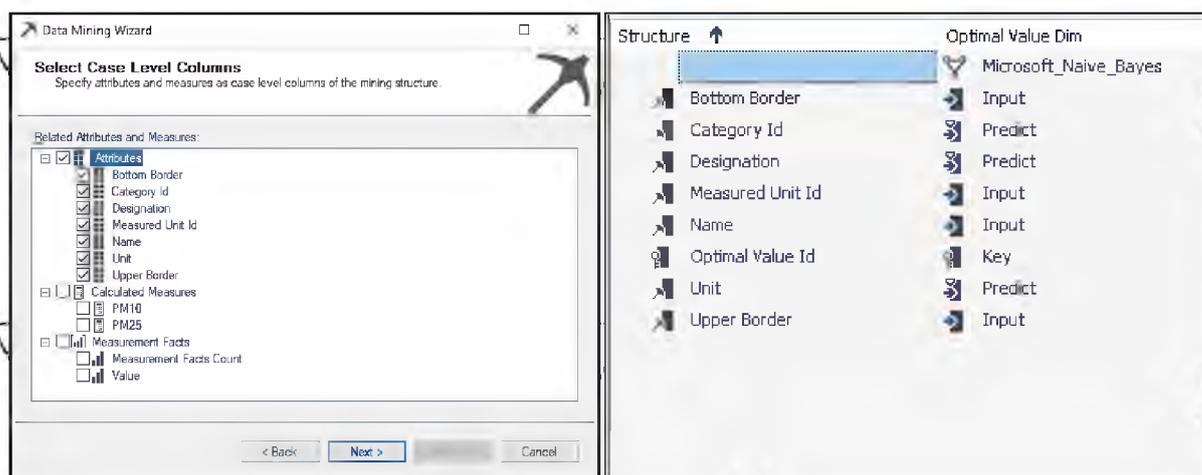


Рис. 32 Вибір стовпців для аналізу та прогнозування

На основі створеної структури для аналізу було сформовано мережу залежностей (рис. 33).

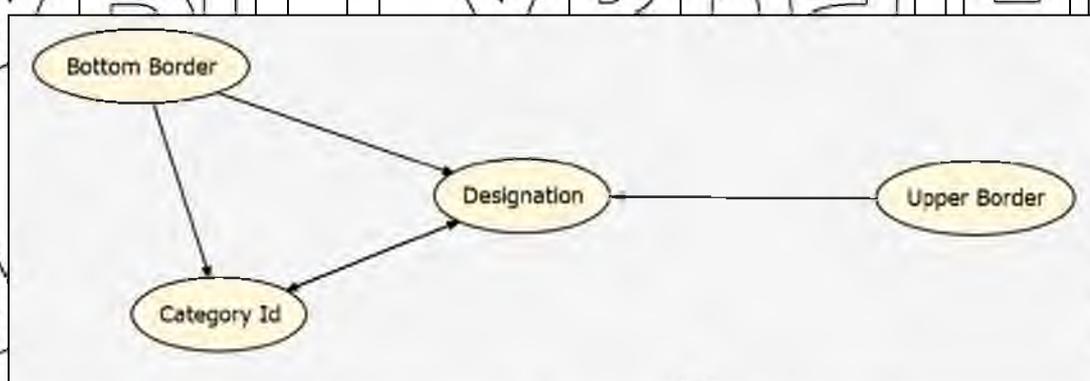


Рис. 33 Розгорнута мережа залежностей

На основі побудованої мережі можна сказати, що категорія стану повітря напряму залежить від граничних значень показників.

3.9 Застосовані алгоритми аналізу даних засобами Data Mining

3.9.1 Часові ряди. Для аналізу показників твердих частинок PM₁₀ та PM_{2.5}

необхідно створити обчислювальну міру, яка розраховує середнє значення. На рис. 34–35 показано створення виразу за допомогою мови запитів MDX [23]

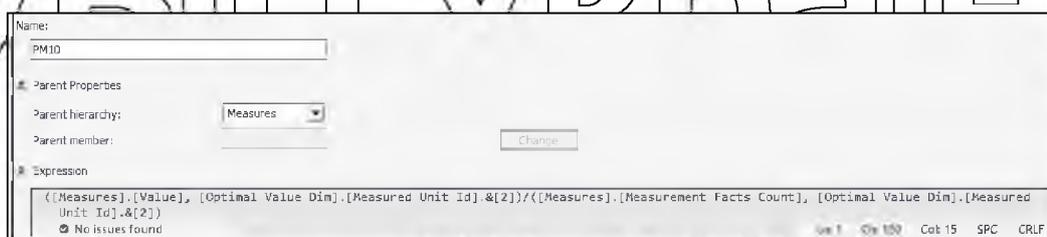


Рис. 34 Створення виразу обчислюваної міри PM₁₀

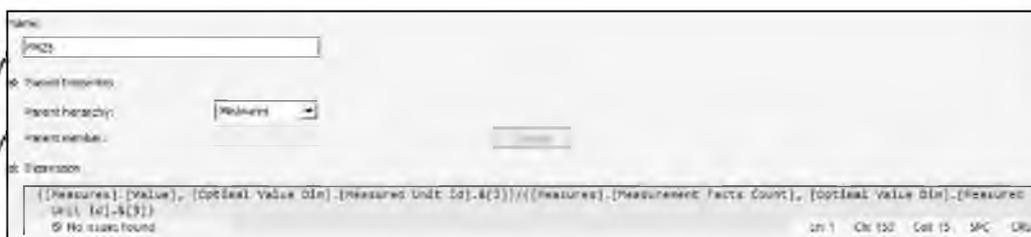


Рис. 35 Створення виразу обчислюваної міри $PM_{2.5}$

Наступним кроком необхідно створити спеціальну структуру для аналізу під назвою «Microsoft Time Series» як показано на рис. 36.

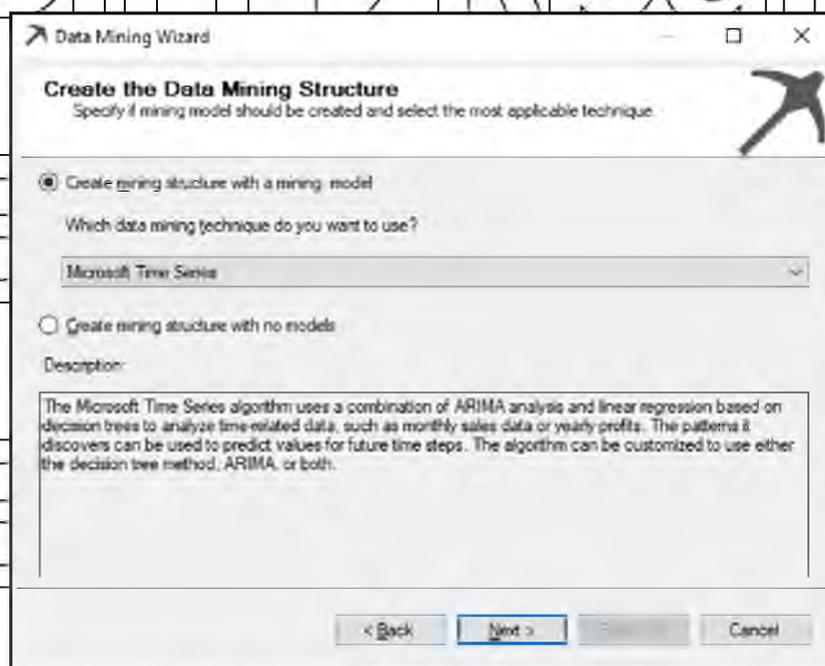


Рис. 36 Вибір алгоритму часових рядів як методу інтелектуального аналізу на основі сформованого кубу

Далі необхідно обрати вимір, на основі якого буде проводитись інтелектуальний аналіз (рис. 37).

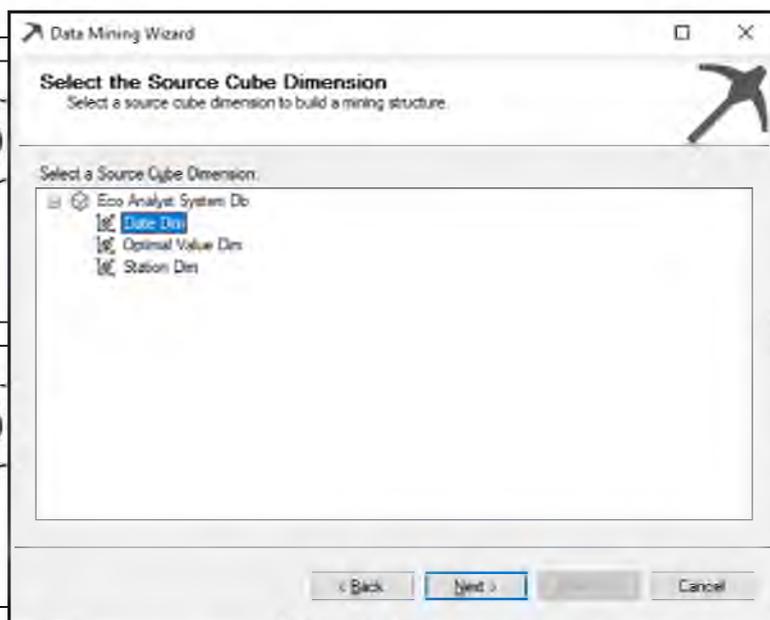


Рис. 37 Вибір виміру кубу

Далі обираємо ключ структури аналізу, вхідні та прогнозовані параметри (рис. 38-39).

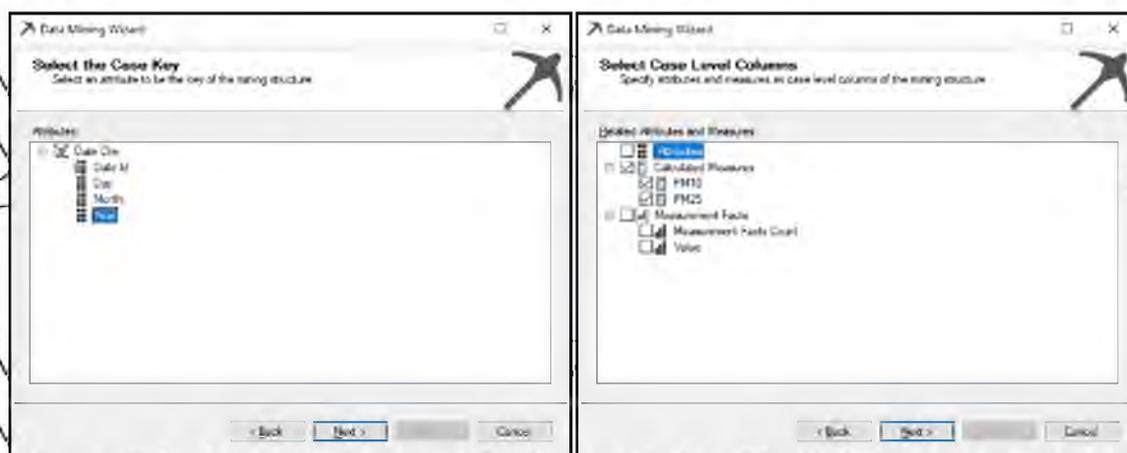


Рис. 38 Налаштування структури аналізу часового ряду

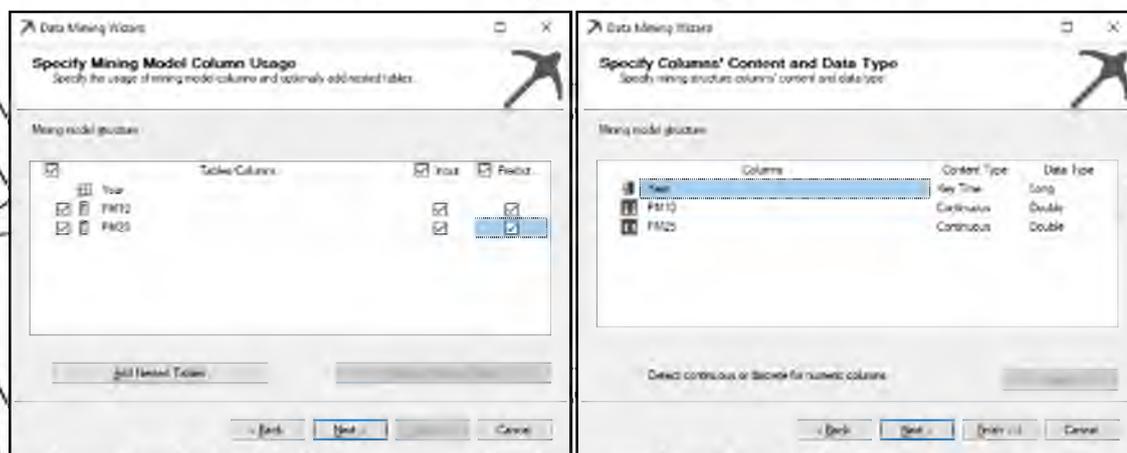


Рис. 39 Налаштування структури аналізу часового ряду

3.9.2 1-Rule алгоритм. OneR алгоритм (One Rule) – не простий, але точний алгоритм класифікації, який генерує одне правило для кожного предиктора даних, а потім вибирає правило з найменшого загальною помилкою як своє «єдине правило». Щоб створити правило для предиктора, необхідно створити частотну таблицю для кожного предиктора щодо цілі [24].

Для реалізації алгоритму 1-Rule було виділено 2 класи задовільного ("Добре") та незадовільного вмісту діоксиду сірки у повітрі ("Погано"), показник яких визначається у порівнянні з середнім показником стану повітря за весь період дослідження.

Спочатку проведено вибірку даних, де визначено середнє значення вмісту діоксиду сірки та класифікацію (рис. 42).

```

declare @avgStationsValue float;
SELECT @avgStationsValue = AVG(mf.average_value)
from measurement_facts mf
where mf.measured_unit_id = 8;
select sd.station_id as '[Номер станції]',
concat(sd.region, ' ', sd.city, ' ', sd.street, ' ', sd.building_number) as '[Адреса]',
dd.date_id as '[Дата]',
ROUND(mf.average_value, 2) as '[Середнє значення]',
CASE
    WHEN ROUND(mf.average_value, 2) <= @avgStationsValue
    THEN 'Добре'
    ELSE 'Погано'
END as '[Класифікація]'
from measurement_facts mf
inner join date_dim dd on mf.date_id = dd.date_id
inner join station_dim sd on mf.station_id = sd.station_id
inner join measured_unit_dim mud on mud.measured_unit_id =
mf.measured_unit_id
where mud.measured_unit_id = 8
and mf.station_id = '14410';

```

Рис. 42 Реалізація вибірки середнього значення та класифікацію за станцією

Наступним кроком було визначення класифікації вмісту діоксиду сірки та її ймовірність відносно середнього показника серед заданих фактів.

3.9.3 Алгоритм кластеризації. Алгоритм кластеризації – це алгоритм сегментації або кластеризації, який повторює випадки в наборі даних, щоб згрупувати їх у кластери зі схожими характеристиками. Ці групи корисні для вивчення даних, виявлення аномалій у даних і створення прогнозів.

Моделі кластеризації ідентифікують зв'язки в наборі даних, які не можуть бути отримані логічно за допомогою випадкового спостереження. Наприклад, люди, які добираються на роботу на велосипед, зазвичай не живуть далеко від

місяця роботи. Однак алгоритм може знайти інші характеристики велосипедистів, які не такі очевидні [25].

Алгоритм кластеризації відрізняється від інших алгоритмів інтелектуального аналізу даних, таких як алгоритм дерев прийняття рішень, тим, що не потрібно призначати передбачуваний стовпець, щоб мати можливість побудувати модель кластеризації. Алгоритм кластеризації навчає модель виключно на основі зв'язків, які існують у даних, і на основі кластерів, які ідентифікують алгоритм.

На основі розгорнутого кубу створюємо нову структуру інтелектуального аналізу даних, використовуючи алгоритм кластерного аналізу (рис. 43-44).

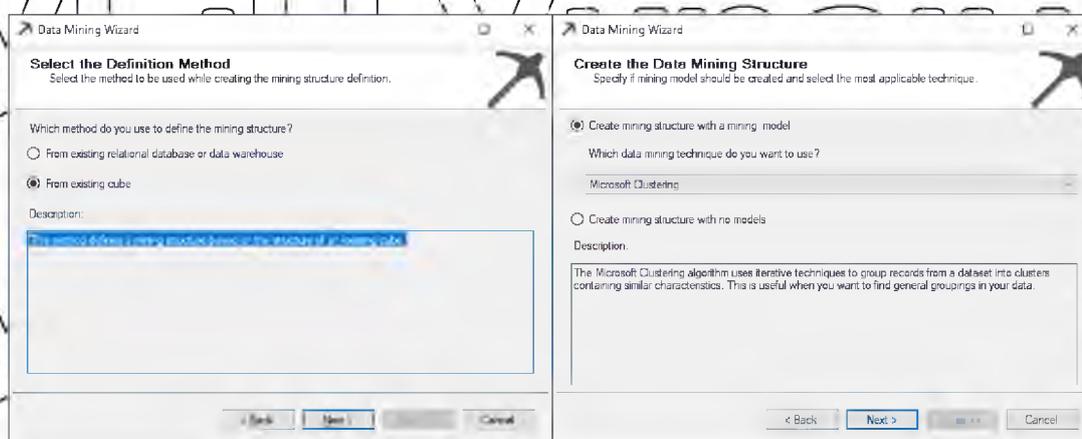


Рис. 43 Вибір алгоритму для інтелектуального аналізу

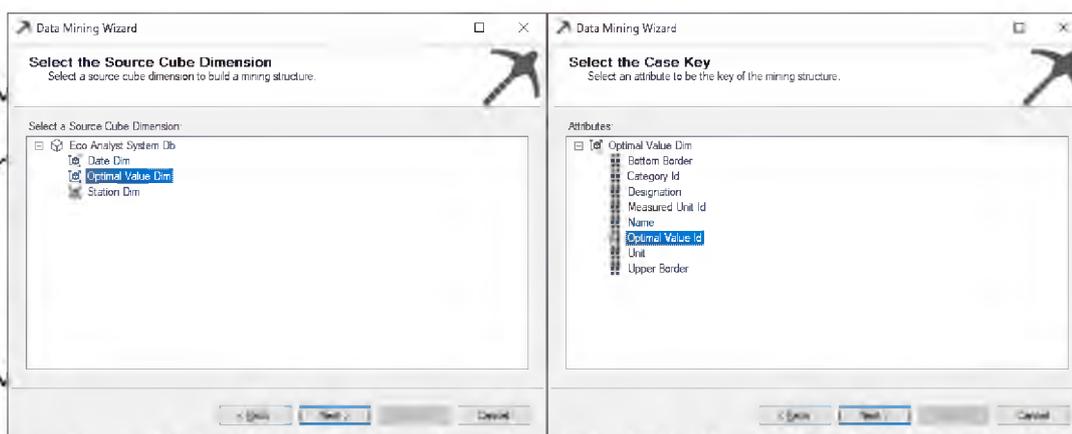


Рис. 44 Вибір виміру, на основі якого буде проводитись аналіз

Наступним кроком обираємо вимір, на основі якого буде проводитися аналіз (рис. 45). Визначаємо поля та їх типи значень (рис. 45).

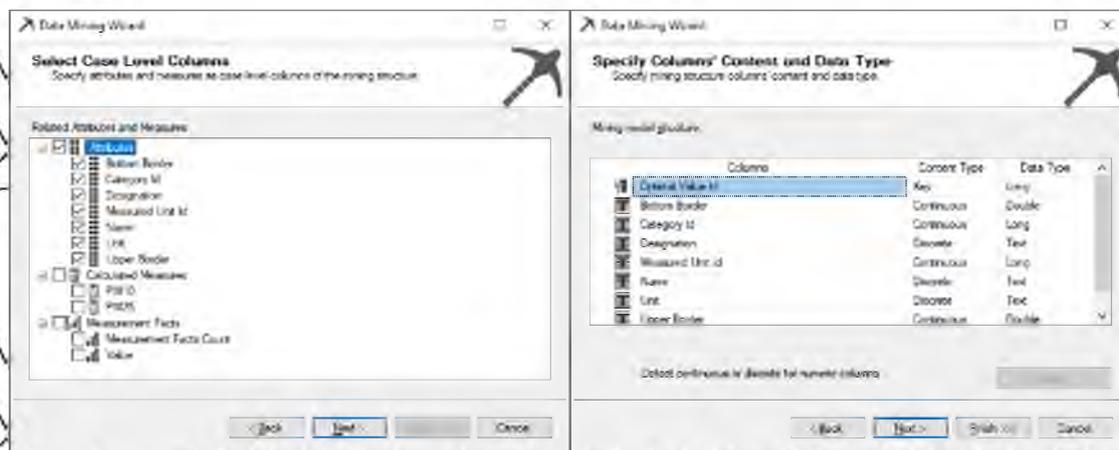


Рис. 45 Визначення полей та я їх типі значень

На основі розгорнутої структури було отримано модель, яка наведена на

рис. 46.

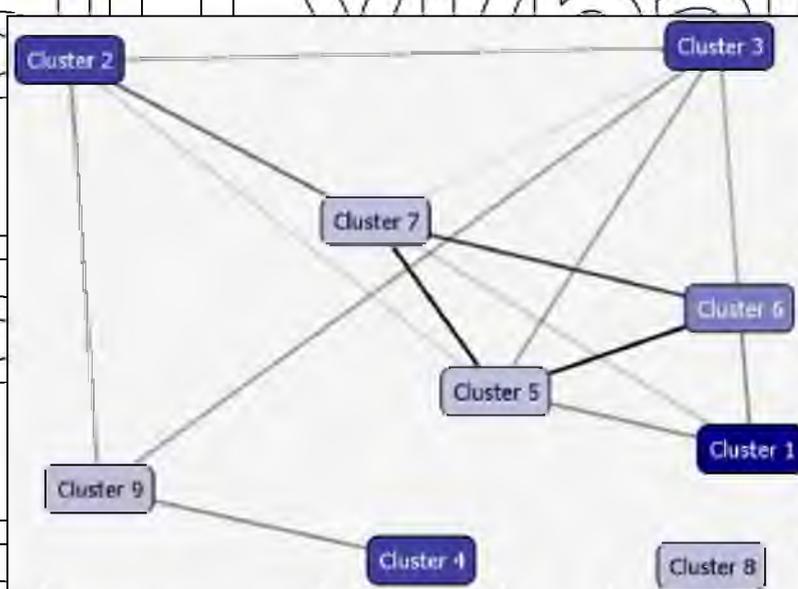


Рис 46 Модель кластеризації

Аналізуючи діаграму на рис. 46 можна зробити висновок, що найбільше даних зосереджено у кластері 1 ("Cluster 1"). Наступними є кластер 2 ("Cluster 2"), кластер 3 ("Cluster 3") та кластер 4 ("Cluster 4").

4 РЕЗУЛЬТАТИ ДОСЛІДЖЕННЯ

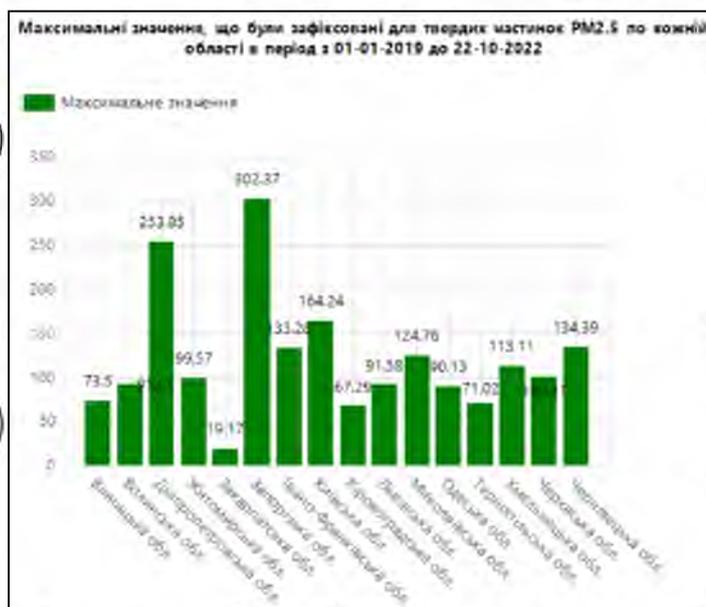
4.1 Результати дослідження на основі оперативного аналізу

Розглянемо результати дослідження, що були отримані за допомогою інструментів SSRS. Першим кроком було сформовано звіт та графік, які демонструють максимальні значення твердих частинок $PM_{2.5}$, що були зафіксовані в областях України за період часу з 01-01-2019 по 22-10-2022. На рис. 48 відображено звіт, на рис. 47 відповідний графік.

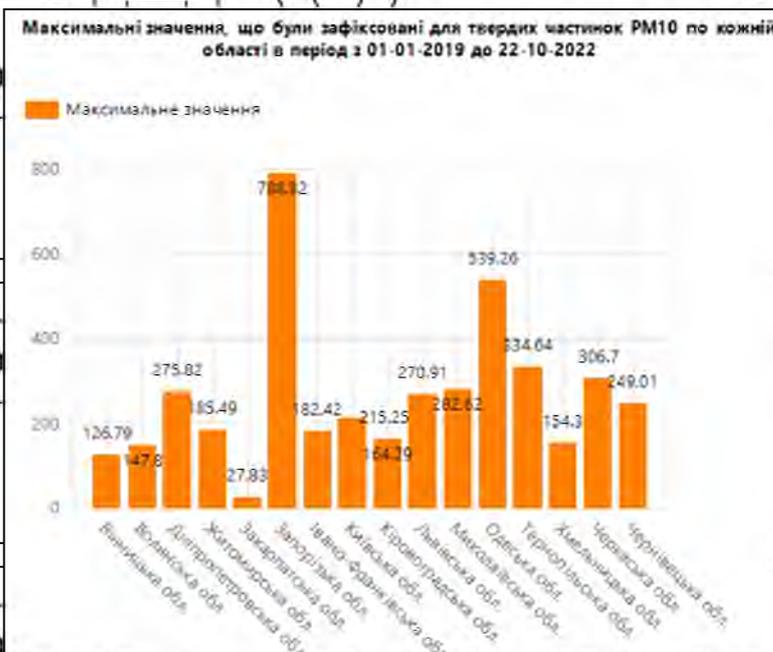
Область	Максимальне значення	Категорія
Вінницька обл.	73.5	Дуже шкідливий
Волинська обл.	91.61	Дуже шкідливий
Дніпропетровська обл.	253.85	Дуже тяжкий
Житомирська обл.	99.57	Дуже шкідливий
Закарпатська обл.	19.17	Помірний
Запорізька обл.	302.37	Дуже тяжкий
Івано-Франківська обл.	133.28	Дуже тяжкий
Київська обл.	164.24	Дуже тяжкий
Кіровоградська обл.	67.29	Дуже шкідливий
Львівська обл.	91.58	Дуже шкідливий
Миколаївська обл.	124.76	Дуже тяжкий
Одеська обл.	90.13	Дуже шкідливий
Тернопільська обл.	71.02	Дуже шкідливий
Хмельницька обл.	113.11	Дуже тяжкий

Рис. 47 Звіт максимальних значень $PM_{2.5}$ за областями

Аналогічний звіт та графік було створено для вмісту твердих частинок PM_{10} , що відображені на рис. 48 та рис. 50 відповідно.

Рис. 48 Графік максимальних значень PM_{2.5} за областями

Область	Максимальне значення	Категорія
Волинська обл.	126.79	Дуже шкідливий
Рівненська обл.	147.8	Дуже шкідливий
Дніпропетровська обл.	275.62	Дуже токсичний
Житомирська обл.	185.49	Дуже токсичний
Закарпатська обл.	27.83	Умовний
Запорізька обл.	788.62	Дуже токсичний
Івано-Франківська обл.	192.12	Дуже токсичний

Рис. 49 Звіт максимальних значень PM₁₀ за областямиРис. 50 Графік максимальних значень PM₁₀ за областями

Висновки

На графіках відображено, що максимальне зафіксоване середньодобове значення твердих частинок $PM_{2.5}$ спостерігається в таких областях, як Запорізька (302.37), Дніпропетровська (253.85) та Київська (164.24). Найменші показники спостерігаються в таких областях, як: Закарпатська (19.17), Кіровоградська (67.29), Тернопільська (71.02), Вінницька (73.5).

Для твердих частинок PM_{10} максимальні значення були зафіксовані в таких областях, як: Запорізька (788.82), Одеська (539.26), Тернопільська (334.64), Черкаська (306.7). Мінімальні значення були зафіксовані в Закарпатській (27.83), Вінницькій (126.79) та Волинській (147.8) областях.

Наступним кроком розглянемо звіт та графік (рис. 51), які відображають скільки разів було зафіксовано середньодобові значення твердих частинок $PM_{2.5}$ значення яких відносяться до шкідливого рівню на певній станції за весь час (починаючи з 2019 р.).



Рис. 51 Графік та звіт кількості вимірювань $PM_{2.5}$, що відносяться до шкідливого рівню за областю

Розглянемо аналогічні дані стосовно твердих частинок PM_{10} на рис. 52.



Рис. 52 Графік та звіт кількості вимірювань PM₁₀, що відносяться до шкідливого рівню за областю

Висновки

На звітах бачимо, що сумарна кількість досліджень, що зафіксували шкідливий рівень значень твердих частинок для PM_{2.5} дорівнює 111, для PM₁₀ –

277. Далі розглянемо звітну інформацію, що відображає категорії оптимальних значень та кількість вимірювань середньодобових значень діоксиду сірки (SO₂), що відповідають кожній категорії оптимальних значень за певною станцією (рис.

53).



Рис. 53 Графік та звіт кількості вимірювань до категорій оптимальних значень для діоксиду сірки

Аналогічні дані були сформовані для чадного газу (CO), що продемонстровані на рис. 54.

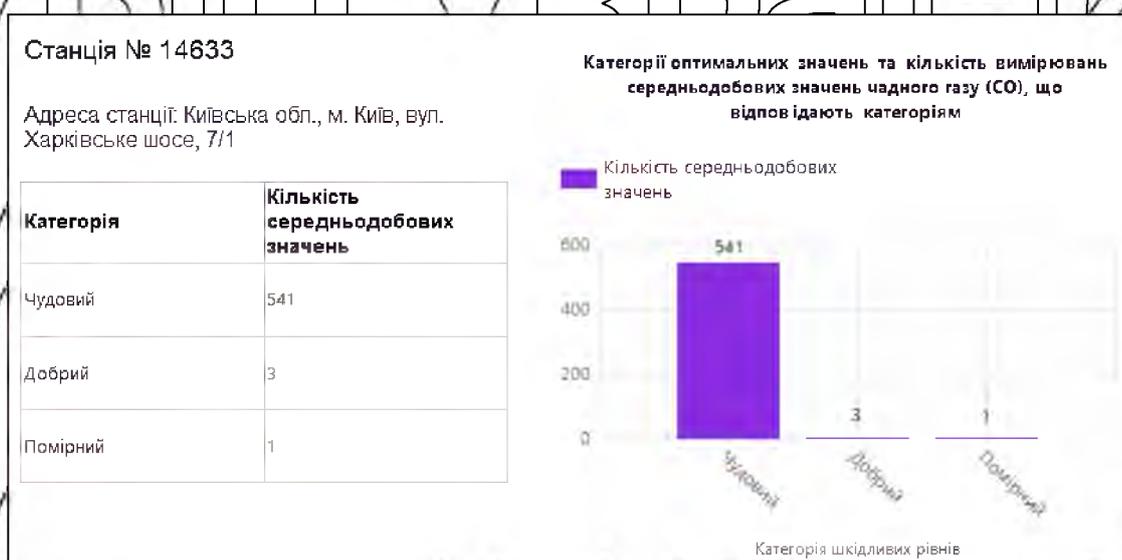


Рис. 54 Графік та звіт кількості вимірювань до категорій оптимальних значень для чадного газу

Висновки

Виходячи із наведених графіків можна зробити висновок, що стан повітря із вмістом діоксиду сірки на станції № 16743 помірний, спостерігаємо 57 випадків вимірювань, коли середньодобове значення відповідало шкідливому рівню.

Проїдежну картину спостерігаємо на станції № 14633 із чадним газом. Графіки показують, що ситуація із вмістом чадного газу довол. чудова за весь час проведення вимірювань.

4.2 Результати дослідження на основі інтелектуального аналізу

Окрім проведення аналізу на основі інструментів SSRS також було проведено інтелектуальний аналіз засобами Data Mining. Розглянемо результати застосованих алгоритмів, а саме: часові ряди та 1-Rule.

4.2.1 Алгоритм 1-Rule. Для реалізації алгоритму 1-Rule було виділено 2 класи задовільного ("Добре") та незадовільного вмісту діоксиду сірки у повітрі ("Погано"), показник яких визначається у порівнянні з середнім показником стану повітря за весь період дослідження.

Спочатку проведено вибірку даних, де визначено середнє значення вмісту діоксиду сірки та класифікацію (рис. 55-56).

```

declare @avgStationsValue float;
SELECT @avgStationsValue = AVG(mf.average_value)
from measurement_facts mf
where mf.measured_unit_id = 8;
select sd.station_id
      concat(sd.region, ' ', sd.city, ' ', sd.street, ' ', sd.building_number) as '[Номер станції]',
      dd.date_id as '[Дата]',
      ROUND(mf.average_value, 2) as '[Середнє значення]',
      CASE
        WHEN ROUND(mf.average_value, 2) <= @avgStationsValue
          THEN N'Добре'
        ELSE N'Погано'
      END as '[Класифікація]'
from measurement_facts mf
  inner join date_dim dd on mf.date_id = dd.date_id
  inner join station_dim sd on mf.station_id = sd.station_id
  inner join measured_unit_dim mud on mud.measured_unit_id =
      mf.measured_unit_id
where mud.measured_unit_id = 8
and mf.station_id = '14410';

```

Рис. 55 Реалізація вибірки середнього значення та класифікацію за станцією

[Station]	[Address]	[Date]	[Average Value]	[Classification]	
1	1228	Azovstalska Street, 4, Mariupol	2019-08-26	280,07	Poor
2	1228	Azovstalska Street, 4, Mariupol	2019-08-27	139,72	Poor
3	1228	Azovstalska Street, 4, Mariupol	2019-08-28	11,27	Good
4	1228	Azovstalska Street, 4, Mariupol	2019-08-29	5,33	Good
5	1228	Azovstalska Street, 4, Mariupol	2019-08-30	5,43	Good
6	1228	Azovstalska Street, 4, Mariupol	2019-08-31	4,13	Good
7	1228	Azovstalska Street, 4, Mariupol	2019-09-01	4,91	Good
8	1228	Azovstalska Street, 4, Mariupol	2019-09-02	5,06	Good
9	1228	Azovstalska Street, 4, Mariupol	2019-09-03	6,93	Good
10	1228	Azovstalska Street, 4, Mariupol	2019-09-04	8,58	Good
11	1228	Azovstalska Street, 4, Mariupol	2019-09-05	6,64	Good
12	1228	Azovstalska Street, 4, Mariupol	2019-09-06	3,74	Good
13	1228	Azovstalska Street, 4, Mariupol	2019-09-07	3,58	Good
14	1228	Azovstalska Street, 4, Mariupol	2019-09-08	5,64	Good
15	1228	Azovstalska Street, 4, Mariupol	2019-09-09	6,63	Good
16	1228	Azovstalska Street, 4, Mariupol	2019-09-10	9,19	Good
17	1228	Azovstalska Street, 4, Mariupol	2019-09-11	6,71	Good
18	1228	Azovstalska Street, 4, Mariupol	2019-09-12	6,09	Good
19	1228	Azovstalska Street, 4, Mariupol	2019-09-13	7,54	Good
20	1228	Azovstalska Street, 4, Mariupol	2019-09-14	5,68	Good

Рис. 56 Результат вибірки середнього значення та класифікацію за станцією

Наступним кроком було визначено класифікації вмісту діоксиду сірки та її ймовірність відносно середнього показника серед заданих фактів відповідно до станції (рис. 57) та часового періоду (рис. 58).

[Station]	[Value amount by station]	[Less amount than avg]	[Bigger amount than avg]	[Avg by all stations]	[Avg by station]	[Classification]	[Less than avg probability %]	[Bigger than avg probability %]	
1	1228	896	431	398	31,4	34,52	Poor	55,58	44,42
2	1227	333	161	185	31,4	35,33	Poor	45,65	54,35
3	4152	177	166	21	31,4	15,4	Good	85,14	14,86
4	4220	600	436	164	31,4	27,15	Good	72,67	27,33

Рис. 57 Класифікація відповідно до станції

[Year]	[Month]	[Value amount by year]	[Less amount than avg]	[Bigger amount than avg]	[Avg by all years]	[Avg by year]	[Classification]	[Less than avg probability %]	[Bigger than avg probability %]
2019	8	6	4	2	31,4	74,32	Poor	96,67	23,33
2019	9	30	30	0	31,4	6,76	Good	100	0
2019	10	31	31	0	31,4	5,23	Good	100	0
2019	11	30	20	10	31,4	22,32	Good	96,67	33,33
2019	12	31	31	0	31,4	2,34	Good	100	0
2020	1	31	31	0	31,4	2,01	Good	100	0
2020	2	29	29	0	31,4	1,79	Good	100	0
2020	3	31	31	0	31,4	3,37	Good	100	0
2020	4	30	30	0	31,4	3,96	Good	100	0
2020	5	31	31	0	31,4	2,80	Good	100	0
2020	6	32	32	0	31,4	2,94	Good	100	0
2020	7	67	44	13	31,4	34,02	Good	77,78	22,22
2020	8	62	31	31	31,4	43,33	Poor	96	4
2020	9	62	41	21	31,4	50,77	Poor	96	4
2020	10	65	25	40	31,4	62,1	Poor	38,46	61,54
2020	11	90	31	29	31,4	34,33	Poor	51,67	48,33
2020	12	66	40	26	31,4	26,82	Good	60,81	39,19
2021	1	30	25	28	31,4	29,94	Good	69,39	30,61
2021	2	70	31	28	31,4	41,13	Poor	44,29	55,71
2021	3	80	62	22	31,4	34,69	Poor	66	34
2021	4	95	55	22	31,4	32,13	Poor	67,17	32,83

Рис. 58 Класифікація відповідно до часового періоду

Аналізуючи отримані дані можна виділити наступні правила:

- Станція 4152 має клас вмісту твердих частинок задовільний ("Good") з імовірністю 88,14%.
- Станція 4200 має клас вмісту твердих частинок задовільний ("Good") з імовірністю 72,67%.

- Часовий проміжок з 12.2019 до 06.2020 має клас вмісту твердих частинок задовільний ("Good") з імовірністю 100%.

Можна зробити висновок, що на класифікацію і рівень вмісту твердих частинок впливає час, коли зафіксовано вимір.

4.2.2 Алгоритм часових рядів. Алгоритм часових рядів дозволяє спрогнозувати зміну певних значень, що досліджуються на основі історичних даних. У нашому випадку проводиться прогнозування зміни середньорічного вмісту твердих частинок $PM_{2,5}$ (рис. 59) та PM_{10} (рис. 60) на станції під номером

1131.

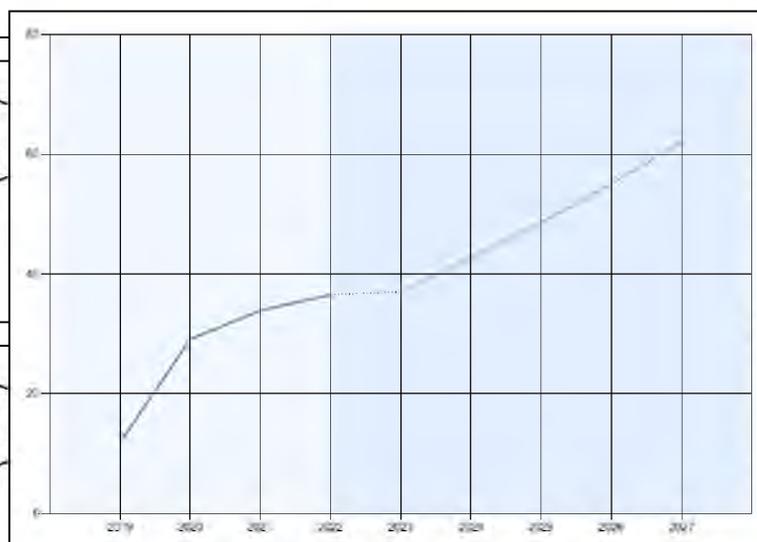


Рис. 59 Сформований графік для параметру PM_{10} у розрізі років

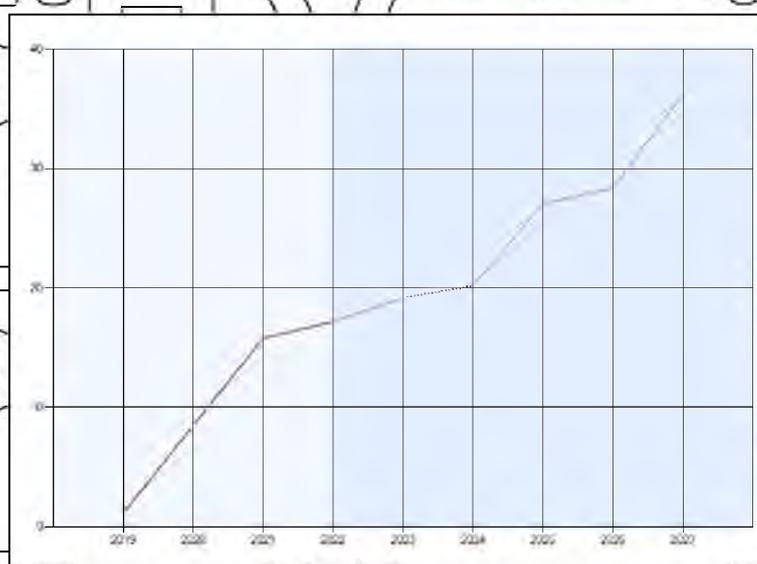


Рис. 60 Сформований графік для параметру $PM_{2.5}$ у розрізі років

Висновки

На основі сформованих графіків бачимо, що починаючи з 2019 року середній показник вмісту твердих частинок (PM_{10} та $PM_{2.5}$) у повітрі стрімко зростає. Можемо зробити висновок, що необхідно приймати відповідні рішення для зменшення кількості вмісту твердих частинок та покращення ситуації в цілому.

ВИСНОВКИ

В результаті виконання магістерської роботи було створено систему аналізу атмосферного повітря на основі сучасних інформаційних технологій таких як OLAP та Microsoft Data Mining, що дозволяє агрегувати великі об'єми даних та проводити дослідження стану атмосферного повітря на основі алгоритмів Data Mining та гнучкого інструментарію для створення звітів. Виконання магістерської роботи було розбито на чотири основні етапи.

На першому етапі було проведено системний аналіз предметної області, визначено основні процеси аналізу стану атмосферного повітря, сформовано технічне завдання та проведено огляд наявних рішень.

На другому етапі було проведено моделювання процесів предметної області, наведено загальні відомості про моделювання бізнес-процесів, створено та описано діаграму прецедентів, послідовності та активності. Отриманий результат надав можливість почати безпосередньо проектувати та розробляти систему.

В рамках третього етапу було розроблено систему аналізу стану атмосферного повітря. Першим кроком було побудовано архітектуру системи, визначено оперативні джерела даних. Наступним – розроблено сховище даних для агрегування аналітичної інформації. Засобами SSAS на основі створеного сховища даних було розгорнуто багатомірний куб. Третім кроком засобами Data Flow було організовано міграцію даних із оперативних джерел інформації до сховища даних. Четвертим – засобами SSRS побудовано основний набір звітів для аналізу. Останнім кроком було зроблено огляд алгоритмів Data Mining для аналізу стану атмосферного повітря.

На четвертому етапі було продемонстровано результати дослідження, що були отримані під час розробки системи. Їх можна поділити на два типи: результати, отримані на основі засобів SSRS та результати досліджень на основі інструментів Data Mining.

Розглянемо результати, отримані засобами SSRS. На першому етапі було сформовано звіт та графік, які демонструють максимальні значення твердих

частинок $PM_{2.5}$ та PM_{10} , що були зафіксовані в областях України у розрізі часу з 01-01-2019 по 22-10-2022.

Було з'ясовано, що максимальне зафіксоване середньодобове значення твердих частинок $PM_{2.5}$ спостерігається в таких областях, як: Запорізька (302.37), Дніпропетровська (253.85) та Київська (164.24). Найменші показники спостерігаються в таких областях, як: Закарпатська (19.17), Кіровоградська (67.29), Тернопільська (71.02), Вінницька (73.5). Для твердих частинок PM_{10} максимальні значення були зафіксовані в таких областях, як: Запорізька (788.82), Одеська (539.26), Тернопільська (334.64), Черкаська (306.7). Мінімальні значення були зафіксовані в Закарпатській (27.83), Вінницькій (126.79) та Волинській (147.8) областях. Великі значення даних показників обумовлені саме великою кількістю промислових підприємств в цих регіонах та великою кількістю автотранспорту на дорогах.

Другий етап полягав у аналізі інформації стосовно того, скільки разів було зафіксовано середньодобові значення твердих частинок $PM_{2.5}$ та PM_{10} , значення яких відносяться до шкідливого рівню на одній із станцій в Запоріжжі за весь час починаючи з 2019 р. З'ясовано, що сумарна кількість досліджень, що зафіксували шкідливий рівень значень твердих частинок для $PM_{2.5}$ дорівнює 111, для PM_{10} – 217, що відповідає помірній ситуації.

На третьому етапі було отримано інформацію, яка відображає категорії оптимальних значень та кількість вимірювань середньодобових значень діоксиду сірки (SO_2), що відповідають кожній категорії оптимальних значень за однією із станцій у Тернопільській області. Виходячи із наведених графіків можна зробити висновок, що стан повітря із вмістом діоксиду сірки на станції № 16743 помірний, спостерігаємо 67 випадків вимірювань, коли середньодобове значення відповідало шкідливому рівню.

Таке саме спостереження було проведено для вмісту чадного газу на станції № 14633 у Київській області. Спостерігаємо протилежну картину. Графіки показують, що ситуація із вмістом чадного газу доволі чудова за весь час проведення вимірювань.

Розглянемо результати отримані засобами Data Mining, а саме за допомогою алгоритмів часових рядів та 1-Rule.

На основі сформованих графіків алгоритмом часових рядів отримали, що починаючи з 2019 року середній показник вмісту тяжких частинок (PM_{10} та $PM_{2.5}$) у повітрі стрімко зростає на обраній станції. Можемо також зробити висновок, що необхідно приймати відповідні заходи для зменшення кількості вмісту твердих частинок та покращення ситуації в цілому.

На основі отриманих результатів, що були отримані після застосування алгоритму 1-Rule було отримано наступні висновки: станція № 4152 має клас вмісту твердих частинок задовільний ("Good") з імовірністю 88,14%; станція № 4200 має клас вмісту твердих частинок задовільний ("Good") з імовірністю 72,67%; часовий проміжок з 12.2019 до 06.2020 має клас вмісту твердих частинок задовільний ("Good") з імовірністю 100%. Можна зробити висновок, що на класифікацію і рівень вмісту твердих частинок впливає час коли зафіксовано вимір.

Підсумовуючи результати розробки системи аналізу атмосферного повітря можемо зробити висновки, що сучасні технології OLAP та Data Mining доцільно використовувати для аналізу стану атмосферного повітря. Вони дозволяють створювати гнучкі засоби аналізу екологічних показників та за допомогою алгоритмів Data Mining прогнозувати зміну цих показників та знаходити взаємодії між тими чи іншими чинниками, що впливають на його зміну. Дані, що є результатом застосування наведеного інструментарію можуть бути використані для прийняття відповідних рішень для покращення стану атмосферного повітря в країні, та, як результат, покращити умови життя громадян в цілому.

СНИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Air pollution: Everything You Need to Know – Розміщено на платформі NRDC.ORG – [Електронний ресурс] – режим доступу:

<https://www.nrdc.org/stories/air-pollution-everything-you-need-know#effects>.

2. Air pollution: Overview – Розміщено на платформі World Health Organization – [Електронний ресурс] – режим доступу:

<https://www.who.int/health-topics/air-pollution>.

3. Проблематика громадського здоров'я – Розміщено на платформі CGZ.VN.UA – [Електронний ресурс] – режим доступу:

http://cgz.vn.ua/problematika-gromadskogo-zdorovya/problematika-gromadskogo-zdorovya_455.html

4. Air pollution and it's control – Розміщено на платформі BYJUS.COM –

[Електронний ресурс] – режим доступу: <https://byjus.com/biology/air-pollution-control/>.

5. Air pollution: A Global Problem – Розміщено на платформі NCAR –

[Електронний ресурс] – режим доступу: <https://gal.ncar.edu/pressroom/features/air-pollution-a-global-problem>.

6. How to calculate our air quality index and why we need it? – Розміщено на

платформі Breeze technologies – [Електронний ресурс] – режим доступу: <https://www.breeze-technologies.de/blog/what-is-an-air-quality-index-how-is-it-calculated/>.

7. SaveEcoBot: Єдиний в Україні екологічний чат-бот – Розміщено на

платформі SaveEcoBot [Електронний ресурс] – режим доступу: <https://www.saveecobot.com/>.

8. What is UML? – Розміщено на платформі Visual Paradigm – [Електронний

ресурс] – режим доступу: <https://www.visual-paradigm.com/guide/uml-unified-modeling-language/what-is-uml/>.

9. Unified Modeling Language: Introduction – Розміщено на платформі Geeks

for geeks – [Електронний ресурс] – режим доступу: <https://www.geeksforgeeks.org/unified-modeling-language-uml-introduction/>.

10. Use case diagram – Розміщено на платформі Tech target – [Електронний ресурс] – режим доступу: <https://www.techtarget.com/whatis/definition/use-case-diagram>.

11. What is a sequence diagram? – Розміщено на платформі Lucidchart – [Електронний ресурс] – режим доступу: <https://www.lucidchart.com/pages/uml-sequence-diagram>.

12. Explore the UML sequence diagram? – Розміщено на платформі IBM – [Електронний ресурс] – режим доступу: <https://developer.ibm.com/articles/the-sequence-diagram/>.

13. What is an activity diagram? – Розміщено на платформі Lucidchart – [Електронний ресурс] – режим доступу: <https://www.lucidchart.com/pages/uml-activity-diagram>.

14. What is an activity diagram? – Розміщено на платформі Visual Paradigm – [Електронний ресурс] – режим доступу: <https://www.visual-paradigm.com/guide/uml-unified-modeling-language/what-is-activity-diagram/>.

15. The importance of Good Software Architecture – Розміщено на платформі DZone – [Електронний ресурс] – режим доступу: <https://dzone.com/articles/the-importance-of-a-good-software-architecture>.

16. What is a deployment diagram? – Розміщено на платформі Visual Paradigm – [Електронний ресурс] – режим доступу: <https://www.visual-paradigm.com/guide/uml-unified-modeling-language/what-is-deployment-diagram/>.

17. What is Analysis Services? – Розміщено на платформі Learn.microsoft.com – [Електронний ресурс] – режим доступу: <https://learn.microsoft.com/en-us/analysis-services/analysis-services-overview?view=asallproducts-allversions>.

18. Introduction to dataflows and self-service data prep – Розміщено на платформі Learn.microsoft.com – [Електронний ресурс] – режим доступу: <https://learn.microsoft.com/en-us/power-bi/transform-model/dataflows/dataflows-introduction-self-service>.

19. What is SQL Server Reporting Services (SSRS)? – Розміщено на

платформі Learn.microsoft.com – [Електронний ресурс] – режим доступу: <https://learn.microsoft.com/en-us/sql/reporting-services/create-deploy-and-manage-mobile-and-paginated-reports?view=sql-server-ver16>.

20. Key performance indicators (KPIs): What they are and how to use them

– Розміщено на платформі Microsoft.com – [Електронний ресурс] – режим доступу: <https://www.microsoft.com/en-us/microsoft-365/business-insights-ideas/resources/what-are-kpis-and-how-to-use-them>.

21. Why a SQL Server BI platform? – Розміщено на платформі

Microsoft.com – [Електронний ресурс] – режим доступу: <https://www.microsoft.com/en-us/sql-server/sql-business-intelligence?rtc=1>.

22. Microsoft Naive Bayes Algorithm – Розміщено на платформі

Learn.microsoft.com – [Електронний ресурс] – режим доступу: <https://learn.microsoft.com/en-us/analysis-services/data-mining/microsoft-naive-bayes-algorithm?view=asallproducts-allversions>.

23. Microsoft Time Series Algorithm – Розміщено на платформі

Learn.microsoft.com – [Електронний ресурс] – режим доступу: <https://learn.microsoft.com/en-us/analysis-services/data-mining/microsoft-time-series-algorithm?view=asallproducts-allversions>.

24. Learn One-Rule algorithm – Розміщено на платформі

Geeks for geeks – [Електронний ресурс] – режим доступу: <https://www.geeksforgeeks.org/learn-one-rule-algorithm/>.

25. Microsoft Clustering algorithm – Розміщено на платформі

Learn.microsoft.com – [Електронний ресурс] – режим доступу: <https://learn.microsoft.com/en-us/analysis-services/data-mining/microsoft-clustering-algorithm?view=asallproducts-allversions>.

26. What are dataflows? – Розміщено на платформі Learn.microsoft.com –

[Електронний ресурс] – режим доступу: <https://learn.microsoft.com/en-us/power-query/dataflows/overview-dataflows-across-power-platform-dynamics-365>.