

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ БІОРЕСУРСІВ
І ПРИРОДОКОРИСТУВАННЯ УКРАЇНИ

Кваліфікаційна наукова праця
на правах рукопису

ШЕВЧЕНКО ДМИТРО ВІТАЛІЙОВИЧ

УДК 004.8:504.5

ДИСЕРТАЦІЯ

**РОЗРОБЛЕННЯ ІНТЕЛЕКТУАЛЬНОЇ ІНФОРМАЦІЙНОЇ ТЕХНОЛОГІЇ
ДЛЯ ОЦІНЮВАННЯ СТАНУ АТМОСФЕРНОГО ПОВІТРЯ**

122 «Комп'ютерні науки»
12 «Інформаційні технології»

Подається на здобуття ступеня доктора філософії

Дисертація містить результати власних досліджень.
Використання ідей, результатів і текстів інших авторів мають посилання
на відповідне джерело Д. В. Шевченко

Науковий керівник
ГОЛУБ Белла Львівна,
кандидат технічних наук, доцент

Київ – 2025

АНОТАЦІЯ

Шевченко Д. В. Розроблення інтелектуальної інформаційної технології для оцінювання стану атмосферного повітря. Кваліфікаційна наукова праця на правах рукопису. Кваліфікаційна наукова праця на правах рукопису.

Дисертація на здобуття наукового ступеня доктора філософії за спеціальністю 122 «Комп'ютерні науки». Національний університет біоресурсів і природокористування України, Київ, 2025.

Дисертація присвячена розробленню інтелектуальної інформаційної технології для оцінювання стану атмосферного повітря з метою забезпечення науково обґрунтованої підтримки прийняття управлінських рішень. Актуальність дослідження зумовлена глобальними екологічними викликами та необхідністю цифрової трансформації систем моніторингу довкілля. За результатами аналізу сучасних методів і технічних рішень щодо контролю якості атмосферного повітря встановлено, що, попри позитивні тенденції, ця сфера в Україні потребує подальшого вдосконалення. Зокрема, необхідним є впровадження уніфікованих методик, розроблення прогностичних моделей та інтеграція з міжнародними системами моніторингу. Відсутність єдиного централізованого сховища екологічних даних ускладнює повноцінний аналіз і прогнозування рівнів забруднення, особливо з урахуванням історичних показників.

Метою дисертаційної роботи є розроблення інтелектуальної інформаційної технології для забезпечення науково обґрунтованої підтримки прийняття управлінських рішень державними органами щодо якості атмосферного повітря.

Наукова новизна отриманих результатів полягає в науковому обґрунтуванні, розробленні та практичній реалізації поєднання трьох ключових технологій – обробки транзакцій (OLTP), інтелектуального аналізу даних (Data Mining) та аналітичної обробки в режимі реального часу (OLAP) для оцінювання стану атмосферного повітря. Це забезпечило повний цикл опрацювання параметрів якості повітря від отримання первинних даних до формування аналітичних висновків і рекомендацій для підтримки управлінських рішень. Удосконалено та

впроваджено методи інтерпретації перевищень нормативних рівнів забруднення на основі ключових показників ефективності (KPI), що підвищує оперативність контролю дотримання нормативів у визначених часових діапазонах. Отримали подальший розвиток теоретичні та прикладні засади інтеграції потоків даних із сенсорів і відкритих джерел у єдине сховище (розроблене вперше), що забезпечує уніфікацію, цілісність та агрегацію екологічних даних для подальшого аналітичного опрацювання.

Практичне значення роботи полягає у тому, що розроблена інформаційна технологія може бути впроваджена у діяльність міських та регіональних служб моніторингу довкілля. Система забезпечує *оперативну оцінку* якості повітря, що сприяє своєчасному інформуванню влади та населення про екологічну ситуацію. Запропоновані аналітичні інструменти (OLAP-звіти, панелі моніторингу, розрахунок KPI) дають можливість керівним органам обґрунтовано приймати рішення щодо заходів з поліпшення стану атмосферного повітря. Результати дослідження апробовано – матеріали дисертації доповідалися на низці міжнародних і всеукраїнських конференцій з інформаційних технологій та екологічної безпеки. Розроблені методи і компоненти системи використано в навчальному процесі при підготовці фахівців з баз даних та екологічного моніторингу.

У вступі подано загальну характеристику дослідження, обґрунтовано актуальність теми, сформульовано мету і завдання, визначено об'єкт, предмет, методи дослідження, викладено наукову новизну та практичне значення результатів, наведено відомості про особистий внесок здобувача, апробацію результатів та структуру дисертації.

Перший розділ присвячено аналізу стану наукових досліджень і технологічних рішень у сфері моніторингу якості атмосферного повітря. Розглянуто основні методики визначення екологічних індексів якості повітря (AQI, CAQI, EAQI), існуючі підходи до збору, обробки й інтерпретації екологічних даних, а також структуру міжнародних інформаційних систем екологічного моніторингу. Проведено порівняльний аналіз вітчизняних та

зарубіжних систем для виявлення їх переваг і недоліків. Встановлено, що наявні рішення характеризуються фрагментарністю джерел даних, відсутністю інтегрованих аналітичних інструментів та обмеженою гнучкістю аналізу. Це підтвердило необхідність створення інтелектуальної інформаційної технології, здатної об'єднувати різноманітні дані та автоматично оцінювати стан атмосферного повітря.

У другому розділі описано розроблення архітектури інформаційної системи моніторингу. Система складається з модулів збору даних, централізованого сховища даних (Data Warehouse), підсистем аналітики та користувацьких інтерфейсів. Обґрунтовано вибір архітектурних підходів на засадах модульності, масштабованості та сумісності з потоковими джерелами даних. Запропоновано багаторівневу модель бази даних із нормалізованою структурою таблиць, яка забезпечує ефективну інтеграцію інформації з відкритих джерел та власних сенсорних вузлів. Розроблено концептуальну та логічну схеми бази даних, визначено ключові виміри і фактні таблиці для реалізації OLAP-аналізу екологічної інформації.

Третій розділ присвячено методам і технологіям збору та попередньої обробки даних про стан атмосферного повітря. Розглянуто підходи до уніфікації форматів даних з різних джерел, автоматизації процесів очищення та відновлення екологічних показників. Детально описано реалізацію процесу ETL (Extract, Transform, Load) – перенесення даних із різноманітних джерел до єдиного сховища. Наведено розроблені алгоритми обчислення агрегованих показників (середніх, мінімальних, максимальних значень) для подальшого аналізу.

У четвертому розділі містяться результати аналітичної обробки та інтерпретації даних про якість повітря. Реалізовано алгоритми розрахунку інтегральних індексів якості повітря (AQI, CAQI, EAQI) з урахуванням методологічних особливостей кожного індексу. Проведено оцінювання точності та оперативності отриманих результатів, проаналізовано продуктивність запропонованої системи. Впроваджено і протестовано методи інтелектуального аналізу даних (Data Mining), зокрема алгоритми кластеризації («K-means») та

класифікації («Наївний Байєс»), для розпізнавання закономірностей і відновленні даних. Розроблено низку багатовимірних OLAP-звітів, що дозволяють аналізувати екологічну інформацію у часовому та просторовому розрізах. Запропоновано систему ключових показників ефективності (KPI) для моніторингу перевищень гранично допустимих концентрацій забруднюючих речовин та оцінки ефективності заходів з покращення якості повітря.

У висновках узагальнено результати дослідження та сформульовано основні наукові й практичні здобутки. Створено інтелектуальну інформаційну технологію, що поєднує методи OLTP, OLAP і Data Mining для комплексного оцінювання стану атмосферного повітря. Розроблено архітектуру трирівневої системи моніторингу з багатовимірною моделлю сховища даних та ETL-процесом для уніфікації, очищення й агрегування показників. Запропоновано методи аналізу перевищень нормативів на основі KPI, а також моделі кластеризації, наївного Байєса й асоціативних правил для виявлення закономірностей і відновлення пропущених даних. Результати можуть бути використані для вдосконалення національної системи моніторингу атмосферного повітря та підтримки управлінських рішень у сфері охорони довкілля.

Ключові слова: екологічна безпека, моніторинг атмосферного повітря, інформаційно-аналітична система, інтелектуальна технологія, інформаційні технології, OLAP, Data Mining, підтримка прийняття рішень, надійність та достовірність даних, людино-машинний інтерфейс.

ANNOTATION

Shevchenko D. V. Development of intelligent information technology for assessing the state of atmospheric air. The qualification scientific work on the rights of the manuscript.

The dissertation on acquisition of Doctor of Philosophy scientific degree of a specialty 122 «Computer Science». National University of Life and Environmental Sciences of Ukraine, Kyiv, 2025.

The dissertation is devoted to the development of an intelligent information technology for assessing the state of atmospheric air in order to provide scientifically grounded support for environmental management decision-making. The relevance of the research is determined by global environmental challenges and the need for digital transformation of environmental monitoring systems. Based on the analysis of existing methods and technical solutions for air quality control, it has been established that, despite positive trends, this field in Ukraine requires further improvement. In particular, it is necessary to introduce unified methodologies, develop forecasting models, and ensure integration with international monitoring systems. The absence of a centralized environmental data repository complicates comprehensive analysis and forecasting of pollution levels, especially when taking historical indicators into account.

The aim of the dissertation is to develop an intelligent information technology to provide scientifically grounded support for decision-making by public authorities regarding atmospheric air quality.

The scientific novelty of the obtained results lies in the theoretical justification, development, and practical implementation of an integrated approach that combines three key technologies – Online Transaction Processing (OLTP), Data Mining, and Online Analytical Processing (OLAP) – for assessing the state of atmospheric air. This approach provides a full cycle of air quality parameter processing, from the acquisition of primary data to the generation of analytical conclusions and recommendations for decision-making support. Methods for interpreting exceedances of regulatory pollution levels based on key performance indicators (KPI) have been improved and implemented, enhancing the efficiency of compliance control within defined time intervals. Theoretical and applied principles for integrating data streams from sensors and open sources into a unified data warehouse (developed for the first time) have been further advanced, ensuring the unification, integrity, and aggregation of environmental data for subsequent analytical processing.

The practical significance of the work lies in the fact that the developed information technology can be implemented in the activities of urban and regional environmental monitoring services. The system provides real-time assessment of air

quality, facilitating timely notification of authorities and the public about the environmental situation. The proposed analytical tools (OLAP reports, monitoring dashboards, KPI calculations) enable decision-makers to adopt well-grounded measures aimed at improving atmospheric air quality. The research results have been tested and presented at numerous international and national conferences on information technologies and environmental safety. The developed methods and system components have also been integrated into the educational process for training specialists in databases and environmental monitoring.

The introduction presents the general characteristics of the research, substantiates the relevance of the topic, formulates the aim and objectives, defines the object, subject, and research methods, outlines the scientific novelty and practical significance of the results, and provides information on the author's personal contribution, the approbation of the results, and the structure of the dissertation.

The first chapter is devoted to the analysis of the state of scientific research and technological solutions in the field of atmospheric air quality monitoring. The main methodologies for determining environmental air quality indices (AQI, CAQI, EAQI), existing approaches to the collection, processing, and interpretation of environmental data, as well as the structure of international environmental monitoring information systems, are examined. A comparative analysis of domestic and foreign systems has been conducted to identify their advantages and limitations. It has been established that existing solutions are characterized by fragmented data sources, a lack of integrated analytical tools, and limited analytical flexibility. These findings confirm the necessity of developing an intelligent information technology capable of integrating heterogeneous data and automatically assessing the state of atmospheric air.

The second chapter describes the development of the architecture of the environmental air quality monitoring information system. The system consists of data collection modules, a centralized data warehouse (Data Warehouse), analytical subsystems, and user interfaces. The selection of architectural approaches is justified based on the principles of modularity, scalability, and compatibility with streaming data sources. A multi-tier database model with a normalized table structure is proposed,

ensuring effective integration of information from open sources and proprietary sensor nodes. Conceptual and logical database schemas have been developed, and the key dimensions and fact tables for implementing OLAP-based environmental data analysis have been defined.

The third chapter is devoted to the methods and technologies of data collection and preprocessing for assessing the state of atmospheric air. The approaches to unifying data formats from various sources, automating the processes of data cleaning and restoration of environmental indicators are considered. The implementation of the ETL (Extract, Transform, Load) process for transferring data from heterogeneous sources into a unified data warehouse is described in detail. The developed algorithms for calculating aggregated indicators (average, minimum, and maximum values) for subsequent analysis are presented.

The fourth chapter presents the results of analytical processing and interpretation of air quality data. Algorithms for calculating integrated air quality indices (AQI, CAQI, EAQI) have been implemented, taking into account the methodological features of each index. The accuracy and efficiency of the obtained results were evaluated, and the performance of the proposed system was analyzed. Methods of intelligent data analysis (Data Mining), including clustering (“K-means”) and classification (“Naive Bayes”) algorithms, were implemented and tested to identify patterns and restore missing data. A series of multidimensional OLAP reports were developed, enabling the analysis of environmental information in temporal and spatial dimensions. A system of key performance indicators (KPI) was proposed to monitor exceedances of permissible pollutant concentrations and to assess the effectiveness of measures aimed at improving air quality.

The conclusions summarize the results of the research and formulate the main scientific and practical achievements. An intelligent information technology has been developed that combines OLTP, OLAP, and Data Mining methods for comprehensive assessment of atmospheric air quality. The architecture of a three-tier monitoring system with a multidimensional data warehouse model and an ETL process for data unification, cleaning, and aggregation has been designed. Methods for analyzing regulatory

exceedances based on KPI, as well as clustering, Naive Bayes, and association rule models, have been proposed to identify patterns and restore missing data. The results can be applied to improve the national air quality monitoring system and support decision-making in the field of environmental protection.

Keywords: environmental safety, atmospheric air monitoring, information-analytical system, intelligent technology, information technologies, OLAP, Data Mining, decision support, data reliability and accuracy, human–machine interface.

СПИСОК ОПУБЛІКОВАНИХ ПРАЦЬ ЗА ТЕМОЮ ДИСЕРТАЦІЇ

**Статті у наукових виданнях,
включених до міжнародних наукометричних баз даних**

Web of Science Core Collection та/або Scopus

1. Holub B., Kyrychenko V., Nikolaienko D., Lendiel M., Shevchenko D., Khomenko A. Algorithmic and information support in atmospheric air quality monitoring systems. Proceedings of the Tenth International Congress on Information and Communication Technology (ICICT 2025). Lecture Notes in Networks and Systems. 2025. Vol. 1441. P. 399–412. *(Holub B. сформулювала концепцію та методологію дослідження; Kyrychenko V. обґрунтував алгоритми розрахунку AQI/CAQI; Nikolaienko D. i Lendiel M. проаналізували нормативно-правову базу України та ЄС; Shevchenko D. спроектував архітектуру й сервер збору даних (Django/Celery, MQTT, API), розробив модель БД та виконав оцінку обсягів даних; Khomenko A. впровадив кешування, попередню агрегацію й дослідив ефективність алгоритмів сортування).*

2. Shevchenko D. V., Holub B. L. Regression analysis as a tool for identifying patterns in atmospheric air monitoring data. Proceedings of the 5th Edge Computing Workshop. 2025. Vol. 4. P. 20–27. *(Shevchenko D. V. розробив методологію регресійного аналізу (OLS), створив аналітичний модуль на Python/StatsModels, провів експерименти та інтерпретацію результатів; Holub B. L. сформулювала*

наукову постановку проблеми й дизайн дослідження, здійснила науковий супровід, валідацію результатів і редагування рукопису).

**Статті у наукових виданнях,
включених до Переліку наукових фахових видань України**

3. Шевченко Д. В., Голуб Б. Л. Моніторинг якості повітря в реальному часі. Науковий журнал «Математичні машини і системи». 2025. № 1. С. 103–112. *(Шевченко Д.В. спроектував і впровадив систему моніторингу в реальному часі (MQTT, Django, Celery), розробив модель БД, реалізував обчислення AQI/CAQI, дашборди та провів експерименти; Голуб Б.Л. сформулювала концепцію й методологію, здійснила науковий супровід, валідацію результатів, огляд літератури та редакційне доопрацювання рукопису).*

4. Шевченко Д. В., Голуб Б. Л. Багатовимірна аналітика екологічних даних: застосування OLAP у системах моніторингу. Науковий журнал «Математичні машини і системи». 2025. № 3–4. С. 54–65. *(Шевченко Д.В. спроектував OLAP-архітектуру сховища (star/snowflake), визначив куби/виміри/факти, реалізував вітрини Measurements і ThresholdExceed, налаштував ETL та попередні агрегації, створив дашборди й KPI у Power BI, провів експерименти та інтерпретацію; Голуб Б.Л. сформувала концепцію та методологію дослідження, здійснила науковий супровід та провела експертне рев'ю OLAP-структур (виміри/факти)).*

5. Шевченко Д. В., Голуб Б. Л. Застосування методів Data Mining для багатовимірного аналізу якості атмосферного повітря на основі екологічних даних. Науковий журнал «Наука і техніка сьогодні». Серія: Техніка. 2025. № 8 (49). С. 1801–1810. *(Шевченко Д.В. спроектував застосування методу Data Mining (підготовка даних, StandardScaler/PCA, K-Means, інтеграція з OLAP), виконав експерименти, візуалізацію та інтерпретацію результатів, підготував рукопис; Голуб Б.Л. здійснила концептуалізацію та дизайн дослідження, верифікацію методів і наукове редагування рукопису).*

Тези наукових доповідей

6. Шевченко Д. В. Розроблення інтелектуальної інформаційної технології для оцінювання стану атмосферного повітря. Продовольча та екологічна безпека в умовах війни та повоєнної відбудови: виклики для України та світу, до 125-річчя НУБіП: Міжнародна науково-практична конференція, м. Київ, 25–26 травня 2023 року: тези доповіді.

7. Шевченко Д. В. Розроблення інтелектуальної інформаційної технології для оцінювання стану атмосферного повітря. Теоретичні та прикладні аспекти розробки комп'ютерних систем 2023: V Всеукраїнська науково-практична конференція студентів і аспірантів «Теоретичні та прикладні аспекти розробки комп'ютерних систем 2023», м. Київ, 2023 року: тези доповіді. С. 10–11.

8. Шевченко Д. В. Розроблення інтелектуальної інформаційної технології для оцінювання стану атмосферного повітря. Інформаційні технології: економіка, техніка, освіта 2023: XIV Міжнародна науково-практична конференція молодих вчених «Інформаційні технології: економіка, техніка, освіта 2023», м. Київ, 26–27 жовтень 2023 року: тези доповіді. С. 240–241.

9. Шевченко Д. В., Голуб Б. Л. Архітектура інформаційного компоненту системи моніторингу якості атмосферного повітря. Інформаційне суспільство: технологічні, економічні та технічні аспекти становлення: Міжнародна наукова інтернет-конференція «Інформаційне суспільство: технологічні, економічні та технічні аспекти становлення», м. Тернопіль, 18–19 січня 2024 року: тези доповіді. № 85. С. 75–78. *(Шевченко Д.В. розробив структуру архітектури інформаційного компоненту; Голуб Б.Л. здійснила наукове консультування та перевірку коректності архітектурних рішень).*

10. Shevchenko D., Golub B. Information service of the atmospheric air quality monitoring system. Proceedings of the IX International Scientific and Practical Conference «Scientific Problems and Options for Their Solution», Bucharest, Romania. 7–9 February 2024. P. 55–60. *(Shevchenko D. спроектував архітектуру системи та реалізував логіку обміну даними; Golub B. здійснила наукове керівництво та валідацію технічних рішень).*

11. Шевченко Д. В. Архітектура для системи моніторингу та аналізу якості атмосферного повітря. Актуальні питання розвитку науки та техніки в умовах глобалізації: Міжнародна науково-практична конференція «Актуальні питання розвитку науки та техніки в умовах глобалізації», м. Боярка, 14 травня 2024 року: тези доповіді. С. 117–119.

12. Шевченко Д. В. Регресійний аналіз як інструмент для виявлення закономірностей у даних моніторингу атмосферного повітря. Глобальні та регіональні проблеми інформатизації в суспільстві і природокористуванні: XII Міжнародна науково-практична конференція «Глобальні та регіональні проблеми інформатизації в суспільстві і природокористуванні», м. Київ, 2024 року: тези доповіді.

13. Шевченко Д. В. OLAP-технології у системах моніторингу та аналізу якості атмосферного повітря. Актуальні питання розвитку науки та техніки в умовах глобалізації: II Міжнародна науково-практична конференція «Актуальні питання розвитку науки та техніки в умовах глобалізації», м. Боярка, 14 травня 2025 року: тези доповіді. С. 86–88.

14. Шевченко Д. В. Багатовимірна аналітика екологічних даних: застосування OLAP у системах моніторингу. Теоретичні та прикладні аспекти розробки комп'ютерних систем 2025: Всеукраїнська науково-практична конференція студентів і аспірантів «Теоретичні та прикладні аспекти розробки комп'ютерних систем 2025», м. Київ, 24 квітня 2025 року: тези доповіді.

ЗМІСТ

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ	15
ВСТУП	16
РОЗДІЛ 1. ЕКОЛОГІЧНИЙ МОНІТОРИНГ ЯКОСТІ АТМОСФЕРНОГО ПОВІТРЯ ЯК ОБ'ЄКТ ДОСЛІДЖЕННЯ	24
1.1. Аналіз предметної області	24
1.2. Стан якості повітря та його динаміка	26
1.3. Індекс якості повітря: визначення та використання	30
1.4. Основні забруднювачі та принципи їх вимірювання	33
1.5. Аналіз існуючих рішень	36
1.6. Стандарти та підходи до оцінювання якості атмосферного повітря	48
1.7. Діаграма прецедентів	55
1.8. Технічне завдання на розробку	57
Висновки до розділу 1	60
РОЗДІЛ 2. МЕТОДИ ТА МОДЕЛІ УПРАВЛІННЯ ЕКОЛОГІЧНИМ МОНІТОРИНГОМ	61
2.1. Алгоритми розрахунку індексів	61
2.2. Алгоритми по дозволеним перевищенням	65
2.3. Регресійна модель	67
2.4. Методи Data Mining	69
2.5. Алгоритм для побудови лінії тренду та інші (Power BI)	75
Висновки до розділу 2	78
РОЗДІЛ 3. ПРОЕКТУВАННЯ СИСТЕМИ	79
3.1. Архітектура системи	79
3.2. Інформаційне забезпечення	94
3.3. Програмне забезпечення	107
Висновки до розділу 3	119

	14
РОЗДІЛ 4. РЕАЛІЗАЦІЯ ТА ВПРОВАДЖЕННЯ СИСТЕМИ	120
4.1. Розгортання системи	120
4.2. Реалізація інтерфейсу користувача	122
4.3. Тестування системи	135
4.4. Аналіз отриманих результатів	137
Висновки до розділу 4	163
ВИСНОВКИ	165
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ	168
ДОДАТКИ	179

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ

AQI – Air Quality Index, індекс якості атмосферного повітря (США).

CAQI – Common Air Quality Index, загальний індекс якості атмосферного повітря (ЄС).

EAQI – European Air Quality Index, європейський індекс якості атмосферного повітря.

ЛОС – Легкі органічні сполуки.

ГДК – Гранично допустима концентрація.

КМУ – Кабінет Міністрів України.

КМДА – Київська міська державна адміністрація.

ВООЗ (WHO) – Всесвітня організація охорони здоров'я (World Health Organization).

EPA – Environmental Protection Agency – Агентство з охорони довкілля США.

ETL – Extract, Transform, Load – процес вилучення, трансформації та завантаження даних.

OLTP – Online Transaction Processing – система оперативної обробки транзакцій.

OLAP – Online Analytical Processing – технологія багатовимірного аналітичного аналізу даних.

DWH – Data Warehouse – сховище даних.

API – Application Programming Interface – програмний інтерфейс взаємодії.

IoT – Internet of Things – Інтернет речей.

MQTT – Message Queuing Telemetry Transport – протокол передавання телеметричних даних.

KPI – Key Performance Indicators – ключові показники ефективності

ВСТУП

Обґрунтування вибору теми дослідження. Забезпечення якості атмосферного повітря є однією з критичних екологічних проблем сучасності. З огляду на глобальні зміни клімату та зростання рівнів забруднення, систематичний екологічний моніторинг стає надзвичайно важливим для попередження негативних наслідків для здоров'я людей та довкілля. Сучасні технології (супутникові знімки, сенсори IoT, мобільні платформи тощо) дозволяють отримувати дані про стан повітря в реальному часі, що значно покращує можливості швидкого реагування на екологічні загрози і розроблення відповідних заходів.

Міністерство цифрової трансформації України сприяє поширенню сучасних інформаційних технологій у системах моніторингу довкілля, гарантуючи прозорість і доступність отриманих даних для громадян.

Проблемна ситуація полягає у фрагментованості джерел інформації про якість атмосферного повітря. Ці джерела (державні мережі спостереження, громадські проекти, відкриті дані тощо) вимірюють різні параметри та використовують неоднакові стандарти оцінки якості. Це унеможливорює надання достовірної та комплексної оцінки. Більше того, відсутність єдиного централізованого сховища накопичених даних ускладнює проведення повноцінного аналізу, а також об'єктивне оцінювання та прогнозування ступеня небезпечності, особливо зважаючи на історичні дані з різних регіонів країни.

Разом з цим, накопичені значні обсяги даних вимагають застосування складних аналітичних методів та моделей. Це мотивує використання інтелектуальних технологій для оцінки, прогнозування та виявлення нових закономірностей між факторами, що впливають на якість атмосферного повітря. Існуючі рішення вже демонструють необхідність інтеграції даних з офіційних та альтернативних джерел для створення комплексної картини стану атмосфери.

Таким чином, актуальність розроблення інтелектуальної інформаційної технології для оцінювання якості повітря зумовлена як глобальними екологічними

викликами та завданнями цифровізації моніторингу, так і потребою в удосконаленні існуючих підходів до аналізу забруднення.

Зв'язок роботи з науковими програмами, планами, темами, грантами. Дисертаційне дослідження виконано в руслі пріоритетних напрямів розвитку науки та техніки України, що стосуються цифровізації довкілля, автоматизації екологічного моніторингу та впровадження інтелектуальних інформаційних систем у природоохоронній сфері. Робота концептуально узгоджується з цілями наукових програм та стратегічних ініціатив у галузі екологічної безпеки, сталого розвитку та цифрової трансформації державного управління.

Дослідження виконані відповідно до планів науково-дослідницької роботи Національного університету біоресурсів і природокористування за темою «Інформаційна та алгоритмічна підтримка інтелектуальних систем в природоохоронній галузі» (номер державної реєстрації №0124U002754).

Мета дослідження. Метою дисертаційної роботи є розроблення інтелектуальної інформаційної технології для забезпечення науково обґрунтованої підтримки прийняття управлінських рішень державними органами щодо якості атмосферного повітря.

Під терміном «інтелектуальна інформаційна технологія» у цій роботі розуміємо реалізацію методів технологій OLTP, OLAP та Data Mining в інформаційних системах, що забезпечує автоматизацію збору, збереження даних в багатовимірних структурах, проведення інтелектуального аналізу даних та аналізу даних в режимі реального часу.

Завдання дослідження. Основними завданнями дисертації є:

1. провести системний аналіз об'єкта дослідження (процес контролю якості атмосферного повітря) та дослідити методи розрахунку індексів, стандарти та підходи до оцінювання якості атмосферного повітря, розробити регресійну модель для виявлення та кількісної оцінки залежностей між параметрами якості повітря та факторами, що на них впливають;
2. сформулювати вимоги до системи екологічного моніторингу атмосферного

повітря з підсистемою аналізу даних та розробити її архітектуру;

3. розробити структуру інформаційного забезпечення системи моніторингу у вигляді реляційної бази даних та сховища даних у вигляді багатовимірного гіперкубу;

4. реалізувати алгоритми збору, накопичення та візуалізації інформації на основі технологій OLTP; розробити алгоритми та реалізувати процеси ETL (отримання, трансформація та завантаження даних) у межах технології OLAP; при цьому використати методи Data Mining (зокрема, кластерний аналіз та модель «Наївного Байєса») для усунення спотворених даних та відновлення пропущеної інформації;

5. використати метод асоціативних правил (як один із методів технології Data Mining) для прогнозування та виявлення нових закономірностей між факторами, що впливають на якість атмосферного повітря;

6. створити OLAP-звіти, визначити та розрахувати ключові показники ефективності (KPI) для проведення багатовимірного аналізу стану атмосферного повітря за часовими та просторовими параметрами, що дадуть змогу підтвердити або спростувати гіпотези, виявлені під час використання технологій Data Mining.

Об'єкт і предмет дослідження. Об'єктом дослідження є процес контролю якості атмосферного повітря засобами екологічного моніторингу. Предметом дослідження є інтелектуальні інформаційні технології для науково обґрунтованої підтримки прийняття управлінських рішень державними органами щодо забезпечення якості атмосферного повітря.

Методи дослідження. У дисертації застосовано комплекс методів, що забезпечують системний підхід до проектування та реалізації інтелектуальних інформаційних технологій. Використано методи системного аналізу для формалізації задачі та проектування архітектури системи. Для обробки даних застосовано технології OLTP і OLAP. Проведено побудову реляційної бази даних для транзакційної обробки, багатовимірної OLAP-моделі для аналізу історичних даних. Підсистема аналізу представлена методами інтелектуального аналізу і аналізу в режимі реального часу. Інтелектуальний аналіз реалізований методами

Data Mining (класифікація алгоритмом “Наївний Байєс”, кластеризація алгоритмом “K-means” тощо). Використано методи візуалізації і розрахунку KPI для представлення результатів моніторингу.

Наукова новизна отриманих результатів. Наукова новизна роботи ґрунтується на цілісному поєднанні підходів поточного й історичного аналізу даних для оцінки якості повітря та полягає у наступному:

- Уперше науково обґрунтовано, запропоновано, апробовано та практично реалізовано поєднання трьох ключових складових (технології обробки транзакцій OLTP, інтелектуальної технології аналізу даних Data Mining та технології аналізу даних у режимі реального часу OLAP) для оцінювання стану атмосферного повітря, що забезпечило повний цикл обробки параметрів якості атмосферного повітря від отримання первинних даних до формування аналітичних висновків та науково обґрунтованих рекомендацій для підтримки прийняття управлінських рішень.

- Удосконалено, апробовано та практично реалізовано нові методи інтерпретації перевищень нормативних рівнів забруднення на основі обчислення ключових показників ефективності (KPI), що розширює можливості сучасного екологічного моніторингу, забезпечуючи оперативну перевірку дотримання нормативів у визначеному часовому діапазоні.

- Отримали подальший розвиток теоретичні та прикладні аспекти реалізації алгоритмів інтеграції реальних потоків даних із екологічних сенсорів та відкритих порталів у єдине сховище даних (структура якого була розроблена вперше), що забезпечило уніфікацію, цілісність і агрегацію екологічних даних для подальшого аналітичного опрацювання.

Практичне значення отриманих результатів. Розроблена технологія може бути застосована у міських і регіональних службах екологічного моніторингу для оперативної оцінки якості повітря та прогнозування забруднень. Вона забезпечує підтримку прийняття рішень у сфері охорони довкілля, формуючи зрозумілу інформацію для влади та громадськості. Використання адаптивної DWH-структури та OLAP-аналітики створює можливість швидко реагувати на

нові типи даних і сценарії, що підвищує гнучкість системи. Розроблені екологічні звіти та показники (KPI), отримані за допомогою системи, можуть інтегруватися в національні та регіональні інформаційні платформи. При цьому моніторинг якості повітря безпосередньо впливає на громадське здоров'я: він інформує населення про рівень забруднення і сприяє зниженню захворюваності, пов'язаної з погіршенням довкілля. Крім того, централізоване зберігання інтегрованих даних забезпечує аналітикам і керівникам можливість обґрунтованого прийняття рішень.

Під час вивчення дисципліни “Технології програмування баз даних” студенти розробляють програмний проект екологічного моніторингу, використовуючи розроблену структуру системи і отримані дані. Розроблено навчальне видання «Методичні вказівки до виконання лабораторних робіт з дисципліни «Технології програмування баз даних».

Особистий внесок здобувача. Автор самостійно виконано проектування архітектури системи, розробку програмних модулів та методів аналітики. Особисто реалізовано схему збору та агрегації даних, побудовано Data Warehouse, запрограмовано обчислення індексів якості повітря і механізми OLAP-звітності. Усі експериментальні дослідження, налаштування програмного забезпечення та аналіз результатів виконано за активної участі здобувача, що підтверджено документально та апробацією окремих результатів.

Апробація результатів дослідження. Результати дисертаційного дослідження були представлені на міжнародних і всеукраїнських науково-технічних конференціях у галузі інформаційних технологій і екологічного моніторингу.

- Міжнародна науково-практична конференція «Продовольча та екологічна безпека в умовах війни та повоєнної відбудови: виклики для України та світу, до 125-річчя НУБіП», 25–26 травня 2023, Київ.

- V Всеукраїнська науково-практична конференція студентів і аспірантів «Теоретичні та прикладні аспекти розробки комп'ютерних систем 2023», 26 квітня 2023 року, НУБіП України, Київ.

- XIV Міжнародна науково-практична конференція молодих вчених «Інформаційні технології: економіка, техніка, освіта 2023», 26–27 жовтень 2023 року, НУБіП України, Київ.

- IX International Scientific And Practical Conference «Scientific Problems And Options For Their Solution», February 7–9 2024, Bucharest, Romania.

- Міжнародна наукова інтернет-конференція «Інформаційне суспільство: технологічні, економічні та технічні аспекти становлення», 18–19 січня 2024, Тернопіль.

- Міжнародна науково-практична конференція «Актуальні питання розвитку науки та техніки в умовах глобалізації», 14 травня 2024, Боярка.

- XII Міжнародна науково-практична конференція «Глобальні та регіональні проблеми інформатизації в суспільстві і природокористуванні», 2024, Київ.

- II Міжнародна науково-практична конференція «Актуальні питання розвитку науки та техніки в умовах глобалізації», 14 травня 2025, Боярка.

- Всеукраїнська науково-практична конференція студентів і аспірантів «Теоретичні та прикладні аспекти розробки комп'ютерних систем 2025», 24 квітня 2025 року, НУБіП України, Київ.

- A Hybrid Conference 10th ICICT 2025, London, 18-21 February 2025.

- 5th Edge Computing Workshop, April 2025.

Публікації. За тематикою дисертації опубліковано наукові праці у фахових виданнях і виданнях що індексуються базою Scopus.

- Holub B., Kyrychenko V., Nikolaienko D., Lendiel M., Shevchenko D., Khomenko A. Algorithmic and information support in atmospheric air quality monitoring systems. Proceedings of the Tenth International Congress on Information and Communication Technology (ICICT 2025). Lecture Notes in Networks and Systems. 2025. Vol. 1441. P. 399–412.

- Shevchenko D. V., Holub B. L. Regression analysis as a tool for identifying patterns in atmospheric air monitoring data. Proceedings of the 5th Edge Computing Workshop. 2025. Vol. 4. P. 20–27.

- Шевченко Д. В., Голуб Б. Л. Моніторинг якості повітря в реальному часі. Математичні машини і системи. 2025. № 1. С. 103–112.
- Шевченко Д. В., Голуб Б. Л. Багатовимірна аналітика екологічних даних: застосування OLAP у системах моніторингу. Математичні машини і системи. 2025. № 3–4. С. 54–65.
- Шевченко Д. В., Голуб Б. Л. Застосування методів Data Mining для багатовимірного аналізу якості атмосферного повітря на основі екологічних даних. Наука і техніка сьогодні. 2025. № 8 (49). С. 1801–1810.

Структура дисертації. Дисертаційна робота складається зі вступу, чотирьох розділів, висновків, списку використаних джерел та додатків. Загальний обсяг дисертації становить 198 сторінок. Робота містить 54 рисунків, 12 таблиць та 119 найменувань у списку літератури.

Перший розділ присвячено аналізу стану наукових досліджень і технологічних рішень у сфері моніторингу якості атмосферного повітря. Розглянуто основні методики визначення екологічних індексів (AQI, CAQI, EAQI), підходи до збору й інтерпретації даних, структуру міжнародних систем екологічного спостереження. Проведено порівняльний аналіз вітчизняних і зарубіжних інформаційних систем моніторингу, зокрема таких, як SaveEcoBot, OpenAQ, AirVisual та ін. На основі узагальнення виявлено недоліки існуючих рішень – фрагментарність джерел, відсутність інтегрованих аналітичних інструментів, обмеження у гнучкості аналізу. Це дозволило обґрунтувати необхідність створення інтелектуальної інформаційної технології, здатної об'єднувати відкриті та власні дані, забезпечуючи автоматизовану оцінку стану атмосферного повітря.

У **другому** розділі розроблено архітектуру інформаційної системи, яка поєднує модулі збору даних, сховище даних (Data Warehouse), аналітичні компоненти та інтерфейси користувача. Обґрунтовано вибір архітектурного підходу, що базується на принципах модульності, масштабованості та сумісності з потоковими джерелами даних. Запропоновано багаторівневу модель бази даних із нормалізованими таблицями, яка забезпечує ефективну інтеграцію інформації з

відкритих джерел (SaveEcoBot, КМДА) та власних MQTT-серверів. Розроблено концептуальну і логічну схеми даних, а також визначено ключові виміри та фактичні таблиці для реалізації OLAP-аналітики.

Третій розділ присвячено методам і технологіям збору та попередньої обробки даних про стан атмосферного повітря. Розглянуто підходи до уніфікації форматів даних, автоматизації процесів очищення, фільтрації та нормалізації показників. Деталізовано процес ETL (Extract, Transform, Load), який забезпечує перенесення інформації з різномірних джерел до централізованого сховища. Запропоновано алгоритми обчислення агрегатних значень для подальшої аналітики. Описано механізми верифікації даних та забезпечення їхньої достовірності, що є критично важливим для коректного аналізу екологічних процесів. Запропоновані рішення інтегровані в архітектуру інтелектуальної системи та забезпечують стабільне надходження якісних даних для аналітичних підсистем.

У **четвертому** розділі представлено аналітичну частину роботи, спрямовану на обробку та інтерпретацію екологічних даних. Реалізовано алгоритми розрахунку індексів AQI, CAQI та EAQI з урахуванням їхніх методологічних особливостей. Виконано аналіз отриманих результатів із точки зору точності, стабільності та продуктивності системи. Впроваджено та протестовано методи Data Mining (“K-Means”, “Наївний Байєс” та інші), що використовуються для класифікації стану повітря, виявлення закономірностей і прогнозування динаміки забруднень. Побудовано багатовимірні OLAP-звіти, які дозволяють виконувати аналіз даних у часовому і просторовому вимірах. Запропоновано систему КРІ-показників для моніторингу перевищень нормативних рівнів забруднення та ефективності заходів з покращення стану повітря.

РОЗДІЛ 1

ЕКОЛОГІЧНИЙ МОНІТОРИНГ ЯКОСТІ АТМОСФЕРНОГО ПОВІТРЯ ЯК ОБ'ЄКТ ДОСЛІДЖЕННЯ

1.1. Аналіз предметної області

Атмосферне повітря є невіддільним та життєво важливим компонентом екосистеми нашої планети. Воно виконує критичні функції: забезпечує киснем для дихання, підтримує тепловий і водний баланс, є середовищем перенесення речовин та перебігу хімічних реакцій. Хоча за обсягом основним компонентом повітря є азот (~78%), саме кисень (~21%) життєво необхідний для аеробного метаболізму клітин людини, тварин і багатьох мікроорганізмів.

Атмосферне повітря – це газова оболонка, що оточує Землю; воно складається з азоту (78%), кисню (21%), аргону (0,9%), вуглекислого газу (~0,04%) та слідів інших газів. Додатково повітря містить водяну пару зі змінною часткою (прибл. 0–4%) і аерозольні частинки, які впливають на погоду та якість повітря. Атмосфера відіграє ключову роль у колообігу речовин – води, вуглецю, азоту та інших елементів. Зокрема, вуглекислий газ є субстратом фотосинтезу, у результаті якого рослини утворюють органічні речовини та повертають у повітря кисень.

Забруднення атмосферного повітря слід розуміти як надходження в атмосферу сторонніх речовин або енергії, що змінюють її природний хімічний склад і фізичні властивості. Такі зміни можуть негативно впливати на здоров'я людини, тваринний і рослинний світ, а також на функціонування екосистем і навколишнє середовище в цілому. Сьогодні ця проблема набула глобального масштабу та розглядається міжнародною спільнотою як один із ключових чинників екологічної кризи.

Джерела забруднення мають антропогенний і природний характер. До антропогенних відносяться промислові викиди, транспортні вихлопи, спалювання вугілля та нафтопродуктів, інтенсивне сільське господарство із застосуванням добрив та пестицидів. Природними чинниками є вулканічна активність, лісові

пожежі, пилові бурі, біогенні процеси. Взаємодія цих факторів призводить до постійного накопичення забруднюючих домішок у тропосфері, особливо в урбанізованих зонах.

Вплив забрудненого повітря на людину є значним та багатовекторним. Доведено, що підвищені концентрації оксидів азоту, сірчаного газу, чадного газу та дрібнодисперсних частинок ($PM_{2.5}$, PM_{10}) зумовлюють розвиток респіраторних та серцево-судинних захворювань, послаблюють імунітет, підвищують ризик онкологічних патологій і скорочують середню тривалість життя. Найбільш вразливими є діти, люди похилого віку та особи з хронічними хворобами. Для тварин і рослин забруднення означає зниження життєздатності, пригнічення росту, порушення відтворення та зменшення біорізноманіття.

Наслідки забруднення відчутні і в масштабі навколишнього середовища: кислотні дощі пошкоджують ґрунти, водойми та лісові масиви, а підвищені концентрації парникових газів є рушійною силою глобального потепління й зміни клімату. Таким чином, якість атмосферного повітря безпосередньо впливає як на добробут нинішніх поколінь, так і на довгострокову екологічну стабільність планети.

Поняття якості атмосферного повітря визначається концентраціями шкідливих домішок у повітрі, яким дихає людина. Основними забруднювачами є оксиди азоту та сірки, вуглекислий газ, озон, вуглеводні та тверді частинки (пил, сажа, аерозолі). Їхня концентрація формується під впливом джерел викидів (промисловість, транспорт, енергетика, сільське господарство), метеорологічних умов (вітер, температура, вологість) та географічних факторів (урбанізованість, топографія, висота).

Важливим аспектом є економічні наслідки забруднення. Витрати охоплюють лікування хвороб, викликаних низькою якістю повітря, заходи зі зменшення викидів і реабілітації територій, а також втрати від зниження продуктивності праці та аграрного виробництва. Зниження якості повітря негативно впливає і на туристичну привабливість регіонів. Інвестиції в поліпшення якості атмосферного

повітря – розвиток технологій зменшення викидів, альтернативна енергетика, створення зелених зон – мають не лише екологічну, а й економічну доцільність.

Узагальнюючи, можна стверджувати, що стан атмосферного повітря є вирішальним для здоров'я та добробуту населення, функціонування екосистем і глобальної кліматичної стабільності. Це підкреслює необхідність створення сучасних систем екологічного моніторингу, здатних забезпечувати своєчасний аналіз, прогнозування та прийняття ефективних рішень щодо зменшення забруднення.

1.2. Стан якості повітря та його динаміка

Стан якості атмосферного повітря є важливим індикатором екологічної безпеки держави та рівня комфортності життя населення. В останні десятиліття проблема чистоти повітря набула особливої актуальності у зв'язку з глобальними процесами урбанізації, зростанням індустріального навантаження та змінами клімату. За оцінками Всесвітньої організації охорони здоров'я, понад 92% населення світу проживає в регіонах, де рівень забруднення повітря перевищує рекомендовані безпечні межі [1]. Це прямо впливає на здоров'я та середню тривалість життя людей, що підтверджує важливість проблеми у глобальному масштабі.

Для України ситуація є ще більш загрозливою. За даними міжнародних досліджень, країна займає перше місце серед 120 країн світу за рівнем смертності, пов'язаної із забрудненням повітря – у середньому 120 смертей на 100 000 населення. Такий показник відображає не лише екологічні проблеми, але й системні труднощі у сфері державного екологічного контролю та впровадженні ефективних заходів щодо зменшення викидів.

Найбільший рівень забруднення фіксується у великих містах та промислових центрах. Зокрема, у Києві близько 85% усіх викидів припадає на автомобільний транспорт, що є головним джерелом оксидів азоту, оксиду вуглецю, твердих частинок ($PM_{2.5}$, PM_{10}) та летких органічних сполук [2]. У столиці

регулярно фіксуються перевищення гранично допустимих концентрацій (ГДК) таких речовин, як NO_2 , формальдегід, фенол та дрібнодисперсний пил. Хронічний характер забруднення підтверджується ще у 2016 році, коли середньодобові концентрації діоксиду азоту та формальдегіду систематично перевищували норми.

До інших критичних регіонів належать індустриальні області – Донецька, Дніпропетровська, Запорізька, де високий рівень забруднення зумовлений роботою металургійних комбінатів, шахт, енергетичних підприємств. У південних регіонах ситуацію ускладнюють пилові бурі та посушливі умови, що сприяють поширенню твердих частинок у повітрі.

За даними Державної служби України з питань безпеки харчових продуктів та захисту споживачів, у 2020–2022 роках спостерігалися суттєві перевищення нормативних показників у Києві та Чернігові (таблиця 1.1). У столиці рівень $\text{PM}_{2.5}$ у 2022 році перевищував норму у 13 разів, PM_{10} – у 9 разів, озон (O_3) – у 2 рази, NO_2 – у 1,5 раза. У Чернігові відповідні показники становили 5, 3, 1,5 та 1,2 раза відповідно [3].

Таблиця 1.1

Рівні забруднення атмосферного повітря в Україні у 2020-2022 р.

Рік	Місто	$\text{PM}_{2.5}$	PM_{10}	O_3	NO_2
2020	Київ	11х	8х	1.8х	1.4х
2021	Київ	12х	8.5х	1.9х	1.45х
2022	Київ	13х	9х	2х	1.5х
2020	Чернігів	4.5х	2.5х	1.3х	1.1х
2021	Чернігів	4.8х	2.8х	1.4х	1.15х
2022	Чернігів	5х	3х	1.5х	1.2х

Примітка. *Побудовано за даними: Державна служба України з питань безпеки харчових продуктів та захисту споживачів. Дані моніторингу атмосферного повітря в Україні за 2020-2022 роки.

Як видно з таблиці, у всіх випадках спостерігається перевищення нормативних значень концентрацій забруднюючих речовин. У Києві рівень забруднення атмосферного повітря є вищим, ніж у Чернігові. При цьому у 2022 році рівень забруднення атмосферного повітря в обох містах погіршився порівняно з 2020 роком.

Так, у Києві у 2022 році концентрація пилу дрібнодисперсного ($PM_{2.5}$) перевищувала нормативне значення у 13 разів, що є дуже високим показником. Концентрація пилу дрібнодисперсного (PM_{10}) також була високою і перевищувала нормативне значення у 9 разів. Концентрація озон (O_3) перевищувала нормативне значення у 2 рази, а концентрація азот діоксиду (NO_2) – у 1,5 раза.

У Чернігові у 2022 році концентрація пилу дрібнодисперсного ($PM_{2.5}$) перевищувала нормативне значення у 5 разів, що також є високим показником. Концентрація пилу дрібнодисперсного (PM_{10}) перевищувала нормативне значення у 3 рази. Концентрація озон (O_3) перевищувала нормативне значення у 1,5 раза, а концентрація азот діоксиду (NO_2) - у 1,2 раза.

Ця тенденція є відображенням зростаючих викликів, з якими стикається Україна у контексті забруднення атмосферного повітря. У 2020 та 2021 роках також спостерігалася висока концентрація забруднюючих речовин, особливо в міських районах, що свідчить про сталість і серйозність проблеми. Ці дані підкреслюють необхідність розробки та впровадження ефективних заходів щодо зниження рівня забруднення атмосферного повітря, включаючи модернізацію промислових підприємств, підвищення екологічних стандартів для транспортних засобів та посилення контролю за викидами шкідливих речовин.

Ці дані свідчать про те, що стан якості атмосферного повітря в Україні є незадовільним і має тенденцію до погіршення. Це негативно впливає на здоров'я людини, тварин і рослин, а також на навколишнє середовище в цілому.

Динаміка змін якості атмосферного повітря в Україні зумовлена низкою взаємопов'язаних чинників. Вони мають як природний, так і антропогенний характер, але спільним є те, що всі вони підсилюють негативний тренд –

погіршення стану довкілля та зростання ризиків для здоров'я населення. Узагальнені результати впливів наведені у таблиці 1.2.

Таблиця 1.2

Основні фактори, що впливають на забруднення повітря в Україні

Фактор	Суть впливу	Наслідки
Динаміка змін	Стійке зростання концентрацій забруднювачів протягом останніх років	Кумулятивний ефект, довготривалі ризики для здоров'я та довкілля
Урбанізація	Зростання міст, транспортне навантаження, «теплові острови»	Утворення приземного озону, затримка домішок у повітрі
Воєнні чинники	Руйнування об'єктів, пожежі, використання генераторів	Викиди продуктів згоряння, додаткове забруднення у містах
Кліматичні зміни	Підвищення температур, хвилі спеки, пилові бурі	Зростання приземного озону, поширення аерозолів

Як видно з таблиці, ключовими чинниками, що визначають сучасну динаміку забруднення атмосферного повітря в Україні, є урбанізація, індустриальне навантаження, воєнні дії та кліматичні зміни. Вони не лише підвищують рівень концентрацій забруднюючих речовин, але й формують складний багатофакторний фон, у якому класичні методи контролю виявляються недостатніми. Це обґрунтовує потребу у впровадженні сучасних інформаційно-аналітичних систем, здатних забезпечувати точний моніторинг, прогнозування та своєчасне реагування на екологічні загрози.

Водночас ситуацію ускладнює те, що в Україні досі спостерігається відставання у впровадженні нормативів, гармонізованих з європейськими стандартами. Хоча у 2022 році було оновлено Закон України «Про охорону атмосферного повітря» та прийнято низку постанов, спрямованих на підвищення ефективності екологічного моніторингу, чинні нормативи щодо гранично допустимих концентрацій окремих забруднюючих речовин залишаються м'якшими, ніж рекомендації Всесвітньої організації охорони здоров'я та вимоги

Директиви 2008/50/ЄС [4, 5]. Це створює ризик, коли навіть офіційно «допустимі» рівні забруднення можуть становити небезпеку для здоров'я населення.

Таким чином, відставання у впровадженні європейських стандартів ускладнює формування сучасної політики у сфері охорони довкілля та підкреслює потребу у створенні нових інформаційно-аналітичних систем моніторингу, здатних не лише фіксувати перевищення національних ГДК, а й оцінювати якість повітря за міжнародними критеріями. Це дозволить забезпечити більш повну інтеграцію України у європейський екологічний простір і підвищити ефективність рішень у сфері захисту довкілля.

1.3. Індекс якості повітря: визначення та використання

Розрахунок індексу якості повітря базується на усереднених показниках концентрацій забруднювальних речовин за певний проміжок часу, які отримуються шляхом інструментальних вимірювань у мережі моніторингу або за допомогою моделей атмосферної дисперсії [6, 7]. Таким чином, поєднання концентрації речовини та тривалості її впливу формує дозу, що визначає можливий ефект для здоров'я людини. Оцінка цих ризиків ґрунтується на результатах багаторічних епідеміологічних досліджень.

Індекси AQI та CAQI мають нормалізований характер і забезпечують можливість порівняння рівнів забруднення у різних регіонах та країнах, навіть якщо методики вимірювання чи кількість станцій відрізняються. Такий підхід дозволяє інтерпретувати дані у стандартизованій формі та робить їх придатними для використання як у наукових дослідженнях, так і у сфері публічної політики.

Для зручності інтерпретації результати зводяться до інтегральної шкали – індексу якості повітря (AQI). Індекс поділяється на діапазони, кожен з яких має відповідний ідентифікатор, кольорове кодування та супровідні рекомендації для населення щодо рівня небезпеки. Завдяки цьому складні вимірювальні дані стають зрозумілими широкій громадськості й можуть використовуватися як інструмент прийняття рішень у повсякденному житті.

Значення AQI суттєво коливається залежно від інтенсивності викидів та метеорологічних умов. Наприклад, у години «пікового» транспортного навантаження спостерігається зростання концентрацій шкідливих речовин. Водночас погодні явища, такі як антициклони, температурні інверсії чи слабкий вітер, сприяють застою повітря й накопиченню забруднювачів у приземному шарі атмосфери. Це призводить до формування локальних зон із високим рівнем забруднення.

У випадках, коли прогнозується суттєве погіршення якості повітря, органи охорони здоров'я та громадські організації зазвичай вживають превентивних заходів. Вони можуть попереджати вразливі групи населення (літніх людей, дітей, осіб із респіраторними та серцево-судинними хворобами) про необхідність обмеження перебування на відкритому повітрі, оголошувати «дні дій» для скорочення викидів (наприклад, дні відмови від використання автомобілів), а також рекомендувати застосування захисних масок.

У більшості країн світу для розрахунку AQI використовуються усереднені дані з моніторингових станцій за різні часові інтервали (зазвичай 1, 8, 24 години). До індексу входять концентрації ключових речовин, що найсильніше впливають на здоров'я людини: дрібнодисперсних частинок ($PM_{2.5}$, PM_{10}), діоксиду азоту (NO_2), озону (O_3), діоксиду сірки (SO_2) та оксиду вуглецю (CO). На основі епідеміологічних досліджень кожному діапазону концентрацій присвоєно категорії небезпеки та кольорові маркери.

У Європейському Союзі для інформування населення використовується також Common Air Quality Index (CAQI), який охоплює подібні забрудники, але передбачає окремі шкали для погодинних і добових значень [8]. Це дозволяє більш оперативно відображати зміни якості повітря у містах та враховувати як короточасні, так і тривалі впливи.

Міжнародна практика демонструє, що використання індексів сприяє:

- підвищенню обізнаності населення про ризики забруднення повітря;
- запровадженню системи превентивних заходів (наприклад, «дні без автомобіля» чи попередження для вразливих груп населення);

- інтеграції екологічних даних у систему охорони здоров'я та міського управління;
- розвитку мобільних застосунків і онлайн-платформ для візуалізації показників у режимі реального часу.

Окрім інформування громадськості, індекси якості повітря відіграють важливу роль у формуванні екологічної політики та управлінні довкіллям. Вони застосовуються для оцінки ефективності заходів зі зменшення викидів, аналізу сценаріїв розвитку урбанізованих територій та інтеграції екологічних критеріїв у стратегії сталого розвитку міст і регіонів.

У сучасних умовах AQI та CAQI дедалі частіше інтегруються у цифрові сервіси – мобільні застосунки, онлайн-карти та інформаційні панелі, що дозволяє відстежувати якість повітря у режимі реального часу та робить ці дані доступними широкому колу користувачів. Водночас в Україні різні системи моніторингу (державні та громадські) нерідко застосовують відмінні методики розрахунку (наприклад, AQI за стандартами EPA та CAQI за європейськими підходами). Це ускладнює порівняння показників і підкреслює необхідність створення уніфікованої національної методики.

Адаптація системи AQI/CAQI до українських умов супроводжується низкою викликів, серед яких – різноманітність джерел даних (державні пости та громадські сенсори), відсутність єдиної національної методики розрахунку індексів та потреба гармонізації з директивами ЄС. У Європейському Союзі такі індекси застосовуються на уніфікованій основі, що забезпечує порівнянність результатів між країнами та дозволяє інтегрувати їх у загальноєвропейські екологічні інформаційні системи [8]. В Україні ж система моніторингу перебуває у процесі трансформації: поряд із державною мережею стаціонарних постів усе більшого значення набувають громадські ініціативи, зокрема мережі EcoCity, КМДА та SaveEcoBot [9, 10, 11]. Ці проєкти вже застосовують методологію AQI/CAQI, забезпечуючи відкритість даних та можливість їх інтеграції у міжнародні ресурси. Водночас ключовим завданням залишається розробка уніфікованої національної

методики оцінювання якості повітря, яка враховуватиме специфіку українського моніторингу та поступово наблизить його до європейських практик.

1.4. Основні забруднювачі та принципи їх вимірювання

Аналіз основних забруднювачів є ключовим елементом системи екологічного моніторингу. Вибір контрольованих параметрів базується на їхньому впливі на здоров'я населення, екосистеми та клімат. До найбільш поширених і небезпечних відносять діоксид сірки (SO_2), оксиди азоту (NO_2), оксид вуглецю (CO), озон (O_3) а також дрібнодисперсні тверді частинки ($\text{PM}_{2.5}$ та PM_{10}).

Діоксид сірки (SO_2) – безбарвний газ із різким запахом, який характеризується високою розчинністю у воді та здатністю утворювати кислотні аерозолі. Основними антропогенними джерелами SO_2 є спалювання викопного палива (вугілля, нафти, мазуту) на теплоелектростанціях та у промислових котельнях, а також металургійне виробництво. Природними джерелами виступає вулканічна активність. Потрапляння SO_2 в атмосферу має серйозні екологічні та медичні наслідки: він подразнює слизові оболонки, провокує загострення бронхіальної астми, бере участь у формуванні кислотних дощів, що призводять до деградації ґрунтів та пошкодження лісових екосистем. Для контролю концентрацій SO_2 застосовують низку методів. Стандартним у мережах автоматизованих станцій є метод ультрафіолетової флуоресценції, який забезпечує високу точність. Для мобільних і низьковартісних рішень використовуються електрохімічні сенсори, що дають змогу інтегрувати вимірювання в IoT-системи. Орієнтовні дослідження також можуть проводитися за допомогою колориметричних методів, однак вони характеризуються нижчою точністю..

Тверді частинки ($\text{PM}_{2.5}$, PM_{10}) – це суміш твердих і рідких аерозолів, зважених у повітрі. Їх класифікують за аеродинамічним діаметром: PM_{10} - частинки діаметром до 10 мкм, тоді як $\text{PM}_{2.5}$ мають розмір до 2,5 мкм і становлять найбільшу небезпеку для здоров'я людини. Завдяки своїй мікроскопічній величині

дрібнодисперсні частинки здатні проникати глибоко у легені, потрапляти у кровотік та викликати системні порушення.

Джерела твердих частинок є як антропогенними, так і природними. Серед основних антропогенних чинників виділяють спалювання палива у промисловості та на транспорті, роботу теплоелектростанцій, виробничі процеси, а також інтенсивне будівництво. До природних джерел відносять пилові бурі, лісові пожежі, вулканічні виверження та біогенні процеси.

Вплив $PM_{2.5}$ та PM_{10} на здоров'я населення надзвичайно серйозний. Багаторічні дослідження підтверджують їхній зв'язок із розвитком серцево-судинних і респіраторних захворювань, зростанням кількості випадків астми, а також підвищенням загальної смертності у містах з високим рівнем забруднення. Визнано, що саме $PM_{2.5}$ є одним із головних чинників зниження середньої тривалості життя у мегаполісах.

Основним референтним підходом є гравіметричний метод (зважування на фільтрах), що забезпечує високу точність, але є трудомістким і ресурсозатратним. Для автоматизованих систем контролю часто застосовується метод бета-абсорбції, який дозволяє безперервно визначати масу частинок. У розподілених і громадських мережах найбільш поширені оптичні методи (лазерна розсіювальна фотометрія), які забезпечують швидкі та відносно недорогі вимірювання.

Діоксид азоту (NO_2) – червонувато-бурий газ із різким запахом та сильними окисними властивостями, один із головних забруднювачів міського повітря. Основні джерела його утворення – автомобільні вихлопи, теплоелектростанції та промислові котельні; природними джерелами виступають грози й біогенні процеси у ґрунті. Діоксид азоту подразнює дихальні шляхи, спричиняє загострення астми, знижує імунітет, а також бере участь у фотохімічних реакціях, що ведуть до утворення приземного озону – ключового компонента смогу.

Для контролю концентрацій NO_2 застосовуються різні методи. Еталонним вважається хемілюмінесцентний метод, що ґрунтується на реакції NO з озоном і вимірюванні інтенсивності випромінювання. У мобільних та громадських системах поширені електрохімічні сенсори, які забезпечують оперативний

моніторинг, тоді як колориметричні методи використовуються переважно для орієнтовних вимірювань, але мають нижчу точність.

Оксид вуглецю (CO) – безбарвний і без запаху газ, відомий високою токсичністю. Основними джерелами є процеси неповного згоряння палива – транспортні вихлопи, робота теплоелектростанцій, побутових генераторів, а серед природних – лісові пожежі та вулканічна активність. CO має здатність зв'язуватися з гемоглобіном у крові, утворюючи карбоксигемоглобін, що блокує перенесення кисню й призводить до гіпоксії, небезпечної для життя навіть при короткочасному впливі.

Контроль концентрацій CO здійснюється переважно за допомогою інфрачервоної спектроскопії, яка дозволяє проводити неперервний моніторинг у стаціонарних системах. Для портативних приладів і громадських IoT-мереж все частіше застосовуються електрохімічні сенсори, що забезпечують доступний та оперативний контроль рівнів цього газу.

Озон (O₃) належить до вторинних забруднювачів, оскільки не викидається безпосередньо, а утворюється в результаті фотохімічних реакцій між оксидами азоту (NO₂) та леткими органічними сполуками (ЛОС) під дією сонячного випромінювання. Його концентрації найчастіше зростають у літній період у містах з інтенсивним транспортним рухом. У високих концентраціях озон є небезпечним: він погіршує функції дихальної системи, знижує імунну відповідь, шкодить рослинам, пригнічуючи фотосинтез, та спричиняє руйнування матеріалів, зокрема полімерів і гумових виробів.

Моніторинг рівнів озону зазвичай здійснюється за допомогою УФ-фотометрії – це «золотий стандарт», який використовується на більшості автоматизованих станцій завдяки високій точності та стабільності. Для мобільних і громадських мереж часто застосовуються електрохімічні сенсори. Хоча вони менш точні та чутливі, їхня доступність і простота експлуатації роблять їх важливим інструментом для оперативного контролю.

Інструментальні методи контролю та типи станцій. Системи моніторингу якості повітря базуються на поєднанні стаціонарних постів,

мобільних лабораторій та розподілених IoT-сенсорів. Такий багаторівневий підхід дозволяє поєднувати високу точність вимірювань із широким просторовим покриттям та оперативністю.

- Стационарні автоматизовані пости оснащуються референтними приладами (УФ-флуоресценція, хемілюмінесценція, гравіметричні аналізатори, УФ-фотометрія), що відповідають міжнародним стандартам (EU Directive 2008/50/EC, WHO). Вони забезпечують безперервний контроль концентрацій основних забрудників і паралельно фіксують метеорологічні параметри. Недоліки – висока вартість, складність обслуговування та обмежене просторове покриття.
- Мобільні лабораторії розміщуються на автомобілях чи контейнерних модулях і використовуються для оперативного моніторингу в проблемних районах, поблизу промислових об'єктів або під час надзвичайних ситуацій (пожежі, аварії). Вони забезпечують гнучкість і можливість швидкого реагування, проте поступаються стаціонарним станціям у довгостроковій стабільності.
- IoT-сенсори та громадські мережі – базуються на недорогих електрохімічних і оптичних датчиках (лазерна фотометрія для РМ, електрохімія для $\text{NO}_2/\text{CO}/\text{O}_3$). Прикладами є Arduino-платформи, спеціалізовані модулі (SDS011, MQ-серія, BME280) та інші IoT-рішення.

Таким чином, оптимальна система екологічного моніторингу поєднує всі три рівні: стаціонарні пости як джерело еталонних даних, мобільні станції для гнучких перевірок, а IoT-мережі – для формування детальної просторової картини розподілу забруднень у містах і сільських регіонах.

1.5. Аналіз існуючих рішень

Розвиток систем екологічного моніторингу є глобальною тенденцією, яка відображає зростаючу увагу до проблеми забруднення атмосферного повітря. У різних країнах світу сформувалися власні підходи до організації таких систем,

проте всі вони мають спільні цілі – забезпечення доступу до оперативної та достовірної інформації, виявлення критичних зон та формування основи для прийняття управлінських рішень.

Залежно від рівня технічного розвитку та інституційної підтримки, системи моніторингу відрізняються масштабом, методологією та форматом представлення даних. Міжнародні приклади демонструють високий рівень інтеграції з державними структурами, використання прогностичних моделей та відкритих даних, тоді як українські рішення часто розвиваються у форматі громадських ініціатив і характеризуються гнучкістю, але й певними обмеженнями щодо точності та стандартизації.

Для комплексного аналізу доцільно розглянути як міжнародний досвід, так і українські практики, аби визначити сильні сторони та слабкі місця, які можуть бути враховані при створенні або модернізації національної системи моніторингу.

1.5.1. Міжнародні системи моніторингу якості повітря. AirNow Fire Map. Сервіс є частиною національної екологічної ініціативи США AirNow [12], яка функціонує під егідою Агентства з охорони довкілля США (EPA) у співпраці з Національною службою океанічних і атмосферних досліджень (NOAA), Національною лісовою службою та іншими організаціями. Цей сервіс спеціалізується на відображенні якості повітря під час пожеж, поєднуючи дані офіційних станцій моніторингу з інформацією від низьковартісних сенсорів PurpleAir (рис. 1.1). Візуалізація здійснюється на інтерактивній карті [13], де відображаються активні пожежі, димові шлейфи та рівні індексу якості повітря (AQI). Користувач має змогу переглядати як поточні показники, так і часові тренди, що дозволяє оцінювати динаміку поширення диму та зміни концентрацій забруднювачів упродовж певного періоду.

Система забезпечує високу оперативність та доступність даних для населення, поєднуючи офіційні вимірювання з даними громадських сенсорів. Її ключова перевага полягає в орієнтації на кризові ситуації: під час масштабних пожеж AirNow Fire Map дозволяє швидко оцінити рівень небезпеки та потенційний вплив на здоров'я. Крім того, інтерфейс зручний у користуванні, а

інтеграція з мобільними застосунками AirNow робить систему доступною широкій аудиторії.

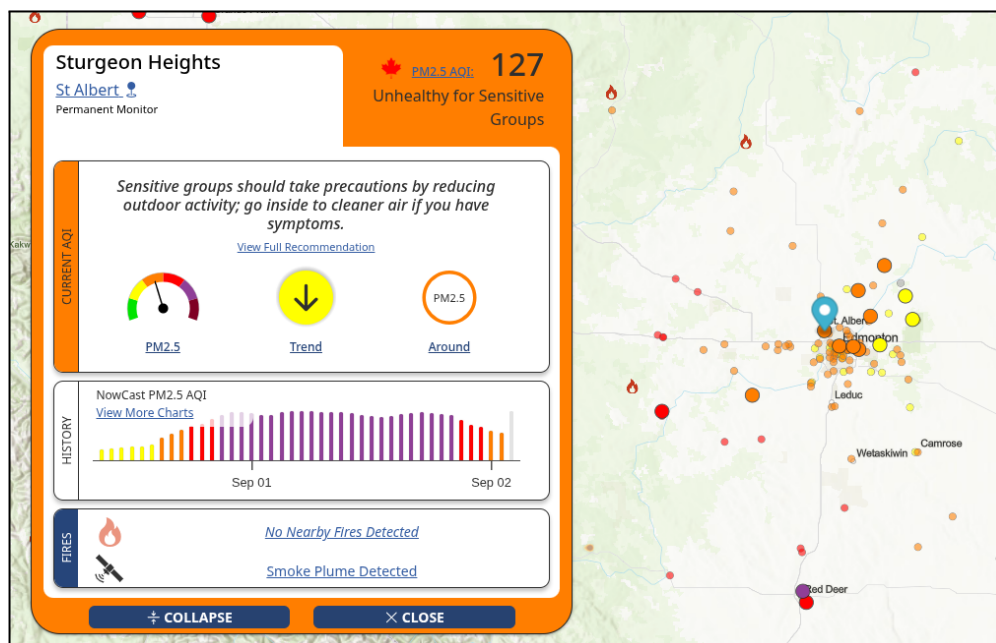


Рис. 1.1. Карта якості повітря AirNow

Основний акцент зроблений саме на відстеженні пожеж і диму, що звужує функціонал порівняно з комплексними системами моніторингу довкілля. Водночас різниця в точності між референтними станціями та низьковартісними сенсорами PurpleAir потребує додаткового калібрування, що може впливати на достовірність локальних даних.

OpenAQ. Глобальна платформа з відкритим доступом до даних про якість повітря, яка об'єднує інформацію з державних моніторингових мереж, громадських сенсорів і дослідницьких проєктів у більш ніж 70 країнах світу. Сервіс функціонує як агрегатор, надаючи уніфіковану базу даних, яка доступна для перегляду через інтерактивний вебінтерфейс або для завантаження за допомогою API [14]. Місія платформи – підвищення прозорості у сфері екологічного моніторингу та сприяння глобальним науковим і громадським ініціативам.

Система агрегує дані з тисяч станцій та сенсорів по всьому світу, забезпечуючи їхню стандартизацію у єдиному форматі. Це дозволяє інтегрувати

інформацію з різних джерел і робить дані більш придатними для аналізу. OpenAQ підтримує доступ через API, що дає можливість дослідникам і розробникам будувати власні інструменти візуалізації та аналітики. В основі роботи платформи лежить open-data підхід, орієнтований на наукову спільноту, органи влади та громадські організації.

Головною перевагою OpenAQ є відкритий доступ до даних, що забезпечує прозорість і безкоштовність використання. Платформа дозволяє працювати з великими масивами як історичних, так і поточних даних, що важливо для досліджень динаміки забруднення (рис. 1.2). Створення єдиної бази для міжнародних порівнянь робить сервіс цінним інструментом у глобальних екологічних дослідженнях. Крім того, дані активно застосовуються в освітніх і громадських проєктах, сприяючи підвищенню екологічної обізнаності.

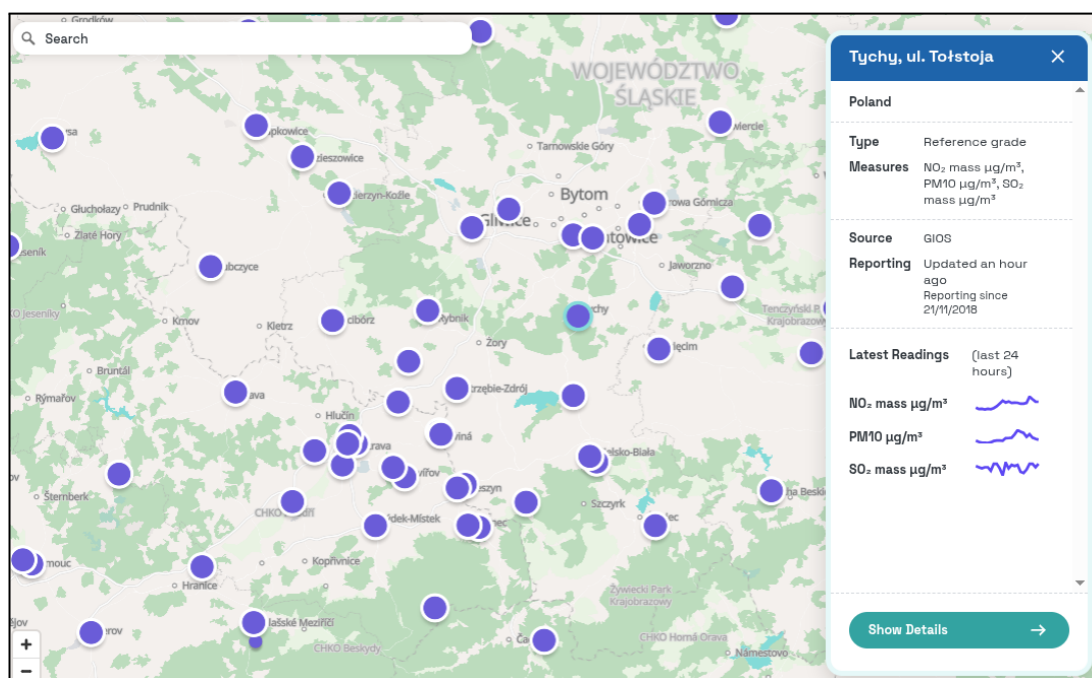


Рис. 1.2. Карта якості повітря OpenAQ

Слід зазначити, що OpenAQ не має власних сенсорів і виконує роль лише агрегатора. Якість і точність даних повністю залежать від першоджерел, якими можуть бути як державні служби, так і локальні ініціативи з різним рівнем надійності. У країнах із нерозвиненою інфраструктурою обсяг даних може бути

обмеженим або нерепрезентативним. Крім того, вебінтерфейс платформи є досить простим і поступається за гнучкістю візуалізації національним або спеціалізованим сервісам.

AQICN(World Air Quality Index Project) – це незалежна глобальна ініціатива, яка забезпечує вільний доступ до даних про якість повітря в реальному часі більш ніж у 100 країнах світу. Проєкт функціонує як агрегатор, поєднуючи державні офіційні спостереження, локальні сенсори та супутникові дані. Крім моніторингу, система пропонує прогнози якості повітря на 8 днів, що робить її унікальною серед відкритих платформ [15].

Платформа використовує індекс AQI відповідно до методології Агентства з охорони довкілля США (EPA), застосовуючи алгоритм NowCast, який дозволяє відображати оперативні зміни концентрацій забруднювачів у короткі часові інтервали. Система інтегрує дані з різних джерел, включно з офіційними національними мережами моніторингу, громадськими станціями та супутниковими спостереженнями. Візуалізація здійснюється у вигляді інтерактивної карти світу, де в реальному часі відображаються показники AQI для конкретних міст і регіонів (рис. 1.3).

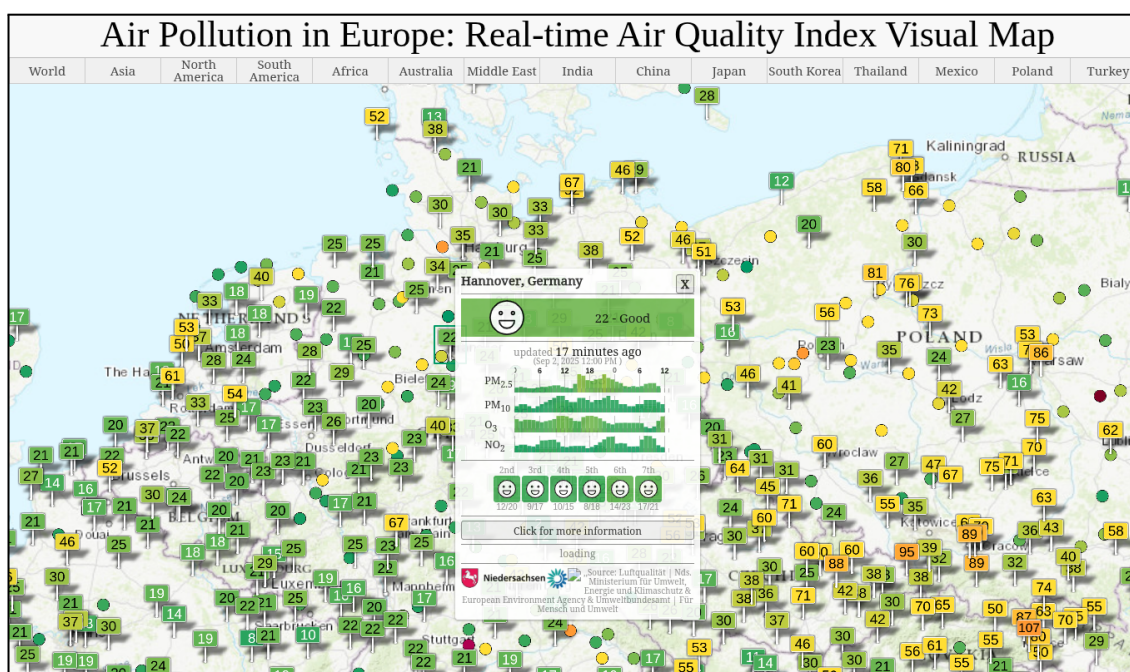


Рис. 1.3. Карта якості повітря AQICN

Серед основних переваг AQICN варто відзначити глобальне охоплення та єдину стандартизовану шкалу для різних країн, що забезпечує просте порівняння даних. Додатковою цінністю є 8-денні прогнози, що базуються на моделях атмосферної дисперсії та метеорологічних даних (рис. 1.4). Інтерфейс платформи є зручним і багатомовним, що робить сервіс доступним широкій аудиторії. Крім того, AQICN підтримує мобільні застосунки та API, завдяки чому дані активно використовуються дослідниками, журналістами та громадськими організаціями.

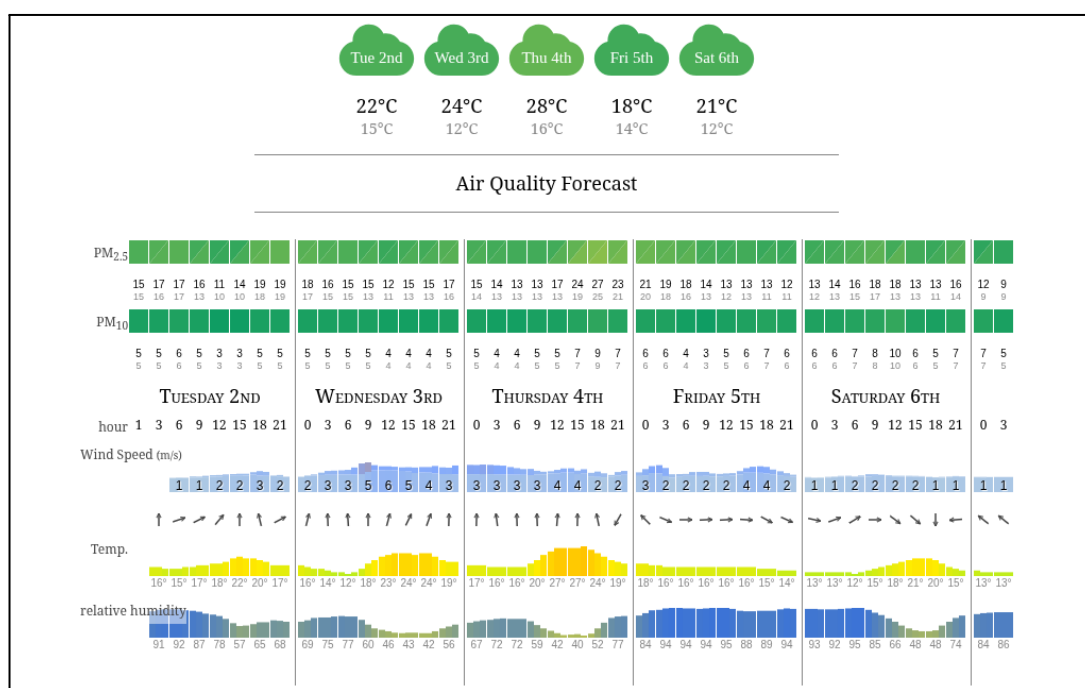


Рис. 1.4. Прогнозування якості повітря AQICN

Попри значні можливості, якість даних залежить від наявності та щільності локальних станцій у конкретному регіоні. У країнах із недостатнім розвитком моніторингової інфраструктури дані можуть бути неповними або мати затримки. Використання алгоритму NowCast робить інформацію дуже оперативною, але водночас може призводити до значних коливань значень, що потребує обережної інтерпретації. Також проєкт не завжди забезпечує повну прозорість щодо джерел калібрування громадських сенсорів, що іноді викликає зауваження наукової спільноти.

1.5.2. Українські системи моніторингу якості повітря. Після розгляду міжнародних практик варто зосередитись на національних ініціативах. В Україні система моніторингу атмосферного повітря перебуває у стадії активного розвитку як на державному, так і на громадському рівнях. Останні роки відзначаються появою нових автоматичних станцій і сенсорних мереж, що дозволяють отримувати дані про екологічний стан у режимі реального часу. Типові станції обладнані датчиками для вимірювання температури, вологості, атмосферного тиску, концентрацій вуглекислого газу та дрібнодисперсних твердих частинок ($PM_{2.5}$ та PM_{10}).

Особливістю сучасних українських систем є орієнтація на доступність даних для населення. Переважна більшість платформ надає результати вимірювань через вебінтерфейс, інтерактивні карти та мобільні застосунки. Деякі сервіси впроваджують додаткові канали інформування: інформаційні банери в публічних місцях, а також чат-боти для швидкого отримання повідомлень у месенджерах.

ASM Kyiv. Система моніторингу якості повітря в Києві (ASM Kyiv) створена у 2019 році Департаментом захисту довкілля та адаптації до зміни клімату КМДА [16]. Її основна мета – надання мешканцям столиці достовірної інформації про стан атмосферного повітря та забезпечення можливості своєчасного реагування на екологічні ризики.

Мережа ASM Kyiv включає понад 50 автоматизованих станцій, розташованих у різних районах міста. Датчики вимірюють концентрації ключових забруднювачів ($PM_{2.5}$, PM_{10} , NO_2 , SO_2 , CO , O_3) та передають дані на центральний сервер. Оброблена інформація публікується на офіційному вебсайті (рис. 1.5) у зрозумілому для населення форматі – із відображенням індексу якості повітря (CAQI) та поточних концентрацій речовин.

До ключових переваг ASM Kyiv належить офіційний статус системи та її інтеграція в міську інфраструктуру. Це забезпечує високий рівень довіри до отриманих даних і можливість їх використання у процесах міського управління. Важливою особливістю є регулярне оновлення інформації у режимі, близькому до

реального часу, що робить систему актуальною для мешканців столиці. Крім того, ASM Kyiv пропонує простий і доступний вебінтерфейс із візуалізацією показників якості повітря, зрозумілою для широкої аудиторії.

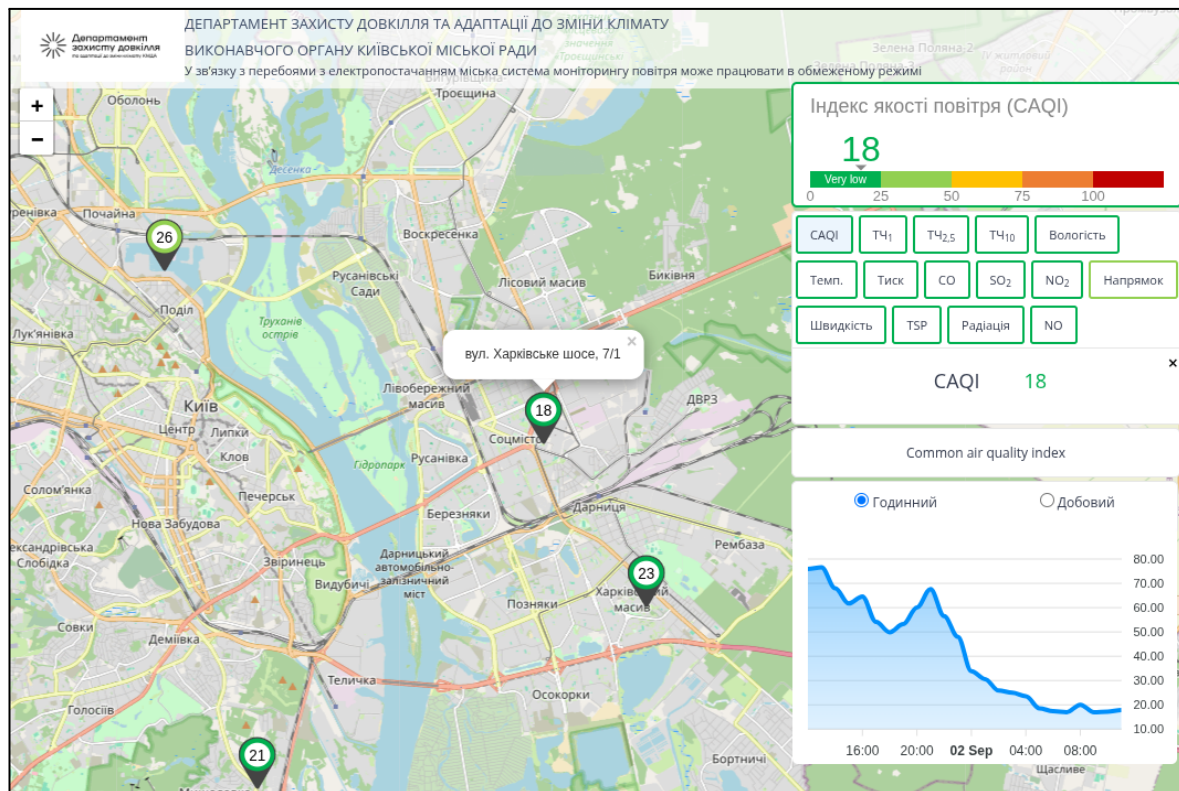


Рис. 1.5. Карта якості повітря ASM Kyiv

Попри значні переваги, система має низку обмежень. Насамперед мережа датчиків поки що не охоплює повністю територію Києва, що знижує репрезентативність даних у певних районах. Точність вимірювань може зменшуватися під впливом несприятливих погодних умов, зокрема під час опадів або сильного вітру. Крім того, ASM Kyiv більше орієнтована на інформування населення, ніж на проведення глибокої аналітики, що обмежує її застосування у наукових дослідженнях.

SaveEcoBot. Громадська ініціатива, започаткована у 2018 році групою активістів, які прагнули підвищити прозорість та доступність даних про якість повітря в Україні [9]. Система поєднує власну мережу датчиків із даними державних, муніципальних і міжнародних платформ (зокрема Airly, PurpleAir,

AirMQ), формуючи єдиний відкритий ресурс. Основна мета – забезпечення громадського контролю за станом атмосфери та підвищення обізнаності населення.

Мережа SaveEcoBot охоплює десятки міст України та складається з компактних сенсорів, здатних вимірювати концентрації низки забрудників, серед яких: NO, NO₂, CO, SO₂, O₃, PM_{2.5}, PM₁₀ та інші. Дані автоматично передаються на сервер, де проходять попередню обробку й публікуються у відкритому доступі. Користувачі можуть переглядати показники через вебпортал (рис. 1.6), інтерактивну карту, інформаційні банери та чат-боти у месенджерах, що робить сервіс максимально зручним.

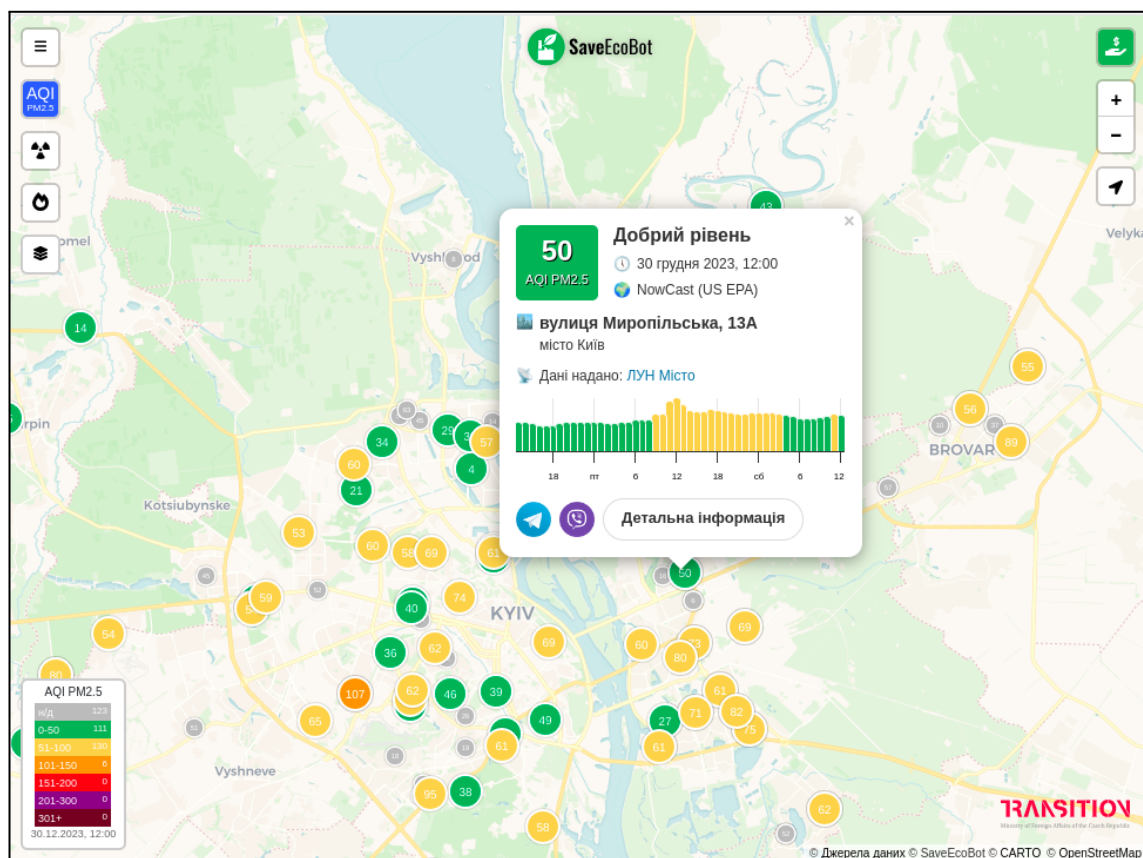


Рис. 1.6. Карта якості повітря SaveEcoBot

Головною перевагою системи є її громадський характер і відкритість: усі дані доступні в режимі реального часу й можуть використовуватися як населенням, так і науковцями чи органами влади. SaveEcoBot інтегрує інформацію

з різних платформ, що забезпечує більш комплексне охоплення території. Важливою рисою є можливість швидкого розширення мережі за рахунок додавання нових автономних датчиків у різних містах та громадах.

Разом із тим система має низку обмежень. Зокрема, дані від низьковартісних сенсорів поступаються за точністю вимірювання показникам стаціонарних референтних станцій, а калібрування залишається важливим завданням. Крім того, кількість власних датчиків поки що недостатня для повного охоплення всіх регіонів країни. Це знижує репрезентативність даних у віддалених чи малонаселених районах.

EcoCity. Одна з найбільших в Україні громадських мереж моніторингу атмосферного повітря, створена у 2018 році групою ентузіастів [10]. Система ґрунтується на волонтерській участі та відкритості даних: будь-який користувач може встановити датчик і під'єднати його до загальної мережі. Наразі EcoCity охоплює сотні станцій у різних регіонах країни та забезпечує вільний доступ до даних через вебсайт і відкритий API. Завдяки простому інтерфейсу користувачі можуть оцінювати стан повітря у своєму місті або навіть конкретному районі (рис. 1.7).

У документі *«Radiation and Smog Alarm. Настанови та принципи оповіщення населення про якість повітря, радіаційну та хімічну небезпеку»* [17] представлено підхід до формування Українського індексу якості повітря (UAQI), який використовується EcoCity для автоматизованого інформування населення про рівень забруднення повітря, а також про можливі радіаційні чи хімічні ризики. Наразі метод UAQI має статус тимчасової технічної настанови, розробленої ГО «Фрі Ардуіно» у рамках проєкту EcoCity, і ще не затверджений на державному рівні. Його розрахунки базуються на даних громадських сенсорних мереж (IoT), що вимагають регулярного калібрування та перевірки точності. Для підвищення надійності й подальшої інтеграції в державні системи моніторингу важливим завданням є верифікація алгоритмів UAQI через зіставлення їхніх результатів із даними офіційних референтних станцій.

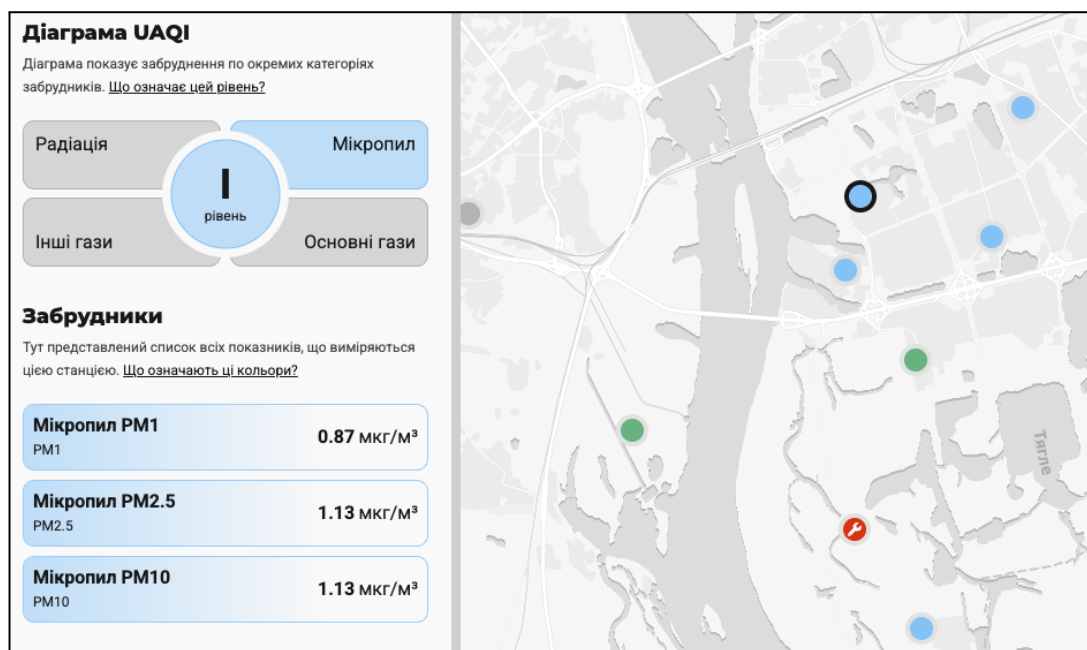


Рис. 1.7. Карта якості повітря EcoCity

Ключовою перевагою системи є громадський характер ініціативи – вона незалежна від державних структур і орієнтована на відкритість. Мережа постійно розширюється, що підвищує просторове покриття та деталізацію даних. Додатковою сильною стороною є відкритий API, який робить EcoCity цінним ресурсом для науковців, аналітиків та розробників екологічних застосунків. Система також інтегрує оповіщення через вебресурс та чат-боти, роблячи дані максимально доступними для широкої аудиторії.

У великих містах мережа ще не охоплює всі райони, що знижує репрезентативність результатів. Крім того, основний акцент зроблено на базових показниках, тоді як глибша аналітика потребує додаткових досліджень та порівняння з даними офіційних референтних станцій. EcoCity є прикладом ефективної волонтерської ініціативи, що зробила екологічні дані доступними для широкої аудиторії та запровадила інноваційний підхід у вигляді UAQI.

ЛУН Air. Відносно нова система моніторингу якості повітря в Україні, запущена у 2022 році громадською організацією ЛУН Місто в умовах війни [18]. Її головна мета – оперативно надавати населенню зрозумілу й доступну інформацію про стан повітря для захисту здоров'я та підвищення екологічної обізнаності.

Мережа налічує понад 100 сенсорів, розміщених у різних регіонах України. Вони вимірюють базові параметри забруднення, насамперед концентрації твердих частинок ($PM_{2.5}$, PM_{10}), діоксиду азоту та інших речовин, необхідних для розрахунку індексу якості повітря (AQI).

Відображення даних у режимі реального часу робить інформацію максимально актуальною та придатною для повсякденного використання (рис. 1.8). Важливим аспектом є відкритий доступ до даних, що дозволяє залучати їх для медіа – та освітніх проєктів, а також для підвищення громадської обізнаності.

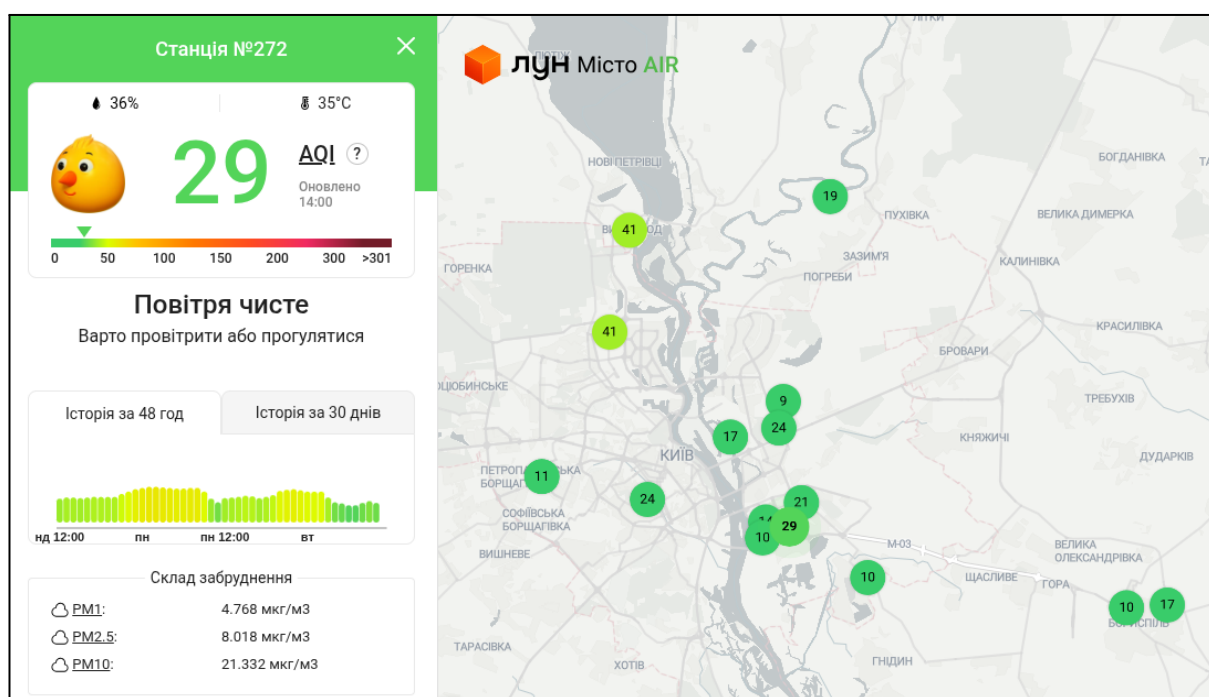


Рис. 1.8. Карта якості повітря ЛУН Air

Водночас система надає переважно базову інформацію про якість повітря, що ускладнює проведення глибших досліджень або ідентифікацію конкретних джерел забруднення. По-друге, кількість сенсорів ще недостатня для повного покриття всієї території країни, що впливає на репрезентативність. По-третє, у методології досі використовуються спрощені стандарти оцінювання, які не завжди враховують останні наукові дані щодо впливу забруднення на здоров'я.

Українські системи моніторингу атмосферного повітря демонструють помітний прогрес у розвитку – мережа сенсорів розширюється, а дані стають

доступними у режимі реального часу. Характерними рисами є відкритість інформації, залучення громадськості та використання сучасних технологій. Це сприяє підвищенню екологічної обізнаності населення та формує основу для прийняття управлінських рішень на локальному рівні.

Водночас у порівнянні з міжнародними платформами, українські ініціативи поступаються у сфері стандартизації методик, точності вимірювань і наявності прогнозних моделей. Переважна більшість систем орієнтована на швидке відображення поточних показників, але не забезпечує достатньої аналітичної глибини для комплексних досліджень чи моделювання. Використання спрощених або застарілих стандартів іноді обмежує можливість інтеграції з європейськими та глобальними екологічними інфраструктурами.

Отже, попри позитивну динаміку, ця сфера в Україні потребує подальшого вдосконалення: впровадження уніфікованих методик, підвищення точності сенсорів, розбудови прогнозних моделей і більш тісної інтеграції з міжнародними системами. Це дозволить підняти моніторинг повітря на якісно новий рівень і зробити його дієвим інструментом як для науки, так і для практичного управління довкіллям.

1.6. Стандарти та підходи до оцінювання якості атмосферного повітря

Оцінювання якості повітря неможливе без чіткої системи стандартів і нормативів, що визначають допустимі рівні забруднення для різних речовин. Ці стандарти є базою для державного контролю, міжнародних порівнянь та практичного інформування населення про потенційні ризики. На їх основі будуються інтегральні індекси, які дозволяють узагальнювати дані моніторингу та зробити їх зрозумілими для широкої аудиторії.

1.6.1. Нормативи концентрацій і дозволени перевищення. Для оцінки рівня забруднення атмосферного повітря у світі застосовують гранично допустимі концентрації (ГДК) та інші порогові показники, які визначають допустимі середньодобові, середньомісячні та максимально-разові концентрації. Такі

нормативи встановлюються для того, щоб оцінювати як короткотривалі пікові впливи, так і довготривале надходження забруднювачів до організму людини.

Згідно з міжнародною практикою (рекомендації ВООЗ 2021 року, Директива 2008/50/ЄС) та українським регламентом (Постанова КМУ №827), вимірювання проводяться з урахуванням верхніх і нижніх порогів оцінювання [19]. Це дозволяє визначати не лише перевищення граничних величин, але й проміжні діапазони концентрацій, що вимагають підвищеної уваги.

ВООЗ у 2021 році оновила свої рекомендації, суттєво знизивши допустимі рівні для $PM_{2.5}$, NO_2 та озону. Це відображає сучасні наукові дані про шкоду здоров'ю навіть при низьких концентраціях.

ЄС (директива 2008/50/ЄС) застосовує підхід «верхнього» та «нижнього» порогу оцінювання, що визначає, чи потрібно посилювати моніторинг у конкретному регіоні.

У США діють стандарти NAAQS (National Ambient Air Quality Standards), які встановлює EPA. Вони передбачають як короткострокові (1-годинні, 8-годинні) ліміти, так і довгострокові (річні середні), та використовуються для розрахунку AQI.

Ці пороги визначають межі верхнього та нижнього рівнів оцінювання концентрацій основних забруднювальних речовин у повітрі, які слугують базою для розрахунку індексів якості повітря (AQI, CAQI, EAQI) та подальшого інтелектуального аналізу даних.

Водночас українська державна система моніторингу поки що не охоплює повний спектр показників, визначених міжнародними стандартами. Це стосується насамперед регулярного вимірювання $PM_{2.5}$ та озону. Така ситуація ускладнює пряме зіставлення результатів із міжнародними базами даних і підкреслює потребу модернізації технічної бази автоматизованих станцій та методів оцінки.

Зокрема, у таблиці 1.3 наведено порогові значення для твердих часток PM_{10} та $PM_{2.5}$, що характеризують запиленість повітря. Ці частинки здатні тривалий час утримуватись у повітрі, проникати в дихальні шляхи людини та негативно впливати на стан здоров'я населення.

Таблиця 1.3

Пороги оцінювання для твердих часток (PM₁₀/PM_{2.5})

Пороги оцінювання	Середнє значення PM ₁₀ за 24 години	Середнє значення PM ₁₀ на рік	Середнє значення PM _{2.5} на рік
Верхній поріг оцінювання	70 % ГВ (35 мкг/м ³) не має бути перевищено понад 35 разів у будь-який календарний рік)	70 % ГВ (28 мкг/м ³)	70 % ГВ ГВ (17 мкг/м ³)
Нижній поріг оцінювання	50 % ГВ (25 мкг/м ³) не має бути перевищено понад 35 разів у будь-який календарний рік	50 % ГВ (20 мкг/м ³)	50 % ГВ (12 мкг/м ³)

Примітка. *Побудовано за даними: Постанова КМУ №827

У таблиці 1.4 подано нормативи для діоксиду сірки (SO₂) та діоксиду азоту (NO₂) – ключових газоподібних забрудників, небезпечних для здоров'я людини та рослинності. Обидві речовини є показниками інтенсивності антропогенного навантаження, насамперед від викидів транспорту та промислових підприємств. Надмірні концентрації SO₂ та NO₂ спричиняють подразнення дихальних шляхів, посилюють кислотність опадів і негативно впливають на рослинність, тому контроль їх рівнів має критичне значення для екологічної безпеки міських територій.

Таблиця 1.4

Пороги оцінювання для SO₂ та NO₂

Пороги оцінювання	Охорона здоров'я			Захист рослинності	
	SO ₂	NO ₂		SO ₂	NO ₂
		Щодинна ГВ	Щорічна ГВ		
Верхній поріг	60% 24-годинної ГДК (75мкг/м ³) не має бути перевищено більш як три рази у рік)	50% ГВ (140 мкг/м ³) не має бути перевищено більше 18 разів у рік)	80% ГВ (32 мкг/м ³)	60 % зимового критичного рівня (12 мкг/м ³)	80 % критичного рівня (24 мкг/м ³)

Нижній поріг	40 % 24-годинної ГДК (50мкг/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 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)
--------------	--	--	---	--	--

Примітка. *Побудовано за даними: Постанова КМУ №827

У таблиці 1.5 узагальнено пороги оцінювання для чадного газу (СО) та озону (О₃), визначені з урахуванням цільових показників захисту здоров'я населення та екосистем. Чадний газ є одним із найбільш поширених продуктів неповного згоряння палива та характеризує рівень забруднення повітря транспортними викидами. Озон, навпаки, є вторинним забрудником фотохімічного походження, концентрація якого зростає під впливом сонячного випромінювання та наявності попередників.

Обидва показники є чутливими маркерами екологічного стану атмосфери, тому їх порогові значення враховуються при комплексній оцінці якості повітря та визначенні рівнів екологічного ризику.

Таблиця 1.5

Пороги оцінювання для СО та О₃

Пороги оцінювання	СО	Цільовий показник	О ₃
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
Нижній поріг	50 % ГВ (5 мкг/м ³)	Охорона рослинності	АОТ 40 (розраховується із значень за одну годину) 80 мкг/м ³ за травень-липень із використанням значень, вимірюваних 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

Примітка. *Побудовано за даними: Постанова КМУ №827

Таким чином, система нормативів є ключовим інструментом для оцінювання стану атмосферного повітря, оскільки задає чіткі межі допустимого впливу забруднювачів на здоров'я населення та довкілля. Вона дозволяє не лише фіксувати факти перевищення встановлених значень, але й визначати рівень екологічних ризиків для окремих територій. На основі таких критеріїв здійснюється державний контроль, плануються заходи із запобігання надмірному забрудненню, а також формується база для довгострокових стратегій у сфері охорони довкілля та громадського здоров'я.

1.6.2. Індекси якості повітря. Індекси якості повітря (Air Quality Index, AQI) належать до показників, що дозволяють уніфіковано оцінювати стан атмосферного повітря. Вони ґрунтуються на концентраціях ключових забруднювачів, але відрізняються між собою як за методикою розрахунку, так і за пороговими значеннями. Найбільш поширеними є три системи: AQI (США), CAQI (ЄС) та EAQI (оновлений європейський стандарт).

AQI. Розроблений Агентством з охорони довкілля США (EPA), AQI застосовується для інформування населення у форматі «nowcast» [20]. Він враховує концентрації $PM_{2.5}$, PM_{10} , NO_2 , SO_2 , O_3 та CO . Для кожного забруднювача обчислюється субіндекс, після чого формується загальне значення AQI. Діапазони та порогові точки показані в таблиці 1.6.

Таблиця 1.6

Діапазони точок зупину AQI

CO	NO ₂	O ₃	O ₃	PM ₁₀	PM _{2.5}	SO ₂	SO ₂	AQI	
C low - C high (avg)								I low - I high	Category
0-4.4 (8-hr)	0-53 1-hr	-	0-54 8-hr	0-54 24-hr	0-9 24-hr	0 - 35 (1-hr)	-	0 - 50	Excellent
4.5-9. 4 8-hr	54 - 100 (1-hr)	-	55 - 70 (8-hr)	55 - 154 (24-hr)	9.1 - 35.4 (24-hr)	36 - 75 (1-hr)	-	51 - 100	Fine
9.5 - 12.4 (8-hr)	101 - 360 (1-hr)	125 - 164 (1-hr)	71 - 85 (8-hr)	155 - 254 (24-hr)	35.5 - 55.4 (24-hr)	76 - 185 (1-hr)	-	101 - 150	Moderate

CO	NO ₂	O ₃	O ₃	PM ₁₀	PM _{2.5}	SO ₂	SO ₂	AQI	
12.5 - 15.4 (8-hr)	361 - 649 (1-hr)	-	86 - 105 (8-hr)	255 - 354 (24-hr)	55.5 - 125.4 (24-hr)	186 - 304 (1-hr)	-	151 - 200	Poor
15.5 - 30.4 (8-hr)	650 - 1249 (1-hr)	205 - 404 (1-hr)	106 - 200 (8-hr)	355 - 424 (24-hr)	125.5 - 225.4 (24-hr)	-	305 - 604 (24-hr)	201 - 300	Very Poor
30.5 - 50.4 (8-hr)	1250 - 2049 (1-hr)	405 - 604 (1-hr)	-	425 - 604 (24-hr)	225.5 - 500.4 (24-hr)	-	605 - 1004 (24-hr)	301 - 500	Severe

CAQI. Застосовується в європейських країнах для оцінки повітря у містах [21]. Він орієнтований на NO₂, O₃, SO₂, PM_{2.5} та PM₁₀, використовуючи погодинні та добові концентрації. Система дозволяє швидко оцінювати як короточасні, так і середньотривалі впливи забруднення, забезпечуючи можливість порівняння результатів між різними регіонами та часовими періодами.

Діапазони та порогові точки для класифікації рівнів якості повітря наведено в таблиці 1.7.

Таблиця 1.7

Діапазони точок зупину CAQI

NO ₂	O ₃	PM ₁₀	PM _{2.5}	SO ₂	CAQI	
C low - C high (avg)					I low - I high	Category
0 - 100 (1-hr)	0 - 60 (1-hr)	0 - 50 (1-hr)	0 - 25 (1-hr)	0 - 100 (1-hr)	0 - 25	Excellent
101 - 200 (1-hr)	61 - 120 (1-hr)	51 - 100 (1-hr)	26 - 50 (1-hr)	101 - 200 (1-hr)	26 - 50	Fine
201 - 400 (1-hr)	121 - 180 (1-hr)	101 - 200 (1-hr)	51 - 100 (1-hr)	201 - 350 (1-hr)	51 - 75	Moderate
401 - 600 (1-hr)	181 - 240 (1-hr)	201 - 400 (1-hr)	101 - 200 (1-hr)	351 - 500 (1-hr)	76 - 100	Poor
600 - 2049 (1-hr)	241 - 604 (1-hr)	401 - 604 (1-hr)	201 - 500.4 (1-hr)	501 - 1004 (1-hr)	101 - 500	Very Poor

EAQI. Сучасніший підхід, представлений Європейським агентством з довкілля (ЕЕА) у 2017 році [22]. Його мета – гармонізація оцінки якості повітря в усіх країнах ЄС. Використовуються концентрації PM_{2.5}, PM₁₀, NO₂, SO₂ та O₃, але

акцент зроблено на довгострокових середніх значеннях. Діапазони та порогові точки показані в таблиці 1.8.

Таблиця 1.8

Діапазони точок зупину EAQI

NO ₂	O ₃	PM ₁₀	PM _{2.5}	SO ₂	EAQI	
C low - C high (avg)					I low - I high	Category
0.0 - 40.0 (1-hr)	0.0 - 50.0 (1-hr)	0.0 - 20.0 (24-hr)	0.0 - 10.0 (24-hr)	0.0 - 100.0 (1-hr)	0.0 - 50.0	Excellent
41.0 - 90.0 (1-hr)	61.0 - 100.0 (1-hr)	21.0 - 40.0 (24-hr)	11.0 - 20.0 (24-hr)	101.0 - 200.0 (1-hr)	51.0 - 100.0	Fine
91.0 - 120.0 (1-hr)	101.0 - 130.0 (1-hr)	41.0 - 50.0 (24-hr)	21.0 - 25.0 (24-hr)	201.0 - 350.0 (1-hr)	101.0 - 150.0	Moderate
121.0 - 230.0 (1-hr)	131.0 - 240.0 (1-hr)	51.0 - 100.0 (24-hr)	26.0 - 50.0 (24-hr)	351.0 - 500.0 (1-hr)	151.0 - 200.0	Poor
231.0 - 340.0 (1-hr)	241.0 - 380.0 (1-hr)	101.0 - 150.0 (24-hr)	51.0 - 75.0 (24-hr)	501.0 - 750.0 (1-hr)	201.0 - 300.0	Very Poor
341.0 - 1000.0 (1-hr)	381.0 - 800.0 (1-hr)	151.0 - 1200.0 (24-hr)	76.0 - 800.0 (24-hr)	751.0 - 1250.0 (1-hr)	301.0 - 500.0	Severe

Ключова відмінність між цими індексами полягає у їхньому підході до оцінювання якості повітря. AQI (США) більш орієнтований на оперативне попередження населення і враховує короткотермінові коливання концентрацій, особливо дрібнодисперсних частинок PM_{2.5}, що є найбільш небезпечними для здоров'я. SAQI (ЄС) спирається на традиційні європейські нормативи й зазвичай використовується у міських системах моніторингу для щоденного інформування населення. Натомість EAQI (оновлений стандарт ЄС) створений для гармонізації оцінки якості повітря на міжнародному рівні, з більшою увагою до довгострокових ризиків і можливості порівняння даних між різними країнами.

Важливо розуміти, що індекси якості повітря – це не лише сухі числові значення, а й безпосередньо орієнтири для населення. У сприятливих умовах (добра чи задовільна якість) ризик для здоров'я мінімальний, проте зі зростанням

забрудненості навіть помірні перевищення можуть становити небезпеку для уразливих груп – дітей, літніх людей та осіб із хронічними захворюваннями. При поганій або дуже поганій якості повітря негативний вплив відчуває вже все населення: від дискомфорту та загострення респіраторних проблем до підвищення ризику серцево-судинних ускладнень. У надзвичайно критичних ситуаціях рівень забруднення може становити небезпеку навіть при короткочасному перебуванні надворі, тому офіційні рекомендації передбачають обмеження будь-якої активності на відкритому повітрі.

1.7. Діаграма прецедентів

У процесі моделювання системи моніторингу атмосферного повітря важливо не лише визначити її архітектуру та технічні складові, але й окреслити функціональні вимоги та взаємодію між користувачами й системою. Для цього застосовуються діаграми прецедентів (use case diagrams), що дозволяють візуально відобразити ролі основних акторів та перелік сценаріїв їхньої взаємодії з системою. Такий підхід робить вимоги більш зрозумілими, структурує функціонал і створює основу для подальшої реалізації та тестування програмного забезпечення.

Діаграма прецедентів (use case diagram) відображає взаємодію користувачів із модулем системи моніторингу стану атмосферного повітря. Вона дозволяє чітко ідентифікувати вимоги до розробки програмного забезпечення та забезпечити узгодженість між технічним рішенням і потребами кінцевих користувачів.

На рис. 1.9 представлено діаграму прецедентів, яка відображає ключові функціональні взаємодії в системі моніторингу атмосферного повітря.

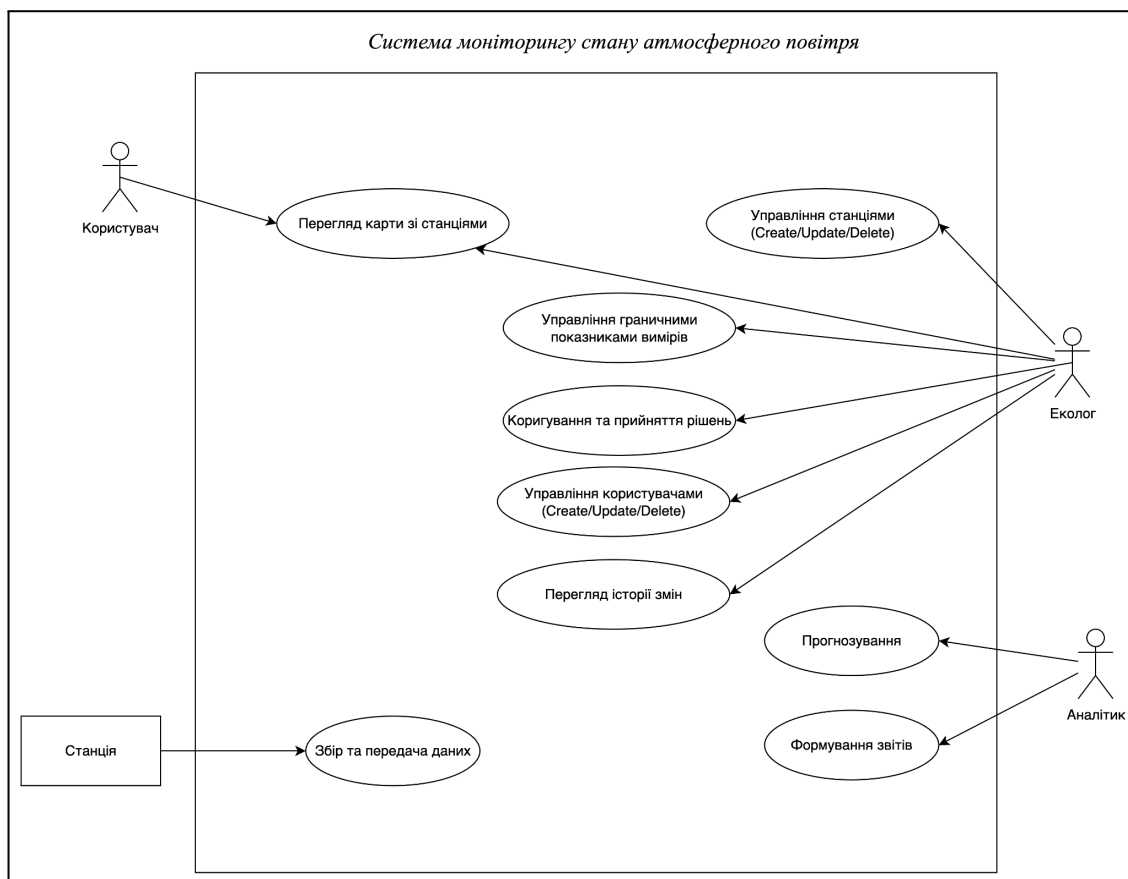


Рис. 1.9. Діаграма прецедентів

У ній виокремлено чотири основні групи акторів: *Користувач*, *Еколог*, *Аналітик* та *Станція*. Кожна з цих ролей виконує специфічні завдання й відповідає за певні процеси, що у сукупності забезпечують коректне функціонування системи.

- *Користувач* взаємодіє із системою на рівні перегляду даних. Його основне завдання – отримання інформації про стан повітря через інтерактивну карту, графіки та інші засоби візуалізації.
- *Еколог* відповідає за управління системою: налаштовує станції, визначає граничні показники вимірів, управляє обліковими записами користувачів та контролює історію змін. Крім того, він має повноваження коригувати результати й приймати управлінські рішення.
- *Аналітик* зосереджений на перетворенні даних у знання: формує звіти, здійснює аналітичну обробку та прогнозує рівень забруднення для підтримки стратегічних рішень.

- *Станція* виступає технічним актором, що забезпечує збір первинних екологічних даних та їхню передачу до центральної системи для подальшої обробки.

Таким чином, діаграма прецедентів відображає повний цикл взаємодії – від збору даних сенсорами до їхнього аналізу та використання для прийняття управлінських рішень і інформування населення. Вона дає змогу розглядати систему не лише як набір окремих компонентів, а як цілісне інтегроване середовище, у якому поєднано технічні засоби вимірювання, адміністративне управління та науково-аналітичну обробку інформації.

Завдяки такому підходу можна простежити, як дані проходять увесь шлях – від моменту їхнього виникнення на рівні вимірювальних станцій до інтерпретації результатів і формування управлінських рішень. Це підкреслює важливість кожного учасника процесу: від сенсора, що збирає дані, до аналітика, який створює прогноз, і адміністратора чи еколога, що приймають практичні рішення на основі цих результатів. У центрі ж цього ланцюга залишається користувач, для якого система має забезпечити простий та зрозумілий доступ до інформації про стан атмосферного повітря.

1.8. Технічне завдання на розробку

Технічне завдання є основою для подальшого проектування та реалізації системи. Воно визначає постановку задачі, формулює цілі розробки й описує вимоги, яким має відповідати програмний продукт. У цьому підрозділі наведено функціональні, нефункціональні та системні вимоги до інформаційної системи екологічного моніторингу атмосферного повітря, які встановлюють рамки її архітектури, функціоналу та умов експлуатації.

Постановка завдання. Мета розробки – створення інформаційної системи екологічного моніторингу атмосферного повітря, яка у режимі реального часу збирає дані з автоматизованих станцій і відкритих джерел, зберігає їх у базі даних,

обробляє та візуалізує результати для різних категорій користувачів: населення, аналітиків, адміністраторів, органів управління.

Система повинна забезпечувати:

- збір і збереження даних від датчиків різних типів;
- оперативне відображення інформації у вигляді інтерактивних карт та графіків;
- формування звітів та експорт даних (за станцією, параметром і періодом) у зручних форматах;
- аналітику та прогнозування, що підтримують прийняття управлінських рішень.

Основні вимірювані параметри включають: температуру, вологість, тиск, $PM_{2.5}$, PM_{10} , O_3 , NO_2 , SO_2 , CO та індекси якості повітря (AQI, CAQI). Додатково можуть збиратися: PM_1 , PM_4 , CO_2 , радіація, формальдегід, вологість, а також напрямок і швидкість вітру.

Отримані дані концентруються в базі даних, організованій за принципом OLTP для оперативної обробки інформації. Для аналітичних завдань формується сховище даних (OLAP), яке забезпечує багатовимірний аналіз (куби) за параметрами: час, локація, тип забруднювача. Це створює підґрунтя для аналізу індексів (AQI, EAQI, CAQI), побудови трендів і прогнозів, а також для інтеграції у державні та міжнародні екологічні системи.

Вимоги до системи. Для реалізації поставлених цілей інформаційна система повинна відповідати функціональним, нефункціональним та системним вимогам.

Система має забезпечувати такі функціональні вимоги:

- керування користувачами (створення, редагування, автентифікація);
- управління станціями моніторингу (додавання, редагування, видалення);
- перегляд даних у вигляді інтерактивних карт і часових рядів;
- фільтрацію та додавання станцій до обраних;
- встановлення та редагування граничних показників;
- ведення історії змін у системі;

- формування та вивантаження звітів за параметрами та періодами.

Із нефункціональних вимог система повинна мати простий та інтуїтивний інтерфейс, працювати лише за наявності стабільного підключення до інтернету, забезпечувати контроль коректності введених даних і захист інформації відповідно до сучасних стандартів безпеки. Також необхідні відмовостійкість, можливість одночасного доступу багатьох користувачів та доступність даних у режимі реального часу.

Системні вимоги. Програмне забезпечення повинно функціонувати на сучасних операційних системах (Windows, Linux, macOS) з підтримкою Python 3.10+ та необхідних бібліотек. Для локальної розробки й тестування потрібні:

- процесор із частотою не нижче 1.8 ГГц;
- оперативна пам'ять обсягом від 2 ГБ;
- не менше 10 ГБ вільного дискового простору.

У продуктивному середовищі система має бути розгорнута у хмарній інфраструктурі, що включає:

- вебсервер і базу даних, розміщені у хмарі для забезпечення масштабованості та доступності;
- резервне копіювання даних із заданою періодичністю (щодня/щотижня) та можливістю швидкого відновлення;
- використання механізмів безпеки (SSL/TLS, контроль доступу, шифрування даних при зберіганні й передачі);
- підтримку відмовостійкості за рахунок розподілу навантаження та реплікації бази даних.

Для доступу до онлайн-функцій необхідне стабільне інтернет-з'єднання. Інтерфейс системи підтримується сучасними браузерами: Google Chrome, Mozilla Firefox, Opera, Microsoft Edge.

Сформульовані вимоги дозволяють розглядати систему як цілісний інструмент екологічного моніторингу, що поєднує збір і збереження даних, їхню візуалізацію, аналітику та підтримку прийняття рішень. Чітке визначення функціональних і нефункціональних параметрів забезпечує узгодженість між

очікуваннями користувачів і можливостями системи, а системні вимоги задають технічне підґрунтя для її стабільної роботи у хмарному середовищі. Це створює основу для наступного етапу - розробки архітектурних рішень і моделей, які реалізують поставлені цілі.

Висновки до розділу 1

Аналіз предметної області засвідчив багатофакторну природу забруднення (антропогенні та природні джерела) і його медико-біологічні, екологічні та економічні наслідки. Оцінка сучасного стану в Україні виявила систематичні перевищення концентрацій забруднювачів. Цей стан оцінюється за допомогою індексу якості повітря (AQI), який є інтегрованим показником та довів свою ефективність як інструмент стандартизованої інтерпретації та виявлення ризиків. Однак, в українських умовах зберігається потреба уніфікації методик розрахунку, гармонізації з директивами ЄС і ширшого охоплення показників у державній мережі.

Системний аналіз об'єкта дослідження дозволив визначити, що це складна багаторівнева система, в якій поєднуються стаціонарні референтні пости, мобільні лабораторії і розподілені IoT-сенсори.

За результатом аналізу існуючих методів і технічних рішень щодо реалізації моніторингу якості атмосферного повітря, визначено, що, попри позитивну динаміку, ця сфера в Україні потребує подальшого вдосконалення: впровадження уніфікованих методик, розбудови прогнозних моделей і більш тісної інтеграції з міжнародними системами.

Враховуючи постанови і директиви КМУ, сформульоване технічне завдання, яке обґрунтовує створення інтелектуальної інформаційної технології моніторингу. Ця технологія побудована на основі методів OLTP, OLAP та Data Mining, і реалізує інтеграцію даних з різних джерел, аналітичну обробку, прогнозування з метою підтримки прийняття управлінських рішень у сфері екологічної безпеки.

РОЗДІЛ 2

МЕТОДИ ТА МОДЕЛІ УПРАВЛІННЯ ЕКОЛОГІЧНИМ МОНІТОРИНГОМ

2.1. Алгоритми розрахунку індексів

У попередньому розділі були розглянуті міжнародні та українські стандарти якості повітря, їхні нормативи та індекси, що визначають граничні концентрації забруднювачів та правила їх оцінювання. Однак самі нормативи ще не дають повної відповіді на питання: *наскільки небезпечним є поточний стан атмосфери чи які тенденції спостерігаються у довгостроковій перспективі.*

Наступним етапом стає моделювання процесів, яке забезпечує перехід від «сирих» даних сенсорів та автоматизованих станцій до узагальнених показників і аналітичних моделей. Воно дає можливість:

- інтегрувати виміряні концентрації у зручні для сприйняття індекси;
- визначати факти перевищення нормативів (добових, річних);
- виявляти закономірності й тренди у багатовимірних даних завдяки регресійному аналізу та методам Data Mining;
- будувати прогностичні сценарії та формувати KPI для екологічного управління;
- представляти результати у вигляді інтерактивних візуалізацій, доступних як спеціалістам, так і широкій громадськості з подальшими прийняттями рішень.

Індекси якості повітря є інтегральними показниками, що дозволяють швидко оцінити стан атмосферного повітря на основі концентрацій ключових забруднювачів. Їхнє основне завдання – перетворити складні екологічні дані на зручні для розуміння значення та категорії, які відображають рівень небезпеки для здоров'я населення.

Основне завдання індексів полягає в тому, щоб перекласти складні результати вимірювань (наприклад, десятки мікрограмів на кубічний метр для

PM_{2.5} чи діоксиду азоту) у шкалу категорій – від «доброї» до «небезпечної» якості повітря.

Дані для розрахунку надходять із автоматизованих станцій моніторингу або IoT-сенсорів у вигляді часових рядів з певною періодичністю (залежно від конфігурації системи). Перед обчисленням індексу виконується попередня обробка, що включає фільтрацію аномальних значень (видалення шумових або некоректних даних), нормалізацію показників до єдиних одиниць вимірювання та усереднення за потреби (наприклад, якщо використовується кілька сенсорів на одній станції). Технічно вони формуються через обчислення субіндексів для кожного забруднювача, після чого визначається агреговане значення індексу – зазвичай як максимальне серед усіх субіндексів, що гарантує відображення найгірших умов у певний момент часу.

Індекси враховують як короткострокові пікові перевищення, що становлять безпосередню небезпеку для вразливих груп, так і середньотривалі концентрації, пов'язані з хронічними наслідками впливу забруднювачів.

AQI (*Air Quality Index*). Індекс AQI використовується Агентством з охорони довкілля США (EPA) для уніфікованої оцінки стану атмосферного повітря [23]. Він відображає інтегральний рівень забруднення на основі концентрацій кількох основних речовин: PM_{2.5}, PM₁₀, NO₂, O₃, SO₂ та CO. Для кожного забруднювача окремо розраховується субіндекс, а підсумкове значення AQI визначається як максимальне серед усіх отриманих субіндексів, що гарантує врахування найкритичнішого впливу. Кожен субіндекс AQI обчислюється за формулою (2.1):

$$AQI_i = \frac{(I_{high} - I_{low})}{(C_{high} - C_{low})} \times (C - C_{low}) + I_{low}, \quad (2.1)$$

де:

- C – виміряна концентрація забруднювача;
- C_{high} та C_{low} – межі концентрації для відповідного діапазону AQI (див. таблиці порогів у розділі 1);

- I_{high} та I_{low} – межі субіндексу AQI для відповідного діапазону.

Такий підхід дозволяє виявляти найбільш критичний фактор у конкретний момент часу та забезпечує своєчасне відображення інформації на інтерактивних картах і графіках.

CAQI (*Common Air Quality Index*). Використовується у європейських країнах для оцінки якості повітря у міських і приміських умовах [24]. Він базується на тих самих принципах та ключових забруднювачах ($PM_{2.5}$, PM_{10} , NO_2 , O_3 , SO_2), але граничні значення для розрахунку відрізняються від американського AQI та орієнтовані на європейські нормативи.

Формула лінійної інтерполяції для CAQI є аналогічною до загальної (див. формулу розрахунку AQI), однак у типовій реалізації шкала ділиться на рівні інтервали по 25 одиниць. Це дозволяє застосувати спрощену формулу для розрахунку субіндексів (2.2):

$$CAQI_i = I_{low} + \frac{(C - C_{low}) \times 25}{(C_{high} - C_{low})}, \quad (2.2)$$

де:

- C – виміряна концентрація забруднювача;
- C_{high} та C_{low} – верхня та нижня межі концентрації речовини у відповідному нормативному діапазоні;
- I_{low} – нижня межа субіндексу CAQI для конкретного діапазону концентрацій.

CAQI менш поширений у міжнародній практиці, ніж американський AQI, і не завжди забезпечує порівнюваність результатів між країнами поза ЄС. Крім того, спрощений підхід із рівними інтервалами (0-25-50-75-100) робить систему інтуїтивно зрозумілою, але може знижувати чутливість у зонах низького чи високого забруднення порівняно з більш детальними шкалами.

EAQI (*European Air Quality Index*). Розроблений Європейським агентством з довкілля (EEA) у 2017 році для уніфікації оцінки якості повітря в усіх країнах ЄС [25]. Алгоритм орієнтований на ті ж самі ключові забруднювачі, що й AQI та CAQI, але використовує інші діапазони «точок зупину», наближені до актуалізованих рекомендацій ВООЗ та європейської директиви 2008/50/ЄС. На відміну від AQI, EAQI приділяє більше уваги довгостроковим середнім концентраціям, роблячи його більш придатним для порівняння між країнами.

Для кожного забруднювача обчислюється субіндекс EAQI за формулою лінійної інтерполяції (2.3):

$$EAQI_i = I_{low} + \frac{(I_{high} - I_{low})}{(C_{high} - C_{low})} \times (C - C_{low}), \quad (2.3)$$

де:

- C – виміряна концентрація;
- C_{high} та C_{low} – межі діапазону концентрацій;
- I_{high} та I_{low} – межі субіндексу для цього інтервалу.

Загальний індекс EAQI визначається як максимальне значення серед усіх субіндексів. Таким чином, «найгірший» забруднювач у даний момент визначає рівень небезпеки для населення.

EAQI працює з погодинними та добовими усередненнями (наприклад, для O_3 застосовується 1-годинний інтервал, тоді як для $PM_{2.5}$ і PM_{10} – добовий) та більш чутливий до тривалого впливу забруднень, ніж CAQI, і водночас менш «тривожний», ніж AQI у США (де $PM_{2.5}$ часто визначає індекс навіть при невеликих перевищеннях).

Алгоритми розрахунку індексів якості повітря (AQI, CAQI, EAQI) базуються на спільному принципі: перетворення вимірних концентрацій основних забруднювачів на інтегрований показник, що відображає рівень небезпеки для здоров'я населення. Для кожного забруднювача обчислюється

субіндекс за допомогою лінійної інтерполяції між нормативними порогами, після чого підсумкове значення визначається як максимальне серед усіх субіндексів.

Різниця між системами полягає у виборі забруднювачів, нормативних меж та шкали категорій. AQI (США) більше орієнтований на короткострокові коливання, SAQI (ЄС) застосовується для оперативного інформування населення, тоді як E AQI (ЄС, 2017) створений для гармонізованої довгострокової оцінки та міжнародного порівняння.

Завдяки такому підходу індекси стають зручним інструментом для комунікації наукових даних у зрозумілій формі, що дозволяє своєчасно реагувати на ризики забруднення та формувати політику у сфері охорони довкілля.

2.2. Алгоритми по дозволеним перевищенням

Для практичної реалізації контролю нормативів та порогів концентрацій забруднювачів важливим є побудова алгоритмів автоматизованої обробки даних. Алгоритм має враховувати специфіку вимірювань [19]: погодинні усереднення для NO_2 та SO_2 , добові середні для $\text{PM}_{2.5}$ і PM_{10} , трьохгодинні максимуми для озону та восьмигодинні середні для CO та O_3 . Таке поєднання дозволяє не лише фіксувати разові перевищення, а й забезпечувати коректне співставлення з граничними величинами та порогами оцінювання, що визначені у нормативних документах (Постанова КМУ №827, Директива 2008/50/ЄС).

На рис. 2.1 показано фрагмент алгоритму первинної обробки інформації, що надходить з автоматичних станцій моніторингу атмосферного повітря. Алгоритм включає етапи перевірки максимальних значень концентрацій для різних забруднювачів ($\text{PM}_{2.5}$, PM_{10} , O_3 , CO , SO_2 , NO_2 тощо) та їх порівняння з нормативними порогами. Це забезпечує можливість своєчасного виявлення перевищень, формування повідомлень для користувачів і подальшу передачу даних для аналітичних модулів системи.

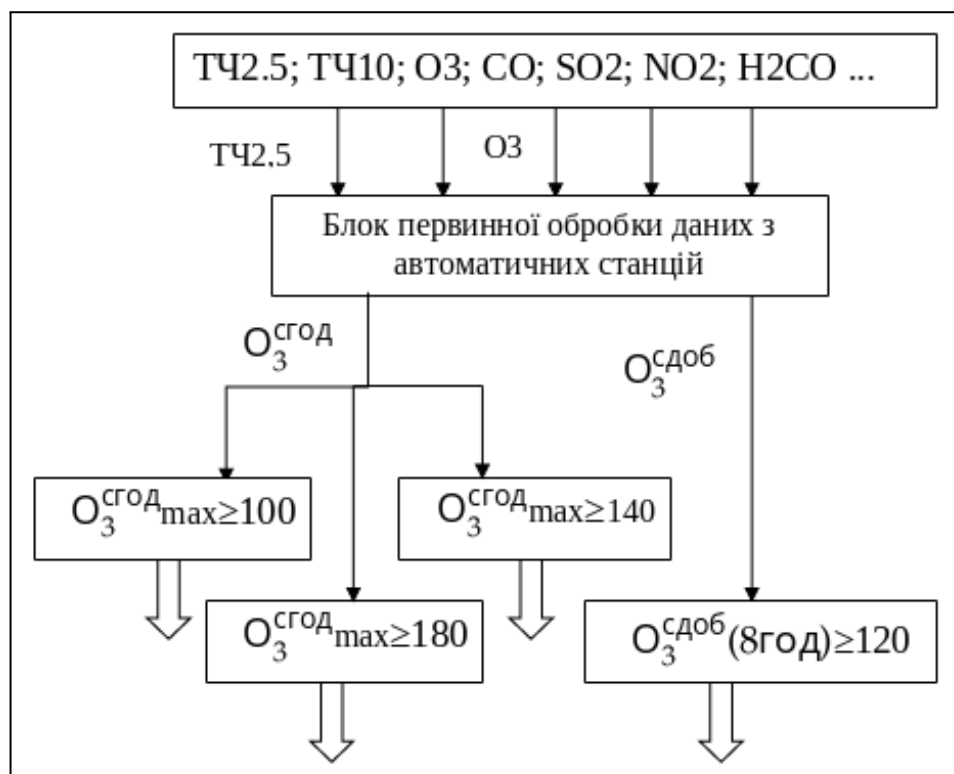


Рис. 2.1. Фрагмент алгоритму обробки інформації, отриманої з автоматичних станцій моніторингу атмосферного повітря

Таким чином, у систему моніторингу інтегрується подвійний підхід: поєднання нормативних значень (гранично допустимих концентрацій і порогів оцінювання) з алгоритмами їхнього автоматичного контролю. Це дозволяє перетворювати «сирі» дані станцій на інформаційно-аналітичний продукт – оцінку ризиків для здоров'я та довкілля, яка є зрозумілою як для органів управління, так і для населення. Це відкриває можливості для формування прогностичних моделей, визначення «гарячих точок» забруднення та своєчасного інформування чутливих груп населення.

Для органів управління та екологічних служб це означає наявність інструменту, що забезпечує доказову базу для ухвалення рішень: від оперативних попереджень до стратегічного планування у сфері охорони довкілля. Для громадськості – це доступ до зрозумілих показників і рекомендацій щодо поведінки у критичних ситуаціях (наприклад, обмеження перебування на відкритому повітрі при високому рівні забруднення).

2.3. Регресійна модель

Для глибшого аналізу і прогнозування доцільно застосовувати регресійні моделі [26, 27, 28], які дають змогу оцінити вплив окремих змінних (метеорологічних параметрів, антропогенних джерел, часових характеристик) на концентрації шкідливих речовин і будувати кількісні залежності.

Регресійний аналіз використовується для виявлення та кількісної оцінки залежностей між параметрами якості повітря та факторами, що на них впливають. У випадку систем екологічного моніторингу це можуть бути як метеорологічні умови (температура, вологість, швидкість вітру, атмосферний тиск), так і антропогенні чинники (транспортна активність, промислові викиди).

Основним методом, що застосовується у даному дослідженні, є Ordinary Least Squares (OLS) [29], який дозволяє побудувати лінійну модель залежності. Метод OLS ґрунтується на розрахунку прямих, які найкраще описують взаємозв'язок між незалежними змінними $X_1, X_2, \dots, X_k$ і залежною змінною Y . Вираз для лінійної регресії (2.4):

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \epsilon, \quad (2.4)$$

де:

- Y – цільова (залежна) змінна;
- $X_1, X_2, \dots, X_k$ – незалежні змінні;
- β_0 – інтерсепт (початкове значення Y , коли всі $X_i = 0$);
- $\beta_1, \beta_2, \dots, \beta_k$ – коефіцієнти регресії, що відображають внесок кожного фактора;
- ϵ – залишок моделі (шум, не пояснена частина).

OLS мінімізує суму квадратів залишків (SSE) за виразом (2.5) метою якого є знаходження значення $\beta_1, \beta_2, \dots, \beta_k$ які мінімізують SSE:

$$SSE = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2, \quad (2.5)$$

де:

- Y_i – фактичні значення залежної змінної;
- $\hat{Y}_i$ – передбачені значення залежної змінної;
- n – кількість спостережень;
- ϵ – залишок, відхилення фактичних значень від передбачених.

Коефіцієнти β обчислюються за виразом(2.6):

$$\beta = (X^T X)^{-1} X^T Y, \quad (2.6)$$

де:

- X – матриця незалежних змінних ;
- Y – вектор значень залежної змінної;
- X^T – транспонована матриця X ;
- $(X^T X)^{-1}$ – обернена матриця.

У моніторингу якості повітря залежною змінною виступає концентрація забруднювача ($PM_{2.5}$, PM_{10} , NO_2 та навіть сам індекс AQI чи CQAI), а незалежними змінними – погодні фактори (температура, тиск, вологість, вітер тощо), часові характеристики (сезонність, добові цикли) та супутні показники (наприклад, озон). Це дозволяє оцінювати внесок кожного чинника у рівень забруднення, будувати короткотермінові прогнози та виявляти аномальні ситуації.

У поєднанні з OLAP-аналітикою модель дозволяє прогнозувати концентрації та оцінювати ключові показники. Таким чином, OLS-регресія виконує роль базового інструменту аналітики й прогнозування у системі

моніторингу: від оцінки впливу факторів до побудови моделей прогнозу якості повітря для управлінських рішень.

2.4. Методи Data Mining

Методи Data Mining є ключовим інструментом для аналізу великих масивів даних у системах моніторингу атмосферного повітря. Вони дозволяють не лише виявляти приховані закономірності, але й будувати прогностичні моделі, які пояснюють, як формуються забруднення та які чинники найбільше впливають на їхній розвиток. На відміну від традиційних статистичних підходів, Data Mining орієнтований на пошук нетривіальних зв'язків і автоматизовану обробку даних, що особливо актуально для екологічних часових рядів, які є великими, динамічними та багатовимірними.

У системі моніторингу атмосферного повітря ці методи дають змогу перейти від простого накопичення даних до створення аналітичних інструментів, здатних прогнозувати зміни та підтримувати управлінські рішення. Для цього використовуються кілька підходів: кластеризація – для групування станцій чи вимірювань зі схожими характеристиками; класифікатор – для оцінки ймовірності певного стану атмосфери на основі доступних параметрів; правила асоціацій – для пошуку прихованих взаємозв'язків між рівнями забруднювачів та погодними умовами.

Разом ці інструменти формують основу інтелектуальної обробки даних, дозволяючи системі перетворювати «сирі» вимірювання на практичні знання та рекомендації для захисту здоров'я населення і довкілля.

2.4.1. Кластеризація. Кластеризація є одним із ключових методів Data Mining, який дозволяє виявляти приховані структури у даних без попереднього знання класів. Для систем моніторингу атмосферного повітря цей підхід особливо цінний, адже дає можливість автоматично групувати дані за схожими характеристиками та виявляти як типові сценарії, так і аномальні періоди забруднення.

У нашому дослідженні [30] було застосовано метод “K-Means”, який є одним із найпоширеніших алгоритмів неконтрольованого навчання (*unsupervised learning*) і широко використовується для розв’язання задач кластеризації [31]. Завдання “K-Means” полягає в пошуку такої конфігурації кластерів, що мінімізує сумарну відстань усіх точок до центрів своїх груп. Алгоритм послідовно оновлює координати центрів та склад кластерів до досягнення стабільності.

У математичному формулюванні мета алгоритму полягає у мінімізації функції (2.7):

$$J = \sum_{k=1}^K \sum_{x_i \in C_k} \|x_i - \mu_k\| \quad (2.7)$$

де:

- K – кількість кластерів,
- C_k – множина точок, що належать кластеру k ,
- μ_k – центр (центроїд) кластера k ,
- x_i – вектор ознак (набір виміряних екологічних параметрів).

Для моделювання використовувалися усереднені значення ключових екологічних показників: $PM_{2.5}$, PM_{10} , NO_2 , SO_2 , вологість та швидкість вітру. Попередньо дані проходили етапи очищення від аномалій, нормалізації та агрегації у стандартизовані часові вікна.

Алгоритм працює ітеративно, виконуючи послідовність кроків: ініціалізація центроїдів, призначення об’єктів до найближчого центру (на основі евклідової відстані), оновлення координат центроїдів та перевірка на стабільність. Для аналізу структури даних обирається число кластерів, яке можна змінювати в процесі дослідження. Це дозволяє виявити приховані шаблони у розподілі екологічних показників без потреби у заздалегідь відомих мітках категорій.

Вибір кількості кластерів є важливим етапом і зазвичай виконується за допомогою методів «лікоть» (*Elbow Method*), силуетного коефіцієнта або інших

метрик якості кластеризації. Це забезпечує баланс між деталізацією та узагальненням результатів.

Разом із тим, метод “K-Means” має обмеження: чутливість до початкового розташування центроїдів, можливість застрягання у локальних мінімумах та необхідність масштабування даних для уникнення домінування показників із більшими числовими діапазонами.

Таким чином, застосування кластеризації у системах моніторингу атмосферного повітря дозволяє не лише структурувати дані та виділяти аномальні сценарії, але й створювати основу для прогнозування та раннього попередження. Водночас ефективність кластеризації значною мірою залежить від коректної підготовки даних та обґрунтованого вибору числа кластерів.

2.4.2. *Наївний Байєсівський класифікатор (Gaussian Naive Bayes).* Класифікатор “Наївний Байєс” належить до методів, які ґрунтуються на теоремі Байєса та припущенні незалежності ознак [32]. У версії “Gaussian Naive Bayes” робиться додаткове припущення, що числові параметри розподілені за нормальним законом, що добре узгоджується з природою багатьох екологічних змінних (наприклад, концентрації забруднювачів, температура, тиск, вологість) [33].

Нормальний розподіл (Gaussian distribution) – це класична статистична модель, за якої більшість значень ознаки зосереджені навколо середнього, а ймовірність відхилення зменшується симетрично в обидва боки. Графік такого розподілу має «дзвоноподібну» форму.

У контексті моніторингу атмосферного повітря багато екологічних параметрів (температура, тиск, вологість, швидкість вітру) часто демонструють близькі до нормального розподіли, особливо коли розглядати усереднені добові значення. Це робить “Gaussian Naive Bayes” природним вибором для моделювання таких даних.

Важливо зазначити, що навіть якщо розподіл не є ідеально нормальним (наприклад, $PM_{2.5}$ або NO_2 мають пікові «відхилення» під час смогу чи викидів), алгоритм зберігає прийнятну точність. Саме тому він широко застосовується як

базова модель (baseline) у дослідженнях: швидкий у реалізації, обчислювально ефективний і достатньо надійний для порівняння з більш складними методами.

Модель використовує наступне формальне представлення (2.8):

$$P(C_k | x) = \frac{P(C_k) \prod_{i=1}^n P(x_i | C_k)}{P(x)}, \quad (2.8)$$

де:

- C_k – клас (наприклад, категорія якості повітря),
- $x = (x_1, \dots, x_n)$ – вектор ознак (екологічні параметри),
- $P(x_i | C_k)$ – ймовірність значення ознаки за умови належності до класу, яка апроксимується гаусівським розподілом (2.9):

$$P(x_i | C_k) = \frac{1}{\sqrt{2\pi\delta_k^2}} \exp\left(-\frac{(x_i - \mu_k)^2}{2\delta_k^2}\right), \quad (2.9)$$

де:

- μ_k – середнє значення,
- δ_k – стандартне відхилення ознаки в класі C_k .

Цей алгоритм є доцільним для класифікації поточного стану атмосфери за багатьма параметрами одночасно. Як цільову змінну можна брати категорію індексу якості повітря, а як предиктори – концентрації забруднювачів ($PM_{2.5}$, PM_{10} , NO_2 , SO_2 , CO) та метеорологічні фактори (температура, вологість, швидкість вітру). Завдяки цьому модель здатна:

- оперативно класифікувати ситуацію (якість повітря);
- виявляти ймовірні сценарії погіршення стану атмосфери за сукупністю факторів;
- формувати основу для системи попереджень населення та підтримки управлінських рішень.

До ключових переваг належать простота реалізації, висока швидкодія, стійкість до «високої розмірності» даних та здатність працювати навіть за відносно невеликих вибірок. Це робить його особливо придатним для систем моніторингу, де важлива як інтерпретованість, так і оперативність обчислень.

Разом із тим метод має і обмеження: він робить сильне припущення про незалежність ознак та їх нормальний розподіл, що не завжди відповідає реальним екологічним даним (наприклад, у випадках пікових викидів чи смогу). Тому “Gaussian Naive Bayes” у нашій системі доцільно розглядати насамперед як базову (стартову) модель, яка забезпечує швидке та надійне отримання результатів і може слугувати відправною точкою для подальшого розвитку методів аналізу.

2.4.3. Пошук асоціативних правил. Асоціативний аналіз є важливою складовою Data Mining, що дає змогу виявляти закономірності у вигляді логічних зв'язків між змінними у багатовимірних даних. У системах екологічного моніторингу цей підхід дозволяє знаходити приховані залежності, наприклад, між підвищенням концентрацій певних забруднювачів та конкретними метеорологічними умовами.

Одним із базових алгоритмів цього напрямку є Apriori [34], який використовується для пошуку частих наборів ознак та генерації правил типу. Вираз який описує пошук правила (2.10):

$$A \Rightarrow B, \quad (2.10)$$

де A і B – множини ознак, що спостерігаються одночасно з певною частотою у вибірці.

У контексті екологічного моніторингу це дозволяє виявляти взаємозв'язки між, наприклад, підвищеними концентраціями певних забруднювачів та метеорологічними умовами

Принцип роботи Apriori спирається на властивість антимонотонності, яка полягає в тому, що якщо певна комбінація ознак зустрічається рідше за заданий

пори́г, то всі її розширення також будуть рідкісними. Це дозволяє значно звузити простір пошуку та прискорити обчислення.

Алгоритм реалізує процес пошуку закономірностей у даних через послідовність чітко визначених етапів. Виконується у два основні етапи:

1. Виявлення частих наборів ознак тих, що мають рівень підтримки (support) вищий за заданий поріг;
2. Генерація асоціативних правил на основі знайдених наборів з оцінюванням їх достовірності (confidence), інформативності (lift) та інших метрик.

Для кількісної оцінки знайдених правил застосовуються такі ключові показники:

- Support (підтримка) – частка записів, у яких одночасно спостерігаються A і B ;
- Confidence (достовірність) – умовна ймовірність появи B за умови A ;
- Lift (підняття) – міра посилення, що відображає наскільки ймовірність B за наявності A вища за очікувану при незалежному розподілі.

Щоб побудувати транзакційний простір без втрати кількісної інформації, дані попередньо бінаризуються. Наприклад, значення концентрацій перетворюються у категорії «high»/«low» на основі порогових величин (медіани або нормативів).

Результати асоціативного аналізу мають практичне значення. Такі правила можна використовувати для створення інтерпретованих моделей оцінки ризику, що дозволяють ідентифікувати умови, за яких ймовірність погіршення якості повітря зростає. Крім того, вони формують основу для систем підтримки прийняття рішень, оскільки забезпечують зрозуміле пояснення виявлених закономірностей.

Завдяки високій інтерпретованості, алгоритм Apriori ефективно доповнює традиційні статистичні методи аналізу, допомагаючи дослідникам не лише описувати наявні дані, а й висувати нові гіпотези про взаємозв'язки у навколишньому середовищі.

Методи Data Mining у системах екологічного моніторингу відіграють важливу роль у трансформації «сирих» вимірювань у знання, придатні для аналізу й прогнозування. Кластеризація дозволяє структурувати дані та виявляти аномалії, “Gaussian Naive Bayes” забезпечує швидку й інтерпретовану класифікацію станів атмосфери, а асоціативний аналіз відкриває приховані взаємозв’язки між параметрами середовища. Використання цих підходів у комплексі створює основу для розробки інтелектуальних систем підтримки прийняття рішень, що здані своєчасно попередити про ризики та сприяти формуванню ефективних екологічних стратегій.

2.5. Алгоритм для побудови лінії тренду та інші (Power BI)

Аналітика у системах моніторингу атмосферного повітря не обмежується лише класифікацією та пошуком закономірностей. Важливим етапом є візуалізація результатів, що робить складні дані зрозумілими та придатними для оперативного використання. Представлення у вигляді трендів, графіків і показників ефективності (KPI) дозволяє оцінювати динаміку забруднення, виявляти критичні відхилення від нормативів і своєчасно реагувати на них.

Для цього доцільно застосовувати BI-інструменти (Business Intelligence), серед яких Power BI є одним із найбільш гнучких і поширених рішень [35]. Він забезпечує:

- інтеграцію з базами даних у реальному часі;
- побудову інтерактивних репортів;
- автоматичне оновлення показників у міру надходження нових даних;
- можливість налаштовувати рівні доступу для різних категорій користувачів (населення, аналітиків, органів влади).

Завдяки BI-підходу технічні параметри, отримані зі станцій моніторингу, стають зрозумілими навіть для нефахівців – вони набувають вигляду інтерактивних трендів, KPI та попереджувальних індикаторів, що дозволяє

ефективно використовувати дані як експертам, так і органам влади чи широкому загалу.

2.5.1. Лінії тренду. Побудова лінії тренду дозволяє виявити загальні закономірності у часових рядах даних про якість повітря, згладжуючи коливання та виділяючи довгострокові тенденції. У Power BI реалізовані інструменти для автоматичного додавання трендових ліній до графіків, що полегшує аналітику навіть для великих масивів даних [36].

Для побудови трендів застосовуються різні математичні підходи, кожен з яких має свої переваги залежно від характеру даних та завдань аналізу. Найчастіше використовуються такі методи як лінійну регресію, ковзне середнє та інші.

Лінійна регресія. Цей метод будує пряму, яка мінімізує суму квадратів відхилень фактичних значень від прогнозованих. Він дозволяє оцінити загальний напрямок змін (зростання чи зниження) показників, наприклад, рівня $PM_{2.5}$ протягом року. Простота та інтерпретованість роблять лінійну регресію базовим інструментом для аналізу трендів.

Ковзне середнє (Moving Average) [37]. Алгоритм ґрунтується на усередненні значень у фіксованому вікні часу (наприклад, 7 днів). Це згладжує короткострокові флуктуації та дозволяє виявити довгострокові закономірності, що є особливо корисним при сезонних коливаннях забруднювачів. Метод часто застосовують у системах моніторингу для виявлення стійких змін, відокремлюючи їх від випадкового шуму.

Інші методи трендової аналітики застосовуються у специфічних випадках. Експоненціальна регресія описує процеси з прискореним зростанням чи спаданням (наприклад, концентрації озону влітку). Поліноміальна регресія дає змогу моделювати складні нелінійні залежності з кількома точками перегину. Логарифмічна регресія підходить для процесів, які спочатку різко зростають, а згодом стабілізуються, наприклад, при пікових викидах.

Для кожного з цих методів у Power BI є вбудовані інструменти, а математичні основи докладно описані у підручниках зі статистики та офіційній документації Microsoft Power BI [38, 39].

2.5.2. *Показники ефективності (KPI)*. Показники ефективності (KPI) у системі моніторингу атмосферного повітря виступають формалізованими метриками, які дозволяють оцінювати не лише поточний стан довкілля, а й динаміку виконання нормативів та екологічних цілей. Розрахунок KPI ґрунтується на кількісній обробці часових рядів [40].

Середньомісячні концентрації $PM_{2.5}$ та PM_{10} обчислюються як середнє арифметичне добових значень за місяць. Вираз який описує розрахунок такого KPI (2.11):

$$KPI_{PM} = \frac{1}{n} \sum_{i=1}^n C_i, \quad (2.11)$$

де C_i – середньодобова концентрація, n – кількість днів у місяці.

Кількість днів із перевищенням допустимих значень – рахується як частка днів, коли C_i більше ГДК. Це дозволяє оцінювати відповідність системи міжнародним стандартам (наприклад, рекомендаціям ВООЗ чи директиві 2008/50/ЄС).

Частка часу у «поганих» категоріях якості повітря вимірюється як відношення кількості днів, коли індекс перебував у зонах «*Poor*» чи «*Very Poor*», до загальної кількості вимірювань за період (2.12):

$$KPI_{bad} = \frac{t_{poor} + t_{very\ poor}}{T}, \quad (2.12)$$

де T - загальна кількість часових інтервалів.

У Power BI ці показники реалізуються у вигляді KPI-візуалізацій, які поєднують числові значення з графічними індикаторами (зелений – норматив

дотримано, жовтий – наближення до критичної межі, червоний – перевищення) [41]. Такий підхід забезпечує швидку інтерпретацію результатів як для дослідників, так і для управлінців.

Висновки до розділу 2

Формалізовані алгоритми розрахунку інтегральних індексів (AQI, CAQI, EAQI) надають можливість оцінити якість повітря в уніфікованому вигляді, використовуючи різні підходи. Їхнє одночасне використання в системі моніторингу дозволяє під різним кутом виявляти залежність якості повітря від забруднювачів.

Запропонований алгоритм контролю дозволених перевищень з урахуванням нормативних часових вікон (погодинні, добові, 8-годинні тощо) забезпечує коректну верифікацію відповідності стандартам і автоматизоване інформування користувачів та органів управління.

Побудована регресійна модель дозволяє оцінити вплив забруднювачів на індекс якості повітря.

Методи Data Mining розширюють можливості аналізу:

- за допомогою методу K-means сформовано просторово-часові патерни та виявлено аномалії, які використані в кластерному аналізі;
- метод Naive Bayes забезпечує швидку інтерпретовану класифікацію станів атмосфери, що використовується для прогнозування;
- метод асоціативних правил дозволяє сформулювати правила які також використовуються для прогнозування;
- інтеграція з BI-інструментами (Power BI) уніфікує представлення результатів через трендові лінії, KPI та індикатори, роблячи аналітику придатною для експертів.

РОЗДІЛ 3

ПРОЕКТУВАННЯ СИСТЕМИ

3.1. Архітектура системи

Проектування системи моніторингу є ключовим етапом переходу від концептуального рівня (індекси, моделі, аналітика) до практичної реалізації. У цьому розділі розглядається архітектура системи, яка включає апаратні вузли (станції, сервери), канали та протоколи передавання даних, а також програмно-інформаційне забезпечення.

Особлива увага приділяється базам даних і сховищам даних, оскільки саме вони забезпечують накопичення, обробку й подальшу інтеграцію великих масивів екологічної інформації. Також розглядатимуться алгоритмічні блоки, що формують ядро системи: від розрахунку індексів та трендів до застосування методів Data Mining для прогнозування та виявлення прихованих закономірностей.

Системи моніторингу атмосферного повітря проектуються як багаторівневі та модульні комплекси [42]. Такий підхід дозволяє їм не лише масштабуватися із зростанням кількості сенсорів, але й легко інтегрувати нові компоненти, джерела даних та алгоритми. Гнучкість архітектури забезпечує роботу в реальному часі навіть за умови значних навантажень.

Однією з ключових вимог є можливість інтеграції з різними джерелами інформації – від стаціонарних станцій і мобільних сенсорних модулів до державних API та сторонніх систем. Це створює передумови для формування єдиної інформаційної платформи, здатної охоплювати просторово розподілені дані та забезпечувати комплексну оцінку стану атмосфери.

Архітектура системи включає апаратні вузли (сенсори, станції, edge-модулі), програмне забезпечення для обробки інформації, канали зв'язку та протоколи обміну, сховища даних для довгострокового збереження та аналітики, а також сервіси візуалізації та підтримки рішень [43]. Такий поділ дозволяє кожному

рівню архітектури виконувати свою спеціалізовану функцію й водночас узгоджено працювати у спільному середовищі.

3.1.1. Компоненти та вузли системи. Системи моніторингу якості повітря забезпечують безперервний збір, аналіз і візуалізацію даних про стан атмосфери, що дозволяє своєчасно виявляти потенційно небезпечні рівні забруднення та приймати відповідні заходи. Проте сучасні реалії вимагають від таких систем нових підходів до обробки великих обсягів інформації в режимі реального часу, гнучкої інтеграції з різними джерелами даних і впровадження інноваційних алгоритмів для точного прогнозування стану повітря [44, 45].

Забезпечення ефективного збору даних вимагає впровадження асинхронних механізмів накопичення інформації з усіх станцій за певними інтервалами. Це досягається шляхом використання підписки брокерів на отримання повідомлень від датчиків через протокол MQTT, що забезпечує надійний зв'язок навіть у складних умовах. Крім того, інтеграція API дозволяє взаємодіяти з іншими джерелами даних, такими як державні служби та екологічні організації, забезпечуючи обмін інформацією та розширення функціональних можливостей системи.

Окремою проблемою є обробка та збереження великого обсягу даних в асинхронному режимі. Інформація, що надходить від датчиків, повинна бути оперативно оброблена, проаналізована та збережена в базі даних для подальшого використання. На цьому етапі застосовуються алгоритми обчислення індексів якості повітря AQI (Air Quality Index), CAQI (Common Air Quality Index) та EAQI (European Air Quality Index), які дозволяють визначати рівні забруднення та прогнозувати їх зміни. Для забезпечення структурованого доступу дані додатково проходять процес ETL (Extract-Transform-Load), у ході якого вони очищуються, нормалізуються й уніфікуються. Результатом є завантаження у сховище даних (data warehouse або data lake), оптимізоване для багатовимірного аналізу (OLAP), що забезпечує швидке агрегування та аналітичну обробку.

Таким чином, розробка систем моніторингу якості повітря вимагає вирішення численних технічних і технологічних завдань, пов'язаних із

забезпеченням стабільної роботи в реальному часі, асинхронного збору та попередньої обробки даних, інтеграції з різними джерелами інформації, а також побудови OLAP-структур як основи для методів Data Mining та прогнозування. Впровадження таких рішень дозволяє створити ефективні та надійні системи, що перетворюють «сирі» екологічні вимірювання на знання, придатні як для прийняття управлінських рішень, так і для інформування громадськості. Діаграма розгортання зображена на рис. 3.1.

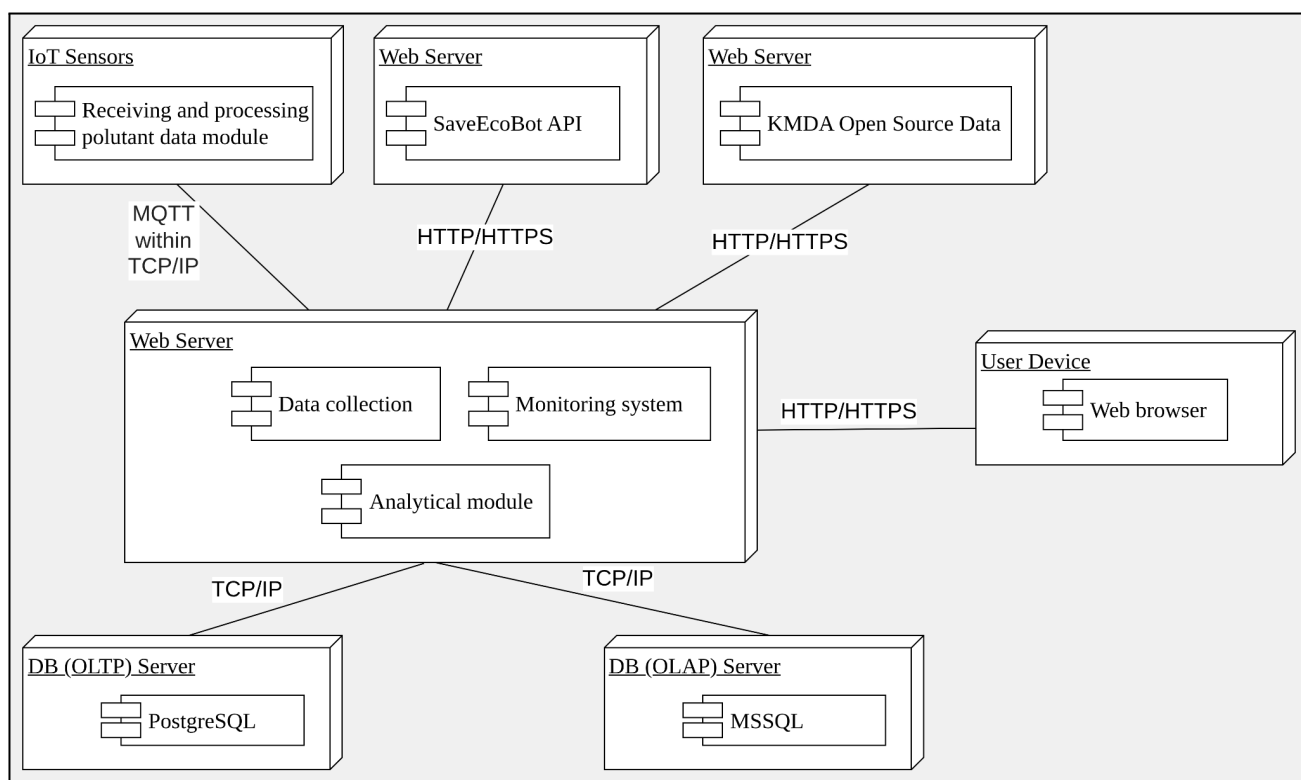


Рис. 3.1. Діаграма розгортання

Архітектура системи моніторингу атмосферного повітря побудована на модульному принципі та включає низку взаємопов'язаних компонентів. У її основі лежать апаратні сенсори й станції, що здійснюють вимірювання концентрацій забруднювачів ($PM_{2.5}$, PM_{10} , NO_2 , SO_2 , O_3), а також параметрів довкілля – температури, вологості та атмосферного тиску. Станції можуть бути як стаціонарними, так і мобільними, що дозволяє поєднувати постійний моніторинг у ключових точках із можливістю оперативного розширення мережі.

Зібрані дані надходять до edge-модулів, де виконуються попередня агрегація, фільтрація та синхронізація часу. Це зменшує навантаження на серверну інфраструктуру та мінімізує ризики втрати інформації при перебоях зв'язку. Передача даних у центральну систему відбувається за допомогою брокерів повідомлень (MQTT, HTTP чи WebSocket), які забезпечують асинхронний та надійний обмін між вузлами.

Центральна система відповідає за приймання, валідацію й нормалізацію інформації. Оброблені дані зберігаються в реляційних базах (PostgreSQL) для транзакційних операцій і в сховищах (data warehouse або data lake) для багатовимірного аналізу та побудови OLAP-структур. Для підвищення продуктивності застосовується Redis – як для кешування проміжних результатів, так і для управління чергами завдань.

Інтеграцію з іншими системами забезпечують API та мікросервіси, що дозволяють обмінюватися даними з державними службами, екологічними платформами, мобільними застосунками та внутрішніми модулями. На основі накопичених даних працюють аналітичні сервіси: вони обчислюють індекси якості повітря (AQI, CAQI, EAQI), застосовують методи Data Mining (кластеризація, класифікація, асоціативні правила) та формують прогностичні моделі.

Результати аналізу відображаються через користувацькі інтерфейси. Для побудови дашбордів, трендів і KPI застосовуються BI-інструменти (Power BI), тоді як інтерактивна взаємодія з даними реалізується у веб і десктоп додатках, розроблених на основі Vue.js. Це дає змогу спеціалістам, органам влади та громадськості оперативно оцінювати стан атмосфери, отримувати сповіщення про перевищення нормативів і формувати управлінські рішення.

Таким чином, система являє собою комплекс взаємопов'язаних вузлів – від сенсорів і брокерів повідомлень до аналітичних модулів і візуалізаційних інструментів. Їхня узгоджена робота забезпечує безперервний моніторинг, масштабованість і здатність інтегрувати нові джерела даних.

3.1.2. Протоколи та формати обміну. У системах моніторингу якості повітря застосовуються кілька мережевих протоколів обміну даними. Зокрема, ключовими для побудови сучасної архітектури є MQTT, HTTP та WebSocket. Кожен із цих протоколів має свою архітектуру та переваги, тому вибір обґрунтовується специфікою задач: збором телеметрії з IoT-сенсорів, інтеграцією з вебсервісами відкритих даних та забезпеченням реального часу для користувацьких інтерфейсів. Нижче розглянуто архітектурні особливості кожного протоколу та мотивацію їх використання у розробленій системі.

MQTT (Message Queue Telemetry Transport) – легковагий мережевий протокол, спеціально створений для телеметрії та IoT-рішень. Архітектура MQTT базується на моделі публікації/підписки: дані передаються через брокер повідомлень, до якого підключаються як видавці (сенсори), так і підписувачі (центральный сервер та інші споживачі) [46]. На відміну від традиційної схеми клієнт-сервер, MQTT-брокер виступає посередником, який приймає повідомлення від пристроїв та розсилає їх усім зацікавленим підписникам відповідних “топиків” (тем). Протокол працює поверх TCP, що забезпечує надійність передачі (контроль цілісності пакетів, упорядкування). MQTT максимально мінімізує службове навантаження в пакеті – заголовки дуже малі, а корисні дані можуть бути бінарними, що особливо важливо для пристроїв з обмеженими ресурсами. Після встановлення з’єднання MQTT залишається постійно відкритим, і через нього може передаватися будь-яка кількість повідомлень в обох напрямках – від сенсора до сервера та назад. Протокол підтримує різні рівні QoS (гарантії доставки) і спеціальні механізми (збережені повідомлення “retained”, повідомлення заповіту “last will” тощо) для забезпечення надійності та відстеження стану клієнтів.

У цілому MQTT спроектований під потреби IoT: він оптимізує енергоспоживання пристроїв та ефективність передачі даних у мережах з високою затримкою чи обмеженою пропускну здатністю [47]. Дослідження показують, що MQTT має менші затримки й накладні витрати проти HTTP, що прямо подовжує час автономної роботи батарей у пристроях. Саме тому в нашій системі

MQTT обрано для передачі даних від власних сенсорів як найбільш придатний протокол для ресурсообмежених модулів.

HTTP (HyperText Transfer Protocol) – універсальний протокол прикладного рівня, першочергово створений для обміну документами в Інтернеті. Він має модель клієнт-сервер: клієнт ініціює запит (GET, POST тощо) до сервера, а сервер повертає відповідь. Сесії HTTP зазвичай короткочасні: кожен запит відкриває з'єднання, отримує відповідь і може закривати з'єднання (хоча можливі й *keep-alive* сесії). В контексті IoT та моніторингу HTTP часто застосовується для надання RESTful API – зручного способу отримання даних сторонніми споживачами. Зокрема, наша система використовує HTTP для інтеграції з відкритими джерелами даних (приклади – API сервісу SaveEcoBot, відкриті API міських екологічних порталів). HTTP-протокол широко підтримується і стандартизований, що спрощує взаємодію з вебсервісами. Архітектурно HTTP поступається MQTT за ефективністю при частих невеликих повідомленнях: для кожного запиту містить значний обсяг службових заголовків, виконує повторну автентифікацію та не зберігає стан між сеансами [48]. На відміну від MQTT, сервер HTTP не може самостійно надсилати дані клієнту без запиту останнього, тому для отримання актуальної інформації потрібне періодичне опитування (polling) або ж використання розширень (довготривалі з'єднання, Server-Sent Events).

Незважаючи на ці обмеження, HTTP залишається незамінним для доступу до відкритих екосистем: практично всі зовнішні екологічні сервіси надають дані через HTTPS API. Тому в нашій реалізації HTTP-взаємодія відповідає за періодичне зчитування даних з зовнішніх джерел та передачу результатів користувачам у разі запитів до вебінтерфейсу. Простота реалізації HTTP-клієнта та сумісність з інфраструктурою Інтернету обґрунтовують вибір цього протоколу для інтеграції зовнішніх даних, попри його дещо нижчу енергоефективність.

WebSocket – протокол поверх TCP, що забезпечує встановлення постійного двобічного з'єднання між клієнтом і сервером. Історично WebSocket розроблено як розширення можливостей веббраузерів: він дозволяє вебдодаткам підтримувати

full-duplex взаємодію з сервером у реальному часі (те, чого стандартний HTTP не може зробити без додаткових прийомів) [49]. Встановлення WebSocket-з'єднання починається зі спеціального HTTP *handshake*, після чого канал переходить у режим сокету, по якому обидві сторони можуть вільно обмінюватися повідомленнями. Архітектурно WebSocket близький до моделі “постійне з'єднання” MQTT, але не диктує семантики повідомлень – це скоріше транспортний механізм. У контексті моніторингу якості повітря WebSocket доцільно застосувати для трансляції даних у режимі реального часу до клієнтських застосунків. Наприклад, вебдашборд або мобільний додаток можуть встановити WebSocket-з'єднання з сервером та отримувати оновлення про показники забруднення негайно після їх надходження, без затримок на опитування. Це істотно підвищує інтерактивність системи: користувач бачить зміну рівня $PM_{2.5}$ чи інших параметрів практично миттєво.

Протокол WebSocket ефективніший за чистий HTTP для стрімких оновлень, оскільки після ручного потиску руки витрати на заголовки мінімальні. Однак у порівнянні з MQTT WebSocket не надає вбудованих можливостей, таких як маршрутизація по темах, черги повідомлень чи QoS-рівні – усі ці функції довелося б реалізовувати вручну поверх WebSocket. Таким чином, MQTT залишається основним протоколом для IoT-девайсів (сенсорів), а WebSocket виконує допоміжну роль для доставки агрегованих даних користувачам.

Механізм передачі даних від IoT-сенсорів через MQTT. Збір телеметричних даних від розподілених IoT-сенсорів до центрального обчислювального вузла реалізовано за допомогою протоколу MQTT та брокерної інфраструктури. Сенсорні модулі на пунктах моніторингу функціонують як MQTT-клієнти, що публікують результати вимірювань у визначені топіки. Повідомлення у форматі JSON містять ідентифікатор сенсора чи станції, часову мітку та значення концентрацій контрольованих забруднювачів ($PM_{2.5}$, PM_{10} , NO_2 тощо). Для кожного типу даних передбачено окремі черги (топіки). Процес зображений на рис. 3.2.

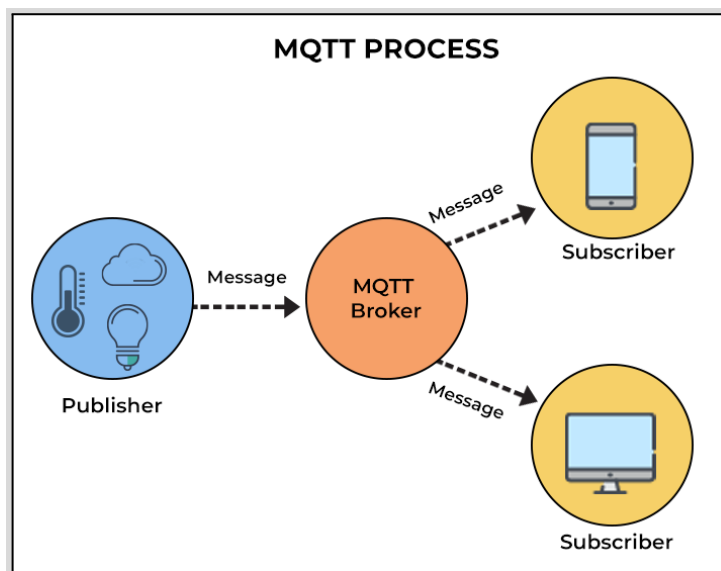


Рис. 3.2. MQTT Process View

Центральний сервер виконує роль підписувача, підтримуючи постійне з'єднання з брокером і отримуючи дані у режимі push. Такий підхід підвищує ефективність і забезпечує надходження інформації майже в реальному часі.

Використання MQTT-брокера забезпечує кілька ключових переваг: декуплінг відправників і одержувачів (сенсори взаємодіють лише з брокером, а не з сервером безпосередньо); масштабованість і буферизація (обробка сотень клієнтів, зберігання останніх повідомлень, повторна доставка при відновленні з'єднання); стійкість до перебоїв зв'язку (автоматичне перепідключення клієнтів, гарантії доставки за рахунок QoS). Зокрема, $QoS=1$ гарантує доставку «принаймні один раз», а $QoS=2$ – «рівно один раз».

Таким чином, схема «MQTT-брокер – центральний сервер» формує надійний і відмовостійкий канал збору телеметрії, що є основою для подальшого збереження, агрегування та аналітичного опрацювання даних у системах інтелектуального екологічного моніторингу.

Інтеграція з відкритими джерелами даних. Окрім власних сенсорних показників, система моніторингу інтегрує інформацію з відкритих зовнішніх ресурсів, зокрема екологічних API та відомчих сервісів. Архітектурно взаємодія побудована на основі вебзапитів: центральний сервер періодично виконує HTTP

GET-звернення до зовнішніх інтерфейсів та отримує екологічні показники у машиночитному форматі (переважно JSON).

Важливим джерелом є SaveEcoBot – незалежна екоплатформа, що агрегує дані з державних, громадських і приватних станцій моніторингу повітря в Україні. Доступ до даних забезпечується через відкритий API, який надає можливість отримувати актуальні концентрації забруднювачів по сотнях станцій. Частота опитування синхронізована з інтервалами оновлення SaveEcoBot (близько 10 хвилин), що дозволяє своєчасно отримувати нові дані без перевантаження сервісу. Отримана інформація (ідентифікатори станцій, координати, часові мітки та виміряні концентрації) підлягає подальшому парсингу, перевірці та збереженню у внутрішній базі даних.

Додатковим джерелом є муніципальні системи моніторингу, наприклад, Київської міської державної адміністрації (КМДА), які публікують дані у форматах JSON або через власні API. Типовий запит повертає набір останніх вимірів для міських станцій (PM_{10} , $PM_{2.5}$, NO_2 тощо) із зазначенням координат та часу. Частота оновлення залежить від конкретної системи й може варіюватися від режиму близького до реального часу (оновлення кожні 10 сек). Відповідно, сервер адаптує графік опитування під характеристики кожного джерела.

Таким чином, інтеграція з відкритими платформами й офіційними ресурсами дозволяє розширити інформаційну базу системи та підвищити повноту й точність оцінки стану атмосферного повітря.

Важливо, що інтеграція з відкритими API проводиться у відмовостійкому режимі. Сторонні сервіси можуть бути часом недоступні чи повільні, тому виклики до них виконуються асинхронно (у фонових завданнях) з тайм-аутами та повторними спробами при помилках. Такий підхід ізоляції зовнішніх запитів гарантує, що тимчасова недоступність, скажімо, API SaveEcoBot, не вплине на безперервність збору даних з власних датчиків. Отримані з різних джерел екодані після валідації уніфікуються (зводяться до спільної структури) і зберігаються у сховищі даних системи для подальшого аналізу та візуалізації.

Порівняння протоколів MQTT, HTTP та WebSocket. Для обґрунтування вибору мережевого протоколу доцільно проаналізувати MQTT, HTTP та WebSocket за ключовими критеріями. Порівняння представлені в таблиці 3.1 дозволяють оцінити ефективність протоколів у контексті збору даних від розподілених сенсорних систем.

Таблиця 3.1

Порівняльна таблиця протоколів обміну даними

Критерій	MQTT	HTTP	WebSocket
Модель зв'язку	Асинхронна модель «видавець-підписувач» через брокер.	Синхронна модель запит-відповідь; клієнт-сервер.	Постійне двостороннє з'єднання після handshake.
Енергоефективність	Висока: мінімальні накладні дані, довге з'єднання без повторних рукопотискань.	Низька: кожен запит створює додаткові накладні витрати.	Середня: потребує ресурсів для підтримки з'єднання, але знижує трафік.
Надійність передачі	Вбудовані QoS (0, 1, 2), підтримка <i>retain</i> ; робота поверх TCP.	Лише TCP-надійність; повтори на рівні додатка.	Залежить від стабільності з'єднання.
Пристосованість до IoT	Максимальна: оптимізований для пристроїв з обмеженими ресурсами та слабкого зв'язку.	Помірна: ефективний для більших повідомлень, але витратний для сенсорних мереж.	Обмежена: зручний для потоків, але ресурсомісткий для сенсорів.
Масштабованість	Висока: брокер маршрутизує повідомлення, легко додавати нових клієнтів.	Обмежена: велика кількість запитів створює навантаження на сервер.	Помірна: потребує утримання відкритих з'єднань, масштабування ускладнене.

MQTT вирізняється високою енергоефективністю та стійкістю до перебоїв завдяки моделі «публікація – підписка» через брокер і підтримці рівнів QoS. Це забезпечує гнучке масштабування та надійність у мережах з великою кількістю IoT-пристроїв.

HTTP застосовується повсюдно, проте його модель «запит – відповідь» створює значні накладні витрати для частих невеликих повідомлень. Протокол є доцільним для періодичної передачі більших обсягів даних, але менш ефективний у середовищі з обмеженими ресурсами.

WebSocket забезпечує повнодуплексний канал зв'язку та низькі затримки, що робить його придатним для потокової передачі. Однак підтримка постійних з'єднань вимагає додаткових ресурсів, а відсутність QoS-механізмів знижує надійність у порівнянні з MQTT.

Порівняння каналів збору даних: власні сенсори (MQTT) vs зовнішні API продемонстровані в таблиці 3.2. Система моніторингу отримує інформацію двома основними шляхами: від власних IoT-сенсорів через MQTT та з відкритих зовнішніх джерел через API. Хоча кінцева мета однакова – отримання актуальних екологічних показників, – ці канали мають суттєві відмінності у способі передачі, частоті оновлення, контролі над даними та вимогах до подальшої обробки.

Таблиця 3.2

Порівняльна таблиця характеристик двох підходів

Характеристика	MQTT (IoT-сенсори)	HTTP API (зовнішні джерела)
Джерело даних	Фізичні сенсори власної мережі; контроль обладнання та калібрування.	Сторонні сервіси (SaveEcoBot, міські системи тощо); залежність від надійності джерела.
Спосіб передачі	Push-модель: сенсор сам відправляє дані через брокер.	Pull-модель: сервер періодично опитує API.
Частота оновлення	Гнучка: від 1-5 хв до подійних тригерів; налаштовується під потреби.	Визначається зовнішнім джерелом (від 10-15 хв до години й більше).
Надійність та контроль	Високий контроль (QoS, буферизація); відмова одного сенсора не впливає на інших.	Залежність від доступності API; можливі затримки чи пропуски; контроль обмежений.
Формат і сумісність	Єдиний визначений формат JSON з уніфікованими полями.	Різні формати (JSON, CSV); потрібна нормалізація й валідація даних.

Обидва канали доповнюють один одного, забезпечуючи комплексну картину якості повітря. MQTT сенсори дають можливість отримувати дані з високою

роздільною здатністю у просторі та часі (локальні виміри в режимі близькому до реального часу). Натомість відкриті API покривають ширший регіон або надають офіційно підтверджені показники, які можуть слугувати еталоном для звірки. З погляду архітектури, система побудована так, що обидва потоки даних сходяться на центральному сервері та далі обробляються спільно. Це потребує уніфікації форматів та синхронізації часових рядів, але в підсумку збагачує інформаційну базу для аналізу. Варто підкреслити, що асинхронний дизайн (рознесення каналів за чергами завдань) дозволив підтримувати ці два джерела незалежно: проблеми в одному не призводять до зупинки іншого. Таким чином, комбінування власних сенсорних вимірювань і відкритих даних підвищує надійність і повноту моніторингу.

Для представлення та передачі екологічних даних у системі використовується формат JSON (JavaScript Object Notation). JSON обрано як основний формат обміну через його простоту, легку читаємість людиною та широке підтримання програмними засобами. Усі повідомлення від сенсорів через MQTT, а також відповіді від зовнішніх API приводяться до JSON-структури, узгодженої з внутрішньою моделлю даних.

Кожне отримане повідомлення проходить перевірку коректності перед тим, як буде використане в аналізі чи збережене. Валідація даних відбувається на двох рівнях: синтаксичному та семантичному. На першому рівні *синтаксичної валідації* перевіряється правильність формату JSON (чи відповідає рядок JSON-стандарту, чи всі дужки і лапки на місці). Далі застосовується *схемна валідація*: структура повідомлення звіряється з очікуваним шаблоном. Якщо якогось критичного поля бракує або тип даних невірний (скажімо, концентрація $PM_{2.5}$ прийшла як текст “N/A” замість числа), повідомлення відхиляється як невалідне і логіється для подальшого аналізу.

Особлива увага приділяється узгодженості часових міток: система синхронізує часові пояси та формати, конвертуючи всі часи до єдиного стандарту (UTC). Це дозволяє коректно об'єднувати дані від різних джерел у хронологічному порядку. Також виконується перевірка на дублікати: якщо те саме

JSON-повідомлення (ідентифіковане по комбінації сенсор+час) уже було отримане, другий екземпляр ігнорується, щоб не подвоювати записи.

Стандартизована структура повідомлень і ретельна валідація гарантують, що до бази даних та аналітичних модулів потрапляють тільки коректні та узгоджені між собою дані. Безпечний обмін у форматі JSON слугує фундаментом для подальших етапів – від побудови бази даних до OLAP-кубів та застосування алгоритмів Data Mining.

3.1.3. Вибір технологій. Вибір технологій напряму впливає з вимог попередніх підрозділів: потоковий збір телеметрії (MQTT/API), асинхронна обробка, зберігання як “сирих” потоків, так і агрегатів, а також масштабована аналітика й візуалізація. Нижче подано обґрунтований стек, який забезпечує надійність, продуктивність і керованість системи.

Серверна платформа: Python / Django. Django обрано як базовий фреймворк завдяки зрілій екосистемі, вбудованій безпеці (захист від SQLi/XSS/CSRF), ORM для швидкої роботи з БД та зручному адмін-інтерфейсу. Django REST Framework (DRF) використовується для побудови зовнішніх і внутрішніх API (автентифікація, пагінація, throttling, версіонування). Python дає змогу напряму використовувати наукові пакети (NumPy/Pandas/scikit-learn) для обчислення індексів (AQI/CAQI/EAQI), регресій і Data Mining [50, 51, 52].

Асинхронна обробка: Celery (та фонові задачі). Для фонових задач і планових завдань (ETL, опитування зовнішніх API, перерахунок індексів, побудова агрегатів) застосовується Celery [53]. Він забезпечує горизонтальне масштабування воркерів, черги з пріоритетами, ретраї та ізоляцію збоїв інтеграцій від вебзастосунку. Celery Beat керує періодикою завдань (від “кожні 5 хв” до CRON-графіків).

Redis використовується як брокер повідомлень Celery та високошвидкісний кеш (TTL, часті зрізи даних, результати проміжних обчислень, токени доступу). Робота в пам'яті мінімізує затримки, що критично для онлайн-дашбордів та алертів.

Для пристроїв/станцій застосовується MQTT (наприклад, EMQX/Mosquitto) із темами, QoS і ретейном; це оптимально для IoT і нестабільних каналів. Для зовнішніх сервісів та клієнтських застосунків – REST API поверх HTTPS; для push-оновлень у вебінтерфейсі – WebSocket/Server-Sent Events.

Бази даних та сховище. Система моніторингу якості повітря працює з різними типами даних: від поточкових «сирих» вимірювань у режимі реального часу до агрегованих показників для аналітики й прогнозування. Тому застосовуються спеціалізовані рішення для OLTP (операційної обробки транзакцій) та OLAP (аналітичної обробки).

PostgreSQL використовується як основна OLTP-база даних [54]. У ній зберігаються нормалізовані довідники, дані про сенсори й станції, події вимірів та метадані для подальшої обробки. PostgreSQL забезпечує високу продуктивність транзакційних операцій і гнучкість роботи зі структурованими даними. Додатково можуть застосовуватися розширення для часових рядів (наприклад, TimescaleDB), які дозволяють ефективно працювати з великими потоками телеметрії та швидко будувати агрегати (годинні чи добові середні значення, тренди, пікові навантаження).

Microsoft SQL Server застосовується як OLAP-платформа для багатовимірного аналізу [55]. У сховище (DWH) завантажуються дані, що пройшли ETL-процес (очищення, трансформація, уніфікація). На їх основі створюються зіркоподібні чи сніжинкові схеми, які забезпечують швидкий доступ до великих масивів даних для побудови звітів, прогнозних моделей і KPI. SQL Server надає розвинуті механізми побудови OLAP-кубів, інтеграції з BI-інструментами (Power BI), а також підтримку складних аналітичних запитів.

Таким чином, PostgreSQL забезпечує оперативну роботу з поточковими даними, а Microsoft SQL Server – аналітичну складову з багатовимірним аналізом і OLAP-кубами. Комбінація цих технологій дозволяє досягти балансу між ефективністю поточного зберігання та глибинними аналітичними можливостями.

Vue.js обрано як основу для реалізації клієнтської частини системи [56]. Цей фреймворк надає реактивність і високу швидкодію, що дозволяє відображати дані

у режимі реального часу без перезавантаження сторінки. Завдяки цьому користувацькі інтерфейси залишаються інтерактивними та зручними, навіть при роботі з великими потоками екологічної інформації, відображаючи тренди, KPI та сповіщення у дашбордах.

Nginx використовується як високопродуктивний вебсервер і зворотний проксі [57]. Він відповідає за ефективну доставку статичного контенту, рівномірний розподіл навантаження між бекенд-сервісами та захист внутрішньої інфраструктури від прямого доступу ззовні. Таким чином, Nginx підвищує продуктивність системи та одночасно виконує роль додаткового рівня безпеки.

Переваги асинхронного підходу на основі Celery. Крім продуктивної роботи вебсервера та зручного користувацького інтерфейсу, критично важливою складовою є оптимізація обробки даних у бекенд-частині. Для збору екологічної інформації з зовнішніх API та виконання фонових операцій у системі застосовується асинхронний підхід з використанням фреймворку Celery. Це розподілена черга завдань для Python, яка дозволяє виконувати довготривалі або періодичні операції поза основним потоком програми.

По-перше, Celery мінімізує блокування основного застосунку. Виклики до зовнішніх API виконуються у фоні, завдяки чому головний сервер не очікує відповіді, а може паралельно обробляти інші події (наприклад, потік MQTT-повідомлень). Це знижує час відгуку на критичні події та підвищує загальну пропускну здатність системи.

По-друге, фреймворк забезпечує масштабованість і паралельність. Завдання можуть виконуватись одночасно на кількох воркерах чи вузлах, що дозволяє ефективно збирати дані з багатьох джерел. У разі збільшення обсягів чи частоти опитування достатньо додати нові воркери, і навантаження буде розподілене горизонтально без зміни бізнес-логіки.

По-третє, Celery підвищує надійність системи завдяки механізмам повторних спроб (retry). Якщо зовнішній API тимчасово недоступний або запит завершився з помилкою, завдання виконується повторно через певний інтервал.

Це ізолює основний застосунок від збоїв та гарантує збереження даних. Додатково Celery веде журнал виконання завдань, що спрощує моніторинг.

По-четверте, за допомогою компонента **Celery Beat** реалізується планове виконання задач із заданою періодичністю. Наприклад, дані SaveEcoBot оновлюються кожні 2 хв, а запити до міського API – кожні 10 хвилин. Такий підхід забезпечує безперервність надходження екологічних показників та їх автоматичну обробку.

У підсумку, асинхронна архітектура на основі Celery гарантує високу продуктивність, масштабованість та стабільність системи навіть за умов пікового навантаження. Це дозволяє додавати нові інтеграції чи складні обчислювальні функції без ризику зниження швидкодії основних компонентів.

3.2. Інформаційне забезпечення

Ефективність системи моніторингу атмосферного повітря визначається не лише якістю вимірювань, а й тим, як організовано зберігання, обробку та використання даних. Щодня станції генерують тисячі записів, що містять концентрації забруднювачів. Такі дані мають різну природу (структуровані, напівструктуровані та «сирі») і надходять у різний час, тому потребують спеціально розробленої інформаційної інфраструктури.

Грамотно спроектоване інформаційне забезпечення забезпечує повний цикл роботи з даними: від оперативного накопичення у транзакційних базах (OLTP) до завантаження у сховища (OLAP) для подальшого аналізу, прогнозування та підтримки прийняття рішень. У цьому контексті критично важливим є не тільки вибір технологій, але й правильна організація DataFlow – процесів вилучення, трансформації та завантаження даних.

Таким чином, підсистема інформаційного забезпечення виконує роль «ядра» системи моніторингу: вона зберігає історію вимірювань, дозволяє будувати багатовимірну аналітику, забезпечує масштабування при зростанні обсягів та інтеграцію з інструментами візуалізації. У наступних підрозділах буде розглянуто

ключові компоненти цієї підсистеми: бази даних, прогнозування обсягів зберігання, сховища даних, DataFlow та програмні засоби.

3.2.1. База даних. Ключовим етапом у побудові систем моніторингу атмосферного повітря є систематичне зберігання, обробка та виконання складних запитів на великих об'ємах даних. Для цих завдань у запропонованій архітектурі використовується система керування базами даних PostgreSQL.

Цей вибір обумовлений низкою переваг. PostgreSQL підтримує розширені аналітичні функції, складні запити та багаторівневий аналіз, що дозволяє створювати гнучкі звіти й отримувати точну статистику з екологічних даних.

Окремою перевагою є розширене індексування, зокрема підтримка просторових індексів через розширення PostGIS. Це забезпечує швидкий пошук і обробку даних, пов'язаних із географічними координатами станцій. PostgreSQL дотримується принципів ACID, що гарантує цілісність даних навіть у випадку високих навантажень чи технічних збоїв. Логічна модель бази даних представлена на рис. 3.3.

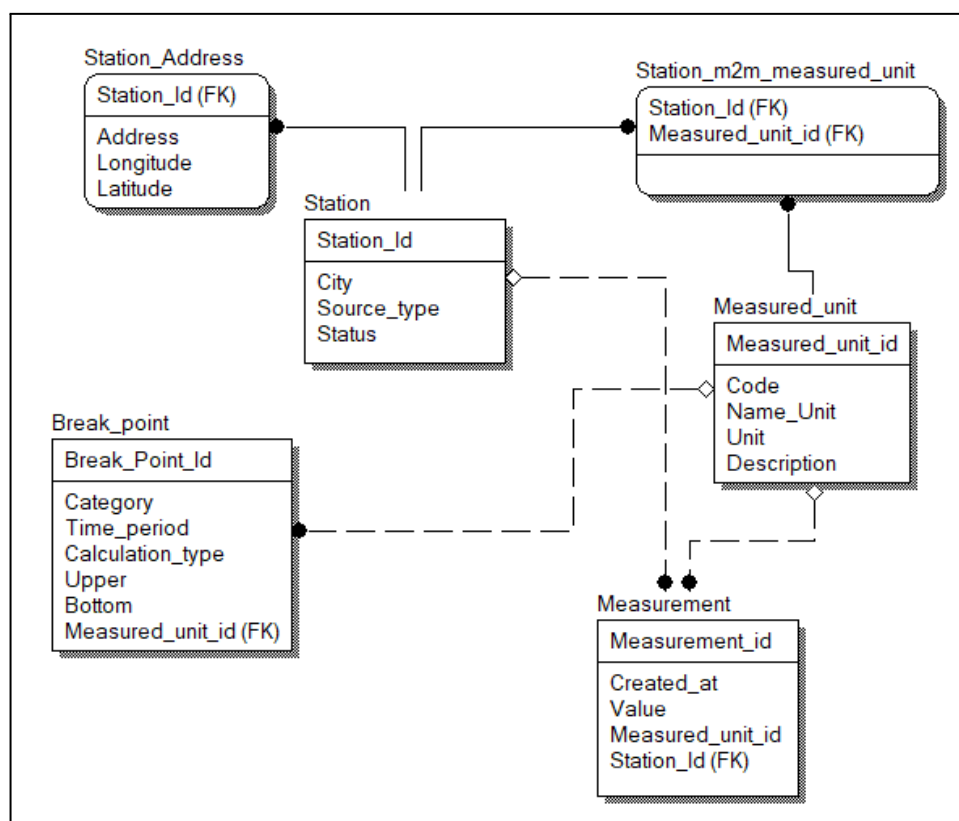


Рис. 3.3. Логічна модель бази даних

Ключовим компонентом системи моніторингу якості повітря є набір взаємопов'язаних таблиць бази даних, які забезпечують зберігання структурованих даних про станції моніторингу, параметри вимірювань, одиниці вимірювання, результати моніторингу та межі забруднення. Ця архітектура даних дозволяє ефективно організувати процеси збору, аналізу та візуалізації екологічних даних.

Центральним елементом є таблиця ***station***, яка містить основну інформацію про всі станції моніторингу:

- *station_id* – унікальний ідентифікатор станції;
- *source_type* – тип джерела даних (КМДА, SaveEcoBot, MQTT);
- *status* – стан роботи (активна / неактивна).

Деталізація адресної інформації зберігається у таблиці ***station_address***, що включає:

- параметри адреси (*street, city, region, postal_code*);
- координати (*latitude, longitude*);
- зв'язок з конкретною станцією (*station_id*).

Для стандартизації вимірювань використовується таблиця ***measured_unit***, яка описує одиниці вимірювання:

- *code* – унікальний код;
- *name* – текстова назва;
- *unit* – скорочене позначення одиниці;
- *description* – додаткова інформація.

Зв'язок між станціями та одиницями вимірювання задається через таблицю ***station_m2m_measured_unit***, що дозволяє гнучко визначати, які параметри відстежує конкретна станція.

Основні екологічні вимірювання накопичуються у таблиці ***measurement***, яка зберігає «сирі» дані:

- *station_id* – станція-відправник;
- *measured_unit_id* – тип параметра;
- *value* – числове значення;

- *created_at* – час отримання даних.

Для класифікації рівнів забруднення використовується таблиця *break_point*, яка містить порогові значення:

- *measured_unit_id* – тип забруднювача;
- *category* – рівень забруднення;
- *upper* і *bottom* – діапазони значень;
- *calculation_type* – тип індексу (AQI, EAQI, CAQI);
- *time_period* – часовий інтервал (годинний / денний).

Завдяки цим таблицям система може не лише зберігати дані у сирому вигляді, але й оперативно класифікувати їх відповідно до екологічних стандартів, будувати індекси якості повітря та генерувати попередження про перевищення нормативів.

Окрім предметних таблиць, що описують екологічні параметри, у базі даних реалізуються й системні таблиці, призначені для забезпечення надійності та контролю за роботою системи. До них належать модулі логування змін, таблиці аудиту доступу та журналів операцій, які фіксують усі критичні дії з даними. Такі структури дозволяють відстежувати історію модифікацій, відновлювати дані у разі збоїв і забезпечувати прозорість роботи системи. Проте в межах цього дослідження основна увага приділяється саме екологічним даним, які становлять ядро інформаційного забезпечення.

Така архітектура бази даних забезпечує гнучкість, масштабованість і точність у роботі з даними. Вона дозволяє інтегрувати нові станції, додавати додаткові параметри вимірювання та розширювати аналітичні можливості системи без кардинальних змін у структурі БД. PostgreSQL у поєднанні з розширенням PostGIS і можливістю інтеграції з OLAP-рівнем утворює надійну основу для всієї інформаційної підсистеми моніторингу.

3.2.2. Прогнозування обсягів зберігання даних. Для ефективної роботи системи моніторингу якості повітря необхідно оцінити, який обсяг даних накопичується під час її роботи, та передбачити методи оптимізації зберігання. Обсяг даних залежить від кількості параметрів, що вимірюються на кожній

станції, і частоти цих вимірювань. Нижче наведено оцінку обсягу зібраних даних та рекомендації щодо скорочення обсягу зберігання без втрати важливої інформації.

В типовій системі відстеження якості повітря станція може збирати дані приблизно по 24 різних параметрах. Основними є близько 10 ключових показників якості повітря (наприклад, концентрації забруднювальних речовин $PM_{2.5}$, PM_{10} , NO_2 тощо), а решта – допоміжні фактори такі як температура, атмосферний тиск та вологість. Додатково система розраховує агреговані індекси якості повітря, що додає ще пару похідних показників. Для спрощення розрахунків далі припустимо, що кожна станція моніторить 10 основних параметрів.

Частота збору даних у системі становить фіксований інтервал, наприклад кожні 10 хвилин, що відповідає 144 вимірюванням на добу для кожного параметра. Таким чином, кожна станція генерує близько 1440 записів на добу (10 параметрів $\times$ 144 вимірювань). Така висока деталізація (дані з інтервалом 10 хвилин) забезпечує точний моніторинг, але призводить до швидкого накопичення великих обсягів даних.

Кожен окремий вимір зберігається в базі даних як окремий запис (рядок таблиці) із фіксованою структурою полів. Основні поля вимірювання та їхній приблизний розмір у базі даних (PostgreSQL) виглядають так:

- *id* – ідентифікатор запису (тип *bigint*, 8 байт)
- *created_at* – мітка часу створення (timestamp with time zone, 8 байт)
- *value* – значення виміряного параметра (double precision, 8 байт)
- *measured_unit_id* – ідентифікатор типу параметра (bigint, 8 байт)
- *station_id* – ідентифікатор станції (bigint, 8 байт)

Сумарний розмір цих полів становить 40 байт даних на один вимір. До цього додаються службові накладні витрати бази даних на зберігання запису (заголовки рядка, вирівнювання тощо) – орієнтовно ще 24 байти на запис. В результаті кожен вимір займає близько 64 байт у сховищі. Це значення можна використати для оцінки загального обсягу даних за певний період.

Виходячи з наведених параметрів (10 вимірювань на станцію, 144 вимірювання кожного на добу, ~64 байт на запис), можна спрогнозувати, як зростатиме обсяг даних у системі при різній кількості станцій та за різні проміжки часу. У таблиці 3.3 наведено приблизну кількість записів та загальний обсяг даних для 1, 10, 100 і 1000 станцій за день, тиждень, місяць і рік роботи системи.

Таблиця 3.3

Прогнозовані обсяги зберігання даних при різній кількості станцій

Кількість станцій	Кількість вимірів за день / Обсяг	Кількість вимірів за тиждень / Обсяг	Кількість вимірів за місяць / Обсяг	Кількість вимірів за рік / Обсяг
1	1440 / 0.05 MB	10080 / 0.38 MB	43200 / 1.65 MB	525600 / 20.05 MB
10	14400 / 0.55 MB	100800 / 3.85 MB	432000 / 16.48 MB	5256000 / 200.50 MB
100	144000 / 5.49 MB	1008000 / 38.45 MB	4320000 / 164.79 MB	52560000 / 1.958 GB
1000	1440000 / 54.93 MB	10080000 / 384.52 MB	43200000 / 1.6093 GB	525600000 / 19.5801 GB

Як видно з таблиці, обсяг даних, що накопичується, стрімко зростає зі збільшенням кількості станцій. Для однієї станції обсяг даних за рік становить приблизно 20 MB, тоді як для 1000 станцій – вже близько 19.6 GB на рік. Така тенденція демонструє, що при розширенні мережі моніторингу або збільшенні тривалості зберігання даних навантаження на базу даних зростатиме пропорційно. Це може призвести до погіршення продуктивності системи.

З часом постійне накопичення детальних даних створює значне навантаження на систему зберігання. Це вимагає застосування спеціальних методів оптимізації, щоб система могла стабільно працювати з великим обсягом історичних даних [44]. Нижче перелічено ключові підходи для зменшення обсягу даних, що зберігаються, без втрати критично важливої інформації.

Методи оптимізації зберігання даних. Важливою складовою стратегії роботи з великими обсягами інформації є оптимізація процесів зберігання.

Ключовим підходом тут виступає баланс між точністю вимірювань і ефективністю використання ресурсів. У перші роки експлуатації система може утримувати як «сирі» дані з високою частотою вимірювань, так і їхні агрегати на рівні години. Це дозволяє проводити глибокий аналіз та перевірку алгоритмів. Водночас у довгостроковій перспективі детальні дані поступово видаляються, а в сховищі залишаються лише денні чи тижневі агрегати. Такий підхід забезпечує компроміс між обсягом накопиченої інформації та потребою в довготривалому аналізі.

Зберігання агрегованих значень дає змогу суттєво зменшити обсяг бази даних без втрати ключових тенденцій. Годинні, денні або тижневі усереднені показники забезпечують необхідний рівень деталізації для виявлення трендів та формування прогнозів. Завдяки цьому довгострокові архіви залишаються компактними, але водночас достатньо інформативними для досліджень та управлінських рішень.

Додаткову ефективність забезпечує поділ функцій між OLTP і OLAP-рівнями. PostgreSQL із партиціюванням таблиць і політиками ретенції використовується для оперативного зберігання «сирих» вимірювань та короткострокового аналізу. Для довгострокових завдань дані завантажуються в OLAP-сховище, де вони представлені у вигляді агрегатів. Це дає змогу будувати багатовимірні куби, досліджувати динаміку показників, формувати звіти та моделі прогнозування.

Таким чином, прогнозування та контроль обсягів даних є критичними факторами для забезпечення стабільності й масштабованості системи моніторингу. Використання OLTP-рівня для поточних операцій у поєднанні з OLAP-рівнем для довгострокового аналізу дозволяє підтримувати як точність у короткостроковій перспективі, так і ефективність при зберіганні історичних даних. Це створює основу для стійкого розвитку системи та її адаптації до зростаючих навантажень.

3.2.3. Сховище даних. У багатовимірному аналізі даних ключовим поняттям є OLAP-куб [58]. Його можна уявити як багатовимірну таблицю, де інформацію можна розглядати не лише у двох вимірах, як у звичайній базі даних, а відразу у

кількох. Це дозволяє аналізувати дані під різними кутами та комбінувати різні ознаки, що особливо важливо для систем моніторингу якості повітря.

Вимір (Dimension) – це аналітична ознака, за якою відбувається розбиття даних. Наприклад, у випадку моніторингу повітря вимірами можуть бути час (день, місяць, рік), місце (місто, район, конкретна станція) або тип забруднювача ($PM_{2.5}$, NO_2 , O_3 тощо). Такі виміри можна деталізувати або, навпаки, узагальнювати: від окремих годин до днів, місяців чи років.

Показник (Measure) – це числове значення, яке відображає сам факт вимірювання. У нашому випадку це може бути концентрація певної речовини, розрахований індекс AQI чи кількість випадків перевищення допустимих норм. Ці значення можуть агрегуватися: наприклад, можна рахувати середні, мінімальні чи максимальні показники за певний період.

OLAP-структури: куби, виміри, показники. В основі багатовимірного аналізу лежить поняття OLAP-куба – логічної моделі даних, що дає змогу одночасно оперувати декількома вимірами для пошуку прихованих закономірностей. Кожен вимір в кубі відповідає певній ознаці, а комірки куба містять числові значення-показники. OLAP-куб можна уявити як гіперпростір: наприклад, у системі моніторингу якості повітря типовий куб «Якість повітря» має виміри Час, Локація та Забруднювач, а кожна комірка містить факт (показник) – середньодобову концентрацію вибраного забруднювача, розрахований індекс AQI/CAQI, максимальне значення за добу чи кількість годин з перевищенням гранично допустимої концентрації (ГДК).

OLAP-системи підтримують ряд базових аналітичних операцій для дослідження кубів даних. Забезпечують деталізацію або агрегування даних по ієрархії виміру (від року до місяця чи дня і навпаки). Такі інструменти дають змогу швидко отримувати відповіді на типові запити аналітики, наприклад:

- показати середньомісячний рівень $PM_{2.5}$ по різних локаціях за останні 5 років;
- визначити дні тижня з піковими концентраціями NO_2 ;
- порівняти середні значення індексів якості повітря у різних сезонах.

Високу швидкодію виконання таких запитів забезпечує попереднє агрегування даних в сховищі та використання багатовимірних індексів. Іншими словами, для OLAP-кубів заздалегідь розраховуються підсумкові показники по популярних комбінаціях вимірів. Часто застосовується збереження популярних зрізів – підмножин даних, що найчастіше використовуються в запитах, у вигляді матеріалізованих подань. Завдяки цьому значно зменшується час обробки повторюваних аналітичних запитів.

Схеми зберігання даних: зіркоподібна та сніжинкова моделі [59, 60]. Для реалізації OLAP-аналітики у сховищі даних важливо вибрати ефективну модель організації таблиць. OLAP-дані зазвичай зберігаються у реляційному сховищі даних за допомогою зіркоподібної або сніжинкової схеми.

Зіркоподібна (star) схема передбачає наявність центральної таблиці фактів, що містить записи із числовими показниками, та набору пов'язаних з нею таблиць вимірів. Запити в схемі «зірка» потребують мінімальної кількості приєднань (JOIN) – факти з'єднуються напряму з кожною таблицею вимірів, – що прискорює виконання аналітичних операцій. На рис. 3.4 показано реалізацію зіркоподібної моделі для вітрини даних вимірюваних концентрацій (*Measurements*).

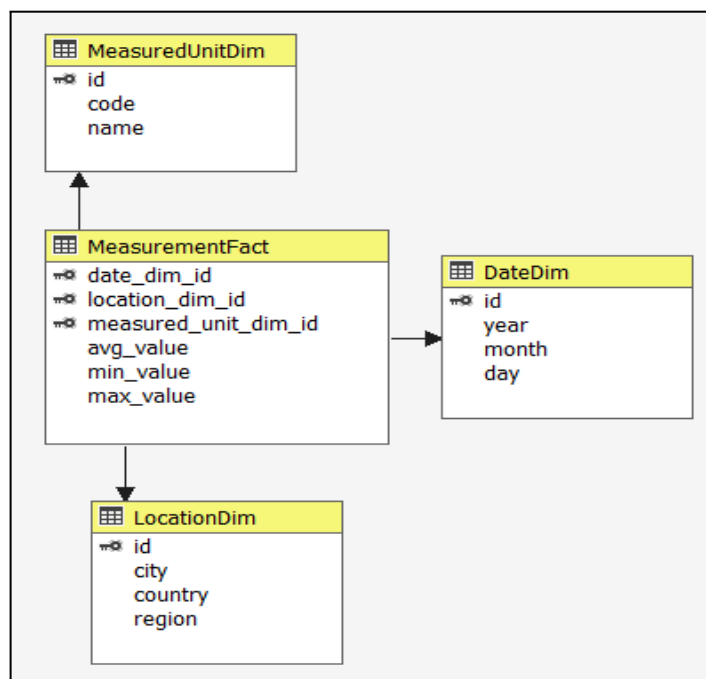


Рис. 3.4. Вітрина даних: *Measurements*

Центральною є фактова таблиця *MeasurementFact*, яка зберігає обчислені статистичні показники: середнє значення (*avg_value*), мінімальне (*min_value*) та максимальне (*max_value*) значення концентрації за певний інтервал. Ця фактова таблиця пов'язана з трьома таблицями вимірів:

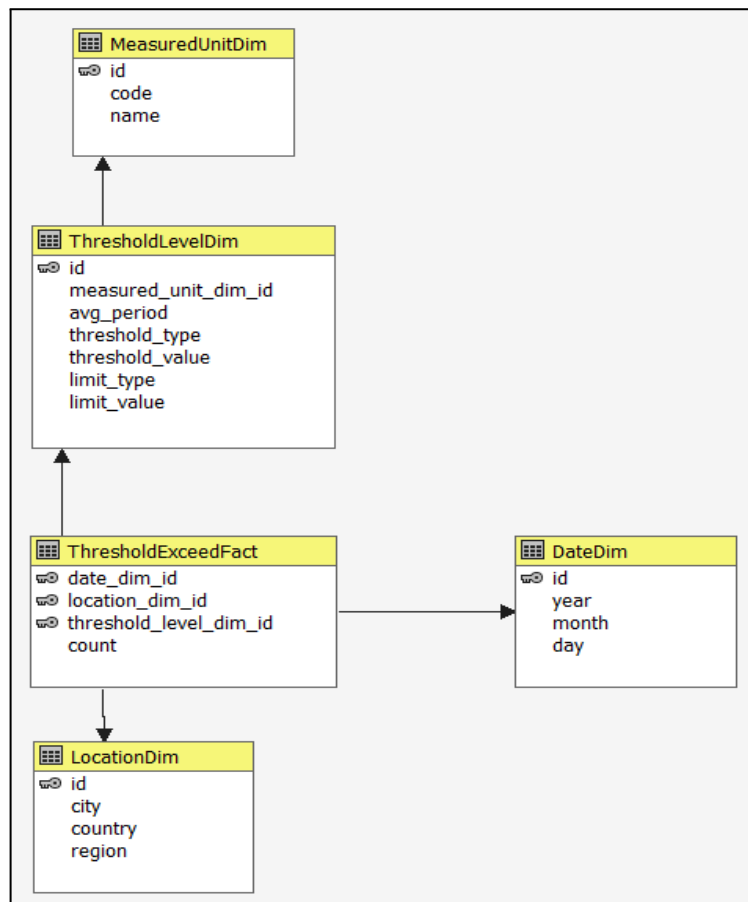
- *DateDim* – календарні атрибути (рік, місяць, день тощо);
- *LocationDim* – географічні дані (місто, регіон, країна);
- *MeasuredUnitDim* – характеристика забруднювача (код параметра, назва, одиниці виміру).

Завдяки такій структурі з'являється можливість гнучко виконувати OLAP-запити по даних якості повітря. Наприклад, система може швидко обчислити середню концентрацію обраного забруднювача за певний період і в заданій локації, а потім дозволити користувачу провести деталізацію від річного показника до помісячного чи до окремих днів. Агреговані значення зберігаються у факт-таблиці, тому отримання результатів здійснюється без складних обчислень “на льоту”.

Поряд із вимірюваннями концентрацій, у екологічному аналізі важливо відстежувати перевищення нормативів шкідливих речовин. Для цього створено окрему вітрину даних *ThresholdExceed*, що базується на фактовій таблиці *ThresholdExceedFact* (рис. 3.5).

Ця таблиця містить записи про випадки перевищення гранично допустимих концентрацій і пов'язана з такими таблицями вимірів:

- *DateDim* – дата (період спостереження);
- *LocationDim* – місце спостереження;
- *ThresholdLevelDim* – параметри екологічних нормативів: тип забруднювача, середній інтервал вимірювання (наприклад, 1 год, 24 год, рік), тип порогу (верхній допустимий чи цільовий рівень), значення порогу та допустима кількість перевищень.

Рис. 3.5. Вітрина даних: *ThresholdExceed*

Така модель дещо відрізняється від класичної «зірки» і ближча до сніжинкової схеми. Зокрема, таблиця *ThresholdLevelDim* опосередковано пов'язана з *MeasuredUnitDim*, тобто деякі атрибути вимірів винесено в окрему довідкову таблицю. Таким чином досягається: уникнення дублювання інформації про забруднювачі в записах порогів, якщо один і той самий параметр має кілька нормативних рівнів для різних періодів. Подібний підхід відповідає принципам схеми «сніжинки», де таблиці розмірів розділені на підпорядковані довідники задля усунення надлишковості даних.

Отже, запропонована модель сховища даних є гібридною, оскільки поєднує переваги сніжинкової та зіркоподібної схем. У результаті вдається збалансувати продуктивність виконання аналітичних запитів із гнучкістю моделі, уникаючи зайвого ускладнення структури без потреби [61]. Таким чином, багатовимірна OLAP-модель на основі кубів і правильно спроектованої схеми зберігання забезпечує ефективну підтримку ухвалення рішень в екологічному моніторингу.

3.2.4. *Потік даних (Data Flow)*. У системах моніторингу атмосферного повітря дані надходять із багатьох джерел і проходять послідовний шлях від збору до аналітичної обробки. Для цього застосовується підхід ETL/ELT (Extract-Transform-Load / Extract-Load-Transform) [62], який подано на рисунку 3.6.

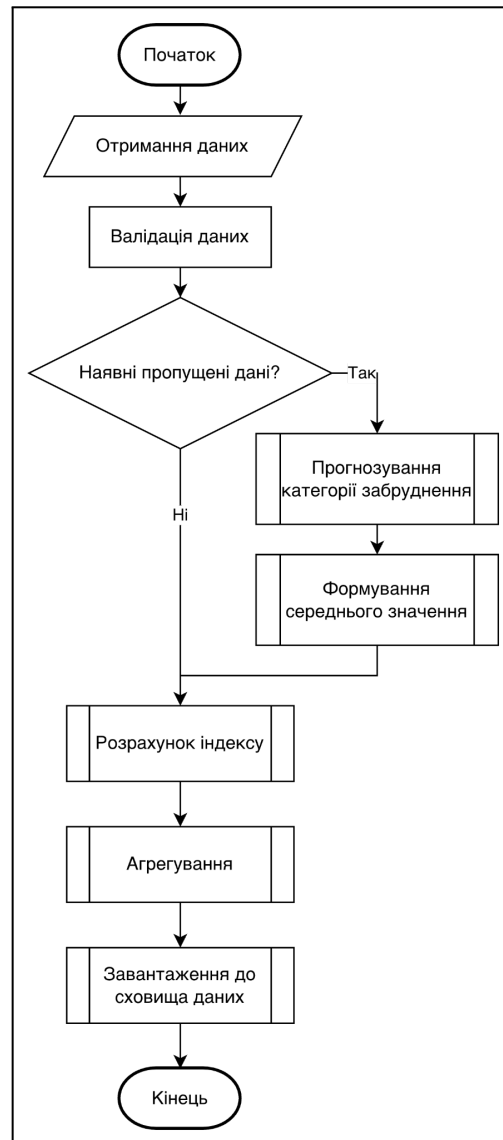


Рис. 3.6. Потік даних (Data Flow)

Етап Extract (вилучення). На першому етапі дані збираються з різномірних джерел. Основним каналом є MQTT-брокери, через які у реальному часі надходить телеметрія від сенсорів. Додатково використовуються API зовнішніх сервісів(наприклад, SaveEcoBot чи КМДА), що дозволяє збагачувати потоки вимірювань референтними значеннями та архівними наборами. До вторинних

джерел належать також системні логи та допоміжні дані, що забезпечують контроль якості та виявлення збоїв у роботі станцій.

Етап Transform (трансформація). На другому етапі здійснюється перетворення даних для інтеграції у спільну аналітичну модель. Етап містить кілька послідовних підпроцесів.

Первинна перевірка якості даних. На цьому кроці відбувається видалення дублікатів, аномальних або фізично неможливих значень, а також приведення форматів дат, часу та координат до уніфікованого стандарту. Це дозволяє забезпечити коректність і сумісність усіх подальших етапів обробки. Таким чином формується єдиний формат даних, придатний для подальшої аналітики.

Перевірка повноти даних. Визначається наявність усіх ключових забруднювачів у межах заданого періоду спостереження. Якщо часовий інтервал має недостатню кількість значень або відсутні окремі параметри, активується механізм інтелектуального заповнення пропусків.

Заповнення пропусків за допомогою методу “Наївного Байєса”. У випадках, коли частина вимірів відсутня, застосовується ймовірнісна модель, що прогнозує категорію якості повітря на основі наявних параметрів. Для кожного відсутнього забруднювача система визначає клас, після чого відповідно до матриці діапазонів обирає характерне (середнє) значення концентрації для цього класу. Таким чином, відновлені значення дозволяють підвищити повноту даних і забезпечити розрахунок індексів якості повітря з більшою точністю та стабільністю.

Обчислення агрегованих показників. Після заповнення даних та розрахунку індексів якості повітря здійснюється формування середніх, мінімальних і максимальних концентрацій за визначені часові інтервали.

Таким чином, система не лише очищує й узгоджує дані, а й підвищує їхню інформаційну повноту завдяки застосуванню інтелектуальних методів відновлення, що забезпечує коректність і надійність результатів подальшого аналізу.

Етап Load (завантаження). Фінальний етап передбачає завантаження агрегованих даних до сховища даних, організованого за моделями «зірки» або

«сніжинки». Це створює основу для багатовимірного аналізу, формування звітів і прогнозування екологічних тенденцій.

Ключову роль у керуванні ETL/ELT процесами відіграє система розподіленої обробки задач Celery. Вона забезпечує:

- асинхронність усіх етапів обробки;
- розподіл завдань між воркерами для підвищення масштабованості;
- планове виконання регулярних задач (через Celery Beat), таких як опитування API кожні 10 хвилин чи щогодинна агрегація даних.

Такий підхід дозволяє систематизувати процеси обробки інформації, забезпечити узгодженість і повноту даних, а також зробити систему придатною для масштабування та подальшого багатовимірного аналізу.

Таким чином, сформований конвеєр обробки даних гарантує безперервність функціонування системи навіть за пікових навантажень. Завдяки поєднанню ETL-підходу з інтелектуальними методами, такими як "Наївний Байєс", дані у сховищі залишаються узгодженими, повними та придатними для глибокої аналітики й прогнозування стану атмосферного повітря.

3.3. Програмне забезпечення

Програмне забезпечення є ключовим компонентом інтегрованої системи моніторингу атмосферного повітря, оскільки воно поєднує інформаційну базу, алгоритмічні методи та інтерфейси взаємодії з користувачем у єдине функціональне середовище. Саме за рахунок програмної реалізації стає можливим забезпечення безперервного збору даних, їх зберігання, обробки, візуалізації та подальшого використання для прогнозування й підтримки управлінських рішень.

Особливість цієї підсистеми полягає у високій складності обробки даних, які надходять у великих обсягах і різних форматах. Програмна частина виконує роль інтегратора: з одного боку, вона забезпечує стабільність роботи з сенсорними потоками даних, а з іншого – дозволяє застосовувати сучасні методи аналітики, регресії, кластеризації та побудови індексів якості повітря. При цьому важливим є

не лише зберігання та аналіз, а й оптимізація процесів через автоматизацію ETL процедур, індексацію та агрегацію даних.

Розроблене програмне забезпечення побудоване модульно: кожен компонент системи відповідає за окремий функціональний блок – від накопичення вимірювань та індексації до реалізації алгоритмів Data Mining і формування візуалізованих звітів. Такий підхід забезпечує масштабованість, розширюваність та гнучкість у використанні, що є критично важливим для систем екологічного моніторингу.

3.3.1. Модуль накопичення даних. Модуль накопичення даних відповідає за збір, обробку та агрегування інформації з мережі датчиків якості повітря. Архітектурно він побудований за принципом конвеєра (pipeline), де дані проходять послідовні етапи: від збору з різних джерел до попередньої обробки та збереження. Кожне джерело даних обслуговується окремим сервісом, що наслідує базовий клас *CollectServiceBase*. Цей базовий клас визначає уніфікований інтерфейс (наприклад, методи на кшталт *process_station*) і забезпечує спільну функціональність для всіх сервісів збору даних. Таким чином, кожен спеціалізований сервіс (як-от *KMDAService* та інші) реалізує власну логіку отримання даних, наслідуючи загальну структуру, задану в *CollectServiceBase*. Це сприяє модульності та розширюваності системи: додавання нового джерела даних зводиться до реалізації нового сервісу за аналогією до існуючих.

У загальному пайплайні збору даних передбачено, що дані з численних станцій збираються паралельно. Для цього використовується окремий компонент – *ProcessingService*, який керує багатопотоковою обробкою. Під час чергового циклу збору *ProcessingService* формує перелік завдань для кожної станції (або групи станцій) та виконує їх одночасно у декількох потоках. Такий підхід дозволяє суттєво підвищити пропускну здатність системи за рахунок паралельного виконання мережевих операцій вводу-виводу. Як зазначається в літературі, потоки є ефективним засобом паралелізації у випадках, коли основним вузьким місцем є очікування I/O-операцій [63]. Іншими словами, поки один потік очікує відповіді від датчика або вебсервісу, інші потоки можуть обробляти дані з інших джерел. Це

забезпечує кращу продуктивність і своєчасне оновлення інформації зі всіх станцій.

Після паралельного збору даних на рівні *ProcessingService* результати від кожного сервісу (наприклад, від *KMDAService* та інших подібних сервісів) агрегуються і передаються далі – у сховище даних або на наступні етапи аналізу. Таким чином, *ProcessingService* і спеціалізовані сервіси збору утворюють єдиний конвеєр, де *ProcessingService* відповідає за асинхронне виконання задач, а сервіси на зразок *KMDAService* – за безпосереднє отримання та форматування даних. Нижче докладніше розглянуто реалізацію цих двох ключових компонентів.

Багатопотокова обробка даних. *ProcessingService* є компонентом, що координує одночасну обробку даних у кількох потоках (рис. 3.7). Його завдання – розподілити множину операцій збору/обробки між потоками, забезпечуючи ефективне використання ресурсів та швидку реакцію системи.

```
class ProcessingService: new *
    def __init__(self): new *
        self.services = {}

    def add(self, key: int, service: CollectServiceBase): new *
        self.services[key] = service

    def _run_process_station(self, station: Station): 1 usage new *
        service = self.services[station.source_type]
        for measured_unit, dataset in service.process_station(station):
            to_create: list = []
            for _time, _value in dataset:
                to_create.append(
                    Measurement(
                        station=station, measured_unit=measured_unit,
                        created_at=_time, value=_value,
                    )
                )
            Measurement.objects.bulk_create(
                to_create, batch_size=len(to_create), ignore_conflicts=True
            )

    @check_performance_time 2 usages (2 dynamic) new *
    def run_for_type(self, source_type: int):
        with ThreadPoolExecutor(max_workers=8) as executor:
            stations = Station.objects.filter(source_type=source_type, status=Station.ACTIVE)
            for station in stations:
                executor.submit(self._run_process_station, station)
```

Рис. 3.7. Багатопотокова обробка даних з різних станцій

У мові Python для реалізації пулу потоків використовується клас *ThreadPoolExecutor* із модуля *concurrent.futures* [64]. Цей інструмент надає високорівневий інтерфейс для виконання викликів у фонових потоках, не блокуючи основний потік програми. *ProcessingService* створює фіксований пул робочих потоків і ставить у чергу завдання – обробку даних з окремих станцій – які виконуються цими потоками паралельно.

Використання *ThreadPoolExecutor* робить реалізацію компактною та надійною: він абстрагує низькорівневі деталі роботи з потоками і дозволяє зосередитися на логіці обробки. Варто відзначити, що кількість робочих потоків (*max_workers*) підбирається відповідно до ресурсів системи та характеру задач. У випадку мережевого вводу-виводу надто великий пул не завадить (адже потоки переважно чекатимуть на відповіді), але оптимальне значення запобігає зайвому переключенню контексту. Завдяки багатопотоковості *ProcessingService* значно прискорює накопичення даних з багатьох джерел одночасно, що є критичним для системи моніторингу в реальному часі.

Обробка даних станцій міського моніторингу. *KMDAService* – це спеціалізований сервіс збору даних, призначений для роботи з джерелом даних Київської міської державної адміністрації (КМДА). Іншими словами, цей компонент відповідає за отримання інформації зі станцій моніторингу якості повітря, що належать до міської мережі. *KMDAService*, як і інші подібні сервіси, успадковує базовий клас *CollectServiceBase*, тому його структура аналогічна: основним методом є *process_station(station)*, який реалізує логіку збору та обробки даних для однієї конкретної станції.

Метод *process_station* у *KMDAService* виконує такі кроки: (1) встановлення з'єднання з джерелом даних (API) для заданої станції, (2) отримання сирих даних (наприклад, у форматі JSON) про поточні показники забруднюючих речовин, (3) парсинг (розбір) отриманих даних та виділення ключових показників, (4) формування структури даних у внутрішньому форматі системи і (опційно) збереження результату у базі даних. Приклад реалізації див. в додатку А.

Дані сенсорів отримуються послідовними запитами за різними інтервалами. Результати конвертуються у формат (datetime, value), після чого для кожної валідної одиниці формується пара: об'єкт *MeasuredUnit* та часовий ряд значень. У підсумку метод повертає список таких пар, придатних для подальшого збереження, аналітики чи візуалізації.

Використання декоратора *@catch_exceptions* забезпечує стійкість до помилок мережевих запитів, а наслідування від *CollectServiceBase* гарантує модульність та можливість розширення системи іншими сервісами збору даних.

Аналогічним чином реалізовані й інші сервіси збору даних у системі. Кожен з них наслідує *CollectServiceBase* переважно відрізняється лише джерелом даних та специфікою обробки. Наприклад, гіпотетичний сервіс *EcoCityService* або інший API-сервіс міг би мати свій власний метод *process_station*, що звертається до відповідного API і парсить дані за власним форматом. Завдяки поліморфізму, *ProcessingService* може викликати метод *process_station* будь-якого з цих сервісів, не турбуючись про деталі реалізації. Це узгоджується з архітектурною логікою: кожен сервіс інкапсулює роботу зі своїм джерелом, а координуючий компонент забезпечує паралельність і синхронізацію виконання.

Збір даних через MQTT. Для приймання телеметричних даних від IoT-станцій можна використати окремий сервіс, що працює за публікаційно-підписною моделлю MQTT.

Загальна схема роботи. MQTT-клієнт (окремий процес/сервіс) підключається до MQTT-брокера та підписується на топіки станцій (наприклад, всі повідомлення від датчиків станцій). Після успішного з'єднання відбувається підписка на потрібні топіки [65]. Брокер MQTT доставляє повідомлення сервісу. Сервіс в колбек-функції *on_message* виконує легку валідацію та нормалізацію даних. Зокрема, декодується *payload* (наприклад, JSON-рядок) та перевіряються ключові поля (чи всі необхідні сенсори надіслали дані, чи дані в допустимих межах тощо). Цей

крок може також включати нормалізацію одиниць виміру або формату даних.

Замість того, щоб одразу записувати дані у базу в межах *on_message*, повідомлення ставиться у чергу для фонової обробки. Це запобігає блокуванню основного потоку MQTT-клієнта довготривалими операціями та підвищує надійність. Найчастіше для цього застосовують брокер завдань на зразок Celery. Наприклад, у колбеці викликається Celery-завдання для обробки отриманого повідомлення. Такий підхід дозволяє швидко делегувати опрацювання даних і негайно продовжити приймати наступні повідомлення. Один або кілька воркерів (процесів) асинхронно витягують повідомлення з черги та виконують запис у базу даних. Воркер може накопичувати кілька записів і виконувати пакетну вставку в БД для підвищення продуктивності.

Як показано на рисунку 3.8, MQTT-клієнт приймає дані від брокера, виконує первинну обробку та передає їх у чергу Celery для асинхронного запису у базу даних.

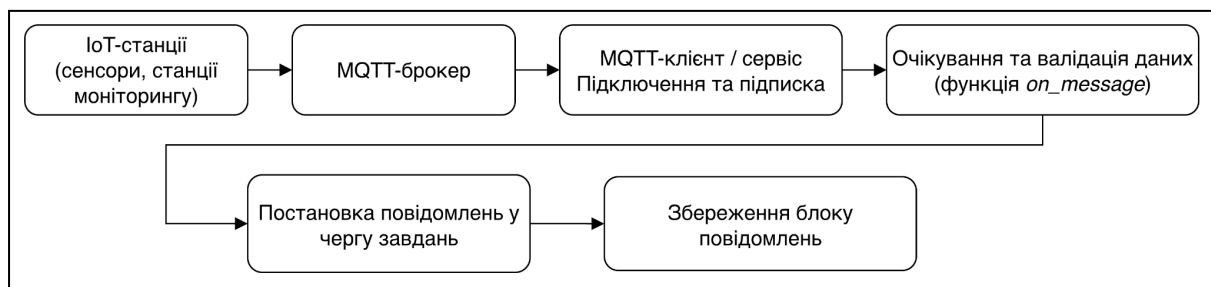


Рис. 3.8. Принцип роботи сервісу MQTT-накопичення даних

Після успішного з'єднання викликається *on_connect*, де здійснюється підписка. Колбек *on_message* спрацьовує при надходженні будь-якого повідомлення на ці топіки. В ньому виконується декодування байтового *payload* і розбір JSON. Далі код витягує ідентифікатор станції з назви топіка (*station_id*)

цього назву забруднювача (*queue*) та здійснює просту перевірку. Підготовлений запис містить всі необхідні дані для збереження.

Найважливіший крок - постановка даних в чергу. Викликається *process_mqtt_sensor_data_task*, що ставить завдання на виконання у чергу Celery. Таким чином, основний MQTT-сервіс швидко повертається до очікування наступних повідомлень, не виконуючи важкої роботи синхронно. Це важливо для високонавантажених сценаріїв, щоб жодне повідомлення не було втрачено через повільну обробку.

Celery-завдання отримує на вхід підготовлений словник *record* (переданий з MQTT-сервісу) і здійснює запис у базу даних. Celery-завдання отримує на вхід підготовлений словник *record* (переданий з MQTT-сервісу) і здійснює запис у базу даних. Запис проводиться всередині транзакції для надійності. При високому навантаженні замість поодинокого створення записів можна використовувати пакетне вставлення (*bulk_create*) для групи записів – для цього воркер може накопичувати записи у пам'яті або застосувати механізм Celery Batches [66]. Зокрема, Celery дозволяє визначити задачу із базовим класом Batches, щоб автоматично буферизувати кілька викликів і передати їх у вигляді списку повідомлень у функцію-завдання. Це оптимізує масовий запис (зменшує кількість транзакцій).

3.3.2. Модуль розрахунку індексів якості повітря. Модуль *AqiOwnCalculator* відповідає за обчислення індексу якості повітря (Air Quality Index, AQI) на основі даних моніторингу атмосфери. Він реалізує методику Агентства з охорони довкілля США для розрахунку AQI, за якою концентрації кількох основних забрудників повітря. Для кожного забрудника модуль обчислює окремий індекс на основі його середньої концентрації, після чого визначає максимальне значення серед цих під-індексів та приймає його як загальний AQI для станції моніторингу. Таким чином, інтегральний показник AQI завжди визначається «найгіршим» показником якості повітря серед наявних забрудників, що відповідає прийнятій практиці інформування про домінуючий вплив забруднення. Реалізацію сервісу *AqiOwnCalculator* див. в додатку Б.

Роль модуля розрахунку AQI в системі полягає в наданні користувачам зрозумілого інтегрованого показника якості повітря в реальному часі. Це дає змогу узагальнити інформацію про рівні різних забрудників в одному числовому індексі та відповідній категорії, що спрощує оцінку стану повітря і потенційного ризику для здоров'я без необхідності аналізувати кілька окремих вимірів. Відповідно до встановлених шкал AQI (наприклад, «добра», «помірна», «нездорова» якість повітря тощо) система може генерувати попередження або рекомендації для населення при перевищенні певних порогових значень. Гнучка архітектура рішення також дозволяє підтримувати альтернативні стандарти індексів якості повітря: окрім реалізації американського AQI, у системі впроваджені аналогічні розрахункові служби для європейських індексів, таких як Common Air Quality Index (CAQI) та оновлений European Air Quality Index (EAQI). Це забезпечує адаптивність платформи до різних нормативних вимог та регіональних стандартів моніторингу якості повітря.

3.3.3. Автоматизація обчислень і ETL процесів. На завершальному етапі реалізовано сервіс для автоматизації розрахунків та перенесення даних до сховища даних. Цей компонент виконує роль ETL процесу (Extract, Transform, Load), який щоденно агрегує результати вимірювань та власних розрахованих індексів і завантажує їх у Data Warehouse. Автоматизація цих обчислень дозволяє регулярно оновлювати аналітичне сховище без втручання людини. Розрахунки (наприклад, середньодобові концентрації, максимальні/мінімальні значення та кількість перевищень встановлених порогів) виконуються на основі даних основної бази даних, після чого результати автоматично переносяться до сховища даних за розкладом (наприклад, раз на добу опівночі).

Реалізацію сервісу *WareHouseExporterService*, який забезпечує автоматизацію щоденних обчислень і виконання ETL-процесів із перенесення даних у аналітичне сховище, наведено у додатку В.

Як видно з коду, спершу здійснюється підготовка та отримання необхідних ідентифікаторів з вимірювальних таблиць сховища, після чого формуються списки записів для вставки у фактні таблиці. Вставка проводиться пакетно для

підвищення продуктивності (методи *cursor.executemany* дозволяють виконати багато вставок за один запит до бази даних). Важливо, що сервіс працює з двома підключеннями до баз даних: оперативної та сховища (маніпуляції через *connections['warehouse']*) [67]. Таким чином, *WareHouseExporterService* виступає мостом між оперативною базою даних і сховищем даних, виконуючи трансформацію та завантаження даних у автоматичному режимі.

Ілюстрація внутрішнього потоку даних з використанням сервісів системи зображена на рисунку 3.9.

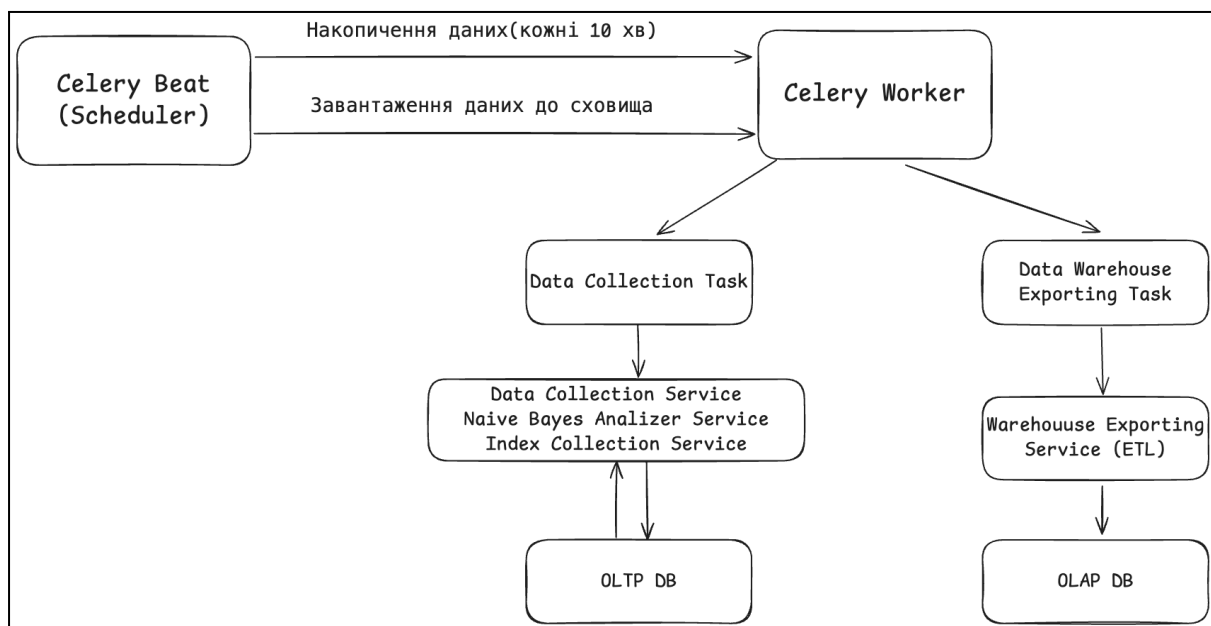


Рис. 3.9. Ілюстрація потоку даних з використанням сервісів системи

На схемі показано, як сервіс збору даних станцій наповнює основну базу даних, після чого сервіс обчислення індексів розраховує значення та зберігає результати назад в основну БД. Warehouse Exporter Service (ETL) запускається за розкладом та зчитує агреговані добові дані з основної БД. Далі він завантажує ці дані до сховища даних для довгострокового зберігання та подальшого аналізу.

Така архітектура забезпечує повний цикл обробки даних: від збору сирих вимірювань із сенсорних мереж до формування багатовимірної аналітичної звітності. Ключовою її перевагою є висока ступінь автоматизації – всі основні процеси (збір, обчислення індексів, агрегація та експорт у сховище) виконуються

за розкладом без потреби у ручному втручанні. Завдяки використанню черг завдань і паралельної обробки система здатна масштабуватися під зростаючі обсяги даних і працювати в режимі близькому до реального часу. У підсумку формується єдина керована інфраструктура, яка не лише накопичує дані, але й забезпечує їх якісну трансформацію, інтеграцію та аналітичну доступність для прийняття управлінських рішень у сфері моніторингу довкілля.

3.3.4. Регресійний аналіз даних забруднення повітря. Для аналізу взаємозв'язків між показниками забруднення повітря було реалізовано окремий модуль регресійного аналізу. Цей модуль дозволяє гнучко вибрати вхідні дані (період спостережень, перелік забруднювачів) через спеціальний інтерфейс користувача. На його основі будується модель залежності одного показника (залежної змінної) від інших (незалежних змінних) із використанням різних методів регресії. Реалізацію модуля *RegressionAnalysisService*, який забезпечує проведення регресійного аналізу даних забруднення повітря з використанням методу найменших квадратів (OLS), наведено у додатку Г.

У поточній реалізації продемонстровано використання класичного методу найменших квадратів (OLS, Ordinary Least Squares), що дозволяє кількісно оцінювати взаємозв'язки між незалежними змінними (наприклад, концентраціями окремих поллютантів) та залежною змінною (зокрема, інтегральним індексом AQI). Такий підхід є базовим інструментом у статистиці та широко застосовується для аналізу часових рядів у сфері екологічного моніторингу.

Архітектура модуля побудована за принципом розширюваності: базовий клас інкапсулює спільну логіку підготовки даних (фільтрація, об'єднання, нормалізація), тоді як підкласи реалізують конкретні типи регресійних моделей. Це дозволяє легко інтегрувати додаткові алгоритми – від поліноміальної та гребеневої регресії до сучасних методів машинного навчання. Такий підхід створює основу для проведення порівняльного аналізу та вибору оптимальної стратегії прогнозування залежно від досліджуваної задачі.

Модуль регресійного аналізу не лише забезпечує базову функціональність прогнозування, але й створює платформу для подальшого розширення й інтеграції складніших моделей у рамках системи моніторингу якості атмосферного повітря.

3.3.5. Класифікація та кластеризація даних. У цьому підрозділі реалізовано поєднання методів класифікації (керованого навчання) та кластеризації (некерованого навчання) для глибокого аналізу підготовлених даних. Зокрема, використовується алгоритм “Gaussian Naive Bayes” для передбачення класу якості повітря за добу, а також алгоритм кластеризації k-середніх (“K-Means”) для виявлення природних груп у даних. Такий підхід дозволяє, з одного боку, навчити модель розпізнавати категорії якості повітря на основі показників забруднення, а з іншого – знайти приховані структури в даних без опори на заздалегідь відомі мітки класів.

Перед навчанням класифікатора здійснюється формування ознак та міток класів. Вихідні виміри (середнє, мінімум, максимум концентрацій забруднювачів) за кожен день і локацію перетворюються у формат “широкої” таблиці: кожен рядок відповідає певній даті (та місту), а стовпці містять значення показників якості повітря ($PM_{2.5}$, PM_{10} , NO_2 , SO_2 , вологість, швидкість вітру тощо). Ключовим цільовим показником є індекс якості повітря CAQI(Common Air Quality Index), на основі середнього значення якого визначається якісний клас дня. Для цього використовується заздалегідь задана шкала бінів та назв класів. Ці категорії відповідають загальноприйнятим рівням індексу якості повітря. Після визначення класу для кожного запису даних формується набір ознак (всі інші показники окрім CAQI) та міток класів, готовий для навчання моделі. Реалізацію модуля *NaiveBayesAnalyzer* наведено у додатку Д.

Для навчання класифікатора використовується алгоритм “Наївний Байєс” у реалізації *GaussianNB* з бібліотеки *scikit-learn* [68, 69]. Цей алгоритм передбачає незалежність ознак між собою та моделює їх розподіл у кожному класі, що спрощує обчислення на основі теореми Байєса. У випадку *GaussianNB* припускається, що значення ознак підпорядковані нормальному (гаусовому) розподілу всередині кожного класу. Модель навчається на тренувальній

підмножині даних (X_{train} , y_{train}) і потім перевіряється на тестовій вибірці (X_{test}). Виведення *classification_report* (див. код вище) показує метрики якості класифікації – точність, повноту та F-міру для кожного класу. Це дозволяє оцінити, наскільки добре алгоритм “Наївний Байєс” навчився розрізняти категорії якості повітря. За потреби, також можуть будуватися додаткові візуалізації для аналізу моделі – наприклад, матриця невідповідностей (*confusion matrix*) для виявлення, які класи найчастіше плутаються між собою, або графіки розподілу значень ознак (*boxplot*, *violinplot*) по класах, щоб зрозуміти внесок різних показників у визначення класу.

Кластеризація даних. Окрім класифікації, над підготовленими ознаками виконано кластерний аналіз, метою якого є виявлення станцій та забруднювачів із потенційними відхиленнями або проблемами у роботі системи моніторингу. Такий підхід дозволяє ідентифікувати групи станцій, що демонструють схожі закономірності вимірювань, та виділити ті, у яких спостерігаються систематичні аномалії чи нестабільна динаміка концентрацій.

Для цього застосовано алгоритм “K-Means” (K-середніх) [70], який розбиває вибірку на наперед задану кількість кластерів k таким чином, щоб об’єкти в межах одного кластера були максимально подібні, а міжкластерна відмінність якомога більшою.

Перед виконанням кластеризації всі числові ознаки були стандартизовані (нормалізовані), що забезпечує рівноправний вплив показників із різними одиницями вимірювання та масштабами.

Для визначення оптимальної кількості кластерів використано метод ліктя (Elbow Method), який базується на аналізі зміни середньоквадратичного відхилення всередині кластерів (Within-Cluster Sum of Squares, WCSS) залежно від k . На графіку «методу ліктя» вибирається така кількість кластерів, після якої подальше збільшення k не призводить до суттєвого зниження внутрішньокластерної дисперсії.

У результаті кожній станції та кожному показнику присвоєно кластерну належність, що дозволило оцінити їхню подібність за профілем забруднень.

Подальший аналіз середніх значень і складу кластерів показав, які станції або окремі забруднювачі відрізняються від типового фону – це вказує на можливі збої сенсорів, локальні джерела викидів або особливості мікроклімату.

Таким чином, кластеризація дозволяє не лише групувати дані за схожістю, а й виступає інструментом виявлення потенційно проблемних точок моніторингу, що є важливим етапом у забезпеченні якості та достовірності екологічних спостережень.

Висновки до розділу 3

Побудована архітектура системи передбачає три взаємодіючі підсистеми:

- підсистема обробки оперативних джерел (СУБД OLTP);
- підсистема збереження інформації (СУБД OLTP та сховище даних);
- підсистема аналізу даних (оперативний аналіз OLAP та інтелектуальний аналіз Data Mining).

Така архітектура забезпечує автоматизацію процесів спостереження, аналізу, прогнозування та підтримки прийняття управлінських рішень.

Важливу роль в системах моніторингу відіграє інформаційне забезпечення, яке представлено реляційною базою даних (для збереження оперативної інформації) та багатовимірним гіперкубом (для накопичення даних, що використовуються для аналізу). Такий підхід дозволив вирішити проблеми управління інформацією, як на етапі транзакційної обробки даних, так і на етапах оперативного та інтелектуального аналізу інформації.

Розроблене програмне забезпечення реалізує такий функціонал:

- збір та збереження даних;
- розрахунок індексів якості повітря;
- реалізація ETL-процесу;
- проведення регресійного аналізу;
- класифікація та кластеризація даних.

РОЗДІЛ 4

РЕАЛІЗАЦІЯ ТА ВПРОВАДЖЕННЯ СИСТЕМИ

4.1. Розгортання системи

Після розробки архітектури, моделей даних та алгоритмів обробки постає завдання практичної реалізації системи моніторингу атмосферного повітря. Цей етап має принципове значення, оскільки саме він демонструє, що запропоновані підходи можуть бути втілені у реальному середовищі та забезпечувати безперервну роботу з великими обсягами даних.

Реалізація охоплює комплекс заходів – від розгортання інфраструктури та налаштування програмного забезпечення до створення користувацького інтерфейсу й організації процесів тестування. Особлива увага приділяється автоматизації збору й обробки даних, інтеграції з базами та сховищами, а також засобам контролю якості та продуктивності.

Як основу інфраструктури обрано хмарні сервіси DigitalOcean та Microsoft Azure [71, 72]. Це надає змогу користуватися керованими рішеннями для баз даних, автоматичними оновленнями, резервним копіюванням і масштабуванням без необхідності самостійно підтримувати інфраструктуру баз даних.

Розгортання системи моніторингу якості атмосферного повітря здійснювалося з урахуванням вимог до продуктивності, масштабованості та доступності даних. У результаті була сформована багаторівнева інфраструктура, яка поєднує застосунок обробки даних, оперативну базу даних та сховище для аналітики. Використання хмарної інфраструктури також спрощує адміністрування та робить систему більш стійкою до непередбачуваних обставин.

Архітектура передбачає використання окремих серверів, кожен з яких виконує власну роль у загальному процесі. Основний сервер призначений для розгортання основного застосунку, створеного на базі Django. Для асинхронної обробки задач використовується система черг Celery, що дає змогу паралельно виконувати обчислення та обробку потоків даних. Веб застосунок запускається за

допомогою зв'язки Gunicorn/Nginx, що гарантує стабільну продуктивність та рівномірний розподіл навантаження.

Оперативна база даних **PostgreSQL** працює як керований кластер DigitalOcean Managed PostgreSQL, що дозволяє делегувати обслуговування (оновлення, реплікацію, бекапи) провайдеру. Цей підхід знижує адміністративне навантаження та збільшує надійність системи (кластер має резервні вузли, автоматичне переключення).

Аналітична підсистема реалізована у вигляді керованого сервісу **Azure SQL Database (SQL Server)**, де зберігаються агреговані дані та виконуються OLAP-запити й Data Mining. Відокремлення аналітичного сховища від операційної бази дозволяє знижувати навантаження на транзакційну систему в режимі реального часу та забезпечує стабільність виконання складних розрахунків і багатовимірного аналізу.

Розгортання сервісів. Вебзастосунок запускається за допомогою Gunicorn, який виконує роль WSGI-сервера, а Nginx забезпечує балансування та обробку статичних файлів. Обчислювальні завдання (збір даних, розрахунок індексів, агрегація) реалізовані у вигляді асинхронних процесів на базі Celery. Воркери Celery працюють у паралельному режимі, що дозволяє одночасно обробляти потоки даних від багатьох станцій. Це критично важливо для систем, що функціонують у режимі реального часу.

Забезпечення безперервної роботи. Управління усіма компонентами (Gunicorn, Celery, Beat) здійснюється за допомогою *Supervisor* [73]. Кожен процес контролюється окремим модулем Supervisor, що дозволяє автоматично перезапускати сервіси у випадку збоїв (рис. 4.1). Додатково застосовується *systemd* для інтеграції із сервісною моделлю операційної системи. Серверні журнали (логи) зберігаються у файловій системі, підлягають ротації та резервному копіюванню. Це дає змогу проводити аудит, відстежувати помилки та аналізувати ефективність роботи системи.

```
[group:eco_city]
programs=eco_city_web,eco_city_worker

[program:eco_city_web]
command=/home/dmytro/eco_city/.env/bin/gunicorn core.wsgi:application --bind 0.0.0.0:8000
directory=/home/dmytro/eco_city
user=dmytro
autostart=true
autorestart=true
stdout_logfile=/var/log/supervisor/eco_city_web.log

[program:eco_city_worker]
command=/home/dmytro/eco_city/.env/bin/celery -A core worker -B --pool solo -l INFO
directory=/home/dmytro/eco_city
user=dmytro
autostart=true
autorestart=true
stdout_logfile=/var/log/supervisor/eco_city_worker.log
```

Рис. 4.1. Конфігурація Supervisor

Таким чином, розгортання системи на базі хмарної інфраструктури DigitalOcean із використанням керованої транзакційної бази даних PostgreSQL та винесенням аналітичного сховища в Azure SQL Database (SQL Server) забезпечило високу надійність, масштабованість і стійкість до збоїв. Поділ на транзакційну та аналітичну підсистеми дав змогу оптимізувати обробку потоків даних у режимі реального часу, водночас гарантуючи можливість виконання складних розрахунків, OLAP-запитів та багатовимірного аналізу.

4.2. Реалізація інтерфейсу користувача

Інтерфейс користувача системи реалізовано у вигляді вебзастосунку, що забезпечує доступ до всіх функціональних можливостей незалежно від операційної системи чи типу пристрою. Такий підхід дозволяє користувачам працювати із системою як з робочих комп'ютерів, так і з мобільних пристроїв, використовуючи лише веббраузер. Цей метод усуває необхідність встановлення окремого програмного забезпечення та спрощує оновлення системи, одночасно підтримуючи актуальність інформації для користувачів.

Принципи дизайну інтерфейсу. При розробці інтерфейсу пріоритетом були простота, зрозумілість і зручність використання. Користувачі мають доступ лише до найбільш релевантних інструментів на кожному етапі роботи, що знижує когнітивне навантаження та сприяє швидкому опануванню системи навіть непрофесіоналами. Структура інтерфейсу побудована за принципом «від загального до детального»: на початковому екрані відображається узагальнена картина стану атмосферного повітря, а при необхідності користувач може перейти до деталізації показників окремих станцій чи параметрів.

Користувач бачить тільки те, що необхідно для виконання поточного завдання. Система об'єднує моніторинг, аналіз та управління в одному середовищі, забезпечуючи єдиний та узгоджений доступ до даних і функцій. Елементи інтерфейсу розподілені на зрозумілі блоки та етапи роботи, що допомагає користувачу не загубитися в широкому наборі функціоналу.

Інтерактивність та візуалізація даних. Інтерактивні компоненти інтерфейсу значно підвищують ефективність аналізу інформації. Реалізовано динамічні графіки, гнучкі фільтри та інші інструменти візуалізації, що відображають просторові та часові закономірності в даних. Це дозволяє користувачам не лише переглядати вже накопичені результати, а й у реальному часі змінювати параметри відображення (наприклад, часовий діапазон, вибір станцій або параметрів), формуючи власні сценарії аналізу. Такий підхід робить взаємодію з системою динамічною та гнучкою, орієнтованою на конкретні потреби кожного користувача.

Адаптивність інтерфейсу. Інтерфейс адаптовано під різні типи пристроїв – від стаціонарних комп'ютерів та ноутбуків до планшетів. Використання адаптивної верстки та сучасних CSS-фреймворків забезпечує коректне відображення та зручність роботи на екранах будь-якого розміру. Завдяки цьому користувачі можуть отримувати доступ до системи незалежно від місця свого перебування та оперативно стежити за даними в будь-який момент, що особливо важливо для моніторингу в режимі реального часу.

Технологічна реалізація. Інтерфейс користувача побудовано на сучасному стеку вебтехнологій, що забезпечує високу продуктивність і масштабованість. На серверному рівні використано Django (DRF) які відповідають за бізнес-логіку, обробку та збереження даних, а також за автентифікацію користувачів. DRF формує RESTful API, через яке клієнтська частина отримує уніфікований доступ до вимірювань, індексів та агрегованих показників. На клієнтському рівні застосовано Vue.js, що дає змогу реалізувати реактивний і модульний інтерфейс.

Компонентний підхід спрощує повторне використання елементів (графіків, карт, таблиць) і забезпечує швидке оновлення відображення даних у режимі реального часу. Архітектурне розділення клієнтської та серверної частин сприяє масштабованості системи та полегшує інтеграцію з іншими сервісами, зокрема з платформами бізнес-аналітики, такими як Power BI, для формування більш глибоких аналітичних звітів.

4.2.1. Авторизація та доступ до системи. Система моніторингу підтримує як відкритий доступ для зовнішніх користувачів, так і авторизований режим для внутрішніх співробітників. Неавторизовані користувачі мають змогу переглядати загальнодоступні модулі – зокрема, базову інформацію про стан атмосферного повітря, інтерактивні графіки та узагальнені карти. Це забезпечує прозорість та публічність результатів екологічного моніторингу.

Для розширеного доступу передбачено модальне вікно авторизації (рис. 4.2), яке викликається при спробі виконати дії, доступні лише внутрішнім користувачам.

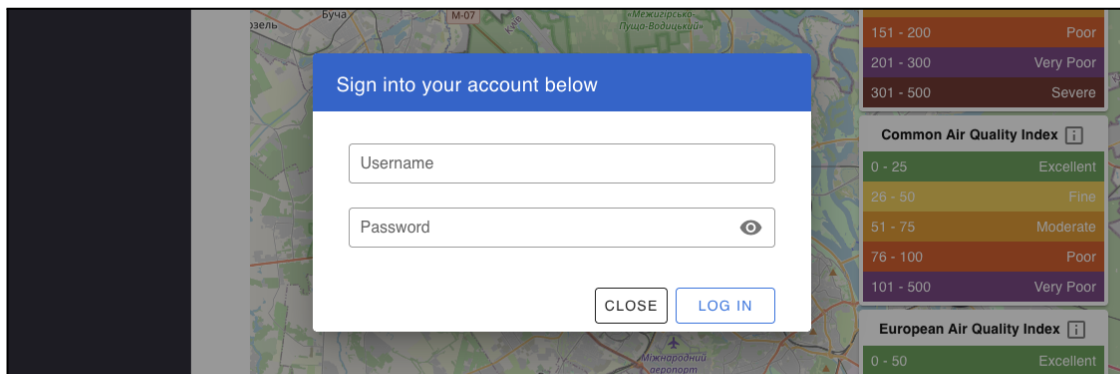


Рис. 4.2. Інтерфейс авторизації користувача

У цьому вікні користувач вводить ім'я облікового запису (username) та пароль. Після успішної автентифікації відкривається розширений інтерфейс, що включає додаткові модулі наприклад, аналітику в реальному часі, інструменти адміністрування станцій та перегляд історії даних.

Така модель поєднує доступність для широкої громадськості з одночасним забезпеченням контролю та безпеки для службових користувачів, дозволяючи розмежовувати права залежно від ролі та рівня відповідальності.

4.2.2. *Головна сторінка системи (дашборд).* Головним елементом інтерфейсу є дашборд (рис. 4.3), який забезпечує швидкий доступ до основної інформації про стан атмосферного повітря. Структура сторінки побудована так, щоб користувач одразу бачив узагальнену картину екологічної ситуації та міг за потреби перейти до детального аналізу.

У центральній частині дашборда розміщена інтерактивна карта з позначенням моніторингових станцій. Кожна станція відображається з урахуванням її географічного розташування та поточного рівня забруднення, що дозволяє швидко оцінити просторовий розподіл показників якості повітря.

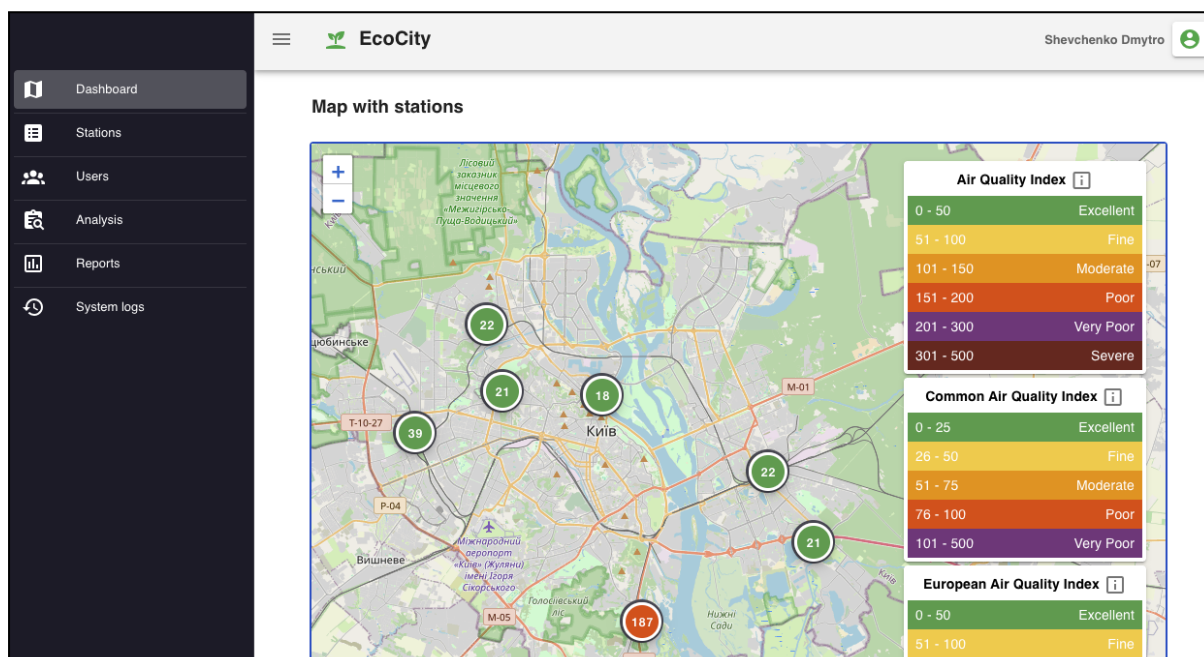


Рис. 4.3. Головна сторінка системи (дашборд) з картою станцій моніторингу

Зліва розташоване навігаційне меню, яке дає змогу переходити між різними модулями системи. Праворуч від карти знаходяться інформаційні картки (рис. 4.4), які відображають індекси якості повітря та категорії забруднення.

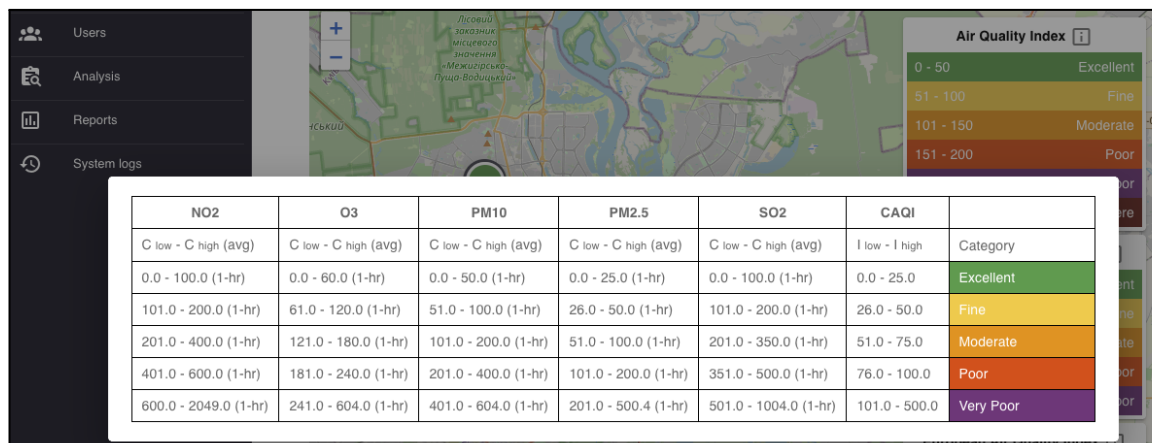


Рис. 4.4. Інформаційні картки індексів та категорій забруднення

Картки мають кольорове кодування, що підвищує наочність і дозволяє оперативно визначати рівень небезпеки для населення. Меню можна згорнути, що забезпечує зручність роботи та оптимальне використання робочої області екрана.

Після вибору конкретної станції на карті користувач отримує доступ до розширеної інформації про неї (рис. 4.5). У вікні відображаються основні характеристики: розташування, адреса, номер ресурсу та перелік вимірюваних параметрів. Ключовим елементом цього модуля є інтегрований аналітичний блок, що дозволяє переглядати історичні дані по станції у графічному вигляді. Користувач може обирати часовий інтервал (день, тиждень, місяць, рік) та метод агрегації (погодинний чи денний). Завдяки цьому забезпечується швидкий доступ до динаміки показників, що дає змогу проводити базовий аналіз безпосередньо у межах системи.



Рис. 4.5. Детальна сторінка станції з історичними даними та графіками

Інтерактивні графіки та фільтри роблять модуль зручним для практичного використання як фахівцями-екологами, так і представниками органів управління. Це дозволяє отримати наочну картину стану атмосферного повітря на конкретній станції та вчасно виявляти потенційні проблеми.

Головна сторінка системи виконує роль інтегрованого інформаційного центру, де поєднано оглядові та аналітичні інструменти. Завдяки поєднанню карти моніторингових станцій, інформаційних карток та можливості деталізованого перегляду даних у розрізі окремих станцій, користувач отримує як загальне уявлення про екологічну ситуацію, так і доступ до поверхневого аналізу. Це забезпечує гнучкість у роботі з даними, сприяє оперативному виявленню

проблемних зон і створює підґрунтя для ухвалення обґрунтованих управлінських рішень.

4.2.3. Управління станціями. Модуль управління станціями призначений для адміністрування та контролю всіх моніторингових точок, підключених до системи (рис. 4.6).

List of stations

Source type
All

Status
All

Sort by
Favorite

+ ADD

Number	City	Address	Source type	Measured units	Status	Actions
016	Kyiv	Kharkivske shose, 7/1	KMDA	20	Active	
017	Kyiv	Architect Verbytsky Street, 26	KMDA	20	Active	
018	Kyiv	Prospect Pravdy, 64G	KMDA	20	Active	
019	Kyiv	Vulytsia Schyseva, 20	KMDA	20	Active	
020	Kyiv	Turivska Street, 28	KMDA	20	Active	
021	Kyiv	Prospect Peremohy, 97	KMDA	20	Active	
022	Kyiv	Kyivska Street, 22	KMDA	20	Active	
001	Dnipro	Haharina Avenue, 43	SaveEcoBot	6	Inactive	
002	Kyiv	Mykhaila Lomonosova Street, 73	SaveEcoBot	6	Inactive	

Рис. 4.6. Список моніторингових станцій із фільтрацією та пошуком

Він надає можливість переглядати перелік станцій, керувати їхніми параметрами та забезпечує простий доступ до інструментів створення й редагування. У базовому представленні відображається список станцій, де користувач може бачити їхній статус, перелік доступних вимірів та іншу службову інформацію. Для зручності роботи передбачені:

- фільтри (за ресурсом або статусом),
- сортування та пошук,
- функція «закріплення улюблених» станцій нагорі списку.

Для додавання нової станції використовується модальне вікно створення (рис. 4.7). Користувач вносить базові параметри, після чого система автоматично присвоює станції унікальний ідентифікатор. Після завершення налаштувань станція починає накопичення даних у базі.

Рис. 4.7. Модальне вікно створення нової станції

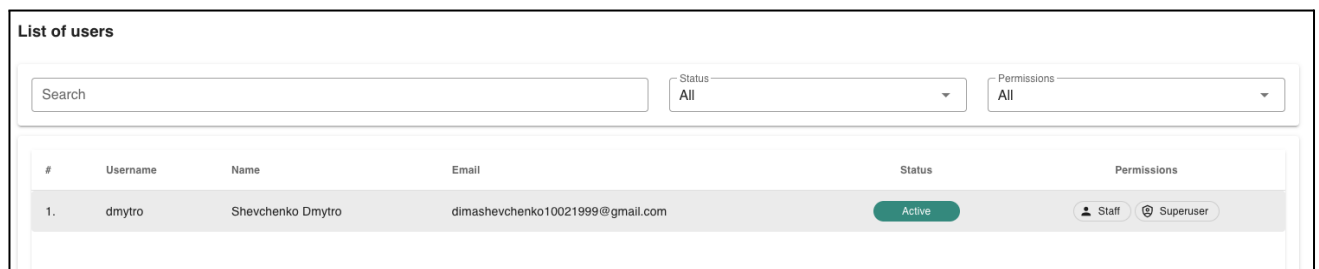
Редагування наявних станцій реалізовано через спеціальну сторінку редагування (рис. 4.8). Тут можна змінювати статус станції, її географічне розташування, перелік вимірюваних параметрів та додаткові характеристики, необхідні для інтеграції з іншими системами (наприклад, ідентифікатор станції у сторонньому сервісі). Передбачено також можливість видалення станцій.

Рис. 4.8. Сторінка редагування параметрів станції

Модуль забезпечує централізований контроль над усіма моніторинговими точками, підключеними до системи. Він об'єднує базові адміністративні функції – перегляд списку станцій, їх пошук, фільтрацію та сортування – з можливістю

створення, редагування та видалення об'єктів. Завдяки цьому забезпечується гнучкість у налаштуванні системи та підтримка її актуальності.

4.2.4. Перегляд користувачів. У системі реалізовано можливість перегляду списку користувачів, що мають доступ до функціоналу (рис. 4.9). Основне управління обліковими записами виконується через адміністративну панель, однак для зручності у вебінтерфейсі передбачено окремий модуль «Користувачі», де відображаються ключові дані про акаунти.



The screenshot shows a web interface titled "List of users". It includes a search bar, a "Status" dropdown menu set to "All", and a "Permissions" dropdown menu set to "All". Below these filters is a table with the following data:

#	Username	Name	Email	Status	Permissions
1.	dmytro	Shevchenko Dmytro	dimashevchenko10021999@gmail.com	Active	Staff Superuser

Рис. 4.9. Список користувачів системи

У списку користувачів показуються базові атрибути: ім'я, логін, роль у системі та статус. Для швидкої роботи реалізовано інструменти фільтрації (за статусом або типом доступу), а також пошук за ключовими параметрами. Це дає змогу оперативно орієнтуватися серед користувачів навіть у масштабних інсталяціях системи.

Модуль дозволяє отримати загальне уявлення про структуру доступів у системі, зберігаючи при цьому централізоване управління через адмін-панель. Тому у разі потреби оновлення паролю, можна звернутися на пошту адміністратора з необхідним запитом.

4.2.5. Аналітичний модуль для регресійного аналізу. У системі реалізовано аналітичний модуль, що забезпечує виконання регресійного аналізу даних із використанням статистичної бібліотеки Statsmodels. В основі покладено метод Ordinary Least Squares (OLS), який є класичним інструментом для оцінки залежностей між змінними. Завдяки цьому модуль дозволяє досліджувати

взаємозв'язки між різними параметрами якості атмосферного повітря та здійснювати кількісну оцінку впливу факторів.

Користувач має можливість гнучко налаштовувати параметри аналізу (рис. 4.10). Зокрема, доступні:

- вибір часових інтервалів (доба, тиждень, місяць, рік та довільний);
- вибір однієї або кількох станцій;
- визначення набору показників забруднювачів та метеорологічних факторів як незалежних змінних.

The screenshot shows a web-based configuration form titled 'Analysis'. It contains several input fields and dropdown menus for setting up a regression analysis. The 'Stations' field is set to '016: Kyiv, Kharkivske shose, 7/1(KMDA) - Active'. The 'Regression type' is 'Ordinary Least Squares'. The 'Period from' is '2024-08-10' and 'Period to' is empty. 'Average type' is 'Hours' and 'Time period' is 'All Day(24 hours)'. The 'Dependent measurement' is 'Common Air Quality Index(Own)'. The 'Independent measurements' field lists 'PM10, PM2.5, Nitrogen dioxide(NO2), Wind speed(WS), Sulfur dioxide(SO2), Radiation'. An 'ANALYZE' button is located at the bottom right of the form.

Рис. 4.10. Налаштування параметрів регресійного аналізу

Подібна конфігурація робить модуль універсальним інструментом, що дозволяє як локальний аналіз (для окремої станції), так і регіональні порівняння.

Результатом роботи модуля є формування набору ключових метрик моделі (рис. 4.11). До них належать:

- частка варіації залежної змінної, поясненої незалежними;
- F-статистика для оцінки загальної значущості моделі;
- р-значення, що підтверджує статистичну значущість оцінок;
- таблиця коефіцієнтів регресії (β) із вказаними стандартними похибками та t-статистиками.

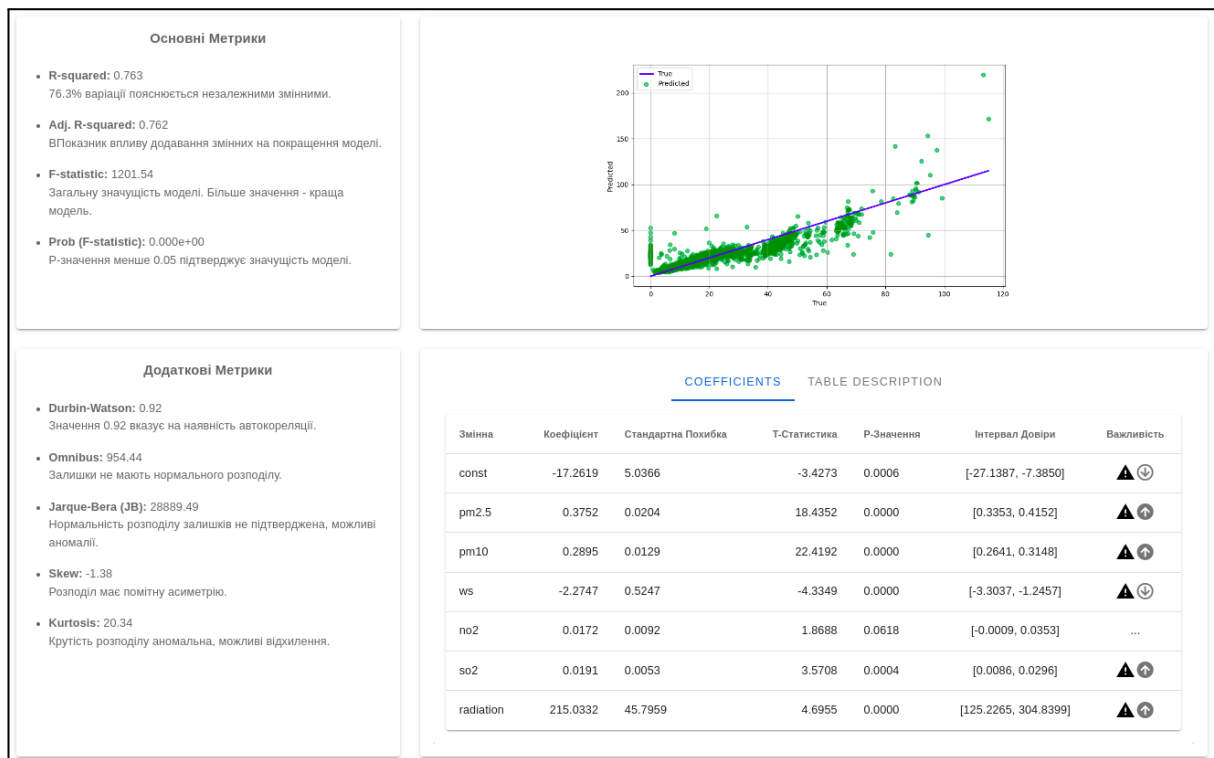


Рис. 4.11. Формат результатів регресійного аналізу

Додатково модуль виконує аналіз залишків, що дозволяє оцінити адекватність моделі та виявити відхилення чи аномалії у даних. Окремо важливою є візуалізація результатів. Користувач може переглядати графік прогнозування, який відображає співвідношення між фактичними та передбаченими значеннями змінної. Це дозволяє візуально оцінити точність моделі та її придатність для практичного прогнозування якості повітря.

Цей аналітичний інструментарій може бути використаний для глибшого розуміння залежностей у даних екологічного моніторингу та може бути розширений іншими регресійними методами, включно з сучасними алгоритмами машинного навчання.

4.2.6. Логування в системі. Важливою частиною забезпечення надійності та прозорості роботи є модуль логування. Усі зміни в системі фіксуються у спеціалізованій таблиці бази даних *History*, що дозволяє відстежувати дії користувачів та автоматичні події на рівні застосунку (рис. 4.12).

System logs

Change types
Field changed, Status changed, Added, Deleted

Period from Period to

#	Model name	Change type	Message	User	Date
85.	station	Added	"#25 - European Air Quality Index(Own)" was added to measured_units	System	2025-01-12 09:01:50
84.	station	Added	"#25 - European Air Quality Index(Own)" was added to measured_units	System	2025-01-12 09:01:49
83.	station	Added	"#25 - European Air Quality Index(Own)" was added to measured_units	System	2025-01-12 09:01:49
82.	station	Added	"#25 - European Air Quality Index(Own)" was added to measured_units	System	2025-01-12 09:01:49
81.	station	Added	"#25 - European Air Quality Index(Own)" was added to measured_units	System	2025-01-12 09:01:49
80.	station	Added	"#25 - European Air Quality Index(Own)" was added to measured_units	System	2025-01-12 09:01:48
79.	station	Added	"#25 - European Air Quality Index(Own)" was added to measured_units	System	2025-01-12 09:01:48
71.	station	Added	"#24 - Common Air Quality Index(Own)" was added to measured_units	System	2024-08-05 16:04:55
70.	station	Added	"#24 - Common Air Quality Index(Own)" was added to measured_units	System	2024-08-05 16:04:54
69.	station	Added	"#24 - Common Air Quality Index(Own)" was added to measured_units	System	2024-08-05 16:04:54

< 1 2 ... 7 >

Рис. 4.12. Системні логи із зафіксованими змінами

Лог містить інформацію про тип об'єкта (наприклад, станція), характер змін, опис події та часову мітку.

Система підтримує різні типи логів, серед яких:

- тип зміни (наприклад, додавання чи редагування об'єкта);
- поле, яке зазнало змін;
- значення які були замінені;
- хто зробив зміну та коли.

Забезпечується аудит усіх важливих операцій та можливість швидкого відновлення історії змін. Це є ключовим для аналітичної достовірності та підтримки управлінських процесів, де важливо знати не лише поточний стан даних, але й динаміку їх змін у часі.

4.2.7. Адміністративна панель системи (Django Admin). Окремим інструментом управління є адміністративна панель, що надається у складі вебфреймворку Django [74]. Це готовий модуль, який не потребує додаткової розробки та дозволяє адміністраторам системи отримувати повний доступ до даних і керувати ними в єдиному середовищі. На рис. 4.13 наведено загальний вигляд адміністративної панелі, яка забезпечує доступ до повної структури даних.

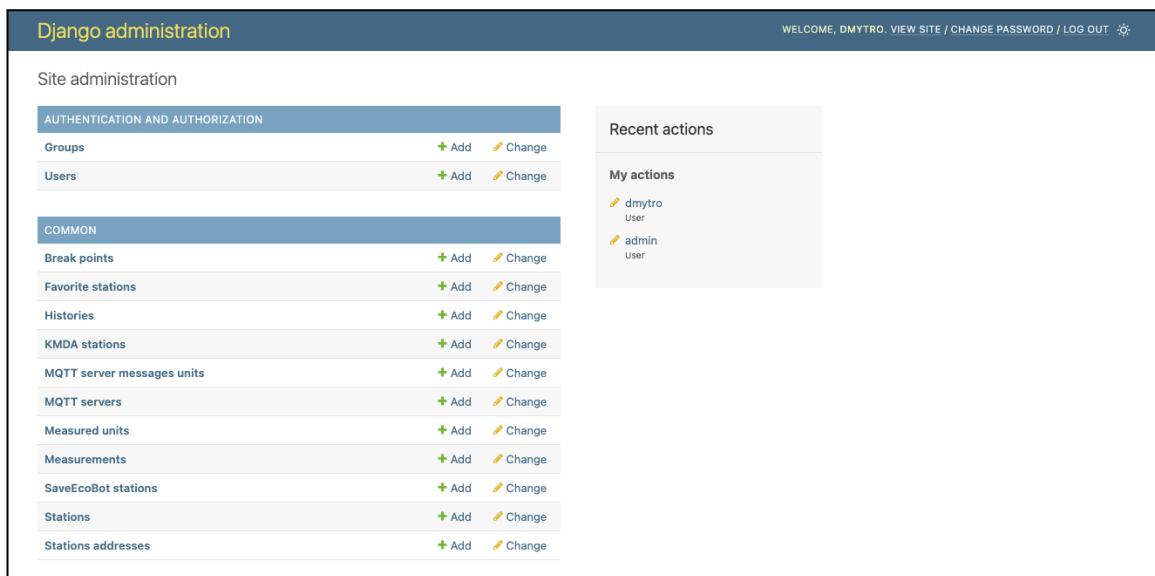


Рис. 4.13. Адміністративна панель Django Admin

Додатково важливо, що адміністративна панель має власний механізм логування змін. Це означає, що всі операції (редагування, створення, видалення об'єктів) зберігаються у внутрішніх логах, що підвищує рівень прозорості та контролю. На рис. 4.14. зображено приклад редагування контрольних точок (*break points*) для розрахунку індексів якості повітря через адміністративну панель. Це дозволяє оперативно вносити зміни до параметрів моделі без необхідності додаткового програмування чи модифікації коду.

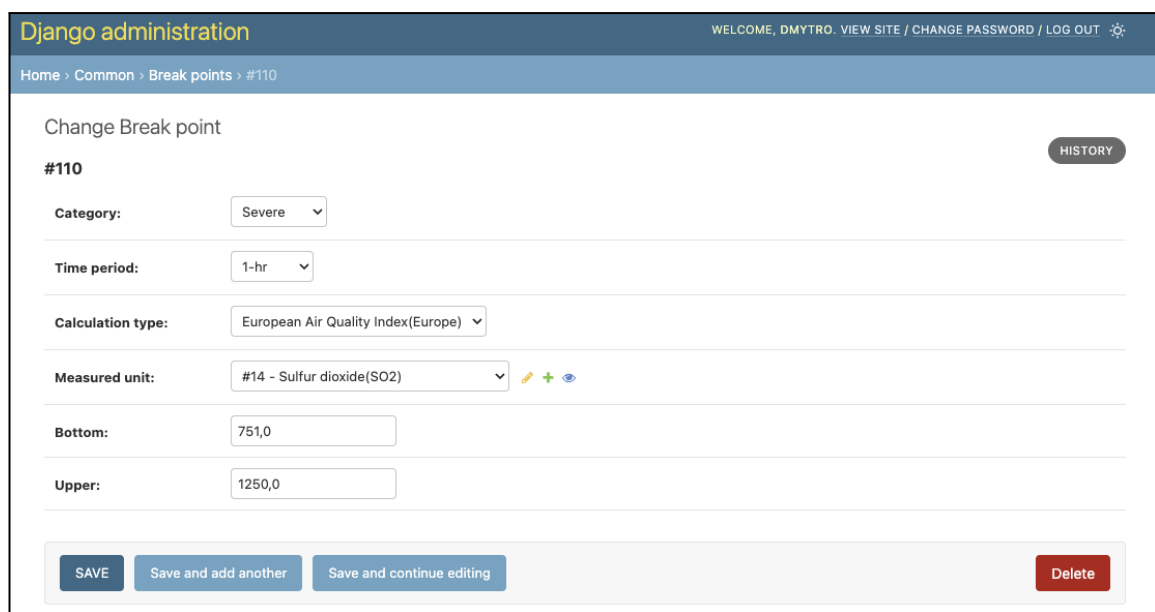


Рис. 4.14. Адміністративна панель Django Admin

Завдяки стандартній архітектурі Django вона інтегрується безпосередньо з моделями бази даних, автоматично генеруючи інтерфейси для управління. Адміністративна панель забезпечує повний набір інструментів для керування даними системи: перегляд та редагування всіх об'єктів (станцій, вимірювань, користувачів, індексів тощо), створення нових записів і оновлення існуючих, а також пошук та фільтрацію за будь-якими полями. Доступ до панелі обмежений і надається лише користувачам з розширеними правами (*superuser*), що гарантує контрольованість і безпеку управління даними. Це створює цілісне середовище, яке поєднує зручність для користувачів та ефективність для адміністраторів, і водночас формує основу для подальшого аналізу результатів роботи системи.

4.3 Тестування системи

Тестування системи здійснювалося комплексно, з урахуванням усіх основних компонентів, сценаріїв її використання та вимог до продуктивності. Метою було не лише перевірити коректність реалізації окремих модулів, а й оцінити узгодженість їхньої роботи у складі єдиної екосистеми.

Модульне тестування. На цьому рівні перевіряється коректність роботи окремих функцій та класів. Особливу увагу приділено обробці повідомлень від сенсорів, валідації отриманих даних, розрахунку індексів якості повітря та операціям із базою даних. Завдяки модульному тестуванню вдалося на ранніх етапах локалізувати помилки та забезпечити стабільність внутрішньої логіки системи.

Інтеграційне тестування. Перевіряється коректність взаємодії між окремими сервісами збору даних (API, MQTT) та їх узгодженість із транзакційною базою PostgreSQL. Тести охоплювали як стандартні сценарії обробки, так і виняткові випадки (наприклад, надходження неповних чи некоректних повідомлень). Це дало змогу підтвердити надійність передачі даних між підсистемами та цілісність збереженої інформації.

Системне тестування. Проводилася повна перевірка працездатності системи в умовах, наближених до реальної експлуатації. У процесі тестування здійснювалися розрахунки індексів якості повітря, експорт даних у сховище (Data Warehouse на базі MS SQL Server) та формування OLAP-запитів. Це дозволило впевнитися у безперервності та коректності всього циклу обробки: від надходження даних від сенсорів до отримання агрегованих результатів у системі бізнес-аналітики.

Валідація інтерфейсу користувача. Було проведено серію мануальних *usability-тестів*, під час яких оцінювалася зручність роботи з вебінтерфейсом, зрозумілість навігації та доступність ключових функцій. Важливим результатом стало підтвердження, що система є однаково придатною як для експертів-екологів, так і для користувачів без спеціальної технічної підготовки. Це відповідає початковим вимогам до доступності та простоти використання.

Тестування продуктивності та масштабованості. У процесі перевірки було підтверджено здатність системи обробляти значні обсяги даних у режимі реального часу завдяки використанню асинхронних механізмів (Celery). Для забезпечення подальшого розвитку передбачено два сценарії масштабування:

- вертикальне масштабування – збільшення ресурсів серверів (CPU, RAM, дискові обсяги);
- горизонтальне масштабування – додавання додаткових серверів і балансування навантаження між серверами.

Тестування охопило всі ключові аспекти системи – від роботи окремих модулів до комплексної взаємодії між ними. Важливо підкреслити, що на всіх етапах розробки й після впровадження система безперервно функціонувала на платному сервері з хмарною інфраструктурою, що забезпечило проведення перевірок у реальних умовах експлуатації, а не на локальних тестових даних. Такий підхід дозволив виявляти та оперативно усувати потенційні проблеми безпосередньо у робочому середовищі, де обробляються дані з реальних сенсорів.

Завдяки цьому тестування відбувалося постійно, у режимі реального часу, що підвищило його практичну цінність та наблизило результати до умов

промислової експлуатації. Система довела свою здатність стабільно працювати під навантаженням, забезпечуючи своєчасну обробку великих потоків даних, підтримку аналітичних обчислень та доступність інтерфейсу користувача.

Отримані результати підтвердили, що розроблена інфраструктура є надійною та стійкою, відповідає вимогам масштабованості, а також може бути розгорнута у виробничому середовищі без необхідності суттєвих доопрацювань. Таким чином, система готова до практичного використання як у наукових дослідженнях, так і для підтримки управлінських рішень у сфері моніторингу якості атмосферного повітря.

4.4. Аналіз отриманих результатів

Розроблена система моніторингу атмосферного повітря не лише забезпечує збір, обробку та збереження даних у режимі реального часу, але й створює основу для проведення наукового аналізу. Попередні розділи були присвячені архітектурі, програмному забезпеченню та технічним аспектам розгортання системи, що дозволяють гарантувати її стабільність і масштабованість. Однак ключовою метою роботи є не лише технічна реалізація, а й виявлення закономірностей у даних, що відкриває нові можливості для розуміння екологічних процесів.

На цьому етапі увага зосереджується на аналітичних методах, що застосовуються до накопичених масивів даних. Використання класичних статистичних підходів (регресійний аналіз), методів Data Mining, алгоритмів асоціативних правил, а також інструментів багатовимірного аналізу (OLAP) дозволяє розкрити приховані взаємозв'язки між параметрами якості повітря та побудувати науково обґрунтовані висновки.

Крім того, проведений аналіз доповнюється розрахунком ключових показників ефективності (KPI), що дає змогу оцінити як стан навколишнього середовища, так і функціональність самої системи. Важливим завданням є порівняння результатів, отриманих різними методами, зокрема через Data Mining

та OLAP-аналітику, що дозволяє підтвердити або спростувати робочі гіпотези та підвищити надійність зроблених висновків.

4.4.1. Регресійний аналіз. Для дослідження взаємозв'язків між основними забруднювачами та інтегральними індексами якості повітря (AQI та CAQI) було застосовано метод Ordinary Least Squares (OLS) з використанням бібліотеки Statsmodels. Аналіз проводився на прикладі станції 016: Київ, Харківське шосе, 7/1 (КМДА) з використанням як погодинних, так і денних середніх значень.

Методика дослідження полягає у застосуванні регресійного аналізу для виявлення залежностей між інтегральними індексами якості повітря (CAQI та AQI) та набором ключових екологічних параметрів. У ролі залежних змінних розглядалися Common Air Quality Index (CAQI) та Air Quality Index (AQI), які виступають інтегральними показниками рівня забрудненості атмосферного повітря та є зручними для порівняння у просторі та часі. Як незалежні змінні у моделі були включені найбільш поширені та значущі для оцінки якості повітря фактори: концентрації $PM_{2.5}$ та PM_{10} (дрібнодисперсний та крупнодисперсний пил), діоксид азоту (NO_2), а також параметри, що відображають фізико-природні умови середовища – рівень радіації та швидкість вітру. Останні було додано для врахування впливу метеорологічних чинників на розсіювання забруднювачів та їх накопичення у приземному шарі атмосфери. Для підвищення точності та багатосторонності аналізу застосовувалися два підходи до агрегації даних: погодинна, яка дозволяє відслідковувати короткочасні зміни та виявляти пікові навантаження, та денна, що відображає усереднені добові тенденції та забезпечує кращу стабільність моделей.

Результати дослідження регресійного аналізу показали, що для інтегрального індексу CAQI найкращі результати були отримані при використанні добової агрегації даних (73% варіації пояснюється незалежними змінними). Це свідчить про високу пояснювальну здатність моделі та стабільність виявлених закономірностей. Водночас погодинний аналіз, хоча й продемонстрував нижчий рівень адекватності (46 %), дозволив детальніше відобразити динаміку змін

протягом доби та виявити пікові значення, що є важливим для оперативного моніторингу.

Аналогічна тенденція простежується і для AQI: денні усереднення забезпечили кращий рівень узгодженості моделі з даними 69%, тоді як погодинний підхід дав нижче значення 40%. Це підтверджує доцільність застосування різних типів агрегації залежно від цілей дослідження: добові моделі – для стратегічного аналізу, погодинні – для виявлення короткочасних коливань.

В обох випадках найвагомішими предикторами залишаються концентрації PM_{10} та $PM_{2.5}$, що закономірно відображає їхній ключовий внесок у формування якості повітря. Водночас вплив NO_2 виявився менш стабільним: у деяких моделях він має значущість, однак не демонструє такої системності, як тверді частки.

Особливу увагу привернув фактор радіації, який у кількох моделях показав статистично значущий вплив. Це може свідчити про його опосередковану роль у процесах формування індексів, можливо, через взаємодію з іншими параметрами середовища.

Ще одне важливе спостереження стосується швидкості вітру (WS). Було виявлено, що у нічний час її вплив на CAQI мінімальний через зниження інтенсивності вітру, тоді як у денний час або при використанні добових середніх спостерігається чітка кореляція між швидкістю вітру та якістю повітря. Це підтверджує роль вітру як ключового фактора розсіювання забруднювачів у приземному шарі атмосфери.

На рисунку 4.15 наведено порівняння фактичних та прогнозованих значень для CAQI та AQI за різних підходів до агрегації даних. Видно, що денна агрегація забезпечує кращу узгодженість між реальними та змодельованими показниками, тоді як погодинні дані відзначаються більшою дисперсією.

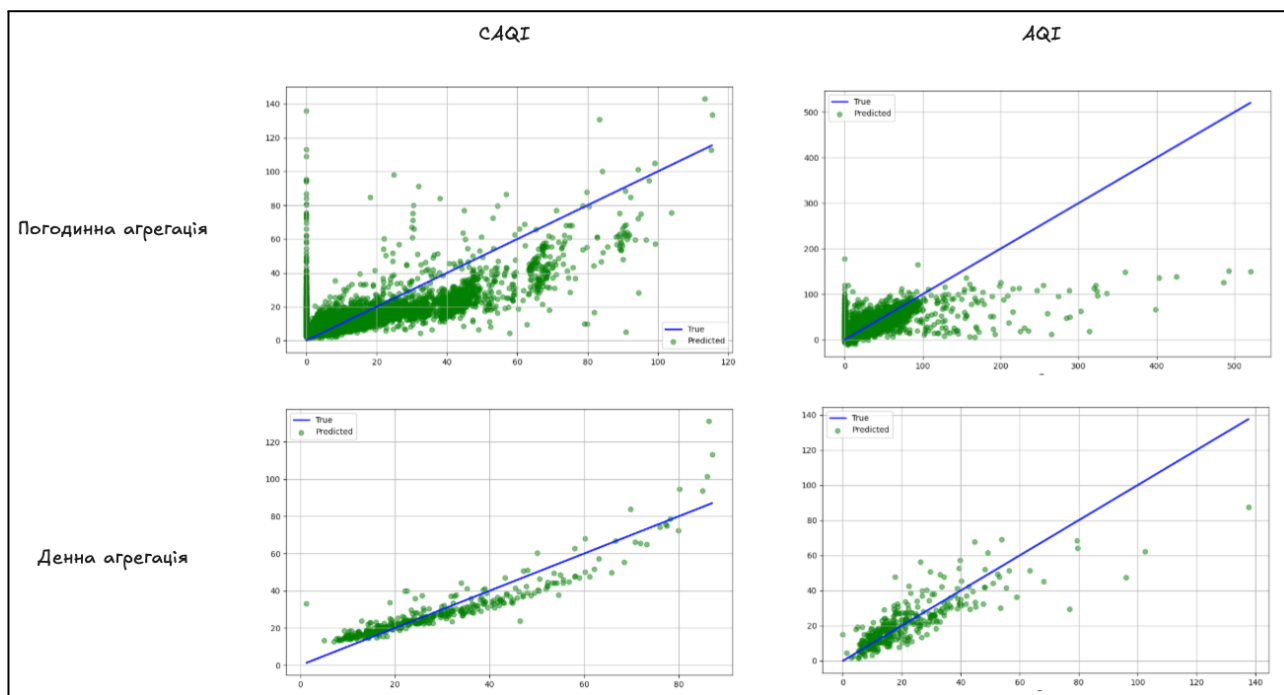


Рис. 4.15. Порівняння фактичних та прогнозованих значень (CAQI, AQI)

Підсумовуючи результати регресійного аналізу, можна відзначити, що як для CAQI так і для AQI більш доцільним є використання денних середніх значень, оскільки це забезпечує стабільніший аналітичний результат. Результати регресійного аналізу для індексів продемонстрували, що точність моделі зростає зі збільшенням часового періоду (інтервалу) та при використанні середніх денних значень.

Таким чином, отримані результати підтверджують доцільність комбінованого підходу, за якого різні індекси аналізуються з урахуванням специфіки їхніх алгоритмів формування, що дозволяє забезпечити більш повне та обґрунтоване уявлення про якість атмосферного повітря.

Водночас доцільним напрямком подальших досліджень є розширення спектра регресійних моделей. Це дозволить порівняти якість прогнозів, підвищити точність моделей та краще адаптувати аналітичні інструменти до різних умов моніторингу.

4.4.2. Кластеризація (K-means). Кластеризація даних за допомогою алгоритму “K-Means” дає змогу виявити приховані закономірності у багатовимірних наборах вимірювань та визначити групи станцій і забруднювачів

із подібною поведінкою. У контексті моніторингу атмосферного повітря цей підхід використовується для виявлення потенційно проблемних станцій або показників, що відхиляються від типових профілів концентрацій, а також для оцінки загальної однорідності даних і якості функціонування системи спостережень.

Оптимальну кількість кластерів визначено за допомогою методу ліктя (Elbow Method), який аналізує зміну внутрішньокластерної дисперсії (*WCSS*) при збільшенні k . Точка “зламу” кривої вказує на раціональне значення k , у нашому випадку – $k = 4$, що забезпечує баланс між точністю та інтерпретованістю результатів (рис. 4.16).

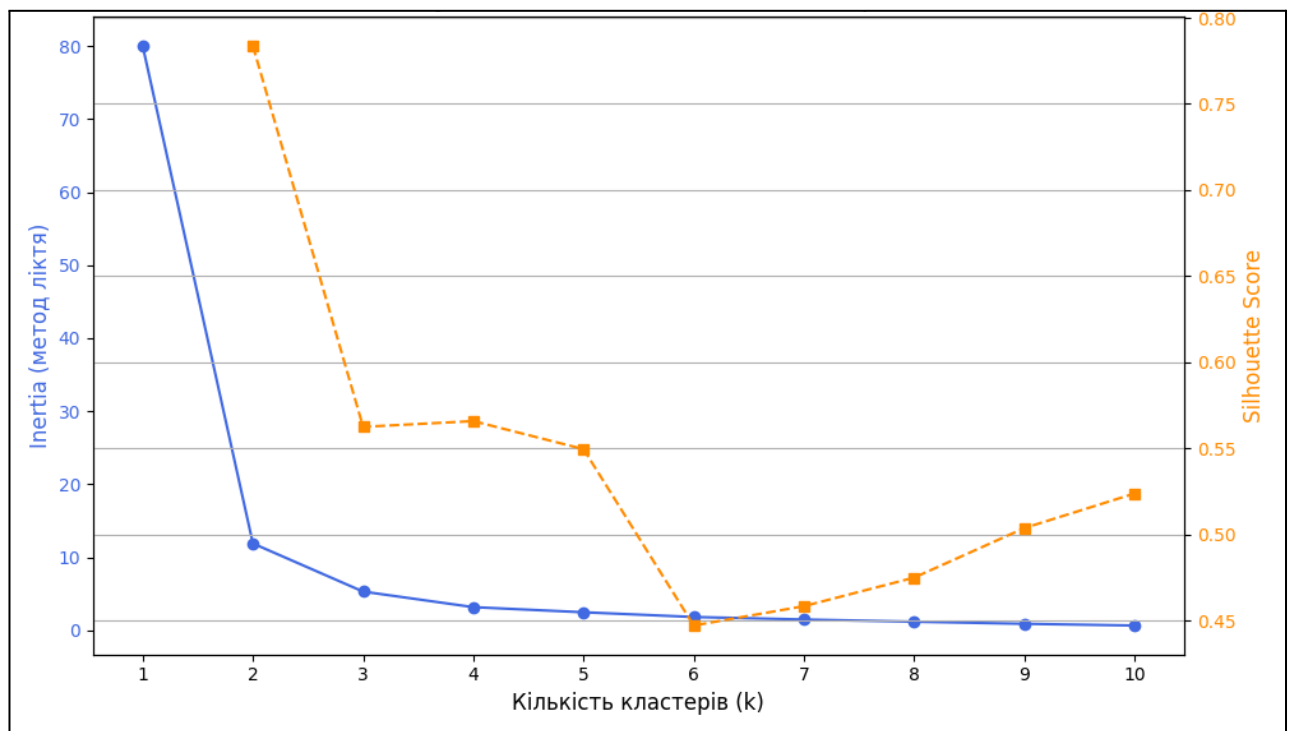


Рис. 4.16. Пошук оптимальної кількості кластерів (Elbow Method)

На рисунку також подано результати оцінки кластеризації за двома метриками. Ліва вісь (синя крива) відображає показник *Inertia*, що характеризує щільність кластерів, тоді як права вісь (помаранчева крива) демонструє *Silhouette Score*, який оцінює якість їх розділення. Спільний аналіз цих двох метрик

дозволяє обґрунтовано визначити оптимальну кількість кластерів, поєднуючи компактність і чіткість групування даних.

У результаті кластерного аналізу було сформовано чотири кластери, що відображають різні типи поведінки станцій за якістю отриманих даних. Більшість станцій характеризуються високим рівнем заповнення та стабільними показниками, проте окремі кластери вказують на аномальні або проблемні випадки. На рисунках 4.17–4.18 наведено приклади розподілу кластерів для двох станцій, які демонструють типові сценарії виявлення відхилень.

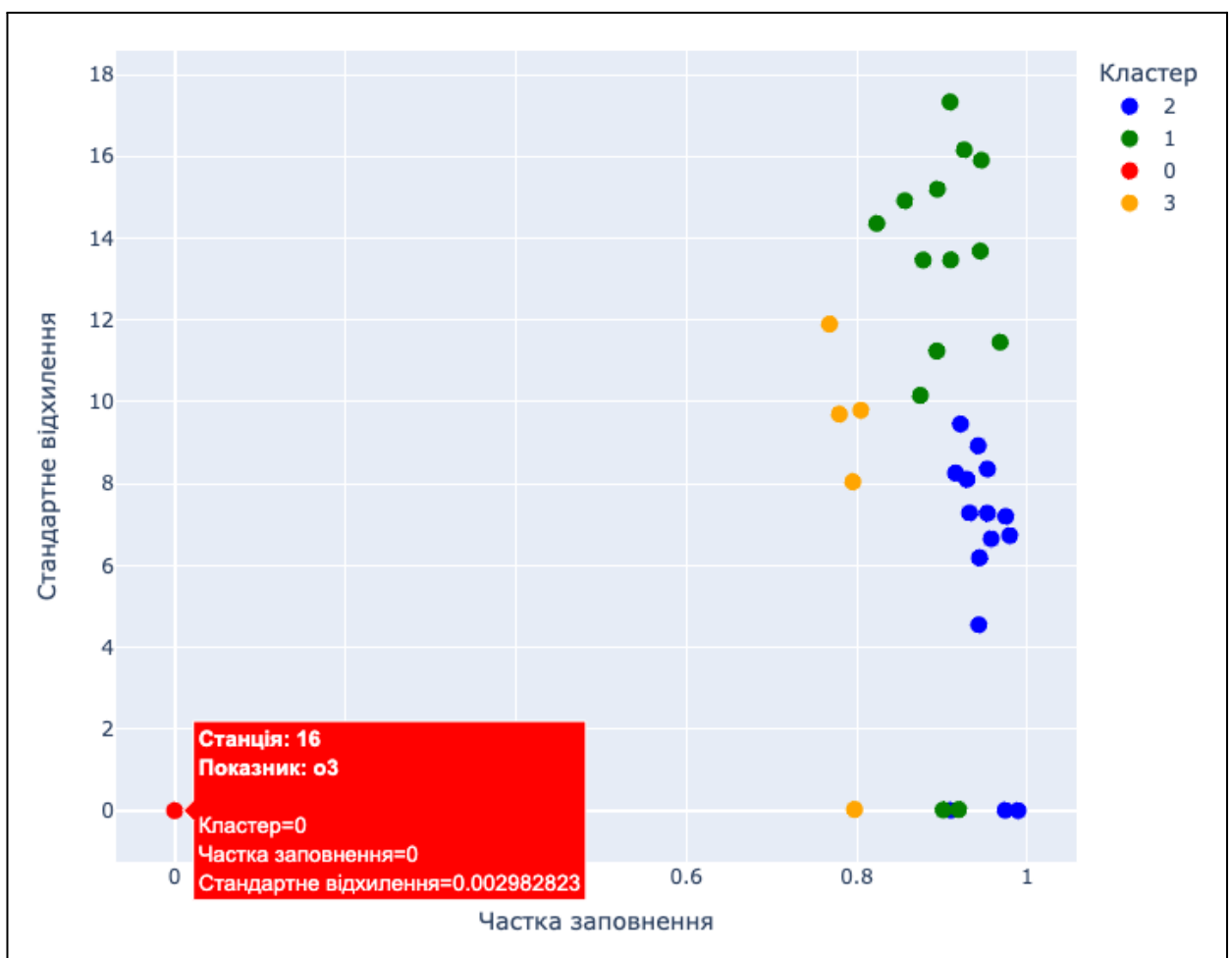


Рис. 4.17. Розподіл кластерів за часткою заповнення (16 станція)

Для станції №16 спостерігається поодинокий випадок, коли показник O_3 потрапляє до окремого кластера (кластер 0), що характеризується повною відсутністю даних та мінімальним стандартним відхиленням. Це свідчить про

несправність сенсора або відсутність накопичення озону на даній локації. Решта забруднювачів належить до кластерів із нормальними характеристиками, що підтверджує локальний характер проблеми.

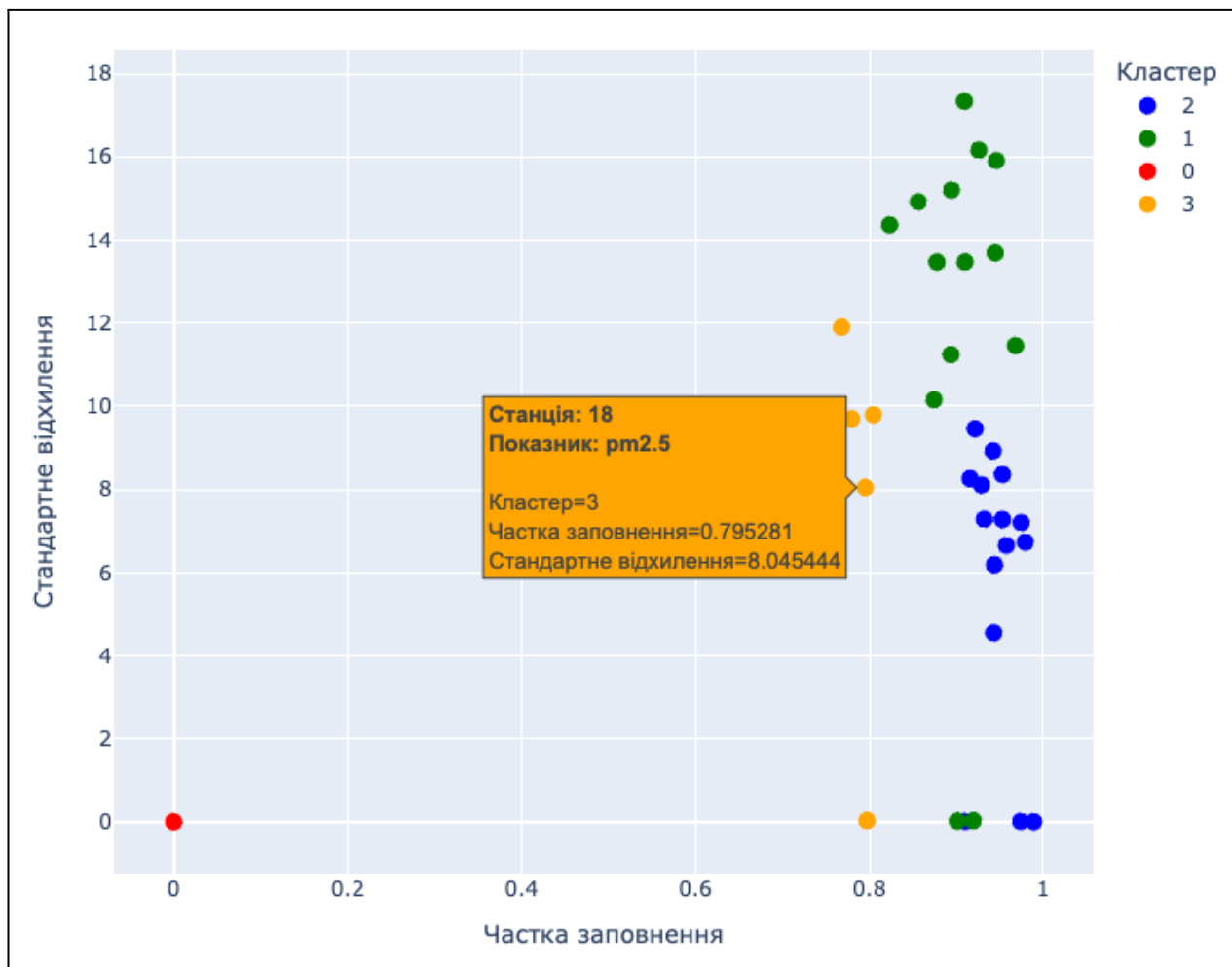


Рис. 4.18. Розподіл кластерів за часткою заповнення (18 станція)

На станції №18 виділено окремий кластер (*кластер 3*), який відображає часткові перебої у роботі станції – значення мають знижену частку заповнення ($\sim 0.75\text{--}0.8$) і підвищену варіативність. Така поведінка типова для нестабільного електропостачання, збоїв зв'язку або нерівномірного надходження телеметрії. Цей кластер вказує на необхідність технічної перевірки обладнання або повторної калібровки сенсорів.

Таким чином, проведена кластеризація дала змогу виявити конкретні станції та забруднювачі з відхиленнями у поведінці вимірювань, що свідчить про

наявність технічних або методичних проблем у їх роботі та є важливим етапом у підвищенні достовірності даних моніторингу атмосферного повітря.

4.4.3. *Наївний Байєс (Naive Bayes)*. Для прикладу реалізації задачі класифікації було обрано індекс CAQI (Common Air Quality Index), що широко використовується в європейській практиці моніторингу якості атмосферного повітря.

Однією з актуальних проблем під час розрахунку індексів є наявність пропущених або спотворених даних, що може суттєво вплинути на оцінку фактичного стану повітря. Тому коректне відновлення таких значень за допомогою статистичних або інтелектуальних методів є важливою передумовою точного та достовірного розрахунку індексів якості повітря.

Разом із тим, запропонований підхід може бути легко розширений і для інших індексів, зокрема, AQI або локальних варіацій, що ґрунтуються на аналогічних параметрах вимірювання.

Для навчання моделі “Gaussian Naive Bayes” було використано ключові ознаки, що характеризують стан атмосферного повітря:

- концентрації твердих частинок $PM_{2.5}$ та PM_{10} ;
- рівні вмісту NO_2 і SO_2 ;
- вологість повітря (humidity);
- швидкість вітру (ws).

Таке поєднання дозволяє комплексно врахувати як хімічні, так і фізичні чинники, забезпечуючи достатню інформативність для задачі багатокласової класифікації.

Теплова карта статистичних екологічних показників. На першому етапі аналізу побудовано теплову карту, яка відображає мінімальні, максимальні та середні значення ключових параметрів (PM_{10} , $PM_{2.5}$, NO_2 , SO_2 , вологість, швидкість вітру) у межах різних категорій індексів (рис. 4.19). Для інтерпретації використано шкалу, що відображає градацію стану повітря від «відмінного» до «дуже поганого» (відповідно до методології CAQI).

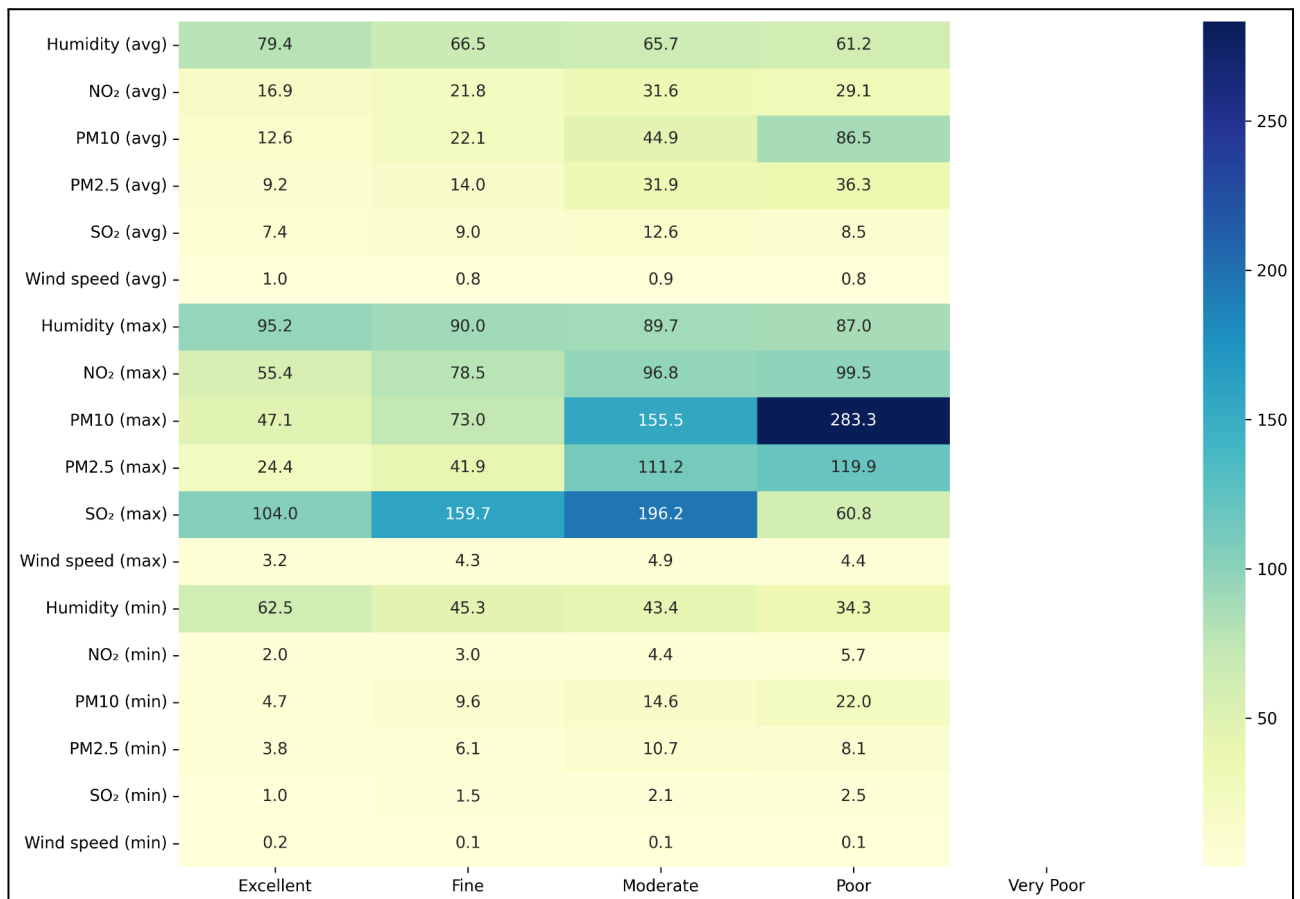


Рис. 4.19. Теплова карта статистичних характеристик екологічних показників у розрізі категорій якості повітря (CAQI)

Такі карти дозволяють візуально простежити закономірності у зміні характеристик залежно від рівня забруднення. Наприклад, простежується чітке зростання концентрацій твердих частинок PM_{10} та $PM_{2.5}$ зі збільшенням рівня індексу, тоді як вологість знижується. Окремі пікові значення можуть свідчити про можливі локальні епізоди інтенсивного забруднення.

Спостерігається чітка тенденція зростання концентрацій твердих часток: PM_{10} підвищується від ~47 у категорії «*Excellent*» до понад 280 у «*Poor/Very Poor*», а $PM_{2.5}$ – до ~220 у найгірших класах.

Для SO_2 характерний локальний сплеск у «*Moderate*» (близько 196), що може свідчити про епізодичні викиди. Вологість поступово зменшується зі зростанням рівня забруднення, тоді як швидкість вітру не має чіткої тенденції, хоча в «*Moderate*» спостерігаються короточасні підвищення, пов'язані з процесами розсіювання домішок.

Матриця помилок. Перед аналізом результатів класифікації доцільно розглянути розподіл даних по класах цільової змінної, оскільки він суттєво впливає на ефективність будь-якої моделі.

На рисунку 4.20 подано візуалізацію структури даних, що демонструє нерівномірність представлення різних категорій індексу якості повітря (CAQI).

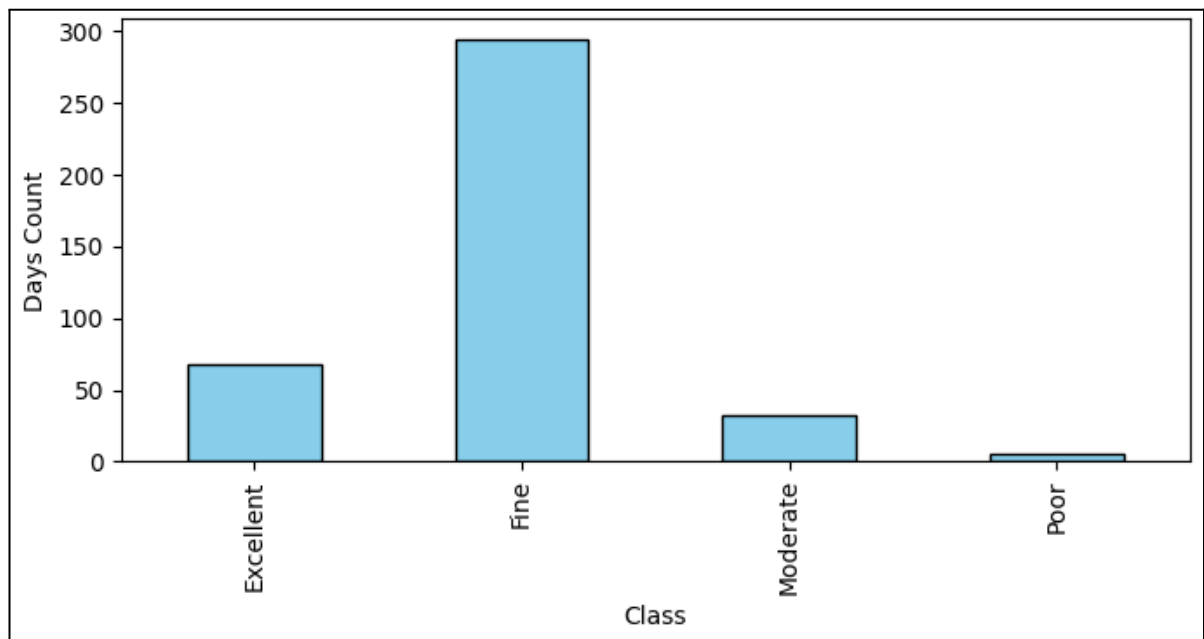


Рис. 4.20. Розподіл кількості даних по класах цільової змінної (CAQI)

Помітно, що окремі класи мають значно більшу кількість спостережень, тоді як інші представлені обмежено. Такий дисбаланс класів призводить до того, що модель краще розпізнає більш численні категорії, забезпечуючи для них вищі значення точності. Це необхідно враховувати під час інтерпретації подальших результатів класифікації та планування можливих методів балансування вибірки у майбутніх дослідженнях.

Для оцінки ефективності класифікації побудовано матрицю помилок (рис. 4.21), яка наочно демонструє розподіл передбачень моделі між реальними та прогнозованими класами.

Вона дозволяє визначити, наскільки точно алгоритм розпізнає кожну категорію якості повітря, а також виявити класи, що найчастіше плутаються між собою.

Таке подання результатів забезпечує глибше розуміння поведінки моделі та її здатності узагальнювати закономірності, виявлені під час навчання на екологічних даних.

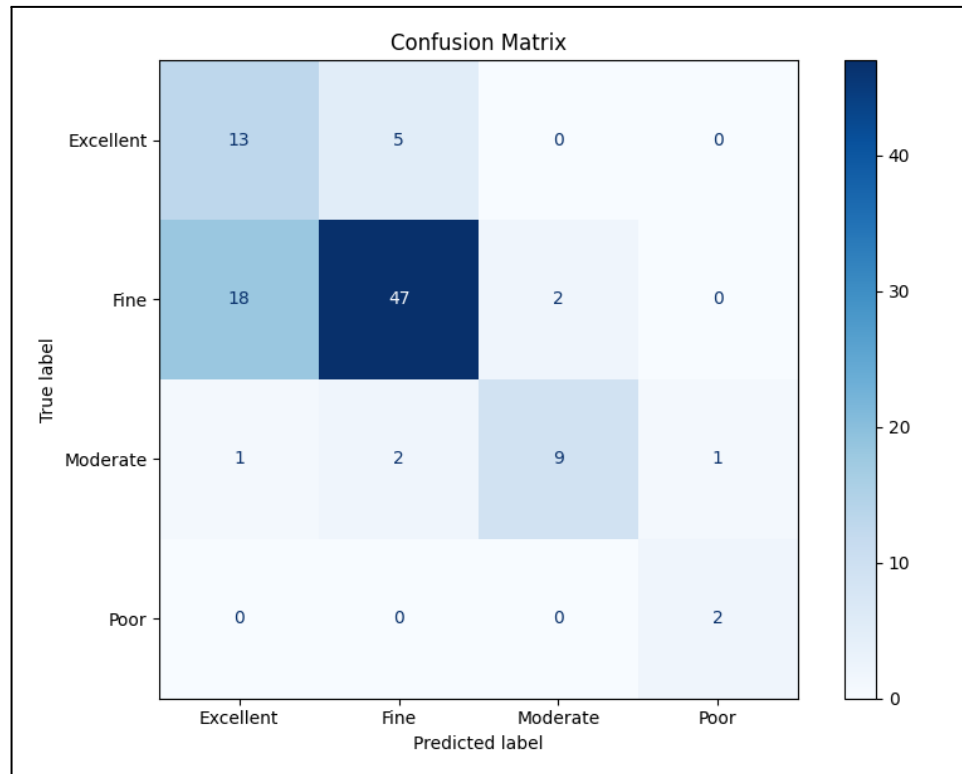


Рис. 4.21. Матриця помилок

Модель найкраще розпізнає клас «Fine», для якого зафіксовано найбільшу кількість правильних класифікацій. Частина прикладів цього класу, однак, помилково віднесена до «Excellent», що вказує на часткове перекриття сусідніх категорій.

Для класу «Excellent» спостерігається зворотна тенденція, що також логічно з огляду на близькість порогових значень.

Класи «Moderate» та «Poor» демонструють стабільні результати, хоча для останнього через обмежену кількість прикладів варіативність класифікації нижча.

Метрики класифікаційної моделі. Після аналізу розподілу прикладів доцільно перейти до оцінки якості самої моделі. Таблиця 4.1 містить основні метрики класифікаційної моделі, побудованої на основі алгоритму “Наївного Байєса”.

Показники точності (precision), повноти (recall) та F1-міри (середнє між precision і recall) для кожного класу дозволяють оцінити, наскільки добре модель розрізняє різні рівні якості повітря.

Таблиця 4.1

Метрики класифікаційної моделі (“Наївний Байєс”)

Клас	Precision	Recall	F1-score	Support
Excellent	0.41	0.72	0.52	18
Fine	0.87	0.70	0.78	67
Moderate	0.82	0.69	0.75	13
Poor	0.67	1.00	0.80	2
	–	–	–	–
Accuracy	–	–	0.71	100
Macro avg	0.69	0.78	0.71	100
Weighted avg	0.78	0.71	0.73	100

Загальна точність (accuracy) становить 0.71, що свідчить про правильну класифікацію близько 71% об’єктів тестової вибірки. Середні значення (macro avg – 0.71; weighted avg – 0.73) демонструють збалансовану ефективність моделі, незважаючи на наявний дисбаланс між класами.

Такі результати підтверджують придатність моделі для базового аналізу стану атмосферного повітря та подальшого використання у системах підтримки прийняття рішень.

Прогнозування на основі вхідних даних. На основі натренованої моделі класифікації реалізовано функціонал прогнозування класу якості повітря для нових вхідних даних (приклад використання зображено на рис. 4.22). Метод дозволяє оцінювати якість повітря як за реальними параметрами, так і за згенерованими сценаріями, що охоплюють типові діапазони значень.

```

excellent_avg = {"pm2.5": 5, "pm10": 10, "no2": 8, "so2": 4, "ws": 3.5, "humidity": 55}
df_excellent = create_df_with_range(excellent_avg, minus_pct=0.1, plus_pct=0.1)
analyzer.predict_quality_class(data=df_excellent)

moderate_avg = {"pm2.5": 25, "pm10": 45, "no2": 40, "so2": 15, "ws": 1.8, "humidity": 65}
df_moderate = create_df_with_range(moderate_avg, minus_pct=0.1, plus_pct=0.1)
analyzer.predict_quality_class(data=df_moderate)

very_poor_avg = {"pm2.5": 90, "pm10": 160, "no2": 120, "so2": 60, "ws": 0.5, "humidity": 85}
df_very_poor = create_df_with_range(very_poor_avg, minus_pct=0.1, plus_pct=0.1)
analyzer.predict_quality_class(data=df_very_poor)

Predicted CAQI class: Excellent
Class probabilities: {'Excellent': 1.0, 'Fine': 'less than 0.001', 'Moderate': 'less than 0.001', 'Poor': 'less than 0.001'}
Predicted CAQI class: Moderate
Class probabilities: {'Excellent': 'less than 0.001', 'Fine': 'less than 0.001', 'Moderate': 1.0, 'Poor': 'less than 0.001'}
Predicted CAQI class: Poor
Class probabilities: {'Excellent': 'less than 0.001', 'Fine': 'less than 0.001', 'Moderate': 'less than 0.001', 'Poor': 1.0}

```

Рис. 4.22. Прогнозування якості повітря за вхідними параметрами (CAQI)

Модель, яка натренована на якісних станціях із повними наборами вимірювань, може застосовуватися для відновлення відсутніх або спотворених даних на інших станціях. Це дозволяє імітувати відсутні показники шляхом прив'язки до прогнозованої категорії та відповідного діапазону концентрацій, що підвищує повноту та узгодженість даних у системі.

Результати демонструють віднесення спостереження до певного класу з відповідним розподілом ймовірностей по всіх можливих категоріях. Така форма представлення є зручною для інтерпретації результатів і може застосовуватись в автоматизованих системах підтримки прийняття рішень у сфері екологічного моніторингу. Крім того, реалізація моделі передбачає гнучке підключення до сценаріїв реального часу або інтерактивних аналітичних інтерфейсів.

Отримані результати класифікації підтверджують здатність моделі не лише прогнозувати рівень забруднення, а й підтримувати процес відновлення пропущених даних, забезпечуючи підвищену точність подальших розрахунків індексів якості повітря.

4.4.4. Пошук асоціативних правил. Для виявлення закономірностей між показниками ми застосували *mlxtend.frequent_patterns (apriori, association_rules)*. У якості прикладу в дисертації наведено результати для індексу CAQI (рис. 4.23), однак побудований підхід є універсальним і може бути розширений також на інші

індекси. Це дозволяє гнучко адаптувати методику до різних сценаріїв аналізу та вимог системи.

Щоб перетворити безперервні вимірювання на бінарні події, для кожної ознаки (усі *avg_value_**, *min_value_**, *max_value_**, крім цільового індексу) виконано бінінг за медіаною:

- *feature_high* = 1, якщо значення > медіани;
- *feature_low* = 0, якщо значення ≤ медіани.

Параметри видобутку:

- *min_support* = 0.10;
- побудова правил за метрикою *confidence* (*min_threshold* = 0.60);
- подальша фільтрація/сортування за *lift*.

Правило 1	
ЯКЩО:	avg_pm10_high, avg_so2_high, max_pm2.5_high, min_humidity_low, min_pm2.5_high
ТО:	avg_humidity_low, avg_pm2.5_high, max_pm10_high, min_pm10_high
Метрики:	support = 0.100 confidence = 0.962 lift = 6.397 leverage = 0.085 conviction = 22.092

Правило 2	
ЯКЩО:	avg_humidity_low, avg_pm2.5_high, max_pm10_high, min_pm10_high
ТО:	avg_pm10_high, avg_so2_high, max_pm2.5_high, min_humidity_low, min_pm2.5_high
Метрики:	support = 0.100 confidence = 0.667 lift = 6.397 leverage = 0.085 conviction = 2.687

Правило 3	
ЯКЩО:	avg_value_no2_low, min_value_humidity_low, min_value_pm10_low
ТО:	avg_value_humidity_low, avg_value_pm10_low, avg_value_pm2.5_low, min_value_no2_low, min_value_pm2.5_low
Метрики:	support = 0.104 confidence = 0.755 lift = 5.280 leverage = 0.084 conviction = 3.502

Рис. 4.23. Фрагмент знайдених правил з ключовими метриками

Support (підтримка) – показує, наскільки часто правило зустрічається у вибірці. *Confidence* (достовірність) – відображає, з якою ймовірністю виконується права частина правила (В), якщо виконується ліва (А).

Lift (підсилення) – показує, наскільки частіше подія В відбувається за умови А, ніж у середньому по вибірці, тобто вимірює силу зв'язку між А і В порівняно зі статистичною незалежністю.

Leverage (виграш) – показує різницю між фактичним і очікуваним рівнем співпадіння подій.

Conviction (переконливість) – характеризує ступінь довіри до правила, тобто, наскільки рідко спостерігається його порушення.

Найсильніші знайдені правила (за значенням *lift* – 6.3–6.4) демонструють стійку комбінацію таких ознак: підвищені концентрації твердих частинок $PM_{2.5}/PM_{10}$ та газоподібних забруднювачів (NO_2 , SO_2), поєднані з низькими значеннями вологості (*humidity_low*) та наявністю «підсилюючих» пікових (максимальних або мінімальних) значень за добу.

Такі поєднання свідчать про те, що спільна поява цих характеристик у даних спостерігається значно частіше, ніж можна було б очікувати за умови статистичної незалежності. Це не обов'язково означає наявність причинно-наслідкових зв'язків, однак дозволяє сформулювати інтерпретовані патерни для експертного аналізу, розробки правил тригерів та формування гіпотез. Наприклад, комбінація «низька вологість + високі концентрації РМ» може вказувати на високу ймовірність появи підвищених піків NO_2/PM_{10} .

Отримані результати є чутливими до обраної стратегії бінаризації ознак (у даному випадку – поріг медіани), параметрів *min_support* та *confidence*, а також до кореляції між ознаками. Тому для застосування у виробничих або моніторингових системах доцільно:

- перевіряти стійкість виявлених правил на інших часових періодах або в різних географічних локаціях;
- експериментувати з альтернативними підходами до бінінгу;
- відбирати правила з урахуванням не лише *lift*, але й *leverage*, *conviction* та експертної доцільності.

4.4.5. *OLAP звіти*. Одним із важливих етапів аналізу результатів моніторингу є побудова багатовимірної аналітики на основі технології OLAP (Online Analytical Processing).

Використання OLAP-підходу забезпечує можливість виконання оперативних зрізів, агрегацій і порівнянь екологічних показників у різних вимірах – зокрема, за

часом, локацією або типом забруднювача. Це дозволяє дослідникам і аналітикам швидко отримувати узагальнену інформацію про стан атмосферного повітря та динаміку його змін.

Аналіз динаміки індексу якості повітря. У межах реалізованої системи аналітики побудовано лінійний графік (рис. 4.24), який відображає усереднені значення індексів якості повітря (AQI / CAQI) за обраний часовий період. Дані агрегуються по всіх доступних станціях, що дозволяє оцінити загальну ситуацію в межах регіону або міста.

Таке групування реалізується завдяки структурі OLAP-куба, де виміри об'єднують дані з різних сенсорів, а часовий вимір забезпечує формування середніх показників на рівні дня, тижня або місяця.

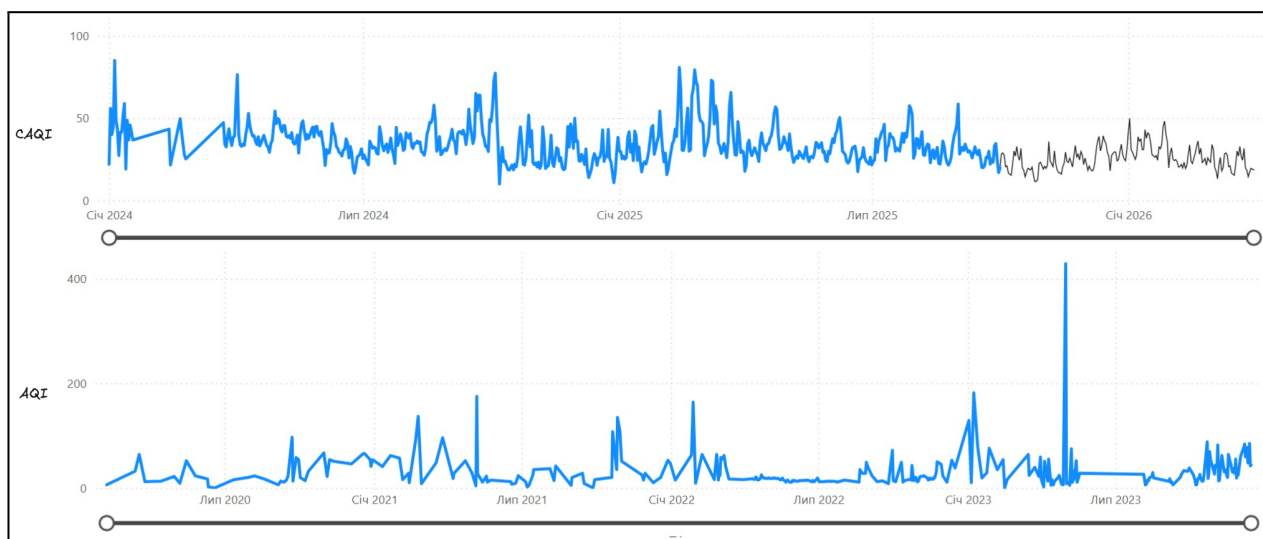


Рис. 4.24. Динаміка AQI та CAQI з трендом і середніми значеннями

У графіку реалізовано можливість візуалізації лінії тренду, що обчислюється автоматично (наприклад, методом ковзного середнього або поліноміальної регресії). Це дозволяє простежити загальні тенденції змін якості повітря у часі – виявити сезонні коливання, а також довготривалі зміни, пов'язані з екологічними ініціативами чи зростанням антропогенних навантажень.

Водночас аналіз показав, що для індексу AQI спостерігається нерівномірність накопичення історичних даних – наявні розриви та періоди з

неповним охопленням спостережень. Це ускладнює побудову коректної лінії тренду, оскільки вибірка є статистично нестабільною.

Натомість для $CAQI$, який базується на більш повному наборі сенсорних показників, побудова трендових моделей є коректною та репрезентативною, що робить цей індекс більш придатним для моніторингових досліджень і аналітичних прогнозів.

Аналіз мінімальних, середніх і максимальних значень AQI , $CAQI$ та $EAQI$. Ще одним важливим інструментом аналітики в межах реалізованої OLAP архітектури є лінійна діаграма з мультисерією (рис. 4.25).

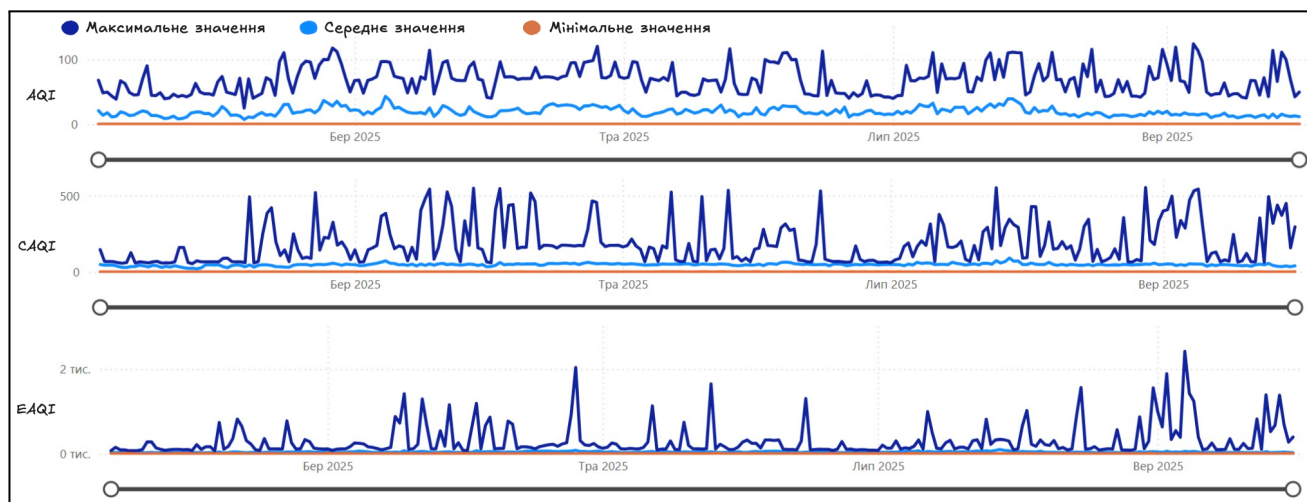


Рис. 4.25. Мінімальні, середні та максимальні значення AQI , $CAQI$ та $EAQI$

Графік відображає мінімальні, середні та максимальні значення індексів якості повітря – AQI , $CAQI$ та $EAQI$ у часовому розрізі. Розрахунок кожного з індексів виконано на основі власних реалізацій алгоритмів, розроблених відповідно до офіційних методик оцінювання якості атмосферного повітря.

Особливо цінним є одночасне відображення кількох індексів, оскільки це дозволяє провести порівняльний аналіз реакцій різних моделей оцінювання.

Так, у більшості випадків спостерігається узгоджена поведінка $CAQI$ та $EAQI$, що підтверджує їх спільну основу у європейських нормах. Водночас AQI може демонструвати більш різкі коливання – зокрема, через відмінності у вагових коефіцієнтах забруднювачів (особливо $PM_{2.5}$ та O_3) і методах нормалізації.

Додатково в системі реалізовано лінійну діаграму з мультисерією, що відображає динаміку зміни індексів якості повітря – AQI, CAQI та EAQI – у часовому розрізі (рис. 4.26).

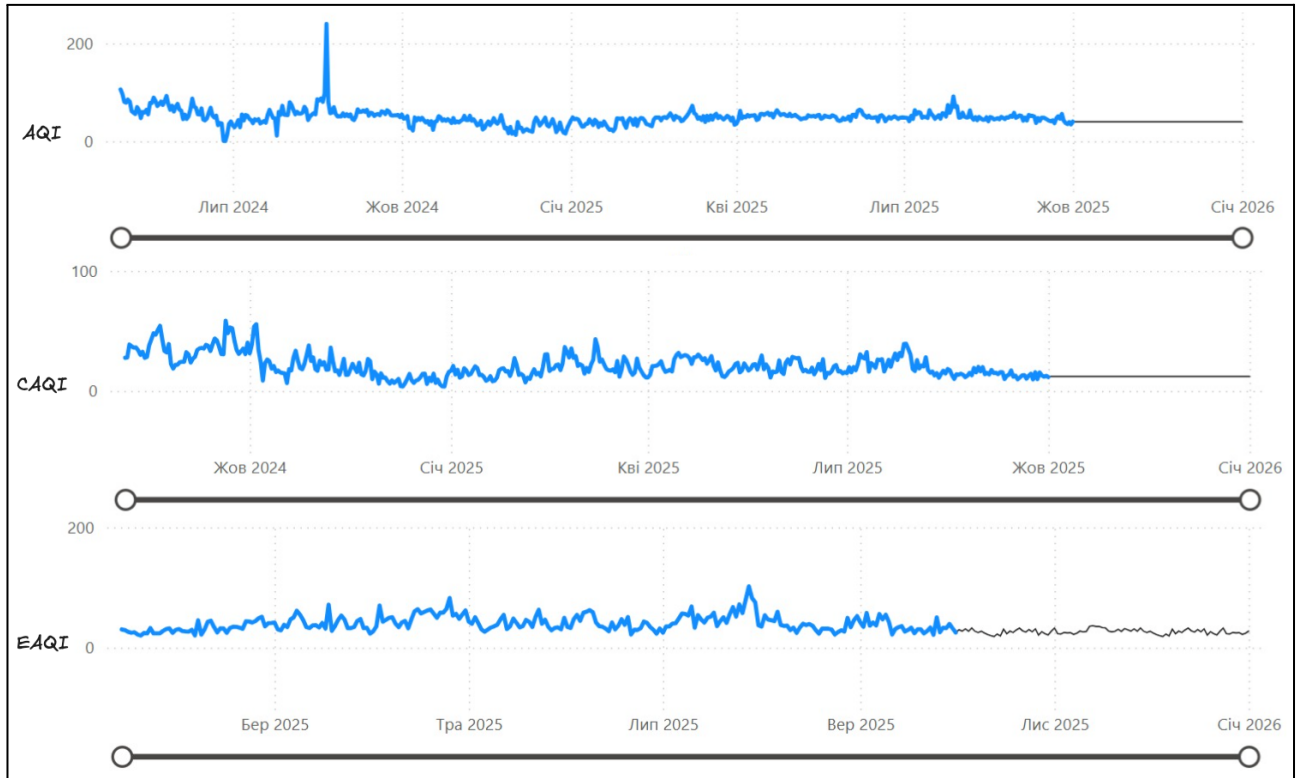


Рис. 4.26. Динаміка AQI, CAQI та EAQI з трендом і середніми значеннями

На графіку передбачено можливість вмикання лінії тренду, яка обчислюється автоматично (методом ковзного середнього або поліноміальної регресії). Це дозволяє оцінити сезонні коливання та довготривалі тенденції, наприклад, покращення якості повітря після впровадження екологічних заходів чи погіршення через збільшення транспортного навантаження.

Такий формат візуалізації дає змогу швидко ідентифікувати періоди аномальних змін, сезонні тренди або локальні піки забруднення, а також оцінити ступінь чутливості різних індексів до окремих екологічних факторів. Це робить мультисерійну діаграму ефективним інструментом для моніторингу, порівняльного аналізу та моделювання тенденцій якості атмосферного повітря.

Аналіз усереднених річних значень AQI та CAQI. Одним із ключових інструментів у межах OLAP-аналітики є стовпчаста діаграма середньорічних значень індексів якості повітря (рис. 4.27–4.28). Візуалізація реалізована на основі багатовимірної структури даних, що забезпечує можливість динамічного групування та порівняння показників за роками, а також деталізації за окремими видами забруднювачів.

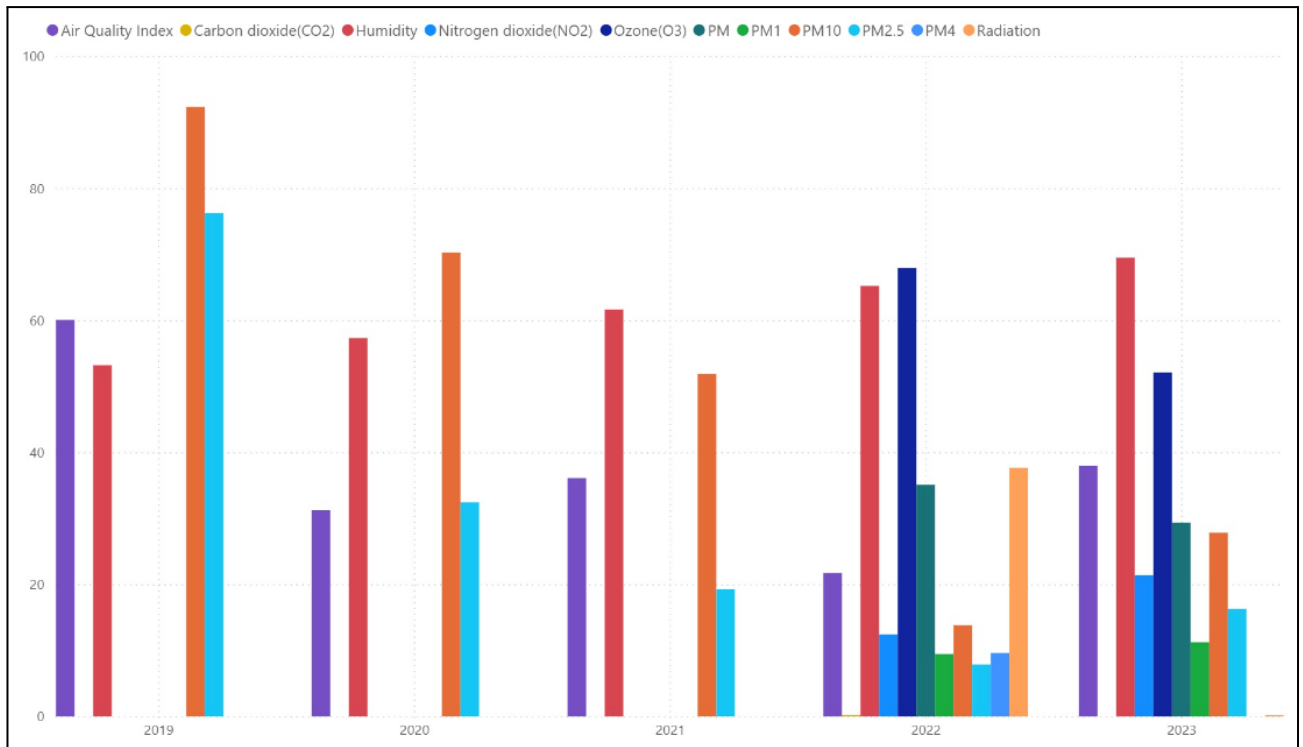


Рис. 4.27. Середньорічні значення індексу AQI (2019–2023)

Кожна діаграма побудована з використанням OLAP-агрегацій за вимірами «Рік», «Тип забруднювача» та «Індекс якості повітря», що дозволяє здійснювати:

- порівняльний аналіз річних середньозважених значень індексів;
- оцінку внеску окремих забруднювачів;
- виявлення аномальних або перехідних періодів.

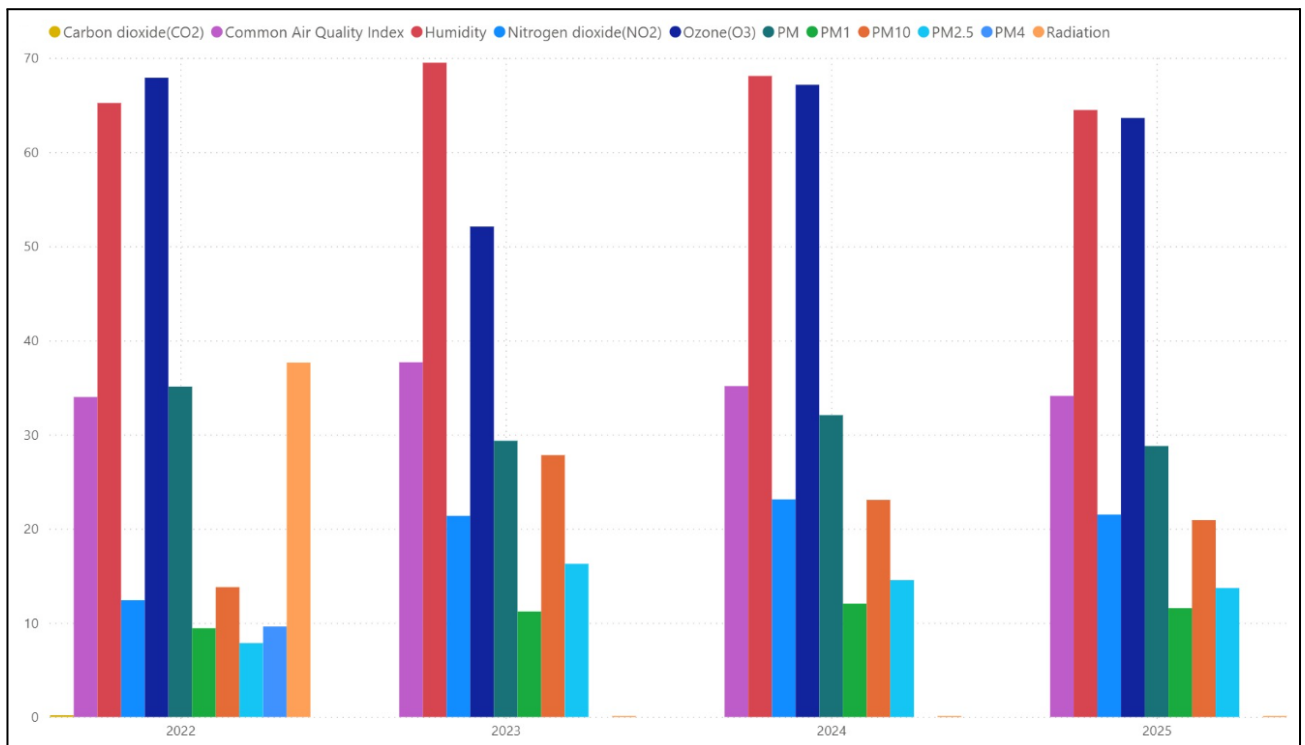


Рис. 4.28. Середньорічні значення індексу CAQI (2022–2025)

Такі діаграми не лише відображають середні значення за роками, а й демонструють вклад кожного полютанта у загальну структуру індексу – завдяки використанню формату *stacked bar*. Це дає змогу оцінити, які фактори є визначальними у формуванні якості повітря в окремі роки, а також простежити їх еволюцію.

Візуалізація у форматі OLAP дозволяє миттєво виявляти такі відхилення, а також співставляти значення індексів із супутніми параметрами. Завдяки цьому можна визначати, чи мають коливання природний (сезонний) або антропогенний (техногенний) характер, що особливо важливо для екологічного моніторингу в урбанізованих регіонах.

Перегляд кластерів. Для підтвердження результатів кластеризації проведено візуальний аналіз у середовищі OLAP, що дозволив дослідити часові закономірності пропусків даних у межах виділених кластерів. Зокрема, для станції №18 та показника PM_{10} , який належить до *кластера 3* (кластер із частковими перебоями у роботі), виявлено характерні інтервали відсутності вимірювань (рис. 4.29).

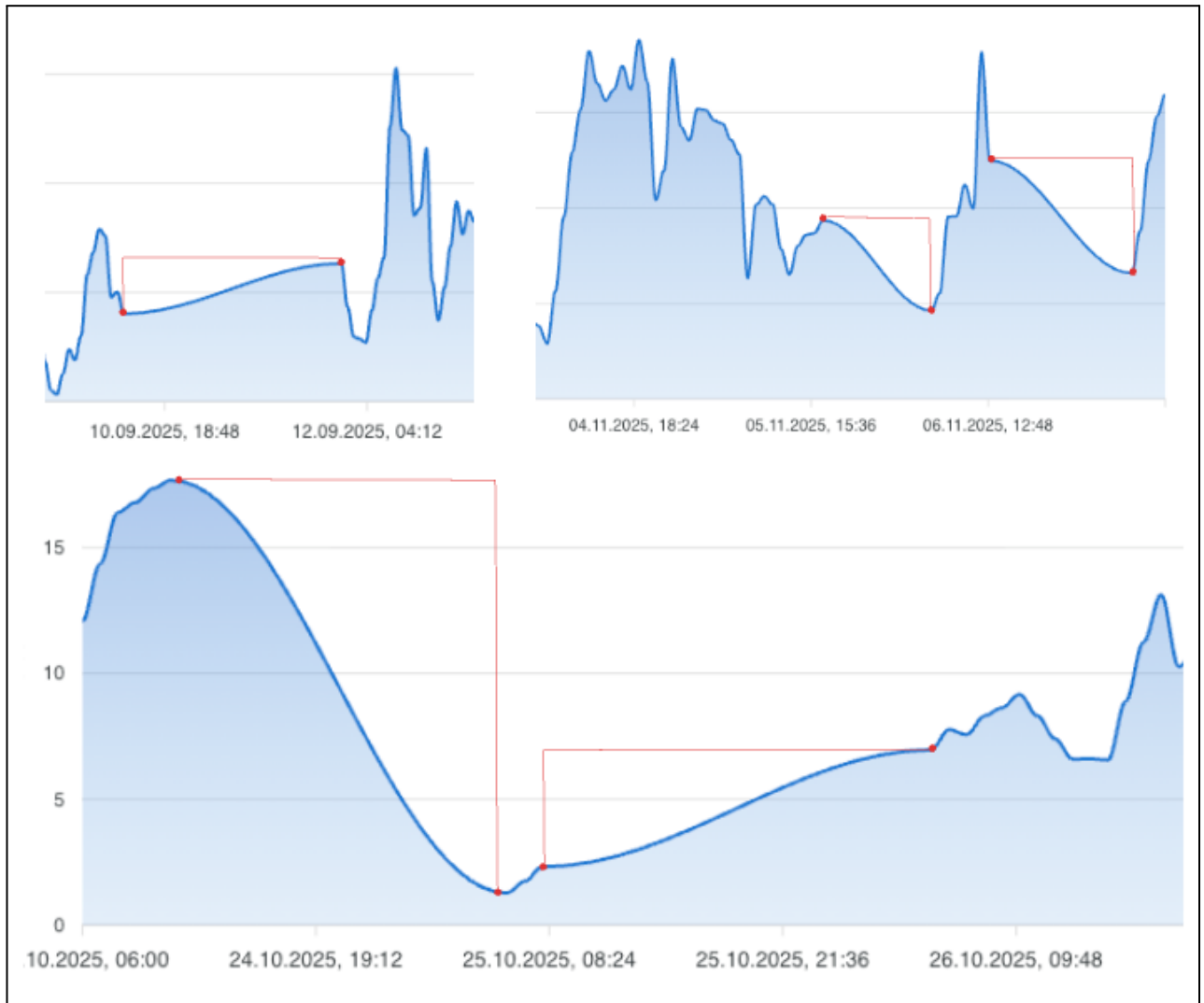


Рис. 4.29. Періоди пропущених даних (18 станція – PM_{10}).

З аналізу часових зрізів видно, що пропуски даних мають чітко виражений періодичний характер та тривалий термін. Такі інтервали можуть бути зумовлені технічними перебоями у роботі станції, відключеннями живлення. У поодиноких випадках вони можуть також збігатися з негативними метеорологічними умовами, що ускладнюють стабільний збір даних.

На рисунку представлено результати узагальнення середньорічних концентрацій озону (O_3) за 2024 рік для всіх станцій спостереження. Більшість станцій демонструють подібні рівні середніх концентрацій, що підтверджує стабільність вимірювань та коректність роботи сенсорів.

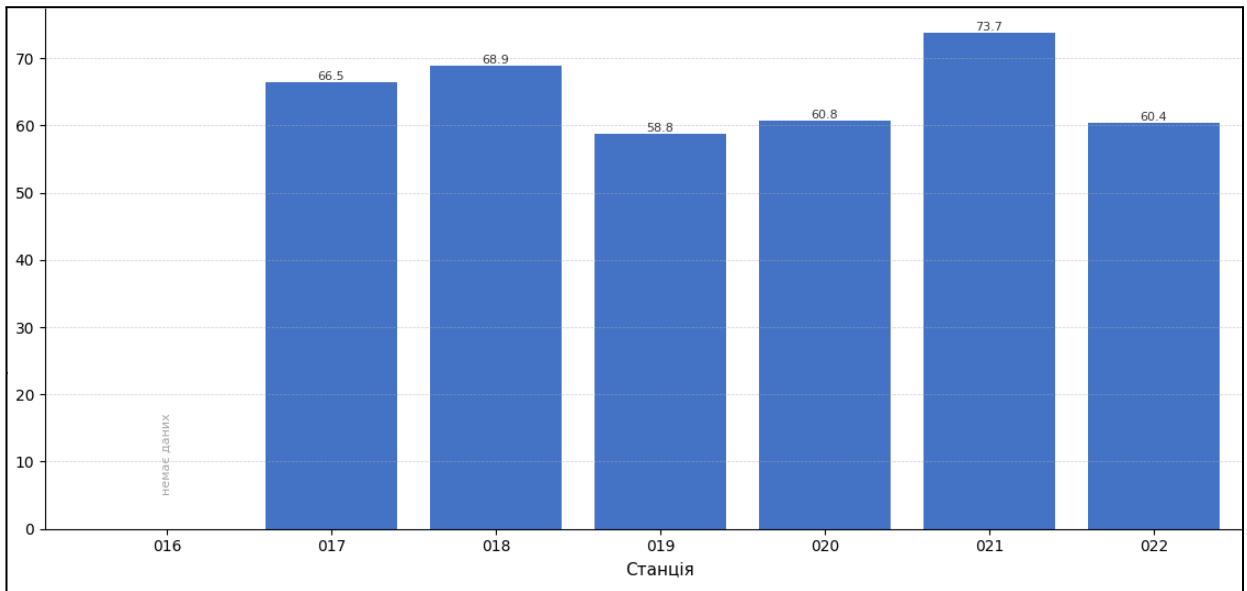


Рис. 4.30. Середньорічні концентрації O₃ по станціях (2024)

Водночас станція №16 чітко виділяється відсутністю значень, що узгоджується з результатами кластерного аналізу, де вона була віднесена до *кластера 0* – групи з повною відсутністю даних по озону. Така ситуація свідчить про тривалу несправність сенсора або відсутність вимірювань за цим параметром і підкреслює важливість контролю якості даних у системі моніторингу.

Отримані результати підтверджують ефективність кластерного підходу, який дозволяє ідентифікувати періоди нестабільної роботи станцій та забезпечує аналітичне підґрунтя для подальшої діагностики і підвищення надійності моніторингової системи.

Таким чином, сформовані кластери можуть використовуватися як основа для відбору найбільш надійних та стабільних станцій, дані з яких застосовуються для навчання моделей прогнозування. Це, у свою чергу, дозволяє використовувати перевірені станції для відновлення пропущених значень на менш якісних або нестабільних точках спостереження, забезпечуючи більш точну і повну картину стану атмосферного повітря.

Перевірка достовірності знайдених асоціативних правил. Для верифікації результатів асоціативного аналізу було сформовано спеціальний

аналітичний звіт, який проходить по всіх комбінаціях ознак, виявлених алгоритмом Apriori, та перевіряє їх фактичну спільну появу у вихідних даних.

На основі цих підрахунків автоматично обчислюються ключові метрики. Для кожної пари правил формується візуалізація (рис. 4.31), що дозволяє провести візуальну перевірку закономірностей та підтвердити аналітичні результати, отримані методом Data Mining.

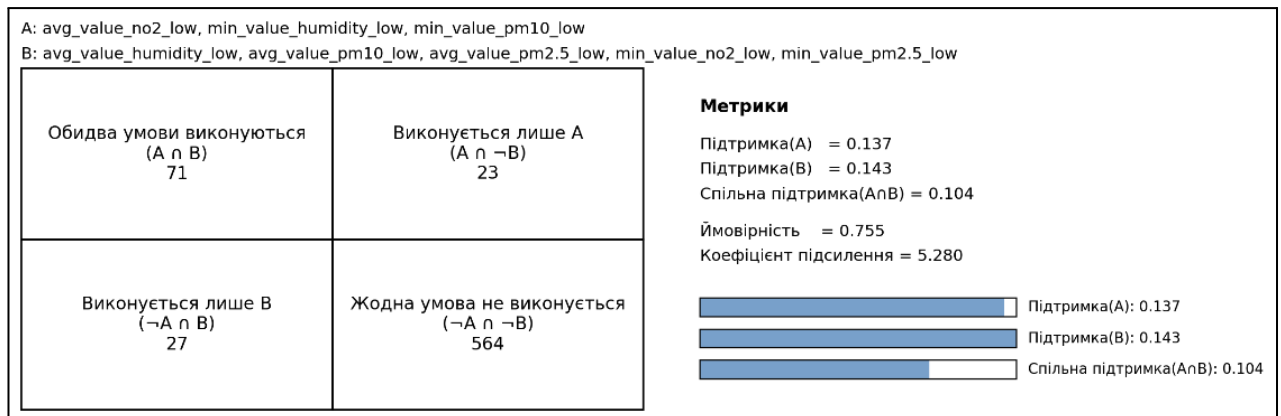


Рис. 4.31. Приклад звіту перевірки асоціативного правила

Такий підхід забезпечує підвищення достовірності інтерпретації виявлених асоціацій та сприяє їх подальшій інтеграції у систему багатовимірного аналізу (OLAP).

Аналіз перевищень нормативних рівнів забруднювачів. Окрім розрахунку індексів якості повітря, важливою складовою аналітичної системи є оцінка частоти та інтенсивності перевищень нормативних концентрацій поллютантів. Такий аналіз дозволяє не лише фіксувати факт забруднення, а й визначати тенденції ризику та оцінювати ефективність екологічних заходів у динаміці.

Основна візуалізація – лінійна діаграма (рис. 4.32), яка відображає динаміку кількості перевищень у часі. Кожна точка на графіку відповідає загальній кількості зафіксованих випадків у певному часовому інтервалі.

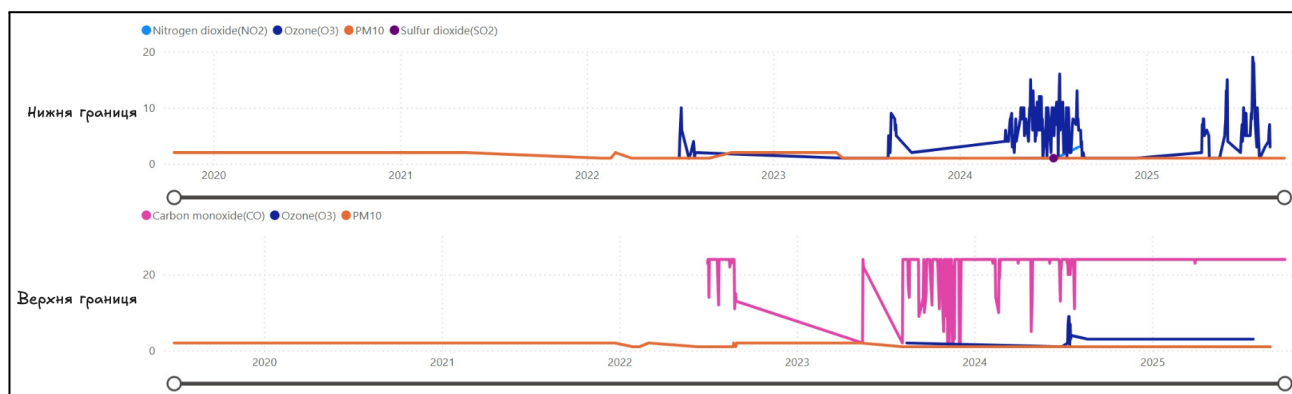


Рис. 4.32. Динаміка кількості перевищень порогових значень

При побудови цієї залежності використовуються дані з факт-таблиці *ThresholdExceedFact*, агреговані за виміром *DateDim* із застосуванням функції *COUNT*. Така форма подання дозволяє виявляти періоди підвищеної екологічної небезпеки, пікові значення та довготривалі тенденції.

Для порівняльного аналізу результатів у міжрічному розрізі побудовано стовпчасту діаграму (рис. 4.33), яка демонструє кількість перевищень за роками. Дозволяє оцінити зміни у загальному стані атмосферного повітря, відслідковувати вплив нових політик у сфері екології, рівень транспортного навантаження чи зовнішні чинники.

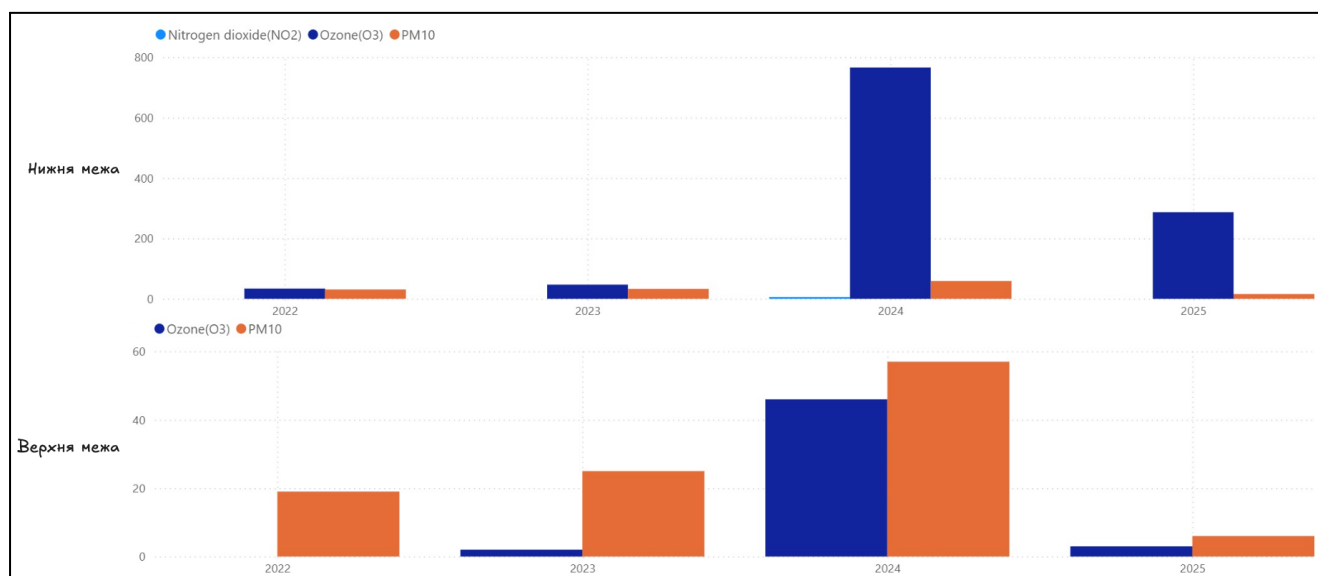


Рис. 4.33. Кількість перевищень порогових рівнів за роками

Аналіз отриманих результатів свідчить, що найвищу кількість перевищень зафіксовано у 2024 році (на момент написання 2025 рік ще не є завершеним). Зростання може бути зумовлене як реальним підвищенням рівня забруднення, так і розширенням мережі моніторингових станцій, що збільшило кількість спостережень. Незалежно від причини, цей факт є важливим сигналом для поглибленого дослідження екологічної ситуації та актуалізації політик контролю викидів.

4.4.6. *KPI*. Для підвищення інформативності моніторингової системи та забезпечення оперативного контролю за екологічними показниками у рамках OLAP-архітектури було реалізовано набір KPI-індикаторів, що ґрунтуються на даних вітрини *ThresholdExceed* (рис. 4.34-4.35).

Такі індикатори дозволяють у реальному часі відстежувати кількість перевищень встановлених нормативів і порівнювати їх із допустимими межами для окремих поллютантів, часових інтервалів та географічних регіонів.

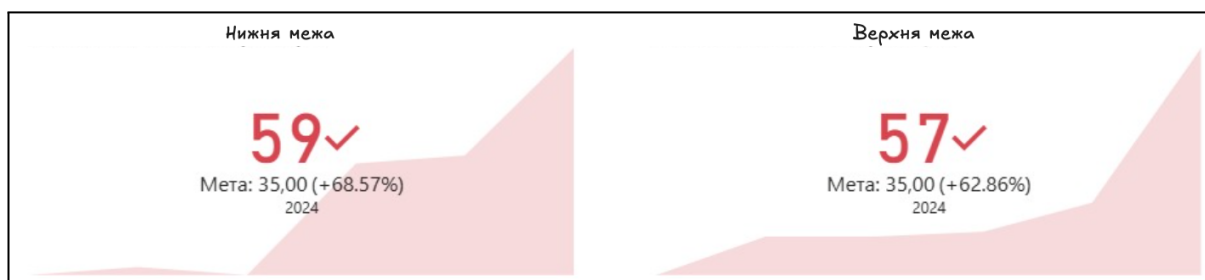


Рис. 4.34. KPI-індикатори перевищень для PM_{10} на 2024 рік

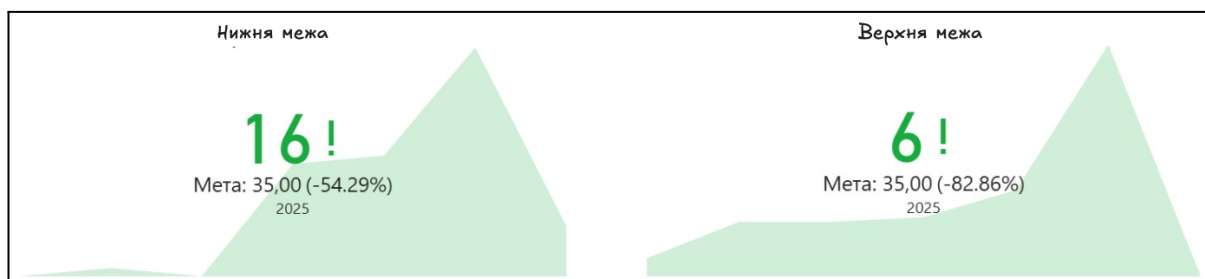


Рис. 4.35. KPI-індикатори перевищень для PM_{10} на 2025 рік

Візуалізація побудована за принципом порівняння двох часових періодів – 2024 та 2025 років, що дозволяє оцінити динаміку виконання нормативів і ступінь наближення поточних значень до граничних.

Зокрема, для параметра PM_{10} використано норматив середньодобової концентрації (24 години) з обмеженням до 35 допустимих перевищень на рік.

У поточному періоді (2025 р.) система зафіксувала:

- Нижня межа: 16 випадків перевищень;
- Верхня межа: 6 випадки перевищень.

Порівняно з 2024 роком, коли ці показники становили 59 і 57 відповідно, простежується позитивна тенденція – зменшення частоти перевищень, що візуально відображається зеленим кольором індикатора (стан «норма»).

Завдяки інтеграції KPI-візуалізацій із OLAP-кубом аналітик може миттєво оцінити, наскільки поточна ситуація відповідає екологічним стандартам, а також виявити потенційні ризики ще до досягнення критичних значень.

Як тільки реальні показники наближаються до граничного ліміту (*threshold*), індикатор автоматично змінює стан – із зеленого на жовтий або червоний, сигналізуючи про необхідність реагування.

У поєднанні з іншими OLAP-звітами (динаміка індексів, річні усереднення, частота перевищень) такі показники створюють замкнену систему екологічної оцінки, що дозволяє проводити стратегічний аналіз стану атмосферного повітря на регіональному рівні.

Проведена багатовимірна аналітика підтвердила ефективність застосування OLAP-технологій у системах екологічного моніторингу для аналізу динаміки індексів якості повітря (AQI, CAQI, EAQI) та контролю за перевищенням нормативних рівнів забруднювачів.

Розроблені звіти й KPI-візуалізації дозволяють виявляти тенденції, сезонні коливання та аномальні зміни, формуючи інформаційну основу для управлінських рішень у сфері довкілля.

Система має масштабований характер – її можна розширювати на інші екологічні показники та нормативи, визначені чинним законодавством,

забезпечуючи створення єдиної інтегрованої платформи аналітики стану навколишнього середовища.

Висновки до розділу 4

Система моніторингу, що представлена, функціонує більше 5 років. Це дозволило отримати більше 10 млн. записів оперативної інформації та близько 50 тис. агрегованих записів в сховищі даних. На основі дослідження цих даних були отримані такі результати.

- *Регресійний аналіз.* Для CAQI більш доцільним є використання погодинних даних, оскільки це дозволяє врахувати короточасні коливання та відобразити динаміку змін упродовж доби. Натомість для AQI найбільш інформативним виявилось застосування денних середніх значень, що відповідає методології його розрахунку та забезпечує стабільніший аналітичний результат.
- *Кластерний аналіз.* У межах проведеного дослідження виконано кластеризацію даних з метою виявлення станцій та забруднювачів із відхиленнями у роботі, неповнотою або аномальною поведінкою вимірювань. Застосування алгоритму “*K-Means*” дозволило згрупувати дані за схожими характеристиками та виокремити кластери, що відповідають стабільним станціям, а також ті, де фіксуються систематичні пропуски або перебої у роботі сенсорів.
- *Метод “Наївного Байєса”.* Після кластерного аналізу було визначено стабільні станції, дані з яких використано для навчання моделі “Наївного Байєса”. Побудовані на цій основі теплові карти відображають середні та максимальні значення параметрів для різних категорій якості повітря, дозволяючи оцінити внесок забруднювачів у формування індексів. Модель забезпечує класифікацію поточного стану атмосфери та відновлення пропущених даних за доступними вимірюваннями, що підвищує точність і надійність оцінки якості повітря.

- Пошук асоціативних правил визначив таке. Найсильніші знайдені правила (за значенням *lift* – 6.3–6.4) демонструють стійку комбінацію таких ознак: підвищені концентрації твердих частинок $PM_{2.5}/PM_{10}$ та газоподібних забруднювачів (NO_2 , SO_2), поєднані з низькими значеннями вологості (*humidity_low*) та наявністю «підсилюючих» пікових (максимальних або мінімальних) значень за добу.
- Побудовані *OLAP-звіти* та значення *KPI* дозволили оцінити результати, отримані методами Data Mining.

ВИСНОВКИ

У дисертаційній роботі вирішено актуальну науково-прикладну задачу розроблення інтелектуальної інформаційної технології для оцінювання стану атмосферного повітря з метою підтримки управлінських рішень у сфері охорони довкілля.

В ході роботи виконано такі теоретичні та експериментальні дослідження.

1. Аналіз предметної області засвідчив багатофакторну природу забруднення (антропогенні та природні джерела) і його медико-біологічні, екологічні та економічні наслідки. Системний аналіз об'єкта дослідження дозволив визначити, що це складна багаторівнева система, в якій поєднуються стаціонарні референтні пости, мобільні лабораторії і розподілені IoT-сенсори. За результатом аналізу існуючих методів і технічних рішень щодо реалізації моніторингу якості атмосферного повітря, визначено, що, попри позитивну динаміку, ця сфера в Україні потребує подальшого вдосконалення: впровадження уніфікованих методик, розбудови прогностичних моделей і більш тісної інтеграції з міжнародними системами. Відсутність єдиного централізованого сховища накопичених даних ускладнює проведення повноцінного аналізу та прогнозування ступеня небезпечності, особливо зважаючи на історичні дані.

2. Формалізовано алгоритми розрахунку інтегральних індексів (AQI, CAQI, EAQI), які надають можливість оцінити якість повітря в уніфікованому вигляді, використовуючи різні підходи. Їхнє одночасне використання в системі моніторингу дозволяє під різним кутом виявляти залежність якості повітря від забруднювачів. Розроблено регресійну модель для виявлення та кількісної оцінки залежностей між параметрами якості повітря та факторами, що на них впливають. Результати регресійного аналізу для індексів CAQI та AQI продемонстрували, що точність моделі зростає зі збільшенням часового періоду (інтервалу) та при використанні середніх денних значень.

3. Відповідно розробленій архітектурі побудована трирівнева система моніторингу, яка забезпечує отримання даних із сенсорів, відкритих API та

метеорологічних джерел, їх очищення, нормалізацію і збереження у реляційній базі даних; збереження інформації у сховище даних (Data Warehouse) з оригінальною багатовимірною моделлю для екологічних показників; оперативний (OLAP) та інтелектуальний аналіз (Data Mining).

4. Розроблено інформаційне забезпечення (СУБД PostgreSQL, MS SQL Server), використовуючи хмарні сервіси DigitalOcean та Microsoft Azure. Програмне забезпечення що реалізує весь необхідний функціонал, реалізовано на мові Python.

5. Розроблено та впроваджено ETL-процес, який забезпечує відбір (extraction), обробку (transformation) та завантаження (loading) даних. На етапі обробки дані узагальнюються (агрегуються) та очищуються. Для усунення проблеми відсутності або спотворення даних запропоновано використання двох методів Data Mining: кластерний аналіз і метод “Наївного Байєса”, точність якого становить 71%.

6. Результати використання асоціативних правил засвідчили, що найсильніші знайдені правила (за значенням *lift* – 6.3–6.4) демонструють стійку комбінацію таких ознак: підвищені концентрації твердих частинок $PM_{2.5}/PM_{10}$ та газоподібних забруднювачів (NO_2 , SO_2), поєднані з низькими значеннями вологості (*humidity_low*) та наявністю «підсилюючих» пікових (максимальних або мінімальних) значень за добу.

7. Для верифікації знайдених правил сформовано спеціальний аналітичний звіт, який проходить по всіх комбінаціях ознак, виявлених алгоритмом Apriori, та перевіряє їх фактичну спільну появу у вихідних даних. Цей звіт підтверджує результати, які отримані в процесі пошуку асоціативних правил.

8. Сформовані ключові показники ефективності дозволяють отримати удосконалені методи інтерпретації перевищень нормативних рівнів забруднення та продемонструвати тренди. На основі отриманих даних, у разі виявлення стійкого перевищення нормативних рівнів, формуються рекомендації для прийняття управлінських рішень.

9. Запропоновано комплексну інтелектуальну інформаційну технологію, що поєднує методи OLTP, OLAP і Data Mining. Вона забезпечує автоматизацію процесів спостереження, аналізу, прогнозування та підтримки прийняття управлінських рішень. Практичне впровадження запропонованої комплексної технології сприятиме подальшому розвитку системи державного моніторингу атмосферного повітря в Україні, створення, функціонування та координацію якої регламентує Постанова КМУ № 827[103].

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Popov O., Iatsyshyn A., Kovach V., Artemchuk V., Kameneva I., Taraduda D., Sobyna V., Sokolov D., Dement M., Yatsyshyn T. Risk Assessment for the Population of Kyiv, Ukraine as a Result of Atmospheric Air Pollution. Journal of Health and Pollution. 2020. Vol. 10, №. 25.
2. IQAir. Air quality in Kyiv. URL: <https://www.iqair.com/us/ukraine/kyiv/kyiv-c>.
3. Моніторинг якості атмосферного повітря: український та міжнародний досвід. Фундація «Відкрите суспільство». URL: https://osf.org.ua/data/blog_dwnl/Analitichna_zapiska_atmosferne_povitrya.pdf.
4. Нормативно-правове забезпечення охорони атмосферного повітря. Державна екологічна інспекція України. URL: <https://dei.gov.ua/post/2321>.
5. Законодавство у сфері охорони атмосферного повітря: досвід європейського союзу та український вимір. URL: http://www.lsej.org.ua/10_2021/152.pdf.
6. WHO Global Air Quality Guidelines: Particulate Matter (PM2.5 and PM10), Ozone, Nitrogen Dioxide, Sulfur Dioxide and Carbon Monoxide. World Health Organization (WHO). Geneva, 2021. 290 с. URL: <https://www.who.int/publications/i/item/9789240034228>.
7. U.S. Environmental Protection Agency (EPA). Technical Assistance Document for the Reporting of Daily Air Quality – the Air Quality Index (AQI). Research Triangle Park, 2018. 22 с. URL: <https://www.airnow.gov/sites/default/files/2020-05/aqi-technical-assistance-document-sept2018.pdf>.
8. Air quality in Europe – 2022 report. European Environment Agency (EEA). Luxembourg: Publications Office of the European Union, 2022. URL: <https://www.eea.europa.eu/en/analysis/publications/air-quality-in-europe-2022>.
9. SaveEcoBot. Поєднує дані про забруднення, забруднювачів та інструменти захисту довкілля. URL: <https://www.saveecobot.com>.
10. Громадський моніторинг якості повітря EcoCity. URL: <https://reborn.eco-city.org.ua>.

11. Автоматизована система моніторингу стану атмосферного повітря міста Києва (ASM Kyiv). URL: <https://asm.kyivcity.gov.ua>.
12. Платформа моніторингу та прогнозу якості повітря Агентства з охорони довкілля США (AirNow). URL: <https://www.airnow.gov>.
13. Карта якості повітря та пожежної активності Агентства з охорони довкілля США (AirNow Map). URL: <https://fire.airnow.gov>.
14. Платформа відкритих даних про якість атмосферного повітря (OpenAQ). URL: <https://explore.openaq.org>.
15. Всесвітня платформа моніторингу якості атмосферного повітря (AQICN). URL: <https://aqicn.org>.
16. Київська міська державна адміністрація (КМДА). URL: <https://kyivcity.gov.ua>.
17. Настанови та принципи оповіщення населення про якість повітря, радіаційну та хімічну небезпеку. URL: <https://cleanair.org.ua/wp-content/uploads/2022/12/cleanair.org.ua-radiation-and-smog-alarm-uaqi-radiation-and-smog-alarm-2022-1.pdf>.
18. Платформа моніторингу якості повітря ЛУН Air. URL: <https://lun.ua/misto/air>.
19. Боголюбов В. М., Голуб Б. Л. Інформаційно-аналітична система оцінювання стану атмосферного повітря. Сталий розвиток – XXI ст.: матеріали VII Міжнародної науково-практичної конференції. Київ, 2021. 527 с. С. 235–246.
20. Індекс якості атмосферного повітря. Агентство з охорони довкілля США (AirNow). URL: <https://www.airnow.gov/aqi>.
21. Common Air Quality Index. URL: <https://doi.org/10.1016/j.scitotenv.2013.10.060>.
22. European Air Quality Index. URL: <https://airindex.eea.europa.eu/AQI/index.html>.
23. Air Quality Index Calculation. URL: <https://www.airnow.gov/aqi/aqi-calculator>.

24. Common Air Quality Index Calculation. URL: <https://airly.org/en/air-quality-index-caqi-and-aqi-methods-of-calculation>.
25. European Air Quality Index Calculation. URL: <https://ecmwf-projects.github.io/copernicus-training-cams/proc-aq-index.html>.
26. Shevchenko D. V., Holub B. L. Regression analysis as a tool for identifying patterns in atmospheric air monitoring data. Proceedings of the 5th Edge Computing Workshop. 2025. Vol. 4. P. 20–27.
27. Enhancing air quality sensor accuracy using machine learning. Scientific Reports, 2024. URL: <https://www.nature.com/articles/s41612-024-00833-9>.
28. Dhanalakshmi M., Radha V. Discretized linear regression and multiclass support vector based air pollution forecasting technique. International Journal of Engineering Trends and Technology. 2022. Vol. 70, №11. P. 315–323.
29. Ordinary Least Squares (OLS). URL: <https://www.datacamp.com/tutorial/ols-regression>.
30. Шевченко Д. В., Голуб Б. Л. Застосування методів Data Mining для багатовимірного аналізу якості атмосферного повітря на основі екологічних даних. Науковий журнал «Наука і техніка сьогодні». Серія: Техніка. 2025. № 8 (49). С. 1801–1810.
31. K-Means Clustering in Python Tutorial. URL: <https://www.datacamp.com/tutorial/k-means-clustering-python>.
32. Naive Bayes Classifier with Scikit-learn Tutorial. URL: <https://www.datacamp.com/tutorial/naive-bayes-scikit-learn>.
33. Gaussian Naive Bayes Algorithm Explained. URL: <https://builtin.com/artificial-intelligence/gaussian-naive-bayes>.
34. Apriori Algorithm Tutorial for Association Rule Learning. URL: <https://www.datacamp.com/tutorial/apriori-algorithm>.
35. Official Power BI Documentation. URL: <https://learn.microsoft.com/en-us/power-bi>.
36. Microsoft Learn – Analytics pane in Power BI Desktop. URL: <https://learn.microsoft.com/en-us/power-bi/transform-model/desktop-analytics-pane>.

37. Ковзне середнє (Moving Average). GoodCrypto – Посібник для трейдерів: прості, експоненційні та зважені ковзні середні (SMA, EMA, WMA). URL: <https://goodcrypto.app/uk/kovzni-seredni-sma-ema-wma-posibnik-dlya-trejderiv-na-prikladi-goodcrypto>.
38. Montgomery D.C. Applied Regression Analysis. – New York : John Wiley & Sons, 2012. 816 p.
39. Microsoft Power BI Documentation. URL: <https://docs.microsoft.com/power-bi>.
40. Ключовий показник ефективності. URL: <https://peopleforce.io/uk/hr-glossary/key-performance-indicator>.
41. Power BI KPI // Microsoft Learn – KPI visual in Power BI. URL: <https://learn.microsoft.com/en-us/power-bi/visuals/power-bi-visualization-kpi>.
42. Toma C., Alexandru A., Popa M., Zamfiroiu A. IoT solution for smart cities' pollution monitoring and the security challenges. Sensors. 2019. Vol. 19, №15, Art. 3401.
43. Шевченко Д. В., Голуб Б. Л. Моніторинг якості повітря в реальному часі. Науковий журнал «Математичні машини і системи». 2025. № 1. С. 103–112.
44. Holub B., Kyrychenko V., Nikolaienko D., Lendiel M., Shevchenko D., Khomenko A. Algorithmic and information support in atmospheric air quality monitoring systems. Proceedings of the Tenth International Congress on Information and Communication Technology (ICICT 2025). Lecture Notes in Networks and Systems. 2025. Vol. 1441. P. 399–412.
45. Setiawan F.N., Kustiawan I. IoT Based Air Quality Monitoring. IOP Conference Series: Materials Science and Engineering. 2018. Vol. 384. 012008.
46. Understanding the differences between MQTT and WebSockets for IoT. URL: <https://www.hivemq.com/blog/understanding-the-differences-between-mqtt-and-websockets-for-iot>.
47. Camacho-León S. Comparative Analysis of Power Consumption between MQTT and HTTP Protocols in an IoT Platform Designed and Implemented for Remote

Real-Time Monitoring of Long-Term Cold Chain Transport Operations. Sensors. 2023. Vol. 23, №10, 4896.

48. MQTT vs HTTP for IoT. URL: <https://www.hivemq.com/blog/mqtt-vs-http-protocols-in-iiot>.
49. MQTT vs WebSocket: Key Differences & Applications. URL: <https://www.emqx.com/en/blog/mqtt-vs-websocket>.
50. Мова програмування Python. URL: <https://www.python.org>.
51. Django is a high-level Python web framework. URL: <https://www.djangoproject.com>.
52. Django REST Framework. URL: <https://www.django-rest-framework.org>.
53. Celery - Distributed Task Queue. URL: <https://docs.celeryq.dev/en/stable>.
54. PostgreSQL: The World's Most Advanced Open Source Relational Database. URL: <https://www.postgresql.org>.
55. Microsoft SQL Server. URL: <https://www.microsoft.com/en-us/sql-server>.
56. Vue.js: Progressive JavaScript Framework. URL: <https://vuejs.org>.
57. Nginx: High-Performance HTTP Server and Reverse Proxy. URL: <https://nginx.org>.
58. Online Analytical Processing (OLAP). URL: <https://aws.amazon.com/what-is/olap>.
59. Understanding star-schema and the importance for Power BI. URL: <https://learn.microsoft.com/en-us/power-bi/guidance/star-schema>.
60. Star schema vs Snowflake schema. URL: <https://www.integrate.io/blog/snowflake-schemas-vs-star-schemas-what-are-they-and-how-are-they-different>.
61. Шевченко Д. В., Голуб Б. Л. Багатовимірна аналітика екологічних даних: застосування OLAP у системах моніторингу. Науковий журнал «Математичні машини і системи». 2025. № 3–4. С. 54–65.
62. Difference Between ETL and ELT. URL: <https://aws.amazon.com/compare/the-difference-between-etl-and-elt>.

63. Розділяй і володарюй. Як працюють процеси в Python. URL: <https://dou.ua/forums/topic/53004>.
64. Паралельність на основі потоків у Python. URL: <https://docs.python.org/uk/3.10/library/threading.html>.
65. Building a real-time IoT sensor data dashboard with Django, MQTT, React, and ESP32. URL: <https://medium.com/@shiva10/building-a-real-time-iot-sensor-data-dashboard-with-django-mqtt-react-and-esp32-7132198f522a>.
66. Celery Batches. URL: <https://celery-batches.readthedocs.io/en/latest>.
67. Multiple databases in Django. URL: <https://docs.djangoproject.com/en/5.2/topics/db/multi-db>.
68. Scikit-learn library. URL: <https://scikit-learn.org/stable>.
69. Gaussian Naive Bayes (GaussianNB). URL: https://scikit-learn.org/stable/modules/generated/sklearn.naive_bayes.GaussianNB.html.
70. Scikit-learn: StandardScaler. URL: <https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.StandardScaler.html>.
71. DigitalOcean (Droplets). URL: <https://www.digitalocean.com/products/droplets>.
72. Microsoft Azure. URL: <https://azure.microsoft.com/en-us>.
73. Supervisor: Process Control System for UNIX. URL: <https://supervisord.org>.
74. Django Admin Site. URL: <https://docs.djangoproject.com/en/5.2/ref/contrib/admin>.
75. An Overview of Experiences Made and Tools Used to Inform the Public on Ambient Air Quality. URL: <https://www.mdpi.com/2073-4433/12/11/1524>.
76. European Environment Agency. Air quality in European cities: 2023 report. 2023, № 9, P. 1–188. URL: <https://op.europa.eu/en/publication-detail/-/publication/f2bf4bce-72ee-11ee-9220-01aa75ed71a1/language-en>.

77. Shevchenko D., Golub B. Information service of the atmospheric air quality monitoring system. Proceedings of the IX International Scientific and Practical Conference «Scientific Problems and Options for Their Solution», Bucharest, Romania. 7–9 February 2024. P. 55–60.
78. Shahid S., Brown D.J., Wright P., Khasawneh A.M., Taylor B., Kaiwartya O. Innovations in Air Quality Monitoring: Sensors, IoT and Future Research. *Sensors*. 2025. Vol. 25, № 7, 2070.
79. Rollo F., Bachechi Ch., Po L. Anomaly Detection and Repairing for Improving Air Quality Monitoring. *Sensors*. 2023. URL: <https://www.mdpi.com/1424-8220/23/2/640>.
80. Azevedo R., Silva J.P., Lopes N., Curado A., Nunes L.J.R., Lopes S.I. Designing an IoT-enabled Data Warehouse for Indoor Radon Time Series Analytics: Proc. of the 17th Iberian Conference on Information Systems and Technologies (CISTI), IEEE. 2022. URL: <https://ouci.dntb.gov.ua/en/works/7WQLQwzl>.
81. Setiawan, F.N.; Kustiawan, I. IoT based Air Quality Monitoring. In Proceedings of the 2018 IOP Conference Series: Materials Science and Engineering, Bandung, Indonesia, 16 November 2017; Volume 384, p. 012008
82. Holovchenko T.O. 2021. Legislation in the field of atmospheric air protection: the experience of the European Union and the Ukrainian dimension. *Juridical scientific and electronic journal*, № 10: 578–581.
83. Rahman M.M., Hossain M.S., Islam M.T., Saha S.K., Paul S., et al. Secure IoT-enabled air quality monitoring system. *Sensors*. 2022. Vol. 22, № 21. Art. 8241.
84. Murena, F. Measuring air quality over large urban areas: Development and application of an air pollution index at the urban area of Naples. *Atmos. Environ.* 2004, №38, P. 6195–6202.
85. Dhanalakshmi & Radha (2022). IoT-based air pollution forecasting with machine learning. URL: <https://arxiv.org/abs/2211.15095>.
86. Why we investigate the composition of the atmosphere. URL: <https://atmosphere.copernicus.eu/why-we-investigate-composition-atmosphere>.

87. Khomenko A. O., Kyrychenko V. V. The efficiency of sorting algorithms with an emphasis on the processing of Ukrainian Cyrillic in Python and Cython. Proceedings of the International Scientific and Practical Conference “Actual Issues of the Development of Science and Technology in the Conditions of Globalization”. Boyarka, 2024. URL: <https://nubip.edu.ua/sites/default/files/u423/materials22.pdf>.
88. Bruno, F.; Cocchi, D. A unified strategy for building simple air quality indexes. *Environ. Metr.* 2002, № 13, P. 243–261.
89. Шевченко Д. В., Голуб Б. Л. Архітектура інформаційного компоненту системи моніторингу якості атмосферного повітря. Інформаційне суспільство: технологічні, економічні та технічні аспекти становлення: Міжнародна наукова інтернет-конференція «Інформаційне суспільство: технологічні, економічні та технічні аспекти становлення», м. Тернопіль, 18–19 січня 2024 року: тези доповіді. № 85. С. 75–78.
90. Обговорення звіту КМДА про результати моніторингу якості атмосферного повітря у місті Києві за 2023 рік. URL: <https://nubip.edu.ua/node/141976>.
91. Understanding Ordinary Least Squares (OLS). URL: <https://medium.com/@dahami/understanding-ordinary-least-squares-ols-and-its-applications-in-statistics-machine-learning-ad2c13681501>.
92. Bishoi, B.; Prakash, A.; Jain, V.A. Comparative Study of Air Quality Index Based on Factor Analysis and US-EPA Methods for an Urban Environment. *Aerosol Air Qual. Res.* 2009, № 9, P. 1–17.
93. Karavas, Z.; Karayannis, V.; Moustakas, K. Comparative study of air quality indexes in the European Union towards adopting a common air quality index. *Energy Environ.* 2020, № 32, P. 959–980.
94. Garcia A., Saez Y., Harris I., Huang X., Collado E. Advancements in air quality monitoring: a systematic review of IoT-based air quality monitoring and AI technologies. *Artificial Intelligence Review.* 2025. Vol. 58, art. 275.

95. Buelvas J., Múnera D., Tobón D.P., Aguirre J., Gaviria N. Data Quality in IoT-Based Air Quality Monitoring Systems: a Systematic Mapping Study. *Water, Air, & Soil Pollution*. 2023. Vol. 234, art. 248.
96. Azevedo R., Silva J.P., Lopes N., Curado A., Nunes L.J.R., Lopes S.I. Designing an IoT-enabled Data Warehouse for Indoor Radon Time Series Analytics. *Proc. of the 17th Iberian Conference on Information Systems and Technologies (CISTI 2022)*, Madrid, Spain, June 22–25, 2022. P. 620.
97. Hassan A.K., Saraya M.S., Ali-Eldin A.M.T., Abdelsalam M.M. Low-Cost IoT Air Quality Monitoring Station Using Cloud Platform and Blockchain Technology. *Applied Sciences*. 2024. Vol. 14, № 13, 5774.
98. Wang Z., Chen Z., Shahid I., Asif Z., Haghighat F. Indoor Air Quality Assessment Through IoT Sensor Technology: A Montreal-Qatar Case Study. *Atmosphere*. 2023. Vol. 16, № 5, 574.
99. Tastan M. Machine Learning-Based Calibration and Performance Evaluation of Low-Cost Internet of Things Air Quality Sensors. *Sensors*. 2025. Vol. 25, № 10, 3183.
100. Ravindra K., Kumar S., Kumar A., Singh S., Mor S. Enhancing accuracy of air quality sensors with machine learning to augment large-scale monitoring networks. *Climate and Atmospheric Science*. 2024. Vol. 7, art. 326.
101. Devi M.S., Rahamathulla V. Air Quality Through IoT and Big Data Analytics. *Advances in Data Science and Management*. 2020. P. 181–187.
102. Ali S.K., Abbood G.H., Asker Z.A., Hussain A. A comprehensive review of sensor technologies in IoT: technical aspects, challenges, and future directions. *Computers*. 2025. Vol. 14, № 8, 342.
103. Постанова КМУ № 827. URL: <https://zakon.rada.gov.ua/laws/show/827-2019-п>.
104. Шевченко Д. В. Розроблення інтелектуальної інформаційної технології для оцінювання стану атмосферного повітря. Продовольча та екологічна безпека в умовах війни та повоєнної відбудови: виклики для України та світу, до 125-річчя НУБіП: Міжнародна науково-практична конференція, м. Київ, 25–26 травня 2023.

105. Шевченко Д. В. Розроблення інтелектуальної інформаційної технології для оцінювання стану атмосферного повітря. Теоретичні та прикладні аспекти розробки комп'ютерних систем 2023: V Всеукраїнська науково-практична конференція студентів і аспірантів «Теоретичні та прикладні аспекти розробки комп'ютерних систем 2023», м. Київ, 2023. С. 10–11.

106. Шевченко Д. В. Розроблення інтелектуальної інформаційної технології для оцінювання стану атмосферного повітря. Інформаційні технології: економіка, техніка, освіта 2023: XIV Міжнародна науково-практична конференція молодих вчених «Інформаційні технології: економіка, техніка, освіта 2023», м. Київ, 26–27 жовтень 2023. С. 240–241.

107. Шевченко Д. В. Архітектура для системи моніторингу та аналізу якості атмосферного повітря. Актуальні питання розвитку науки та техніки в умовах глобалізації: Міжнародна науково-практична конференція «Актуальні питання розвитку науки та техніки в умовах глобалізації», м. Боярка, 14 травня 2024. С. 117–119.

108. Шевченко Д. В. Регресійний аналіз як інструмент для виявлення закономірностей у даних моніторингу атмосферного повітря. Глобальні та регіональні проблеми інформатизації в суспільстві і природокористуванні: XII Міжнародна науково-практична конференція «Глобальні та регіональні проблеми інформатизації в суспільстві і природокористуванні», м. Київ, 2024.

109. Шевченко Д. В. OLAP-технології у системах моніторингу та аналізу якості атмосферного повітря. Актуальні питання розвитку науки та техніки в умовах глобалізації: II Міжнародна науково-практична конференція «Актуальні питання розвитку науки та техніки в умовах глобалізації», м. Боярка, 14 травня 2025. С. 86–88.

110. Шевченко Д. В. Багатовимірна аналітика екологічних даних: застосування OLAP у системах моніторингу. Теоретичні та прикладні аспекти розробки комп'ютерних систем 2025: Всеукраїнська науково-практична конференція студентів і аспірантів «Теоретичні та прикладні аспекти розробки комп'ютерних систем 2025», м. Київ, 24 квітня 2025. С. 122–123.

111. Lysenko V., Lendiel T., Bolbot I., Pavlov S. Mobile system for monitoring plant environment parameters for biogas production. *Machinery & Energetics*. 2023. Vol. 14 (4). P. 111–120.
112. М. І. Лендел, Т. І. Лендел, І. М. Болбот База даних реального часу підсистеми моніторингу процесу вирощування овочевої продукції в теплиці. *Енергетика і автоматика*. 2021. № 4. С. 128–136.
113. О.В.Гудзь, Б.Л.Голуб, О.В.Бушма Тестування системи аналізу даних для технології прогнозування шкідливих речовин. *Математичні машини і системи*, 2021, № 3. С. 113–120.
114. Ganna Weigang, Kateryna Komar Multi-agent modeling of traffic organization in urban agglomerations. *Journal of Transport technologies*. 2024. Vol. 5, № 1. P. 10–22.
115. Weigang G., Komar K. Analysis of the Security of on-Board Information Systems in Vehicles. *Journal of Transport technologies*. Vol. 4, № 1, 2023. P. 62–72.
116. В.М. Триснюк, Д.В. Варавін, Т.В. Триснюк. Геомоделі екологічної безпеки урбанізованого середовища в умовах пандемічних загроз на прикладі м. Києва. *Екологічна безпека та збалансоване ресурсокористування*, № 1(25), 2022. С. 7–13.
117. В.М. Триснюк, Т.В. Триснюк, А.В. Курило, Ю.М. Голован, Є.Ю. Пащенко. Системний аналіз інформаційно-аналітичного забезпечення органів адміністративного управління. *Сучасні інформаційні системи*. Науковий журнал. Том 6, № 2. Харків 2022. С. 37-41.
118. А. Аль-Амморі, Р. М. Іщенко, А. Є. Клочан, О. П. Шкурко, Х. А. Аль-Амморі. Інформаційні технології стохастичної моделі надійності компютеризованих систем для захисту інформації. *ЗВ'ЯЗОК*, № 4, 2024. С. 10-18.
119. Алі Аль-Амморі, Д.В. Дзюбан. Інформаційні технології підтримки прийняття рішень при аналізі надійності на автомобільному транспорті. *Наука і техніка сьогодні*. 2025. № 7 (48) С.1213-1224.

ДОДАТКИ

Алгоритм функціонування сервісу *KMDAService* для збору даних зі станцій міського екологічного моніторингу

```

class KMDADSensorIndicator(BaseModel):
    timestamp: int
    value: float

class KMDADSensor(BaseModel):
    id: Union[int, str]
    sensor: str
    description: str

class KMDASensorsResult(BaseModel):
    id: int
    sensors: List[KMDADSensor]

class KMDAService(CollectServiceBase):
    def __init__(self, station_url: str, station_data_url: str):
        self.station_url = station_url
        self.station_data_url = station_data_url
        self.intervals = ['twenty_minutes', 'hour', 'day']
        self.headers = {'User-Agent': 'Mozilla/5.0 (X11; Linux x86_64)'}

    @catch_exceptions(default=None)
    def _convert_time(self, _time: int):
        timestamp = _time / 1000 # Convert milliseconds to seconds
        return datetime.utcfromtimestamp(timestamp)

    def _get_indicators(self, device_id: str) -> KMDASensorsResult:
        url = f'{self.station_url}?{DEVICE_KEY}={device_id}'
        response = requests.get(url=url, headers=self.headers)
        return KMDASensorsResult(**response.json())

    @catch_exceptions(default=None)
    def _get_indicator_data(
        self, sensor_id: str, interval: str
    ) -> List[Tuple[datetime, float]]:
        url = f'{self.station_data_url}?sensor_id={sensor_id}&interval={interval}'
        response = requests.get(url=url, headers=self.headers)
        return [(self._convert_time(t), v) for t, v in response.json()]

    def process_station(
        self, station: Station
    ) -> List[Tuple[MeasuredUnit, List[Tuple[datetime, float]]]]:
        results = []
        if not station.is_kmda:
            return results

        measured_units = station.measured_units.all()
        measured_units_ids = [x.id for x in measured_units]

```

```

        for kmda_measured_unit in
self._get_indicators(station.station_kmda.identifier).sensors:
            system_pk =
KMDA_MEASUREMENT_UNITS_MAPPER_2.get(kmda_measured_unit.sensor, 0)
            if system_pk not in measured_units_ids:
                continue

            dataset: list = []
            measured_unit = list(filter(lambda x: x.id == system_pk,
measured_units))[0]
            for interval in self.intervals:
                measured_unit_results =
self._get_indicator_data(kmda_measured_unit.id, interval)
                if (measured_unit_results is not None and len(measured_unit_results) >
0):
                    dataset = measured_unit_results
                    break
                else:
                    continue

            if len(dataset) > 0:
                results.append((measured_unit, dataset))

    return results

```

Реалізація програмного модуля *AqiOwnCalculator* для розрахунку індексу якості атмосферного повітря

```

class AqiOwnCalculator:
    """ Calculate own Air Quality Index (AQI - United States) """
    def __init__(self):
        self.measured_units = {
            x.code: x for x in MeasuredUnit.objects.filter(
                code__in=['pm2.5', 'o3', 'no2', 'so2', 'co']
            )
        }
        # ..., etc initializations

    @classmethod
    def _prepare_range_period(
        cls, period: int
    ) -> Tuple[datetime, datetime]:
        """ Prepare range period by period type """
        now = timezone.now()
        if period == BreakPoint.HOUR_1:
            return now - timedelta(hours=1), now
        elif period == BreakPoint.HOUR_8:
            return now - timedelta(hours=8), now
        elif period == BreakPoint.HOUR_24:
            return now - timedelta(hours=24), now

    def _find_break_points(
        self, value: float, break_points: List[BreakPoint],
    ) -> Optional[Tuple[BreakPoint, BreakPoint]]:
        """ Find break points (Measured unit & Aqi) """
        for break_point in break_points:
            if break_point.bottom <= value <= break_point.upper:
                return break_point, list(filter(
                    lambda x: x.category == break_point.category,
                    self.aqi_break_points
                ))[0]
        return None

    @classmethod
    def _calculate_index(

```

```

        cls,
        value: float,
        i_low: float, i_high: float,
        c_low: float, c_high: float,
    ) -> float:
        """ Algorith: https://en.wikipedia.org/wiki/Air\_quality\_index """
        return ((i_high - i_low) / (c_high - c_low)) * (value - c_low) + i_low

def calculate_index(self, source_type: int):
    """ Calculate Air Quality Index(Own) """
    annotate_data = {}
    measured_units_codes = defaultdict(list)
    for code, measured_unit in self.measured_units.items():
        for time_period in self.measured_units_break_points[code].keys():
            annotate_key = f'{code}_{time_period}_avg'
            measured_units_codes[code].append(annotate_key)

            start, end = self._prepare_range_period(time_period)
            annotate_data[annotate_key] = Subquery(
                Measurement.objects.filter(
                    created_at__gte=start, created_at__lte=end,
                    measured_unit=measured_unit,
                    station=OuterRef('id')
                ).values(
                    'station',
                ).annotate(
                    avg=Avg('value'),
                ).values('avg')[:1]
            )

    stations = Station.objects.filter(
        status=Station.ACTIVE,
        source_type=source_type,
        measured_units__code__in=[
            'aqi_own', *self.measured_units.keys(),
        ]
    ).annotate(
        **annotate_data,
    ).prefetch_related(
        'measured_units',
    ).distinct()

    to_create = []

```

```

aqi_own = MeasuredUnit.objects.get(code='aqi_own')
for station in stations:
    station_index = 0
    for measured_unit in station.measured_units.all():
        measured_unit_index = 0
        for time_period, break_points in (
            self.measured_units_break_points[measured_unit.code].items()
        ):
            annotate_key = f'{measured_unit.code}_{time_period}_avg'
            if hasattr(station, annotate_key):
                value = getattr(station, annotate_key)
                if value is None:
                    continue

            found_results = self._find_break_points(
                value, break_points
            )
            if isinstance(found_results, tuple):
                _i, _c = found_results
                _index = self._calculate_index(
                    value, _i.bottom, _i.upper, _c.bottom, _c.upper
                )
                if _index > measured_unit_index:
                    measured_unit_index = _index
            if measured_unit_index > station_index:
                station_index = measured_unit_index
        to_create.append(
            Measurement(
                station=station, measured_unit=aqi_own,
                created_at=timezone.now(), value=station_index,
            )
        )

if to_create:
    Measurement.objects.bulk_create(to_create)

```

Реалізація сервісу *WareHouseExporterService* для автоматизації обчислень та ETL-процесів у системі моніторингу атмосферного повітря

```
class WareHouseExporterService:
    def __init__(self, start: date, end: date):
        self.start_date_time = from_date_to_datetime(start, time.min)
        self.end_date_time = from_date_to_datetime(end, time.max)

    # ..., other internal methods

    @timed_lru_cache(seconds=60 * 10) # 10 minutes
    def _init_threshold_levels_dim(self) -> Tuple[dict, dict]:
        measured_unit_dim = self._init_measured_unit_dim()
        measured_units = {}
        for x in MeasuredUnit.objects.filter(
            code__in=["pm10", "pm2.5", "no2", "so2", "co", "o3"]
        ):
            measured_units[x.code] = x

        threshold_levels = {
            "pm10": [
                [None, "24h", "Upper", 35, "Year", 35],
                [None, "24h", "Lower", 25, "Year", 35],
                [None, "year", "Upper", 28, None, None],
                [None, "year", "Lower", 20, None, None],
            ],
            "pm2.5": [
                [None, "year", "Upper", 17, None, None],
                [None, "year", "Lower", 12, None, None],
            ],
            "no2": [
                [None, "1h", "Upper", 140, "Year", 18],
                [None, "1h", "Lower", 100, "Year", 18],
            ],
            "so2": [
                [None, "24h", "Upper", 75, "Year", 3],
                [None, "24h", "Lower", 50, "Year", 3],
            ],
            "co": [
                [None, "8h", "Upper", 7, None, None],
```

```

        [None, "8h", "Lower", 5, None, None],
    ],
    "o3": [
        [None, "1h", "Upper", 180, None, None],
        [None, "8h", "Lower", 120, "Year", 25],
    ],
}

with connections['warehouse'].cursor() as cursor:
    for code, measured_unit in measured_units.items():
        for threshold_level_fields in threshold_levels[code]:
            measured_unit_dim_id = measured_unit_dim[measured_unit.id]
            _, avg_period, threshold_type, threshold_value, limit_type,
limit_value = threshold_level_fields

            parameters = [avg_period, threshold_type, threshold_value]
            cursor.execute(sql_threshold_level_dim_check,
[measured_unit_dim_id, *parameters])
            if x_results := cursor.fetchone():
                threshold_level_fields[0] = x_results[0]
            else:
                cursor.execute(
                    sql_threshold_level_dim_insert,
                    [measured_unit_dim_id, *parameters, limit_type,
limit_value],
                )
                threshold_level_fields[0] = cursor.fetchone()[0]

    return measured_units, threshold_levels

def _prepare_measurement_fact(self) -> List[dict]:
    queryset: List[dict] = Measurement.objects.filter(
        created_at__gte=self.start_date_time,
        created_at__lte=self.end_date_time,
    ).annotate(
        date=TruncDay('created_at')
    ).values(
        'measured_unit_id', 'station__city', 'date',
    ).annotate(
        avg_value=Avg('value'),
        min_value=Min('value'),
        max_value=Max('value'),
    ).order_by('date')
```

```

return queryset

def _prepare_threshold_exceed_fact(
    self, measured_unit_id: int, avg_period: str
) -> List[dict]:
    annotate_data = {}
    if avg_period == "year":
        annotate_data["date"] = TruncYear('created_at')
    if avg_period == "24h":
        annotate_data["date"] = TruncDay('created_at')
    elif avg_period == "8h":
        annotate_data["date"] = TruncDay("created_at")
        annotate_data["hour_bucket"] = ExpressionWrapper(
            ExtractHour("created_at") / 8, output_field=IntegerField()
        )
    elif avg_period == "1h":
        annotate_data["date"] = TruncHour('created_at')
    else:
        return []

    values = ["station__city", "date"]
    if avg_period == "8h":
        values = ["station__city", "date", "hour_bucket"]

    ordering = ["date"]
    if avg_period == "8h":
        ordering = ["date", "hour_bucket"]

    queryset: List[dict] = Measurement.objects.filter(
        created_at__gte=self.start_date_time,
        created_at__lte=self.end_date_time,
        measured_unit_id=measured_unit_id,
    ).annotate(
        **annotate_data,
    ).values(
        *values,
    ).annotate(
        avg_value=Avg('value'),
    ).order_by(
        *ordering,
    )
    return queryset

```

```

@check_performance_time
def export_measurement_fact(self):
    date_dim = self._init_date_dim()
    location_dim = self._init_location_dim()
    measured_unit_dim = self._init_measured_unit_dim()

    to_insert = []
    for fact_data in self._prepare_measurement_fact():
        date_pk = date_dim[fact_data["date"].date()]
        location_pk = location_dim[fact_data["station_city"]]
        measured_unit_pk = measured_unit_dim[fact_data["measured_unit_id"]]

        to_insert.append((
            date_pk, location_pk, measured_unit_pk,
            fact_data["avg_value"], fact_data["min_value"],
fact_data["max_value"],
            date_pk, location_pk, measured_unit_pk,
        ))

    with connections['warehouse'].cursor() as cursor:
        cursor.executemany(sql_measurement_fact_insert, to_insert)

@check_performance_time
def export_threshold_exceed_fact(self):
    date_dim = self._init_date_dim()
    location_dim = self._init_location_dim()
    measured_units, threshold_levels_all = self._init_threshold_levels_dim()

    for code, threshold_levels in threshold_levels_all.items():
        to_insert = []
        measured_unit = measured_units[code]
        for i in range(0, len(threshold_levels), 2):
            threshold_upper = threshold_levels[i]
            threshold_lower = threshold_levels[i + 1]
            avg_period = (threshold_upper or threshold_lower)[1]

            threshold_upper_counter = defaultdict(lambda: defaultdict(int))
            threshold_lower_counter = defaultdict(lambda: defaultdict(int))
            for fact_data in
self._prepare_threshold_exceed_fact(measured_unit.id, avg_period):
                avg_value = fact_data["avg_value"]
                if avg_value >= threshold_upper[3]:

```

```

threshold_upper_counter[fact_data["date"].date()][fact_data["station__city"]] += 1
    elif threshold_lower[3] <= avg_value < threshold_upper[3]:

threshold_lower_counter[fact_data["date"].date()][fact_data["station__city"]] += 1
    dataset = [
        [threshold_upper_counter, threshold_upper[0]],
        [threshold_lower_counter, threshold_lower[0]],
    ]
    for threshold_counter, threshold_pk in dataset:
        for threshold_date, cities in threshold_counter.items():
            date_pk = date_dim[threshold_date]
            for city, threshold_count in cities.items():
                location_pk = location_dim[city]

                to_insert.append((
                    date_pk, location_pk, threshold_pk,
                    threshold_count,
                    date_pk, location_pk, threshold_pk,
                ))
    if to_insert:
        with connections['warehouse'].cursor() as cursor:
            cursor.executemany(sql_threshold_exceed_fact_insert, to_insert)

```

Реалізація модуля *RegressionAnalysisService* для регресійного аналізу даних забруднення атмосферного повітря

```

class BaseRegressionAnalyzer(ABC):
    regression_type: str = 'default'
    regression_type_name: str = 'Default'

    def __init__(self, measurements_file: str):
        self.measurements_file = measurements_file
        self.measurements_data = pd.read_csv(self.measurements_file)
        self.dependent_variable: str | None = None
        self.independent_variables: list[str] | None = None

    def set_dependent_variable(self, dependent_variable: str):
        self.dependent_variable = dependent_variable

    def set_independent_variables(self, independent_variables: list[str]):
        self.independent_variables = independent_variables

    @abstractmethod
    def do_predicts(self, x, y):
        pass

    def do_regression_analysis(self):
        measurement_data =
self.measurements_data[self.measurements_data['measured_unit'] ==
self.dependent_variable]
        measurement_data = measurement_data[['value', 'timestamp']]

        independent_vars = self.measurements_data[
            self.measurements_data['measured_unit'].isin(self.independent_variables)
        ]

        independent_vars_pivot = independent_vars.pivot(
            index='timestamp', columns='measured_unit', values='value'
        )

        merged_data = pd.merge(measurement_data, independent_vars_pivot,
on='timestamp', how='inner')

```

```

merged_data.rename(columns={'value': self.dependent_variable}, inplace=True)
merged_data = merged_data.dropna()

X = merged_data[self.independent_variables]
Y = merged_data[self.dependent_variable]
X = sm.add_constant(X)

measurement_true = merged_data[self.dependent_variable]
measurement_predicted, regression_summary = self.do_predicts(X, Y)

plt.title(f"True Relationship vs {self.regression_type_name} Predictions")
plt.figure(figsize=(10, 6))

plt.xlabel("True")
plt.ylabel("Predicted")
plt.plot(measurement_true, measurement_true, color="blue", label="True",
linewidth=2)
plt.scatter(measurement_true, measurement_predicted, alpha=0.5, color="green",
label="Predicted")

plt.legend()
plt.grid(True)
image_buffer = io.BytesIO()
plt.savefig(image_buffer, format='png')
image_buffer.seek(0)
plt.close()

return regression_summary, image_buffer

class OLSRegressionAnalyzer(BaseRegressionAnalyzer):
    regression_type: str = 'ols'
    regression_type_name: str = 'Ordinary Least Squares'

    def do_predicts(self, x, y):
        ols_model = sm.OLS(y, x).fit()
        return ols_model.fittedvalues, ols_model

```

Реалізація модуля *NaiveBayesAnalyzer* для аналізу даних забруднення атмосферного повітря

```

class NaiveBayesAnalyzer:
    class_bins_map: dict = {"caqi": [0, 25, 50, 75, 100, 500]}
    class_labels_map: dict = {"caqi": ["Excellent", "Fine", "Moderate", "Poor", "Very
    Poor"]}
    class_measured_units_map: dict = {"caqi": ["caqi", "pm2.5", "pm10", "no2", "so2",
    "ws", "humidity"]}

    def __init__(self, unit='caqi', test_size=0.2, random_state=42):
        self.unit = unit.lower()
        self.test_size = test_size
        self.random_state = random_state
        self.model = None
        self.features = None
        self.df_pivot = None
        self.x_train = None
        self.x_test = None
        self.y_train = None
        self.y_test = None

    def get_pure_data_frame(self):
        unit_codes = self.class_measured_units_map[self.unit]
        placeholders = ', '.join(['%s'] * len(unit_codes))

        sql_query = f"""
        SELECT
            m.date_dim_id,
            d.year,
            d.month,
            d.day,
            m.location_dim_id,
            l.city,
            l.region,
            m.measured_unit_dim_id,
            u.code AS unit_code,
            u.name AS unit_name,
            m.avg_value,
            m.min_value,

```

```

        m.max_value
    FROM MeasurementFact m
    JOIN DateDim d ON m.date_dim_id = d.id
    JOIN LocationDim l ON m.location_dim_id = l.id
    JOIN MeasuredUnitDim u ON m.measured_unit_dim_id = u.id
    WHERE u.code IN ({placeholders});
"""
with connections['warehouse'].cursor() as cursor:
    cursor.execute(sql_query, unit_codes)
    columns = [col[0] for col in cursor.description]
    data_frame = pd.DataFrame(cursor.fetchall(), columns=columns)
return data_frame

def get_quality_class(self, value):
    class_bins = self.class_bins_map[self.unit]
    class_labels = self.class_labels_map[self.unit]
    for i in range(len(class_bins) - 1):
        if class_bins[i] <= value <= class_bins[i+1]:
            return class_labels[i]
    return class_labels[-1]

def get_data_frame_summary(self):
    summary = []
    for cls in self.class_labels_map[self.unit]:
        stats = []
        for feat in self.features:
            vals = self.x_train[self.y_train == cls][feat]
            stats.append({
                'class': cls,
                'feature': feat,
                'mean': round(vals.mean(), 2),
                'std': round(vals.std(), 2)
            })
        summary.extend(stats)

    df_summary = pd.DataFrame(summary)
    return df_summary

def analyze(self,):
    df = self.get_pure_data_frame()
    df["date"] = pd.to_datetime(dict(year=df.year, month=df.month, day=df.day))

    df_values: List[str] = ["avg_value", "min_value", "max_value"]

```

```

self.df_pivot = df.pivot_table(
    index=["date", "city"],
    columns="unit_code",
    values=df_values
).reset_index()
self.df_pivot.columns = [
    "_".join([str(c) for c in col if c])
    for col in self.df_pivot.columns.values
]

self.df_pivot["target_class"] = self.df_pivot[f"avg_value_{self.unit}"].apply(
    lambda x: self.get_quality_class(x) if pd.notnull(x) else None
)
self.df_pivot = self.df_pivot[self.df_pivot["target_class"].notnull()]

self.features = [
    col for col in self.df_pivot.columns
    if col.startswith(
        ("avg_value_", "min_value_", "max_value_")
    ) and not col.endswith(f"_{self.unit}")
]
x = self.df_pivot[self.features].fillna(0)
y = self.df_pivot['target_class']

self.x_train, self.x_test, self.y_train, self.y_test = train_test_split(
    x, y, test_size=self.test_size, random_state=self.random_state
)

self.model = GaussianNB()
self.model.fit(self.x_train, self.y_train)

y_pred = self.model.predict(self.x_test)
report = classification_report(self.y_test, y_pred, zero_division=0)

```

СПИСОК ОПУБЛІКОВАНИХ ПРАЦЬ ЗА ТЕМОЮ ДИСЕРТАЦІЇ

**Статті у наукових виданнях,
включених до міжнародних наукометричних баз даних**

Web of Science Core Collection та/або Scopus

1. Holub B., Kyrychenko V., Nikolaienko D., Lendiel M., Shevchenko D., Khomenko A. Algorithmic and information support in atmospheric air quality monitoring systems. Proceedings of the Tenth International Congress on Information and Communication Technology (ICICT 2025). Lecture Notes in Networks and Systems. 2025. Vol. 1441. P. 399–412. *(Holub B. сформулювала концепцію та методологію дослідження; Kyrychenko V. обґрунтував алгоритми розрахунку AQI/CAQI; Nikolaienko D. і Lendiel M. проаналізували нормативно-правову базу України та ЄС; Shevchenko D. спроектував архітектуру й сервер збору даних (Django/Celery, MQTT, API), розробив модель БД та виконав оцінку обсягів даних; Khomenko A. впровадив кешування, попередню агрегацію й дослідив ефективність алгоритмів сортування).*

2. Shevchenko D. V., Holub B. L. Regression analysis as a tool for identifying patterns in atmospheric air monitoring data. Proceedings of the 5th Edge Computing Workshop. 2025. Vol. 4. P. 20–27. *(Shevchenko D. V. розробив методологію регресійного аналізу (OLS), створив аналітичний модуль на Python/StatsModels, провів експерименти та інтерпретацію результатів; Holub B. L. сформулювала наукову постановку проблеми й дизайн дослідження, здійснила науковий супровід, валідацію результатів і редагування рукопису).*

**Статті у наукових виданнях,
включених до Переліку наукових фахових видань України**

3. Шевченко Д. В., Голуб Б. Л. Моніторинг якості повітря в реальному часі. Науковий журнал «Математичні машини і системи». 2025. № 1. С. 103–112. *(Шевченко Д.В. спроектував і впровадив систему моніторингу в реальному часі (MQTT, Django, Celery), розробив модель БД, реалізував обчислення AQI/CAQI,*

дашборди та провів експерименти; Голуб Б.Л. сформулювала концепцію й методологію, здійснила науковий супровід, валідацію результатів, огляд літератури та редакційне доопрацювання рукопису).

4. Шевченко Д. В., Голуб Б. Л. Багатовимірна аналітика екологічних даних: застосування OLAP у системах моніторингу. Науковий журнал «Математичні машини і системи». 2025. № 3–4. С. 54–65. (Шевченко Д.В. спроектував OLAP-архітектуру сховища (star/snowflake), визначив куби/виміри/факти, реалізував вітрини *Measurements* і *ThresholdExceed*, налаштував ETL та попередні агрегації, створив дашборди й KPI у Power BI, провів експерименти та інтерпретацію; Голуб Б.Л. сформувала концепцію та методологію дослідження, здійснила науковий супровід та провела експертне рев'ю OLAP-структур (виміри/факти)).

5. Шевченко Д. В., Голуб Б. Л. Застосування методів Data Mining для багатовимірного аналізу якості атмосферного повітря на основі екологічних даних. Науковий журнал «Наука і техніка сьогодні». Серія: Техніка. 2025. № 8 (49). С. 1801–1810. (Шевченко Д.В. спроектував застосування методу Data Mining (підготовка даних, *StandardScaler/PCA*, *K-Means*, інтеграція з OLAP), виконав експерименти, візуалізацію та інтерпретацію результатів, підготував рукопис; Голуб Б.Л. здійснила концептуалізацію та дизайн дослідження, верифікацію методів і наукове редагування рукопису).

Тези наукових доповідей

6. Шевченко Д. В. Розроблення інтелектуальної інформаційної технології для оцінювання стану атмосферного повітря. Продовольча та екологічна безпека в умовах війни та повоєнної відбудови: виклики для України та світу, до 125-річчя НУБіП: Міжнародна науково-практична конференція, м. Київ, 25–26 травня 2023 року: тези доповіді.

7. Шевченко Д. В. Розроблення інтелектуальної інформаційної технології для оцінювання стану атмосферного повітря. Теоретичні та прикладні аспекти розробки комп'ютерних систем 2023: V Всеукраїнська науково-практична

конференція студентів і аспірантів «Теоретичні та прикладні аспекти розробки комп'ютерних систем 2023», м. Київ, 2023 року: тези доповіді. С. 10–11.

8. Шевченко Д. В. Розроблення інтелектуальної інформаційної технології для оцінювання стану атмосферного повітря. Інформаційні технології: економіка, техніка, освіта 2023: XIV Міжнародна науково-практична конференція молодих вчених «Інформаційні технології: економіка, техніка, освіта 2023», м. Київ, 26–27 жовтень 2023 року: тези доповіді. С. 240–241.

9. Шевченко Д. В., Голуб Б. Л. Архітектура інформаційного компоненту системи моніторингу якості атмосферного повітря. Інформаційне суспільство: технологічні, економічні та технічні аспекти становлення: Міжнародна наукова інтернет-конференція «Інформаційне суспільство: технологічні, економічні та технічні аспекти становлення», м. Тернопіль, 18–19 січня 2024 року: тези доповіді. № 85. С. 75–78. *(Шевченко Д.В. розробив структуру архітектури інформаційного компоненту; Голуб Б.Л. здійснила наукове консультування та перевірку коректності архітектурних рішень).*

10. Shevchenko D., Golub B. Information service of the atmospheric air quality monitoring system. Proceedings of the IX International Scientific and Practical Conference «Scientific Problems and Options for Their Solution», Bucharest, Romania. 7–9 February 2024. P. 55–60. *(Shevchenko D. спроектував архітектуру системи та реалізував логіку обміну даними; Golub B. здійснила наукове керівництво та валідацію технічних рішень).*

11. Шевченко Д. В. Архітектура для системи моніторингу та аналізу якості атмосферного повітря. Актуальні питання розвитку науки та техніки в умовах глобалізації: Міжнародна науково-практична конференція «Актуальні питання розвитку науки та техніки в умовах глобалізації», м. Боярка, 14 травня 2024 року: тези доповіді. С. 117–119.

12. Шевченко Д. В. Регресійний аналіз як інструмент для виявлення закономірностей у даних моніторингу атмосферного повітря. Глобальні та регіональні проблеми інформатизації в суспільстві і природокористуванні: XII Міжнародна науково-практична конференція «Глобальні та регіональні проблеми

інформатизації в суспільстві і природокористуванні», м. Київ, 2024 року: тези доповіді.

13. Шевченко Д. В. OLAP-технології у системах моніторингу та аналізу якості атмосферного повітря. Актуальні питання розвитку науки та техніки в умовах глобалізації: II Міжнародна науково-практична конференція «Актуальні питання розвитку науки та техніки в умовах глобалізації», м. Боярка, 14 травня 2025 року: тези доповіді. С. 86–88.

14. Шевченко Д. В. Багатовимірна аналітика екологічних даних: застосування OLAP у системах моніторингу. Теоретичні та прикладні аспекти розробки комп'ютерних систем 2025: Всеукраїнська науково-практична конференція студентів і аспірантів «Теоретичні та прикладні аспекти розробки комп'ютерних систем 2025», м. Київ, 24 квітня 2025 року: тези доповіді.