

Вінницький національний технічний університет  
Факультет інтелектуальних інформаційних технологій та автоматизації  
Кафедра системного аналізу та інформаційних технологій

## МАГІСТЕРСЬКА КВАЛІФІКАЦІЙНА РОБОТА

на тему:

**“Інформаційна технологія аналізу та передбачення діоксиду вуглецю за даними громадського моніторингу атмосферного повітря у м. Вінниці”**

Виконав: студент 2 курсу, групи ICT-23м  
спеціальності 126 – «Інформаційні системи та технології»

Губар Олексій ГУБАР

Керівник: д.т.н., проф. каф. САІТ

Мокін Олександр МОКІН  
«05» 12 2024

Опонент: д.т.н., проф. каф. КН

Іванчук Ярослав ІВАНЧУК  
«08» 12 2024 р.

Допущено до захисту

Завідувач кафедри САІТ

Мокін д.т.н., проф. Віталій МОКІН  
«06» 12 2024 р.

Вінниця ВНТУ – 2024 рік

Вінницький національний технічний університет  
Факультет інтелектуальних інформаційних технологій та автоматизації  
Кафедра системного аналізу та інформаційних технологій  
Рівень вищої освіти – другий (магістерський)  
Галузь знань – 12 Інформаційні технології  
Спеціальність – 126 Інформаційні системи та технології  
Освітньо-професійна програма – Інформаційні технології аналізу даних та зображень

ЗАТВЕРДЖУЮ

Завідувач кафедри САІТ

 д.т.н., проф. Віталій МОКІН

«17» 09 2024 року

### ЗАВДАННЯ

#### НА МАГІСТЕРСЬКУ КВАЛІФІКАЦІЙНУ РОБОТУ СТУДЕНТУ

Губару Олексію Михайловичу

1. Тема роботи: «Інформаційна технологія аналізу та передбачення діоксиду вуглецю за даними громадського моніторингу атмосферного повітря у м. Вінниці»,

керівник роботи: Олександр МОКІН, д.т.н., проф. каф. САІТ,  
затверджені наказом ВНТУ від «17» 09 2024 року № 310

2. Строк подання студентом роботи «01» 12 2024 року

3. Вихідні дані до роботи: датасет Kaggle «Air Quality Monitoring from EcoCity»  
(<https://www.kaggle.com/datasets/vbmokin/air-quality-monitoring-from-ecocity>).

4. Зміст текстової частини:

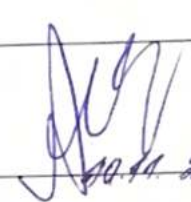
- загальна характеристика об'єкту досліджень;
- побудова моделей та вибір оптимальної для розв'язання поставленої задачі;
- розроблення інформаційної технології передбачення вмісту діоксиду вуглецю у повітрі;
- економічна частина.

5. Перелік ілюстративного матеріалу:

- зображення аномалії;
- зображення датасету аномалій;
- результат передбачення вмісту вуглекислого газу в атмосферному повітрі за валідаційними даними на основі моделі FB Prophe

- побудовані моделі перебачення;
- зображення алгоритму виконання задачі;
- реальні тестові дані вмісту вуглекислого газу та їх прогноз за оптимальною моделлю.

#### 6. Консультанти розділів МКР

| Розділ | Прізвище, ініціали та посада консультанта      | Підпис, дата                                                                                    |                                                                                                 |
|--------|------------------------------------------------|-------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------|
|        |                                                | завдання видав                                                                                  | завдання прийняв                                                                                |
| 4      | Олександр ЛЕСЬКО,<br>к. е. н., проф. каф. ЕПВМ | <br>10.11.24 | <br>30.11.24 |
|        |                                                |                                                                                                 |                                                                                                 |

7. Дата видачі завдання «17» 09 2024 року

#### КАЛЕНДАРНИЙ ПЛАН

| № з/п | Назва та зміст етапу                                                                 | Термін виконання |            | Примітка |
|-------|--------------------------------------------------------------------------------------|------------------|------------|----------|
|       |                                                                                      | початок          | закінчення |          |
| 1     | Загальна характеристика об'єкту досліджень                                           | 20.09            | 30.09      | вик.     |
| 2     | Побудова моделей та вибір оптимальної для розв'язання поставленої задачі;            | 01.10            | 20.10      | вик.     |
| 3     | Розроблення інформаційної технології передбачення вмісту діоксиду вуглецю у повітрі; | 24.10            | 15.11      | вик.     |
| 4     | Економічна частина                                                                   | 16.11            | 25.11      | вик.     |
| 5     | Оформлення матеріалів до захисту МКР                                                 | 26.11            | 30.11      | вик.     |

Студент

Керівник роботи



Олексій ГУБАР

Олександр МОКІН

## АНОТАЦІЯ

УДК 004.09

Губар О. М. Інформаційна технологія аналізу та передбачення діоксиду вуглецю за даними громадського моніторингу атмосферного повітря у м. Вінниці. Магістерська кваліфікаційна робота зі спеціальності 126 – інформаційні системи та технології, освітньо-професійна програма – інформаційні технології аналізу даних та зображень. Вінниця: ВНТУ, 2024. 94 с.

На укр. мові. Бібліогр.: 3 назв; рис.: 48; табл.: 11.

У даній магістерській роботі розглянуто процес аналізу про передбачення вмісту діоксиду вуглецю у певному регіоні. Використано відкриті дані, що містять інформацію про стан повітря навколо станції зчитування якості повітря. В датасеті міститься інформація про рівень забрудненості повітря різним пилом, вміст діоксиду вуглецю у повітрі, температуру, вологість та тиск. Проаналізовано основні закономірності та побудовано ряд моделей з використанням технологій машинного навчання. Вибрано оптимальну модель для передбачення вмісту діоксиду вуглецю. Проведено передбачення вмісту діоксиду вуглецю на основі наявних даних та зроблено висновки щодо ефективності роботи обраної моделі.

Ілюстративна частина складається з 6 плакатів із аномалій, алгоритму виконання задачі та результатів передбачення.

У розділі економічної частини розглянуто питання про доцільність розробки та впровадження інформаційної технології передбачення вмісту вуглекислого газу в атмосферному повітрі м. Вінниці.

Ключові слова: моніторинг якості атмосферного повітря, машинне навчання, передбачення, м. Вінниця.

## **ABSTRACT**

Hubar O. M. Information technology of carbon dioxide analysis and prediction according to the data of public air monitoring in Vinnytsia. Master's qualification work in the specialty 126 - Information systems and technologies, educational and professional program - Information technologies of data and image analysis. Vinnytsia: VNTU, 2024. 94 p.

In Ukrainian. Bibliography: 3 titles; Figures: 48; Tables: 11.

This master's thesis discusses the process of analyzing the prediction of carbon dioxide content in a particular region. Open data containing information about the air condition around an air quality reading station was used. The dataset contains information on the level of air pollution with various dusts, carbon dioxide content in the air, temperature, humidity, and pressure. The main patterns were analyzed and a number of models were built using machine learning technologies. The optimal model for predicting the carbon dioxide content is selected. The carbon dioxide content was predicted based on the available data and conclusions were drawn about the effectiveness of the selected model.

The illustrative part consists of 6 posters of anomalies, the algorithm for solving the problem, and the results of the prediction.

The economic section considers the feasibility of developing and implementing information technology for predicting the content of carbon dioxide in the atmospheric air of Vinnytsia.

Keywords: air quality monitoring, machine learning, prediction, Vinnytsia city.

## ЗМІСТ

|                                                                                               |    |
|-----------------------------------------------------------------------------------------------|----|
| ВСТУП.....                                                                                    | 4  |
| 1. ЗАГАЛЬНА ХАРАКТЕРИСТИКА ОБ'ЄКТУ ДОСЛІДЖЕНЬ .....                                           | 6  |
| 1.1 Аналіз предметної області.....                                                            | 6  |
| 1.2 Стан моніторингу якості атмосферного повітря у м. Вінниця .....                           | 18 |
| 1.3 Вибір датасету для розв'язання задачі.....                                                | 23 |
| 1.4 Вибір оптимальних інформаційних технологій та огляд аналогів .....                        | 26 |
| 1.5 Висновки .....                                                                            | 31 |
| 2 ПОБУДОВА МОДЕЛЕЙ ТА ВИБІР ОПТИМАЛЬНОЇ ДЛЯ РОЗВ'ЯЗАННЯ<br>ПОСТАВЛЕНОЇ ЗАДАЧІ.....            | 32 |
| 2.1 Характеристика веб платформи Kaggle .....                                                 | 32 |
| 2.2 Розвідувальний аналіз даних.....                                                          | 33 |
| 2.2 Побудова моделей часових рядів з використанням Facebook Prophet .....                     | 48 |
| 2.3 Побудова інших моделей машинного навчання .....                                           | 50 |
| 2.4 Вибір оптимальної моделі .....                                                            | 51 |
| 2.5 Висновки .....                                                                            | 53 |
| 3 РОЗРОБЛЕННЯ ІНФОРМАЦІЙНОЇ ТЕХНОЛОГІЇ ПЕРЕДБАЧЕННЯ<br>ВМІСТУ ДІОКСИДУ ВУГЛЕЦЮ У ПОВІТРІ..... | 54 |
| 3.1 Розроблення інформаційної технології передбачення вмісту вуглекислого<br>газу .....       | 54 |
| 3.2 Передбачення тестових даних .....                                                         | 56 |
| 3.3 Висновки .....                                                                            | 57 |
| 4 ЕКОНОМІЧНА ЧАСТИНА.....                                                                     | 58 |
| 4.1 Проведення комерційного та технологічного аудиту науково-технічної<br>розробки .....      | 58 |
| 4.2 Розрахунок узагальненого коефіцієнта якості розробки .....                                | 59 |
| 4.3 Розрахунок витрат на проведення науково-дослідної роботи.....                             | 61 |
| 4.3.1 Витрати на оплату праці .....                                                           | 61 |
| 4.3.2 Відрахування на соціальні заходи .....                                                  | 64 |

|                                                                                                                                    |    |
|------------------------------------------------------------------------------------------------------------------------------------|----|
| 4.3.3 Сировина та матеріали.....                                                                                                   | 64 |
| 4.3.4 Розрахунок витрат на комплектуючі.....                                                                                       | 66 |
| 4.3.5 Спецустаткування для наукових (експериментальних) робіт .....                                                                | 66 |
| 4.3.6 Програмне забезпечення для наукових (експериментальних) робіт ..                                                             | 67 |
| 4.3.7 Амортизація обладнання, програмних засобів та приміщень .....                                                                | 68 |
| 4.3.8 Паливо та енергія для науково-виробничих цілей .....                                                                         | 70 |
| 4.3.9 Службові відрядження.....                                                                                                    | 71 |
| 4.3.10 Витрати на роботи, які виконують сторонні підприємства, установи і організації.....                                         | 71 |
| 4.3.11 Інші витрати.....                                                                                                           | 71 |
| 4.3.12 Накладні (загальновиробничі) витрати.....                                                                                   | 72 |
| 4.4 Розрахунок економічної ефективності науково-технічної розробки при її<br>можливій комерціалізації потенційним інвестором ..... | 73 |
| 4.5 Висновки .....                                                                                                                 | 77 |
| ВИСНОВКИ.....                                                                                                                      | 78 |
| СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ.....                                                                                                    | 80 |
| Додаток А (обов'язковий) Технічне завдання.....                                                                                    | 83 |
| Додаток Б (обов'язковий) Протокол перевірки кваліфікаційної роботи на<br>наявність текстових запозичень.....                       | 86 |
| Додаток В (довідковий) Лістинг програми.....                                                                                       | 87 |
| Додаток Г (обов'язковий) Ілюстративна частина .....                                                                                | 90 |



## ВСТУП

**Актуальність теми.** Актуальність теми передбачення вмісту вуглекислого газу ( $\text{CO}_2$ ) у атмосфері є надзвичайно високою через значні екологічні, економічні та соціальні наслідки.

Перш за все, зростання рівня  $\text{CO}_2$  є однією з головних причин глобального потепління. Передбачення концентрації вуглекислого газу допомагає оцінити темпи підвищення температури, екстремальних погодних явищ і довгострокові зміни клімату.

Друге, що робить передбачення вмісту вуглекислого газу актуальним, – підвищення рівня  $\text{CO}_2$  негативно впливає на екосистеми, зокрема спричиняє окислення океанів, що шкодить морській фауні та флорі, а також впливає на продуктивність сільського господарства.

Третя важливість даної теми полягає в тому, що зростання рівня вуглекислого газу призводить до економічних втрат через погіршення умов ведення сільського господарства, пошкодження інфраструктури внаслідок екстремальних погодних умов, а також підвищення витрат на охорону здоров'я через зростання частоти природних катастроф.

Врешті-решт, в сучасному світі зі збільшенням викидів  $\text{CO}_2$  багато країн взяли на себе відповідальність за зниження викидів парникових газів, за для досягнення цієї мети між різними країнами було підписано низку міжнародних угод, таких як Паризька угода. Передбачення вмісту вуглекислого газу допомагає урядам створювати, оцінювати та вдосконалювати методики та кліматичні політики за для досягнень необхідних цілей та показників.

**Мета і завдання роботи:** Метою даної роботи є підвищення точності передбачення концентрації діоксиду вуглецю в атмосферному повітрі м. Вінниці шляхом розроблення інформаційної технології.

Розроблення інформаційної технології передбачає виконання таких завдань:

1. Аналіз важливості передбачення вмісту вуглекислого газу.



2. Вибір технологій та методів для розв’язування поставленої задачі.
3. Проведення розвідувального аналізу та побудова моделей.
4. Розроблення інформаційної технології передбачення вмісту вуглекислого газу в атмосферному повітрі.

**Об’єктом дослідження** є процес розроблення інформаційної технології аналізу та передбачення вмісту вуглекислого газу в атмосферному повітрі м. Вінниці.

**Предметом дослідження** є методи та інтелектуальні моделі інформаційної технології аналізу та передбачення вмісту вуглекислого газу в атмосферному повітрі м. Вінниці.

**Методи дослідження.** У роботі використано методи розвідувального аналізу для дослідження даних, технології машинного навчання для тренування інтелектуальних моделей для передбачення вмісту вуглекислого газу.

**Новизна одержаних результатів.** Подальшого розвитку набула інформаційна технологія передбачення вмісту діоксиду вуглецю в атмосферному повітрі регіону за рахунок оптимізації інтелектуальних моделей методами машинного навчання, що дозволяє підвищити точність передбачення його концентрації в атмосферному повітрі м. Вінниці.

**Апробація результатів магістерської кваліфікаційної роботи.** Результати роботи подані на всеукраїнську науково-практичну інтернет-конференцію «Молодь в науці: дослідження, проблеми, перспективи» (Вінниця, 2024-2025 рр.).

**Публікації результатів магістерської кваліфікаційної роботи.** Опубліковано тези на Всеукраїнській науково-практичній інтернет-конференції «Молодь в науці: дослідження, проблеми, перспективи» (Вінниця, 2024-2025 рр.) [1].

# 1. ЗАГАЛЬНА ХАРАКТЕРИСТИКА ОБ'ЄКТУ ДОСЛІДЖЕНЬ

## 1.1 Аналіз предметної області

На сьогоднішній день проблема збільшення вмісту вуглекислого газу в атмосфері турбує чи не кожний уряд на планеті. Люди витрачають чи малі кошти за для створення нових методів для боротьби із цією проблемою. Залучаються передові технології, вчені та величезна кількість ресурсів і все для того, щоб зменшити вміст вуглекислого газу у повітрі.

Вивчення вмісту вуглекислого газу ( $\text{CO}_2$ ) в атмосфері є однією з ключових тем в кліматології та екології, що охоплює кілька століть наукових спостережень, відкриттів і розвитку методів вимірювання та передбачення.

Спочатку все почалось з шотландського хіміка Джозефа Блека, що в 1754 році відкрив діоксид вуглецю, спостерігаючи за процесами горіння та випаровування. Винайдений газ він назвав “fixed air” (зв’язане повітря) і почав вивчати його властивості, так він дізнався, що даний газ не підтримує горіння та виділяється під час нагрівання вапняку [2].

Згодом французький хімік Антуан Лавузьє вивів хімічну формулу вуглекислого газу –  $\text{CO}_2$ . Він продемонстрував, що даний газ утворюється під час спалювання вугілля, а також є побічним продуктом дихання живих організмів. Напрацювання Антуана Лавузьє стали основою для розуміння важливості ролі, яку відіграє вуглекислий газ у природних процесах та у житті організмів на планеті [3].

У 1824 році французький фізик Жозеф Фур'є припустив, що атмосфера Землі працює за принципом «теплиці». Він вважав, що частину теплового випромінювання, що випускає поверхня Землі, утримує атмосфера. Таким чином було зроблено перші кроки до розуміння процесу парникового ефекту. Подальші дослідження провів британський фізик Джон Тиндал у 1859 році. Він довів, що вуглекислий газ та водяна пара здатні поглинати тепло в атмосфері планети .

Його напрацювання дали подальший розвиток у вивченні впливу діоксиду вуглецю на клімат та температуру земної поверхні [4].

У 1896 році шведський вчений Сванте Арреніус вперше створив математичну модель, задача якої було продемонструвати вплив підвищення рівня вуглекислого газу на температуру земної поверхні. Він висунув припущення, що збільшення концентрації  $\text{CO}_2$  у повітрі може збільшити температуру на декілька градусів [5]. Вже у 1938 році британський інженер Гай Календар проаналізував перші дані, які показували, що з часів початку індустріальної революції збільшилась кількість вуглекислого газу і температура Землі. Ці дослідження показали, як  $\text{CO}_2$  вплине на розвиток глобального потепління [6].

У 1950-их роках починаюься регулярні спостереження за концентрацією вуглекислого газу у атмосфері. Американський науковець Чарльз Девід Кілінг у 1958 році почав систематичні вимірювання вуглекислого газу в атмосфері на станції Мауна-Лоа на Гавайях. Отримавши відповідні дані він створив графік, який прозвали Кілінговою кривою, даний графік став першим довгостроковим набором даних точних даних про концентрацію  $\text{CO}_2$  і показала постійне зростання рівня вуглекислого газу в атмосфері. Також дана крива підтвердила зв'язок між збільшенням викидів викопного палива та зростанням рівня  $\text{CO}_2$  в атмосфері. Його досягнення також дали підґрунтя для сучасних досліджень викидів парникових газів [7].

Період з 1970-го по 1990-ті був відомим значним проривом у області передбачення та створення математичних моделей вмісту вуглекислого газу. Що вже зараз за допомогою використання методів машинного навчання дає змогу точніше прогнозувати вміст діоксиду вуглецю у повітрі використовуючи різні показники і кліматичні чинники.

Потім в 1990-тих роках відбулося поширення та удосконалення глобальних кліматичних моделей (GCM) дало змогу більш точно прогнозувати рівень викидів  $\text{CO}_2$  та його вплив на клімат. Ці моделі використовуються для оцінки сценаріїв майбутніх змін клімату залежно від викидів парникових газів.

У ХХІ столітті дослідження вмісту діоксиду вуглецю в атмосфері набули значного розвитку завдяки новітнім технологіям, що дозволяють отримувати точні та глобальні дані про рівень  $\text{CO}_2$ . Супутникові вимірювання відіграють важливу роль у цьому процесі, надаючи науковцям інструменти для спостереження за концентраціями діоксиду вуглецю в реальному часі та з високою просторовою роздільною здатністю.

У ХХІ столітті спостерігається значний прогрес у розвитку технологій для вимірювання та аналізу вмісту вуглекислого газу в атмосфері. Розроблено високочутливі лазерні спектрометри, які можуть точно визначати концентрації  $\text{CO}_2$  на різних висотах і в різних частинах планети [8].

Мережа наземних вимірювальних станцій, таких як Mauna Loa Observatory на Гавайях, продовжує відігравати ключову роль у спостереженнях за змінами концентрацій  $\text{CO}_2$ . В ХХІ столітті збільшилася кількість таких станцій, що дозволяє здійснювати вимірювання в різних кліматичних умовах.

Вуглекислий газ є одним з головних парникових газів, який відіграє вирішальну роль у процесі глобальної зміни клімату. Збільшення його концентрації в атмосфері є основною причиною підвищення середньої глобальної температури, що несе в собі серйозні екологічні, економічні та соціальні наслідки.

Зі збільшенням вмісту діоксиду вуглецю в атмосфері також варто і задуматись про парниковий ефект. Хоча  $\text{CO}_2$  не є найпотужнішим парниковим газом (метан і водяна пара мають більший потенціал глобального потепління), на нього припадає найбільша частка антропогенних викидів. Він залишається в атмосфері протягом сотень років і тому є довгостроковим чинником зміни клімату. Одну із ключових ролей у підтримці енергетичного балансу планети відіграє як раз  $\text{CO}_2$  і через збільшення його вмісту в атмосфері відбувається порушення цього самого балансу. На рисунку 1.1 зображено кількість збільшення вмісту вуглекислого газу в атмосфері [9].

## ATMOSPHERIC CARBON DIOXIDE

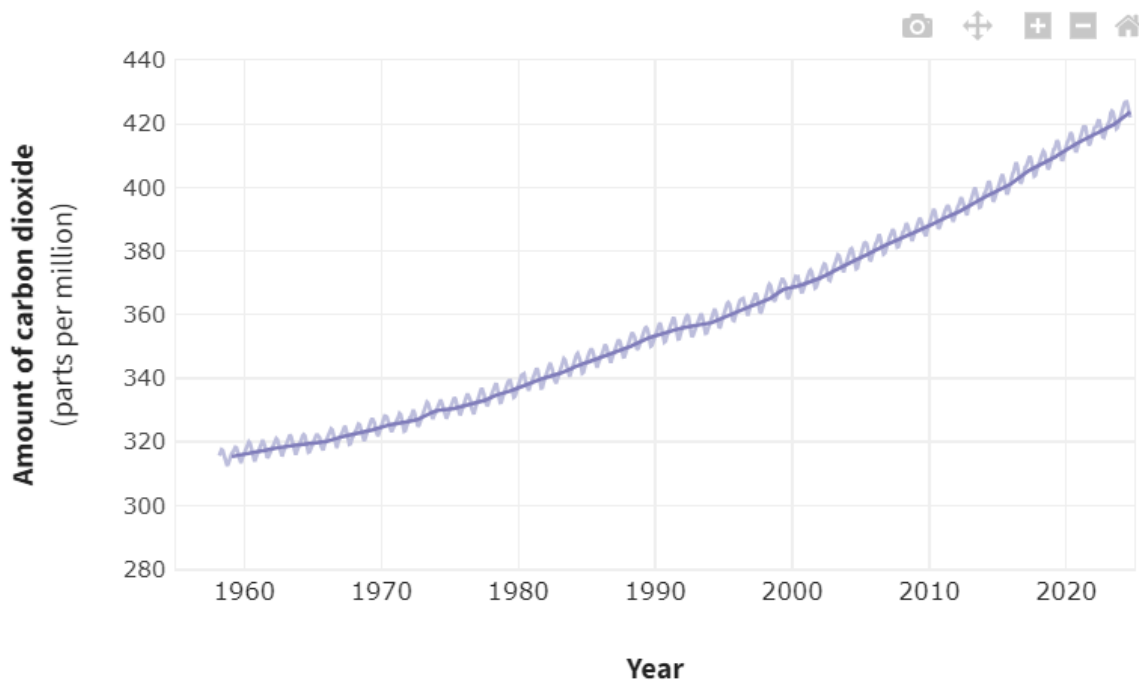


Рисунок 1.1 – Графік зображення збільшення вмісту CO<sub>2</sub> в атмосфері [9]

Також варто зазначити, що зростання глобальної температури земної поверхні також є наслідком зростання вмісту діоксиду вуглецю у повітрі. За останні 150 років вміст вуглекислого газу в атмосфері зріс з приблизно 280 ppm (ppm – це кількість молекул газу на мільйон молекул повітря в цілому) до понад 415 ppm. Це в свою чергу призвело до збільшення середньої глобальної температури на 1,1°C порівняно з доіндустріальним рівнем. Таке підвищення температури особливо відчувається на полюсах де тануть льодовики та в тропічних регіонах, де збільшилась кількість природних катаклізмів. На рисунку 1.2 зображено відхил від середньої норми температури на планеті [10].

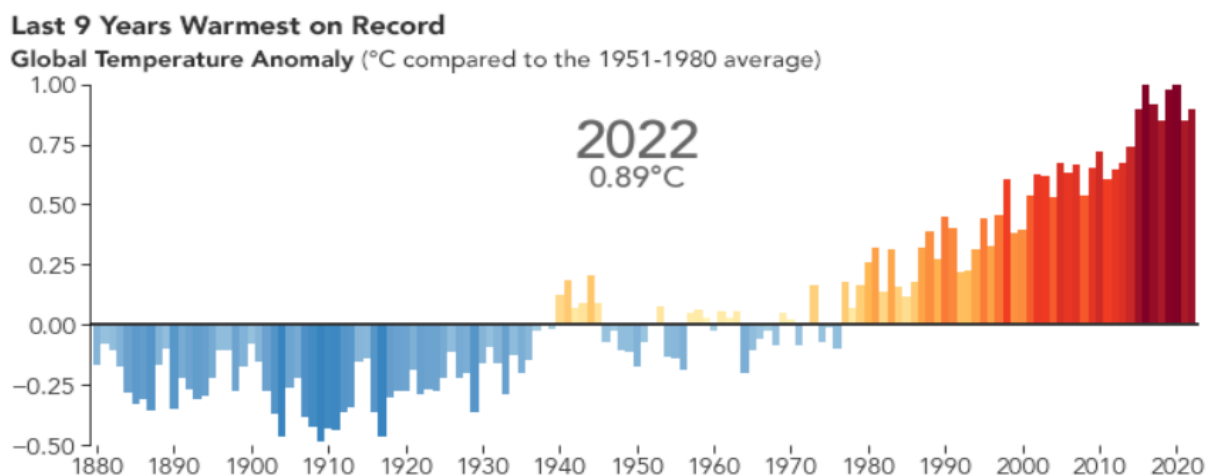


Рисунок 1.2 – Відхилення від норми середньої температури [10]

Відповідно до досліджень груп науковців, останнє десятиліття було найтеплішим за всю історію спостережень, і вони прогнозують, що якщо не скоротити викиди вуглекислого газу у повітрі, то загальна температура збільшиться на 1,5-2°C, що матиме катастрофічні наслідки для усієї екосистеми.

Через танення льодовиків збільшується і рівень світового океану. Вчені прогнозують, що якщо не зменшити рівень викидів діоксиду вуглецю до кінця століття рівень води підвищиться на 0,6 – 1,1 метра. Це в свою чергу призведе до затоплення великих територій, островів і переселення великої кількості людей, а також знищення багатьох екосистем. x

Океани поглинають близько 25-30% викидів CO<sub>2</sub>, що допомагає зменшити концентрацію CO<sub>2</sub> в атмосфері. Однак це також призводить до підвищення кислотності води, оскільки CO<sub>2</sub> реагує з водою, утворюючи вугільну кислоту. За даними вчених, кислотність океанів зросла на 30% з початку індустріальної ери. Підвищена кислотність негативно впливає на морські організми, особливо на корали, молюсків і планктон, які використовують карбонат кальцію для формування своїх скелетів і мушель.

Зникнення коралових рифів і зменшення популяцій морських організмів порушує харчові ланцюги і загрожує всій екосистемі океану.

Також через глобальне потепління багато видів змушенні змінювати свій ареали існування, шукаючи більш сприятливі умови. Види, що проживають в холодному кліматі чи полярних широтах найбільш вразливі до цього, адже їхнє середовище існування просто зникає. Внаслідок стрімкого потепління багато видів не можуть пристосуватися до нових умов і опиняються на межі вимирання. За оцінками вчених, 20-30 % видів можуть зникнути, якщо температура підвищиться більш ніж на 2 °C. На рисунку 1.3 зображено тип та кількість видів, які потерпають від зміни клімату [11].

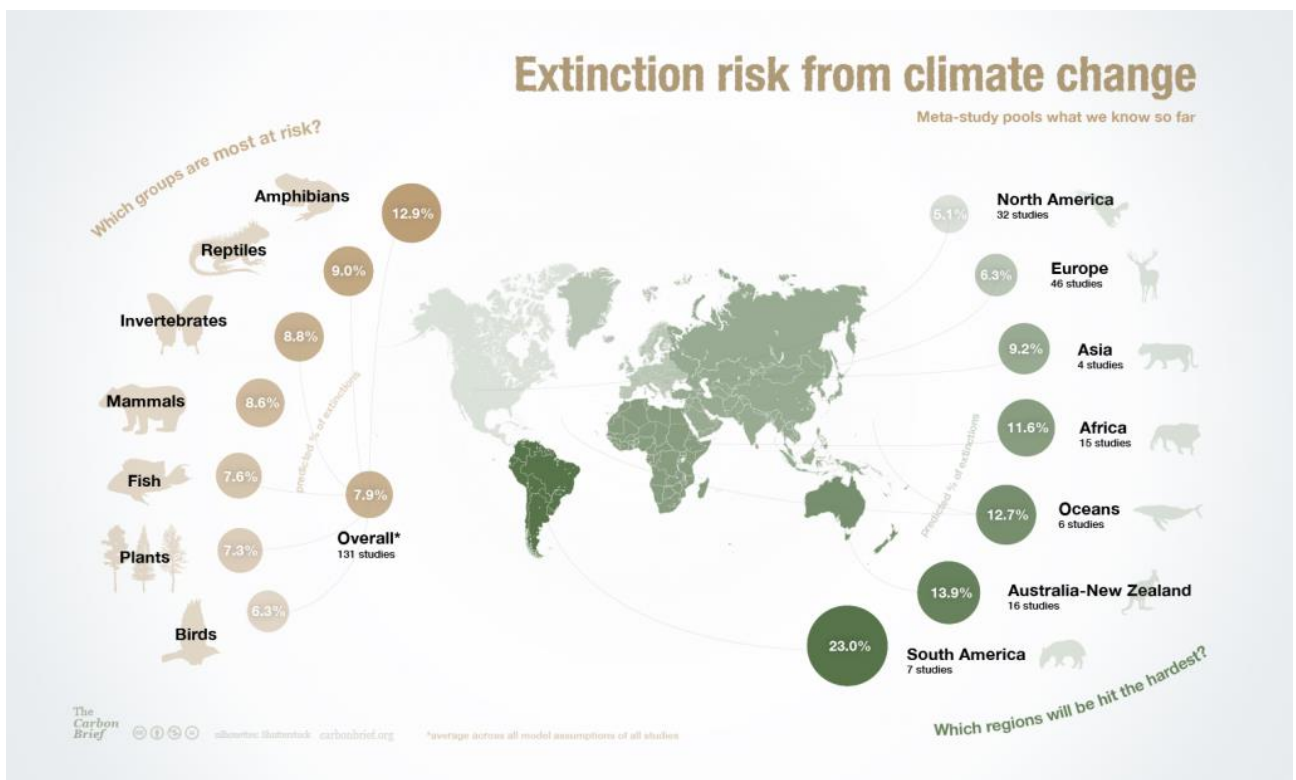


Рисунок 1.3 – Діаграма видів тварини, які потерпають від зміни клімату [11]

Також не будемо забувати і про те, що через збільшення температури і зміни клімату. Підвищення рівня моря і погіршення стану екосистем призводять до переміщення населення, особливо в прибережних і посушливих регіонах. Мільйони людей вже є кліматичними біженцями, і ця проблема буде продовжувати загострюватися. Екстремальні погодні явища, пов'язані з глобальним потеплінням, завдають значних економічних збитків. Руйнування інфраструктури, втрати в сільському господарстві та збільшення витрат на





Про пилові бурі також можна сказати декілька слів, адже у регіонах з низькою вологістю повітря, такі як пустеля Сахара сильні вітри піднімають у повітря тони піску та пилу. Дане явище негативно впливає на якість повітря в регіонах, що знаходять за тисячі кілометрів від пустелі. Можуть досягати територій Європи, зокрема і України. На рисунку 1.5 зображено забрудненість повітря пилом із Сахари. Різними кольорами позначено ступінь забруднення від зеленого, що означає нормальну якість повітря до бордового і фіолетового кольору, що свідчить про високий рівень забруднення при якому не допускається виходити на вулицю без респіраторних масок, потрібно постійно тримати вікна зачиненими та робити вологе прибирання у приміщеннях, допоки рівень забруднення не спаде. [13]

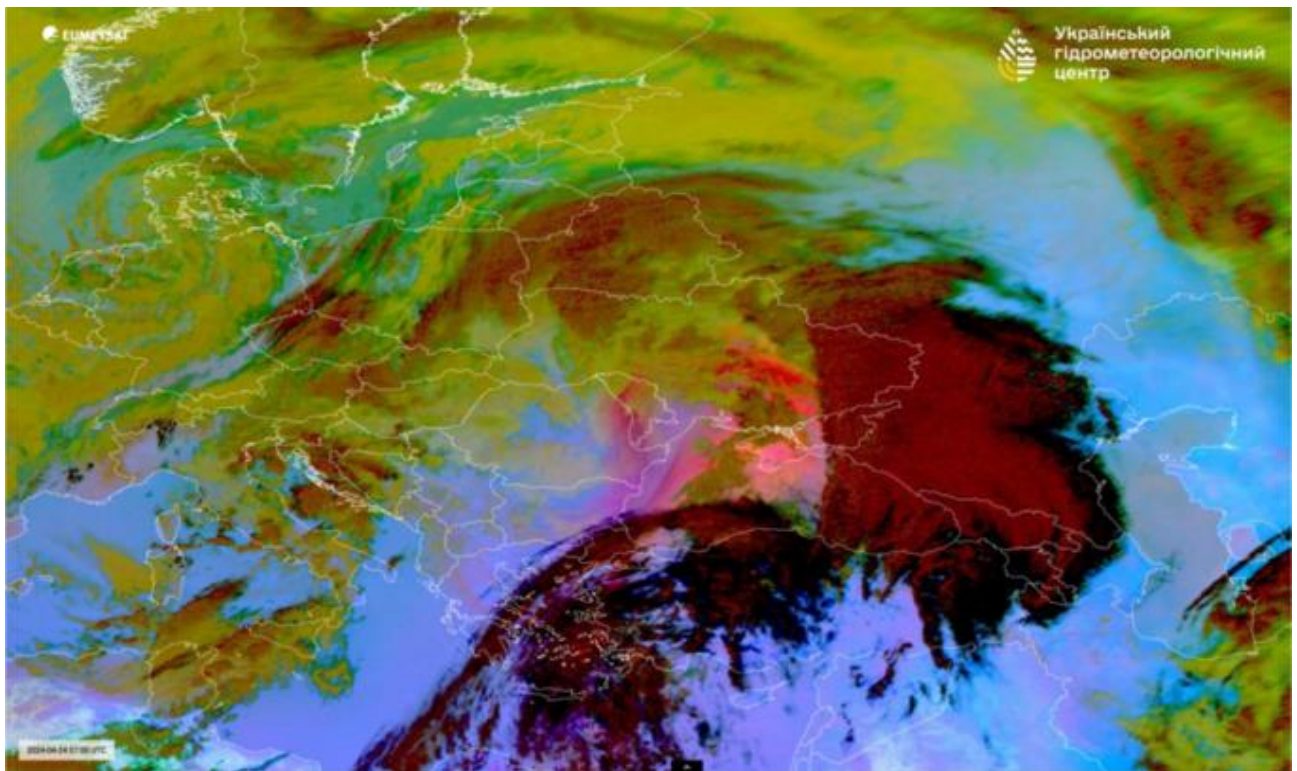


Рисунок 1.5 – Рівень забруднення повітря пилом із Сахари [13]

Пожежі носять також згубний характер впливу на якість повітря, адже спричиненні людиною, або маючи наприклад, природне походження внаслідок аномальної спеки, виділяють у повітря величезну кількість диму, вуглекислого

газу, чадного газу (CO), попіл та інші частинки. Одним із наслідків великих пожеж є наприклад смог, який ми могли спостерігати в сусідніх з Києвом регіонах, коли палали торф'яні родовища. На рисунку 1.6 зображено поширення забруднених частинок у повітрі в наслідок пожеж у чорнобильській зоні відчуження [14].

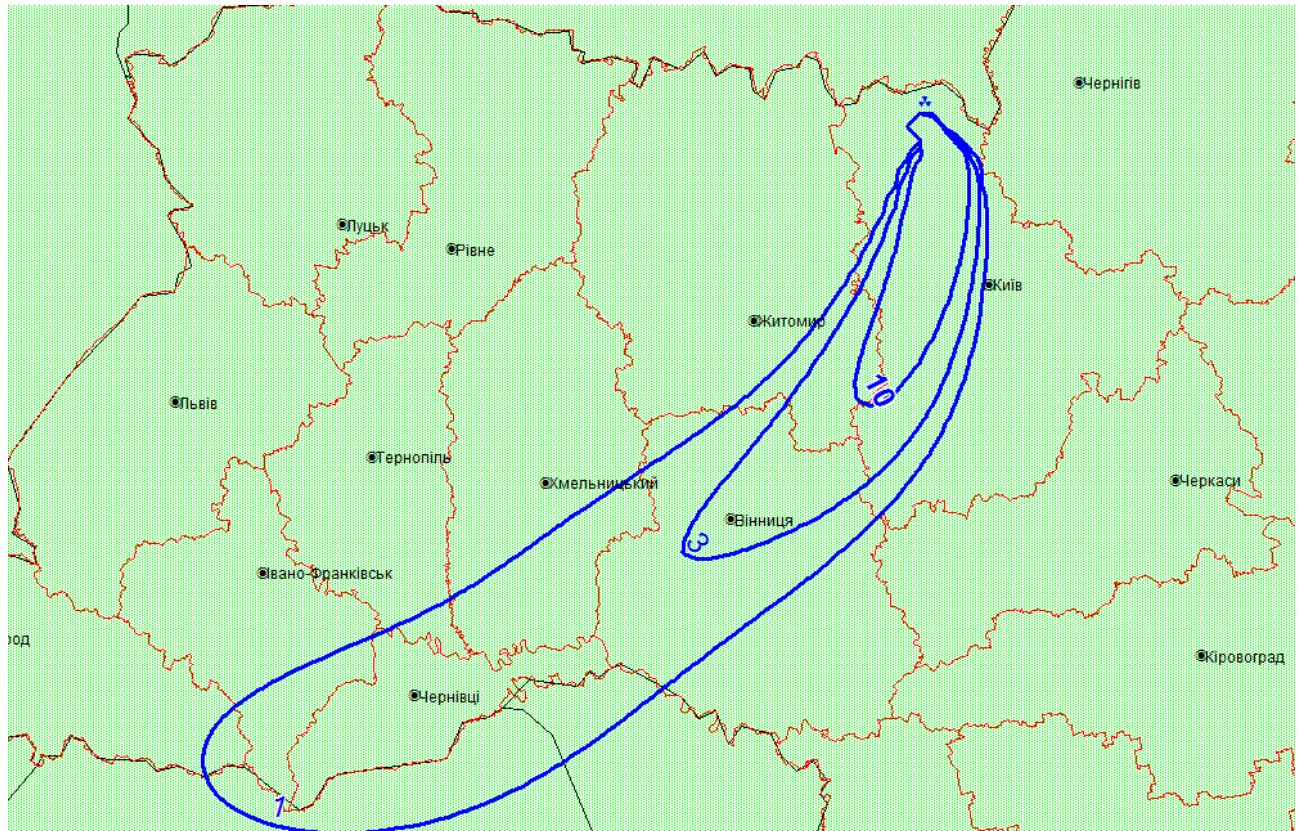


Рисунок 1.6 – Поширення забруднених частинок у повітрі в наслідок пожеж у чорнобильській зоні відчуження [14]

Гниття різного роду рослин чи залишків тварин у болотяних чи закритих водоймах виділяє метан ( $\text{CH}_4$ ) та інші парникові гази. Забруднення даного типу має цілком природний характер, але через масові і неконтрольовані вирубки лісів та осушення різних водоймищ може посилюватись. На рисунку 1.7 зображена мапа концентрації метану в атмосфері, синім кольором позначена найменша концентрація метану, а червоним – найбільша [15].



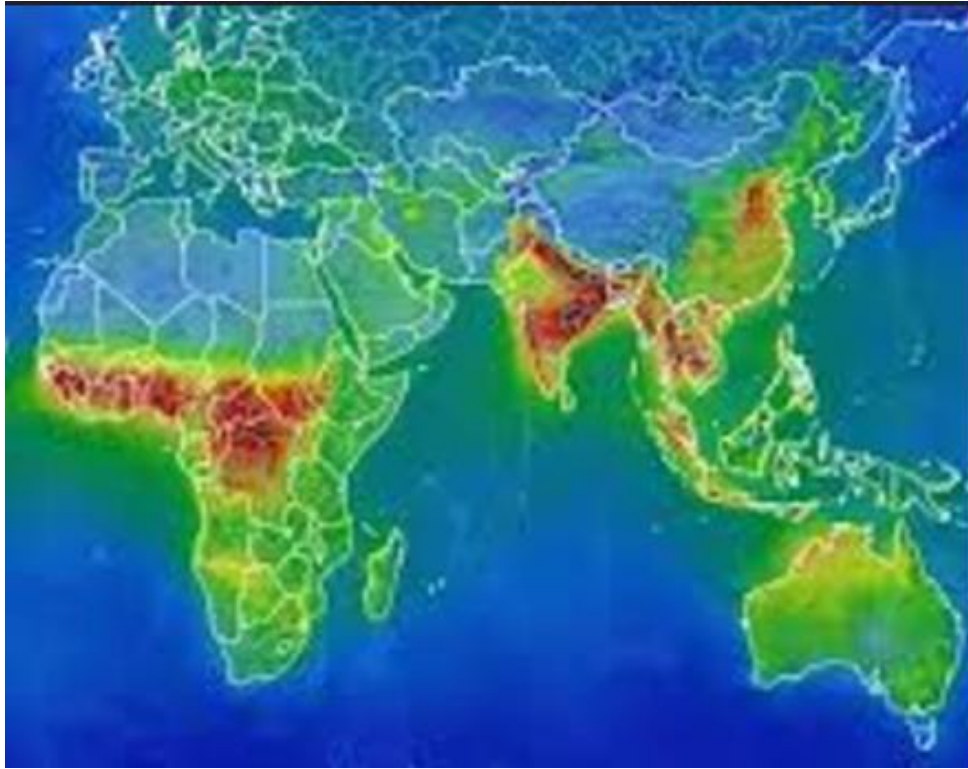


Рисунок 1.7 – Мапа концентрації метану в атмосфері [15]

Також варто згадати, що певні види рослини, такі як хвойні дерева, виділяють леткі органічні сполуки (VOC), самі по собі ці сполуки не несуть у собі сильної загрози природі, проте, вступаючи в сильну хімічну реакцію з забрудненим повітрям, утворює так званий нижній озоновий шар. На відміну від верхнього озонового шару, що захищає нас від ультрафіолетового випромінювання, нижній є сильним забруднюючим фактором, токсичним газом, що здатен нанести шкоди як різним екосистемам, так і нести пряму загрозу здоров'ю людей.

Тепер розглянемо причини забруднення атмосфери через людську діяльність, так звані антропогенні чинники. Перш за все хотілось би поговорити про підприємства. Різного роду заводи та фабрики щорічно викидають в атмосферу оксиди сірки ( $\text{SO}_2$ ), оксиди азоту ( $\text{NO}_x$ ), вуглекислий газ ( $\text{CO}_2$ ), метан та летючі органічні сполуки. Особливо небезпечною є хімічна промисловість, що виробляє досить токсичні продукти і супроводжується такими самими токсичними викидами, серед яких: аміак, діоксини, фосфати.

Транспорт також є чималим чинником забруднення, адже за останні роки кількість різного транспорту, що включають в себе автомобілі, автобуси, літаки й кораблі вони викидають у повітря велику кількість  $\text{CO}_2$ , оксидів азоту, чадного газу й мікрочастинок. Особливу увагу варто звернути на старий транспорт, який не є екологічним по осучасним критеріям або на транспорт з дизельними двигунами, їх викиди є дуже токсичними порівняно із сучасним транспортом. У багатьох великих містах великий транспортний потік та недобросовісне ставлення до стану транспортних засобів є причиною смогу, особливо в години пік.

Хоча зараз багато вкладається у фінансування так званої “зеленої” енергетики, сама галузь є також забруднюючим чинником. Адже електростанції, що використовують вугілля, нафту чи газ, під час спалювання даних видів палива виділяють велику кількість діоксиду вуглецю, метану та оксиду сірки. Також з початку масштабних обстрілів російської федерації і початком відключень електроенергії в Україні масово почались застосовування генераторів, які споживають, описані вище, види палива, що також негативно впливає на екологічний стан.

Сільське господарство теж має негативний вплив на якість повітря. Використовуючи різного виду добрива та пестициди у повітря виділяються аміак та оксиди азоту. Різного виду ферми виділяють метан через відходи. А також фермери після збирання врожаю спалюють залишки культур, що також виділяє у повітря діоксид вуглецю та чадний газ. На рисунку 1.8 зображено мапу забруднення повітря метаном у наслідок функціонування фермерського господарства [16].

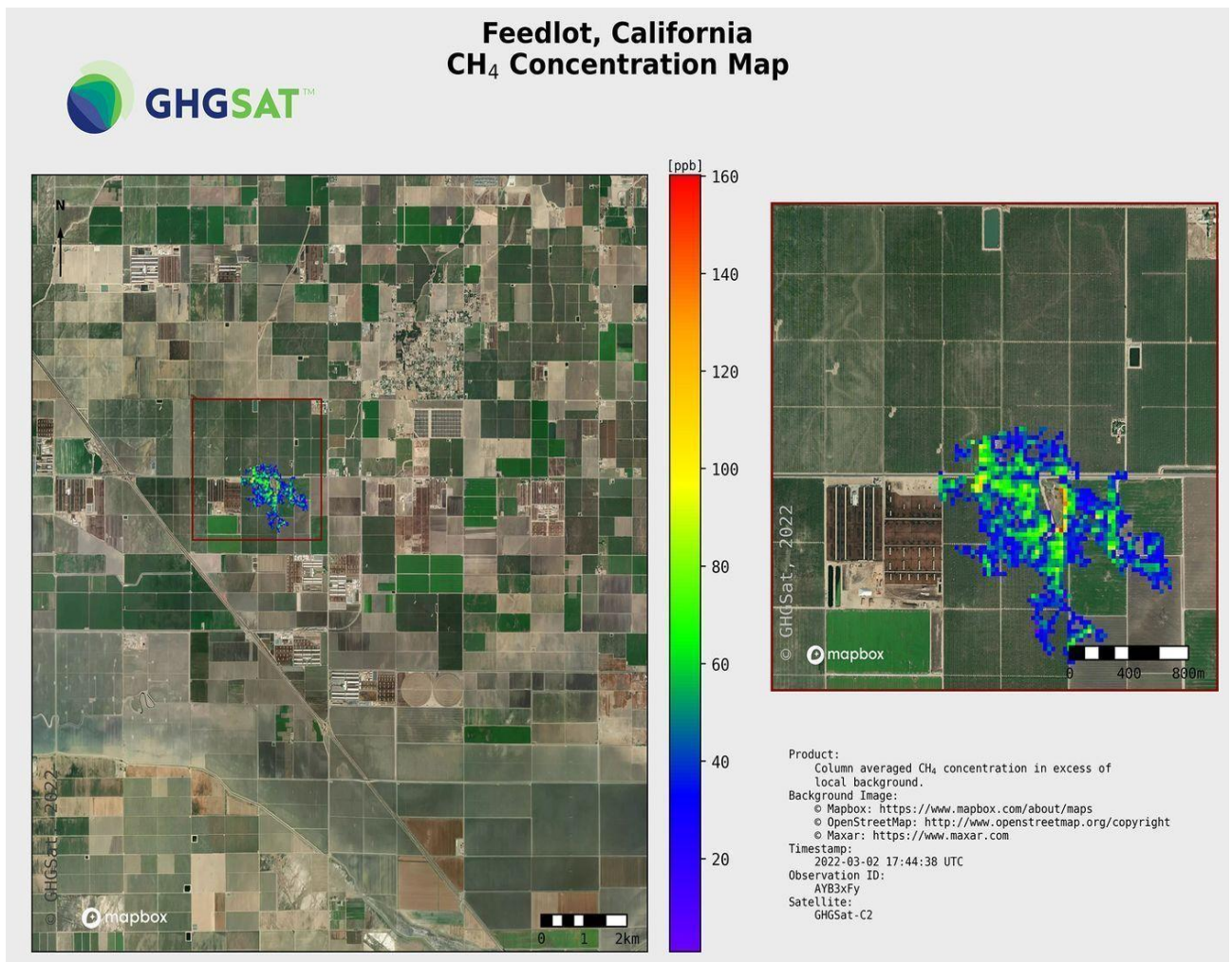


Рисунок 1.8 – Мапа викидів метану фермерського господарства [16]

Використання людьми вугілля чи дров для ведення побутових справ такі як приготування їжі чи опалення приміщень. Також часто в сільських регіонах люди часто спалюють сміття чим створюють ще більшу кількість забруднення повітря діоксидом вуглецю, а якщо серед сміття, яке спалюється, є ще й пластикові відходи чи відходи, що потребують спеціальної утилізації повітря додатково забруднюється токсичними речовини.

Через масову і не контрольовану вирубку лісів зменшується кількість дерев, що можуть поглинати вуглекислий газ, що також призводить і до більшої кількості розкладання органічних речовин, що виділяють додатково парникові гази.

Забруднення повітря може мати найрізноманітніші негативні наслідки. Для здоров'я людей це різні хвороби серцево-судинні та легеневі. Для

довкілля – це кислотні дощі, що завдають шкоди лісам, озерам чи ґрунтам, руйнування озонового шару та зниження врожайності [17].

## 1.2 Стан моніторингу якості атмосферного повітря у м. Вінниця

В місті Вінниця проводиться ціла низка заходів для визначення якості атмосферного повітря. Всю наявну інформацію можна на сайті Вінницької міської ради. Яка включає в себе не тільки обслуговування станцій моніторингу забруднення атмосферного повітря і збирання даних з них, а й оцінка різних підприємств, ступінь забруднення підприємством, атмосферного повітря, кількість зареєстрованих транспортних засобів тощо, дана інформація наведена на рисунку 1.9.

### 1. Джерела забруднення атмосферного повітря

| № | Показник                                                                                                                       | 2016    | 2017    | 2018    | 2019    | 2020    |
|---|--------------------------------------------------------------------------------------------------------------------------------|---------|---------|---------|---------|---------|
| 1 | Загальна кількість підприємств, що здійснюють викиди забруднювальних речовин в атмосферне повітря, од                          | 69      | 83      | 82      | 93      | 91      |
| 2 | Загальна кількість (одиниць) діючих дозволів на викиди забруднювальних речовин в атмосферне повітря, об'єкт якого належить до: | 407     | 465     | 535     | 568     | 617     |
|   | першої групи                                                                                                                   | 3       | 3       | 4       | 5       | 5       |
|   | другої групи та третьої групи                                                                                                  | 404     | 462     | 531     | 563     | 612     |
| 3 | Кількість зареєстрованих транспортних засобів, од                                                                              | 17298   | 16347   | 18630   | 29227   | 26944   |
|   | з них такі, що належать:                                                                                                       |         |         |         |         |         |
|   | юридичним особам, од                                                                                                           | 1678    | 1711    | 1584    | 1925    | 1884    |
|   | фізичним особам, од                                                                                                            | 15620   | 14636   | 17046   | 27302   | 25060   |
| 4 | Протяжність автомобільних доріг, км                                                                                            | 384,326 | 384,326 | 384,326 | 389,316 | 389,316 |
|   | з них з твердим покриттям, км                                                                                                  | 198,727 | 220,171 | 237,271 | 250,326 | 257,126 |
| 5 | Інші джерела забруднення, од                                                                                                   | 65      | 66      | 67      | 69      | 70      |
|   | кількість аеропортів                                                                                                           | -       | -       | -       | -       | -       |
|   | кількість морських/річкових портів                                                                                             | -       | -       | -       | -       | -       |
|   | кількість об'єктів утворення, оброблення та утилізації відходів, місць видалення відходів                                      | 65      | 66      | 67      | 69      | 70      |
| 6 | Природні джерела (за наявності)                                                                                                | -       | -       | -       | -       | -       |

Рисунок 1.9 – Джерела забруднення атмосферного повітря [18]

Також здійснюється виміри забруднюючих речовин, що містяться в атмосферному повітрі, дана інформація наведена на рисунку 1.10.



| Показник                                                                                             | 2016     | 2017     | 2018     | 2019     | 2020     |
|------------------------------------------------------------------------------------------------------|----------|----------|----------|----------|----------|
| Загальний обсяг викидів забруднювальних речовин, т                                                   | 2928,293 | 2954,365 | 2486,330 | 2887,089 | 2663,204 |
| Викиди забруднювальних речовин від стаціонарних джерел (т) всього, у тому числі:                     | 2928,293 | 2954,365 | 2486,330 | 2887,089 | 2663,204 |
| Метали та їх сполуки з них:                                                                          | 4,305    | 4,568    | 1,934    | 2,763    | 2,663    |
| Залізо та його сполуки (у перерахунку на залізо)                                                     | 3,971    | 4,197    | 1,715    | 1,804    | 2,358    |
| Мідь та її сполуки (у перерахунку на мідь)                                                           | 0,005    | 0,008    | 0,005    | 0,702    | 0,063    |
| Хром та його сполуки (у перерахунку на триоксид хрому)                                               | 0,01     | 0,006    | 0,008    | 0,008    | 0,011    |
| Цинк та його сполуки (у перерахунку на цинк)                                                         | 0,001    | -        | 0,001    | 0,001    | 0,003    |
| Алюмінію оксид                                                                                       | 0,062    | 0,058    | 0,046    | 0,043    | 0,043    |
| Манган та його сполуки (у перерахунку на діоксид мангану)                                            | 0,173    | 0,212    | 0,075    | 0,123    | 0,079    |
| Стибій та його сполуки (у перерахунку на стибій)                                                     | 0,08     | 0,080    | 0,08     | 0,080    | 0,080    |
| Ртуть та її сполуки (у перерахунку на ртуть)                                                         |          |          |          |          | 0,002    |
| Кадмій та його сполуки (у перерахунку на кадмій)                                                     |          |          | 0,004    |          |          |
| Речовини у вигляді суспендованих твердих частинок (мікрочастинки та волокна)                         | 323,139  | 296,404  | 230,513  | 244,099  | 262,005  |
| Речовини у вигляді суспендованих твердих частинок більше 2,5 мкм та менше 10 мкм (ТЧ <sub>10</sub> ) | 74,394   | 93,094   | 79,728   | 97,641   | 135,378  |
| Речовини у вигляді суспендованих твердих частинок менше 2,5 мкм (ТЧ <sub>2,5</sub> )                 | 31,28    | 57,167   | 73,241   | 94,728   | 108,365  |
| Сажа                                                                                                 | 107,233  | 24,632   | 6,360    | 13,714   | 10,386   |
| Сполуки азоту з них:                                                                                 | 691,621  | 647,515  | 552,636  | 657,908  | 588,426  |
| оксид азоту (у перерахунку на діоксид азоту) [NO+NO <sub>2</sub> ]                                   | 662,33   | 618,438  | 539,282  | 642,659  | 573,141  |
| азоту (I) оксид (N <sub>2</sub> O)                                                                   | 6,482    | 6,475    | 6,373    | 8,274    | 8,211    |
| аміак                                                                                                | 22,779   | 19,543   | 6,590    | 6,584    | 6,684    |
| азотна кислота                                                                                       | 0,029    | 0,021    | 0,391    | 0,391    | 0,390    |
| Діоксид та інші сполуки сірки з них:                                                                 | 77,435   | 120,291  | 78,602   | 100,999  | 108,704  |
| сірки діоксид                                                                                        | 76,147   | 118,220  | 77,230   | 99,478   | 107,246  |

Рисунок 1.10 – Кількість забруднюючих речовин [18]

Також є інформація про кожну станцію моніторингу повітря, в ній можна дізнатись про номер станції, її місце розташування, забруднюючі компоненти, дані щодо сертифікації, проходження повірки станції тощо, фрагмент даної інформації зображено на рисунку 1.11.



| Місце розташування пункту (адреса, географічні координати) або маршрути – точки відбору | Найменування юридичної особи, якій належить пункт спостереження                                                         | Дата введення в експлуатацію | Перелік забруднювачів | Режим спостережень                      | Метод оцінювання                                        | Дані щодо сертифікації обладнання, приладів (для автоматизованих та напівавтоматизованих пунктів) | Дані щодо процедури перевірки                |
|-----------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------|------------------------------|-----------------------|-----------------------------------------|---------------------------------------------------------|---------------------------------------------------------------------------------------------------|----------------------------------------------|
| <b>Стационарні пункти спостережень</b>                                                  |                                                                                                                         |                              |                       |                                         |                                                         |                                                                                                   |                                              |
| Стационарний пост № 1 – вул. Київська, 25 / 49.241962, 28.478482                        | Вінницький обласний центр з гідрометеорології, відомча приналежність – Державній службі України з надзвичайних ситуацій | -                            | пил                   | Автоматизований та напівавтоматизований | Шляхом порівняння з гранично-допустимими концентраціями | Всі прилади мають свої сертифікати                                                                | Все обладнання проходить своєчасну перевірку |
|                                                                                         |                                                                                                                         |                              | діоксид сірки         |                                         |                                                         |                                                                                                   |                                              |
|                                                                                         |                                                                                                                         |                              | діоксид азоту         |                                         |                                                         |                                                                                                   |                                              |
|                                                                                         |                                                                                                                         |                              | оксид вуглецю         |                                         |                                                         |                                                                                                   |                                              |
|                                                                                         |                                                                                                                         |                              | фтористий водень      |                                         |                                                         |                                                                                                   |                                              |
|                                                                                         |                                                                                                                         |                              | формальдегід          |                                         |                                                         |                                                                                                   |                                              |
|                                                                                         |                                                                                                                         |                              | аміак                 |                                         |                                                         |                                                                                                   |                                              |
| Стационарний пост № 2 – вул. Немирівське шосе, 29 / 49.231392, 28.518974                | Вінницький обласний центр з гідрометеорології, відомча приналежність - Державній службі України з надзвичайних ситуацій | -                            | пил                   | Автоматизований та напівавтоматизований | Шляхом порівняння з гранично-допустимими концентраціями | Всі прилади мають свої сертифікати                                                                | Все обладнання проходить своєчасну перевірку |
|                                                                                         |                                                                                                                         |                              | діоксид сірки         |                                         |                                                         |                                                                                                   |                                              |
|                                                                                         |                                                                                                                         |                              | діоксид азоту         |                                         |                                                         |                                                                                                   |                                              |
|                                                                                         |                                                                                                                         |                              | оксид вуглецю         |                                         |                                                         |                                                                                                   |                                              |
|                                                                                         |                                                                                                                         |                              | фтористий водень      |                                         |                                                         |                                                                                                   |                                              |
|                                                                                         |                                                                                                                         |                              | формальдегід          |                                         |                                                         |                                                                                                   |                                              |

Рисунок 1.11 – Фрагмент списку опису станцій [18]

Також на сайті є інформації про всі лабораторно-аналітичні комплекси (рис. 1.12).

| № з/п                                                                                                | Юридичний статус, форма власності, установа (організація), якій належить лабораторно-аналітичний комплекс/підпорядкування | Кількість працівників | Перелік основного обладнання та приладів, що використовуються для проведення аналізів                                                                  | Дані щодо сертифікації обладнання та приладів | Переліки забруднювальних речовин, що визначаються в пробах         | Метод аналізу              | Процедура верифікації даних               |
|------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------|-----------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------|--------------------------------------------------------------------|----------------------------|-------------------------------------------|
| <b>Лабораторія санітарно-гігієнічних та фізичних досліджень (вул. Малиновського, 11, м. Вінниця)</b> |                                                                                                                           |                       |                                                                                                                                                        |                                               |                                                                    |                            |                                           |
| 1                                                                                                    | Державна установа «Вінницький обласний лабораторний центр Міністерства охорони здоров'я України»                          | 4 чол.                | Ваги ВРЛ-200 св. № 220/262к від 20.08.20р.; Тайфун Р-100, св. №UA/39/200831/1317 від 31.08.20р.; Тайфун Р-20-2, св. №UA/39/201109/2268 від 09.11.20р.; | Сертифікація заводу-виробника                 | Атмосферного повітря: недиференційований за складом пил (аерозоль) | Гравіметричні випробування | ДСТУ ISO/IEC 17025:2006 П7.8/1, Ф7.8/1-23 |

Рисунок 1.12 – Дані однієї із лабораторій [18]

Також є мапа на якій зображуються усі діючі підприємства та кількість допустимих викидів (рис. 1.13).

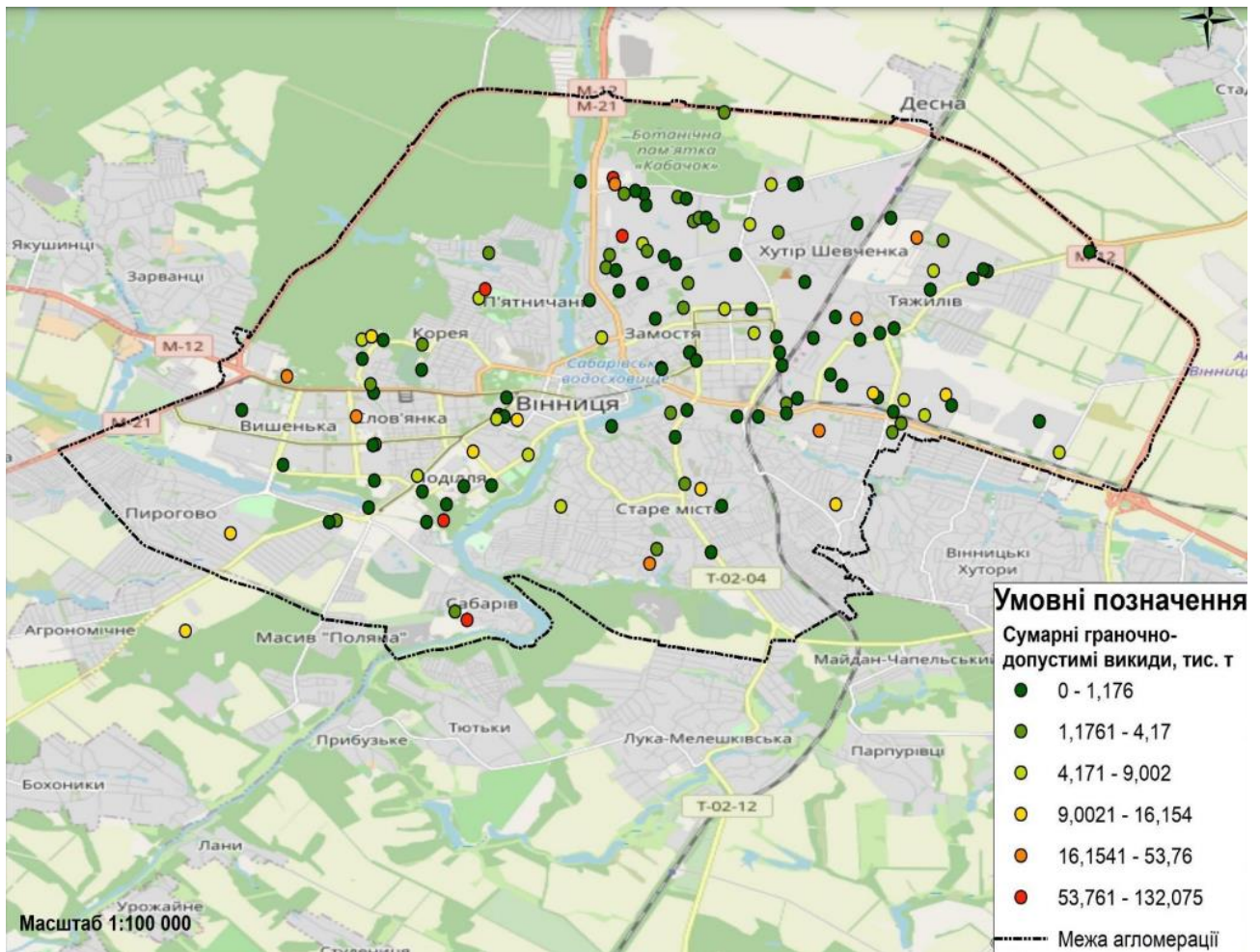


Рисунок 1.13 – Мапа діючих забруднюючих підприємств м. Вінниці [18]

Також кожен користувач може подивитись інформацію про кількість захворювань органів дихання (рис. 1.14), чи наприклад графік середньо місячних концентрацій певної речовини в атмосферному повітрі (рис. 1.15).



Рисунок 1.14 – Кількість людей з хворобами органів дихання [18]

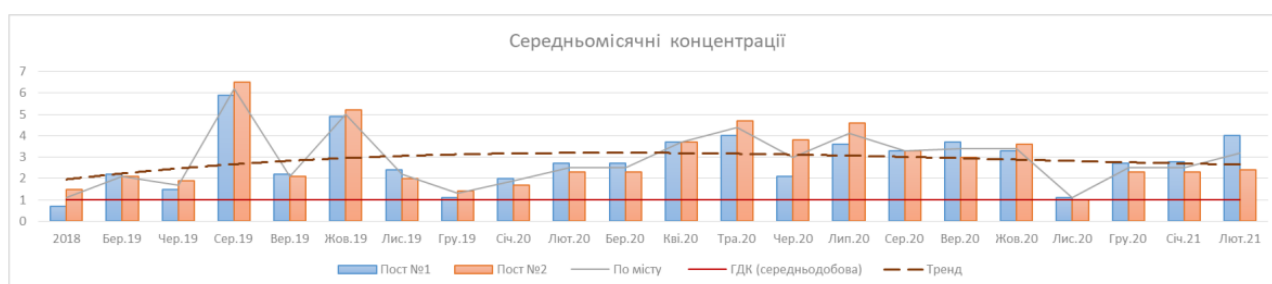


Рисунок 1.15 – Графік середньомісячної концентрації вуглекислого газу [18]

Знання стану атмосферного повітря дозволяє приймати оптимальні управлінські рішення для поліпшення ситуації. Наприклад, з метою зменшення викидів забруднюючих речовин в атмосферне повітря від пересувних джерел для муніципального автобусного парку КП «Вінницька транспортна компанія» закуплено 10 автобусів великої місткості Otokar Kent CNG, також на певних перехрестях було досягнуто зменшення часу простою автомобілів в очікуванні зеленого сигналу шляхом підключення 6 світлофорних об'єктів до автоматизованої системи керування дорожнім рухом. Водночас, у 2022 році було збільшено протяжність велосипедної мережі в межах міста на 1,4 км. Загальна протяжність велосипедної мережі станом на 31.12.2022 року склала 90,7 км.

Також у 2022 році розпочалось і нині продовжується опрацювання питання встановлення індикативних станцій моніторингу атмосферного повітря. Так існує попередня домовленість з компанією BOSCH про надання для користування у тестовому режимі однієї індикативної станції моніторингу якості атмосферного повітря [19].

### 1.3 Вибір датасету для розв’язання задачі

Для вирішення поставленої задачі даної магістерської роботи використовується датасет «Air Quality Monitoring from EcoCity» даний датасет має вибірку всіх даних, що збирають станції моніторингу якості повітря від EcoCity, також в даній роботі буде використовуватись платформа Kaggle. На рисунку 1.16 зображено вигляд опису датасету [20].

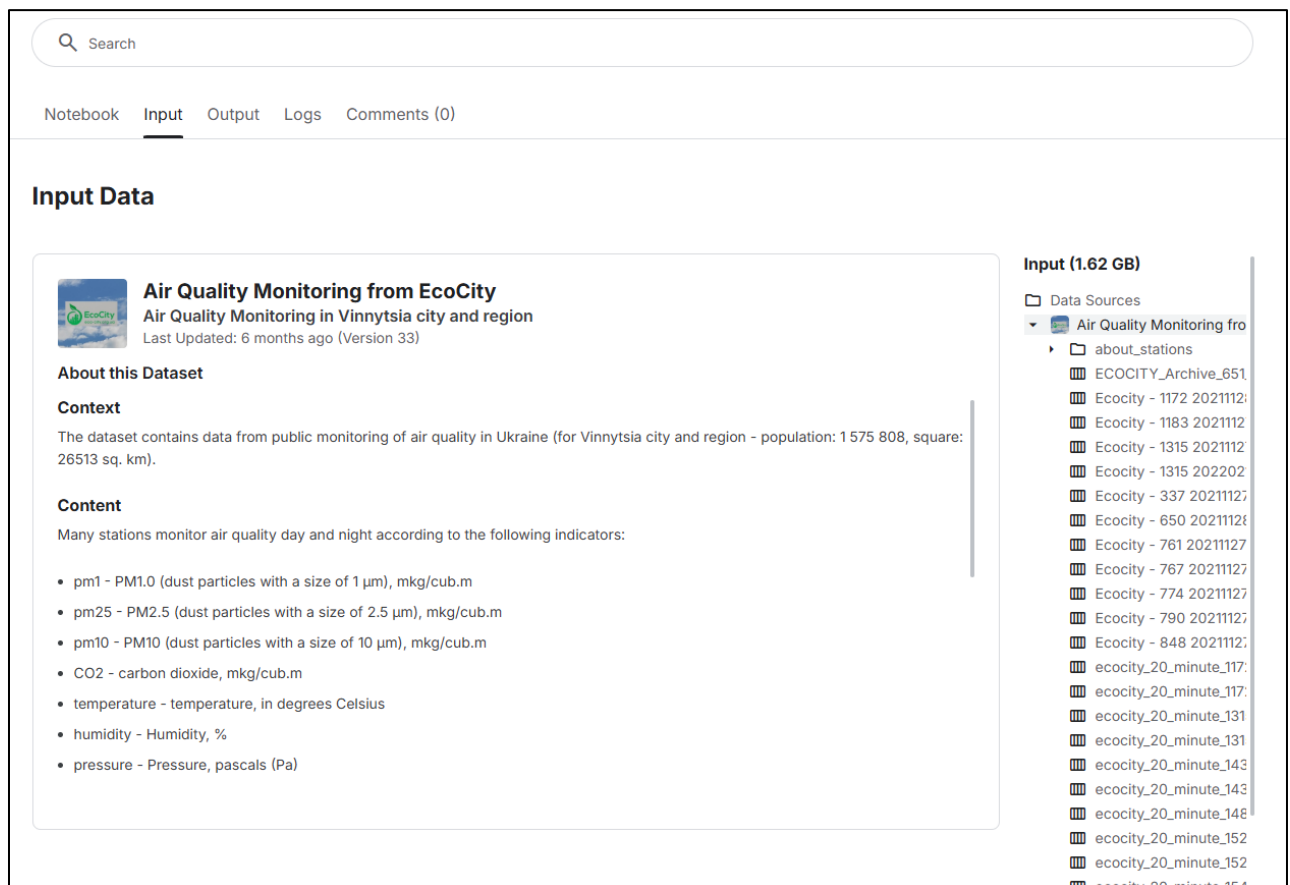


Рисунок 1.16 – вигляд опису датасету [20]

З початку видно, що даний датасет має опис, а саме те, що він використовується для моніторингу якості повітря у вінницькому регіоні, що налічує 1 575 808 населення та площу в 26513 км<sup>2</sup>. Після цього видно параметри, які саме зберігаються в датасеті:

- «pm1» – РМ 1.0 (частинки пилу розміром 1мкм), мкг/м<sup>3</sup>;
- «pm25» – РМ 2.5 (частинки пилу розміром 2,5 мкм), мкг/куб.м;
- «pm10» – РМ 10 (частинки пилу розміром 10 мкм), мкг/куб.м;
- «CO<sub>2</sub>» – вуглекислий газ, мкг/куб.м;
- «temperature» – температура, позначається в градусах Цельсія;
- «humidity» – вологість, позначається у відсотках;
- «pressure» – тиск, вимірюється у паскалях (Па);
- «NH<sub>3</sub>» – аміак, вимірюється в мкг/куб.м;
- «NO<sub>2</sub>» – діоксид азоту, вимірюється в мкг/куб.м;
- «CO» – чадний дим, вимірюється в мкг/куб.м.

Правіше від опису розташовується стовпець, в якому є перелічені всі доступні станції, дані яких додавались в датасет.

Далі розглянемо частину датасету (рис. 1.17)

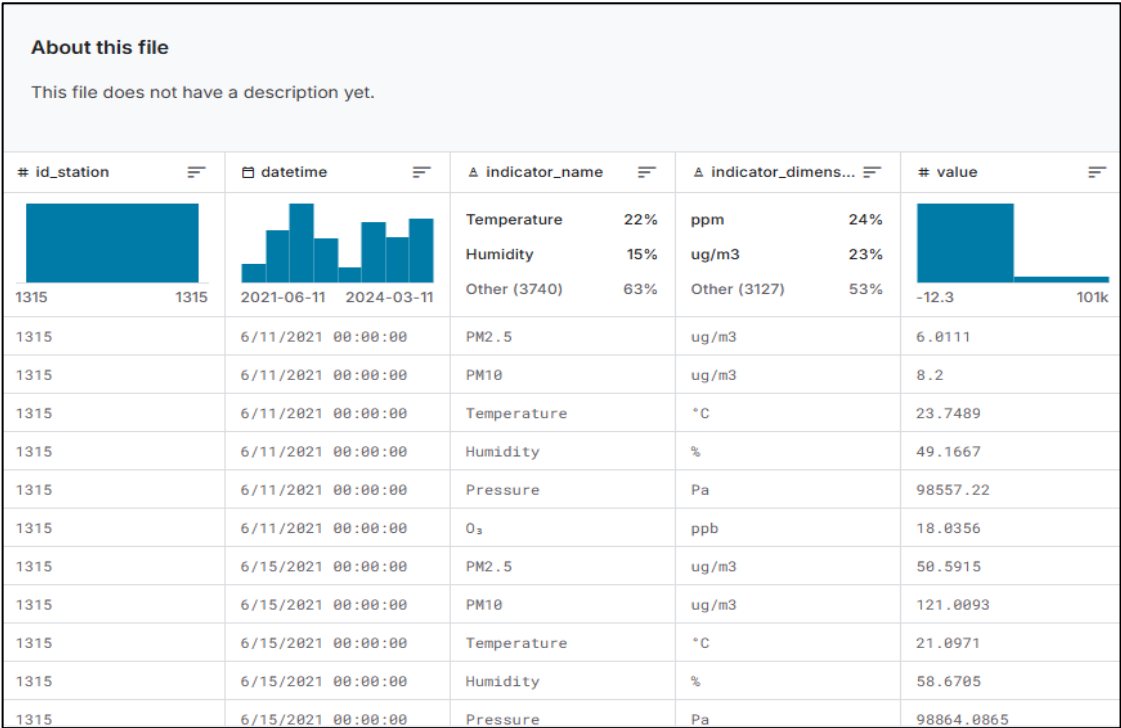


Рисунок 1.17 – Фрагмент датасету [20]

На даному фрагменті видно наступні стовпці:

- «id\_station» – вказує на номер станції;
- «datetime» – вказує на часовий проміжок;
- «indicator\_name» – вказує на назву вимірювального параметру;
- «indicator\_dimension» – вказує на величини вимірювальних параметрів;
- «value» – вказує на значення вимірювальної величини в певний момент часу.

Також є можливість вибрати спрощений варіант демонстрації таблиці (рис. 1.18).

| Detail Compact Column |                    |                |                     |            | 5 of 5 columns |  |
|-----------------------|--------------------|----------------|---------------------|------------|----------------|--|
| # id_station          | datetime           | indicator_name | indicator_dimension | value      |                |  |
| 1315                  | 6/11/2021 13:40:00 | PM2.5          | ug/m3               | 6.0167     |                |  |
| 1315                  | 6/11/2021 13:40:00 | PM10           | ug/m3               | 8.25       |                |  |
| 1315                  | 6/11/2021 13:40:00 | Temperature    | °C                  | 23.7067    |                |  |
| 1315                  | 6/11/2021 13:40:00 | Humidity       | %                   | 49.3667    |                |  |
| 1315                  | 6/11/2021 13:40:00 | Pressure       | Pa                  | 98557.9667 |                |  |
| 1315                  | 6/11/2021 13:40:00 | O <sub>3</sub> | ppb                 | 17.0533    |                |  |
| 1315                  | 6/11/2021 14:00:00 | PM2.5          | ug/m3               | 6          |                |  |
| 1315                  | 6/11/2021 14:00:00 | PM10           | ug/m3               | 8.1        |                |  |
| 1315                  | 6/11/2021 14:00:00 | Temperature    | °C                  | 23.8333    |                |  |
| 1315                  | 6/11/2021 14:00:00 | Humidity       | %                   | 48.7667    |                |  |
| 1315                  | 6/11/2021 14:00:00 | Pressure       | Pa                  | 98555.7267 |                |  |
| 1315                  | 6/11/2021 14:00:00 | O <sub>3</sub> | ppb                 | 20         |                |  |

Рисунок 1.18 – Зображення спрощеного виду датасету [20]

Також є можливість подивитись на мапу відмічених усіх станцій моніторингу якості повітря (рис. 1.19).



## 1.4 Вибір оптимальних інформаційних технологій та огляд аналогів

Розглянемо етапи та їх використання для розв'язання задач машинного навчання:

2. Exploratory Data Analysis (EDA) – етап, що носить назву аналіз даних, його задача дослідити наявні дані, визначити закономірності між даними, пошук аномалій, що можуть спотворити дані і призвести до неправильного результату, пошук кореляцій.

3. Feature Engineering (FE) – даний етап називається інженерія ознак, його задача полягає у покращенні точності і ефективності моделі шляхом подання алгоритму простих і зрозумілих для нього даних. Це можна досягти шляхом аналізу ознак які в нас є, якщо є необхідність, їх додаткового створення, групування чи зміни.

4. Model tuning (налаштування моделі) – в кожній моделі є параметри, задача даного етапу налаштувати модель для досягнення максимально точних результатів для певної задачі. Як приклад можна навести кількість дерев при використанні Random Forest чи кількість шарів нейронної мережі.

5. Forecasting (передбачення) – даний етап є по суті передбаченням прогнозованих даних, для певної задачі на основі введених раніше параметрів. Наприклад, темою даної магістерської роботи є передбачення вмісту вуглекислого газу у атмосферному повітрі.

6. Postprocessing (післяобробка) – даний етап є аналізом результатів навчання моделі, з даних результатів робиться висновок наскільки є точною модель у прогнозах та чи потрібно щось в ній покращити за для кращого результату. Також даний етап включає в себе приведення результатів до легкого сприйняття їх для звичайного користувача, так звана візуалізація (visualization), з використанням різних графіків, діаграм тощо [22].

Розглянемо моделі машинного навчання, які можна застосовувати для передбачення вмісту вуглекислого газу, звичайно, що дані моделі не універсальні і застосовуються в залежності від наявних даних, їх структури та залежності між вибраними даними. Спочатку розглянемо моделі часових рядів:

1. ARIMA – дана модель підходить для передбачення вмісту вуглекислого газу у повітрі. Використовується у короткочасних змінах вмісту  $\text{CO}_2$  та маючи лінійні залежності у часі, а для даних в яких є сезонність використовується SARIMA. [22]

2. ETS (Exponential Smoothing State Space Models) – дана модель добре працює з компонентами, так званими станами, що змінюються з часом, до них належить сезонність та тренди. [22]

3. Prophet – популярна модель, що зручна у використанні та користується великою популярністю, завдяки автоматизації чудово працює з складними чи неповними даними сезонності чи трендових залежностей. Основною перевагою даного методу є легкість у використанні, а також гнучкість у налаштуваннях певних параметрів. Недоліками цього методу є низька точність для часових рядів з сильними нерегулярними коливаннями, а також низькою продуктивністю на великих наборах даних, порівняно з деякими іншими моделями [22]

Далі розглянемо методи класичного машинного навчання:

1. Random Forest – є один із найпопулярніших методів машинного навчання, оскільки він створює великий набір дерев рішень, щоб потім скомбінувати усі прогнози для того, щоб отримати максимально точний та стабільний результат. Дана модель чудово працює з великими обсягами даних та дозволяє визначити важливість впливу тих чи інших ознак на результат прогнозу. Із головних недоліків даного методу є вимога значних ресурсів при обчисленні, а також через використання багатьох дерев рішень може бути важким аналіз результатів [22].

2. XGB Regressor – також поширений метод машинного навчання, принципом роботи даного методу є те, що він намагається зменшувати помилки минулих моделей рішень у нових моделях рішень. Даний метод є швидким при умові великих наборів даних використовуючи власні алгоритми пошуку та обробки, також може застосовуватись в складних задачах регресії завдяки пошуку складних залежностей між даними. Одним з недоліків даного методу є схильність до перенавчання, при умові неправильного налаштування моделі [22].

3. Лінійна регресія (linear regression) – метод, який описує взаємозв'язок між  $n$  фічерами (вхідними даними)  $x_1, \dots, x_n$  і таргетом  $y$  (вихідною ознакою) за допомогою лінійних функцій (рис. 1.20).

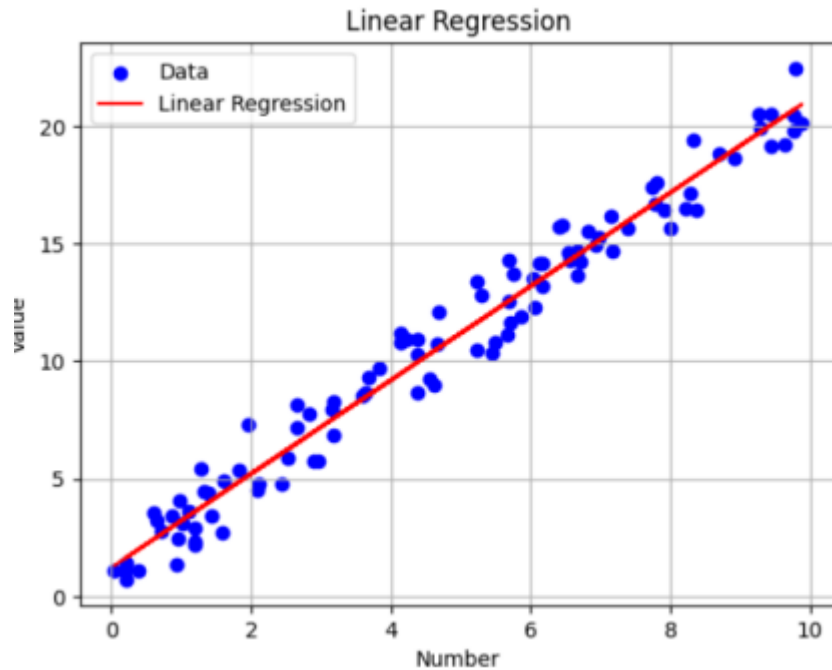


Рисунок 1.20 – Лінійна регресія [22]

4. Логістична регресія (англ. «logistic regression») – статистичний регресійний метод, оснований на логістичній функції, який застосовують для передбачення таргета  $y$  за випадку, коли вхідні змінні  $x_1, \dots, x_n$  є категорійними, тобто можуть набувати фіксовану кількість значень (2–10, рідко більше). Застосовується тільки для класифікаційних задач (рис. 1.21) [22].

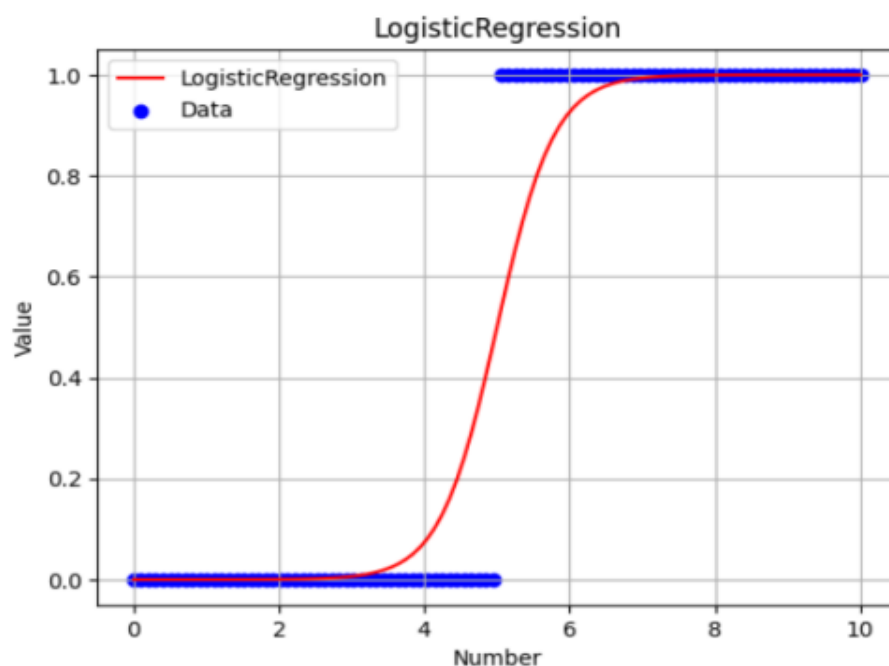


Рисунок 1.21 – Логістична регресія [22]

Метод опорних векторів (опорно-векторних машин) – це метод аналізу даних за допомогою моделі, яка відносить нові дані до однієї чи іншої категорії (класу, кластеру) шляхом проведення найширшого «коридору» між опорними векторами, які є «стінками» сусідніх класів у багатовимірному просторі, і потім відносить дані до кожного з цих класів (рис. 1.22) [22].

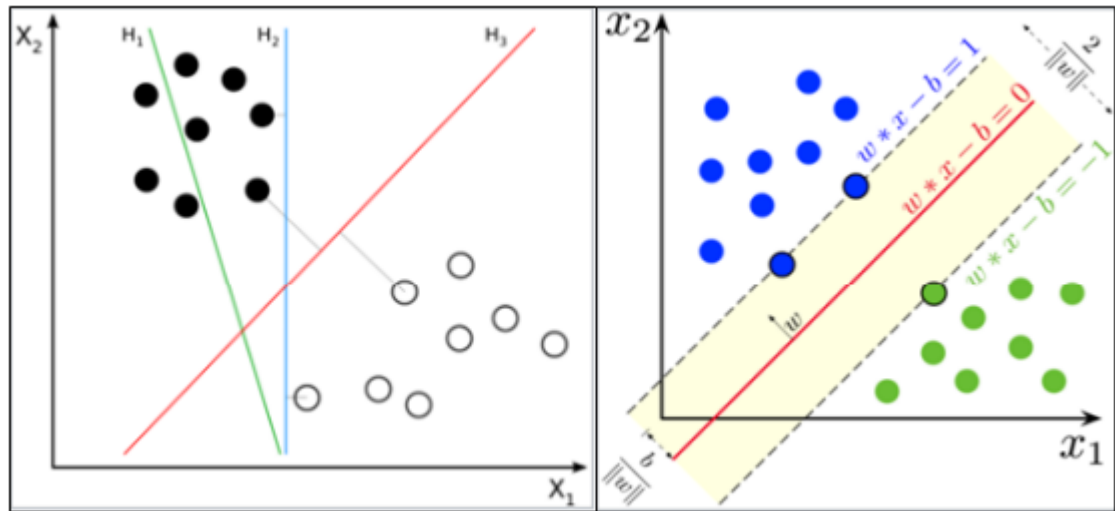


Рисунок 1.22 – Зображення проведення «коридору» між опорними векторами («стінками» цього «коридору») у методі опорних векторів для двовимірної площини за випадку двох вхідних фічерів [22]

Розглянемо аналоги передбачення вмісту вуглекислого газу в атмосферному повітрі. В роботі використовуються різні методи передбачення вмісту  $\text{CO}_2$ , серед яких метод k-найближчих сусідів, штучні нейронні мережі, лінійна регресія, random forest, регресія дерев рішень. Автор даної статті зауважив, що найкраще RMSE (вимірює середню різницю між передбаченими значеннями та фактичними), що дорівнювало 0,129 вдалося досягти використовуючи регресію дерев рішень. В роботі іншого автора відбувалось передбачення за допомогою моделі LSTM, даний автор постановив, що дана модель була найкраща за параметром MAPE (показує середній відсоток абсолютної помилки між прогнозованими та фактичними значеннями) який дорівнював 14,8291% [23].

Кожен з цих підходів має свої переваги і недоліки, і вибір конкретної моделі залежить від характеристик даних, цілей передбачення та доступних обчислювальних ресурсів.

### **1.5 Висновки**

У першому розділі магістерської роботи розглянуто актуальність передбачення вмісту вуглекислого газу у повітрі, описано чинники, що викликають збільшення його вмісту та можливі наслідки для людства та планети. Сформульовано постановку задачі передбачення вмісту вуглекислого газу. Описано ключові параметри датасету, що включають вміст вуглекислого газу, температури, кількості забруднення пилом, вологості та тиску. Дані описаного датасету будуть використовуватись для навчання та тестування моделей передбачення вуглекислого газу. Розглянуто інформаційні технології побудови інтелектуальних моделей, які допоможуть ефективно вирішувати задачі передбачення вуглекислого газу на основі наданих даних.



## 2 ПОБУДОВА МОДЕЛЕЙ ТА ВИБІР ОПТИМАЛЬНОЇ ДЛЯ РОЗВ'ЯЗАННЯ ПОСТАВЛЕНОЇ ЗАДАЧІ

### 2.1 Характеристика веб платформи Kaggle

Kaggle – це онлайн-платформа на якій звичайні користувачі, що просто цікавляться машинним навчанням, аналітики чи дослідники різних даних мають змогу працювати з великою кількістю різних даних. Дана платформа володіє великою кількістю інструментів, ресурсів для навчання та розв'язання задач в області машинного навчання.

Однією із цікавих особливостей даної платформи є те, що різні організації чи фахівці зі збору даних надають користувачам певні датасети. Потім ці користувачі змагаються між собою за для створення кращої моделі передбачення. Дані змагання включають в себе різні види задач такі як: передбачення продажів, аналіз зображення чи обробка природної мови. І як наслідок переможцю змагань можуть запропонувати грошову винагороду або визнання серед спільноти або подекуди навіть і кар'єрного зростання. На рис. 2.1 зображена головна сторінка платформи [24].

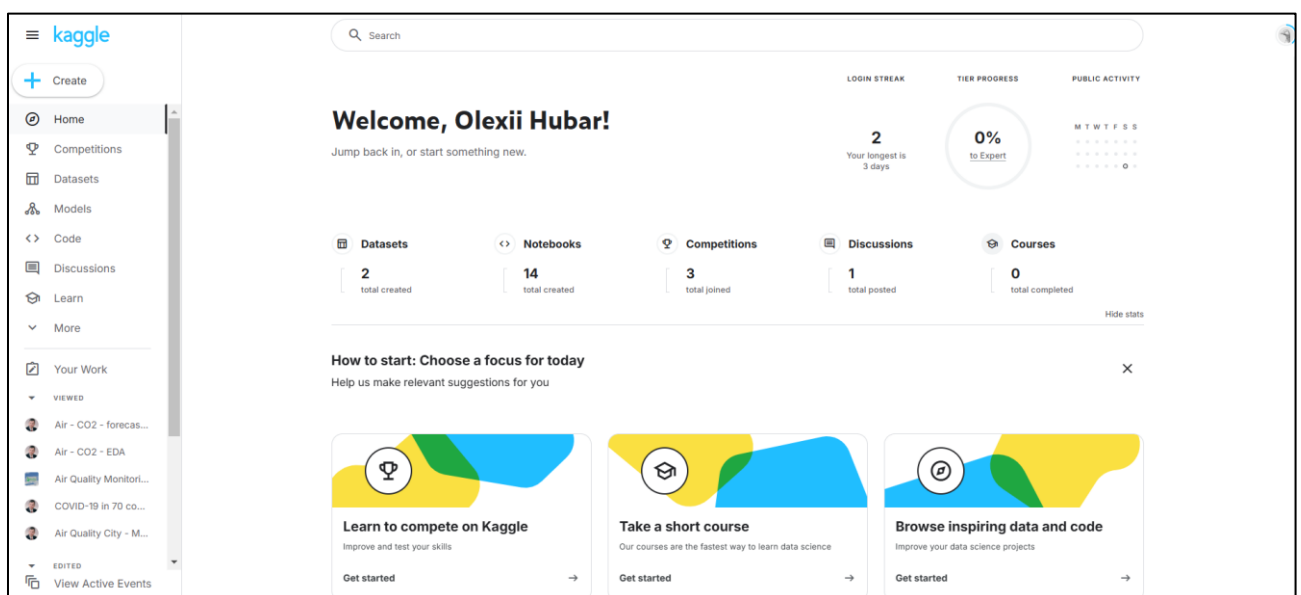


Рисунок 2.1 – Головна сторінка платформи [24]

На даній сторінці у користувача є можливість виконати пошук бажаного датасету. Переглянути змагання в яких користувач приймав участь чи активні змагання. Різні датасети, моделі, форум для обговорення чи навчальні матеріали.

## 2.2 Розвідувальний аналіз даних

Розвідувальний аналіз даних (EDA) є необхідним етапом у процесі машинного навчання та аналізу даних. Він дозволяє зрозуміти структуру, характеристики та взаємозв'язки даних, що є ключовим для побудови ефективних моделей. У даному розділі буде описано етапи та методи, які використовуються для проведення EDA на прикладі датасету про якість повітря.

Перевіримо чи є ряд даних стаціонарним чи ні. Для цього будемо використовувати метод Діка-Фулера. У статистиці тест Дікі-Фуллера перевіряє нульову гіпотезу про наявність одиничного кореня в авторегресійній (AR) моделі часового ряду. Альтернативна гіпотеза відрізняється залежно від версії тесту, але зазвичай є стаціонарною або тренд-стаціонарною. Тест названо на честь статистиків Девіда Дікі та Вейна Фуллера, які розробили його в 1979 році. Існує також розширення тесту Дікі-Фуллера (DF), яке називається розширений тест Дікі-Фуллера (ADF), який усуває всі структурні ефекти (автокореляцію) у часовому ряді, а потім тестує за тією ж процедурою. На рисунку 2.2 зображено програмний код для перевірки на стаціонарність ряду.

```
def check_stationarity(series):
    result = adfuller(series.values)

    print('ADF Statistic: %f' % result[0])
    print('p-value: %f' % result[1])
    print('Critical Values:')
    for key, value in result[4].items():
        print('\t%s: %.3f' % (key, value))

    if (result[1] <= 0.05) & (result[4]['5%'] > result[0]):
        print("\u001b[32mStationary\u001b[0m")
    else:
        print("\x1b[31mNon-stationary\x1b[0m")
```

Рисунок 2.2 – Програмний код для перевірки стаціонарності

На рисунку 2.3 зображено результат перевірки даного ряду на стаціонарність.

```
ADF Statistic: -4.314314  
p-value: 0.000420  
Critical Values:  
    1%: -3.436  
    5%: -2.864  
   10%: -2.568  
Stationary
```

Рисунок 2.3 – Результат перевірки ряду на стаціонарність

Як видно з рисунку 2.3 ряд є стаціонарним. Оскільки важливим для прогнозу є якомога менший порядок його і будемо використовувати.

Проаналізуємо аномалії ряду даних. Аналіз періоду досліджень показав, що, мали місце такі аномалії:

Аномалія, зображена на рисунку 2.4 могла виникнути з декількох причин але враховуючи те, що 2.06. 2023 була п'ятниця і зазначений час був о 18:00, то можна припустити, що дана аномалія виникла через короткий робочий день підприємств поблизу або зменшений потік автотранспорту.

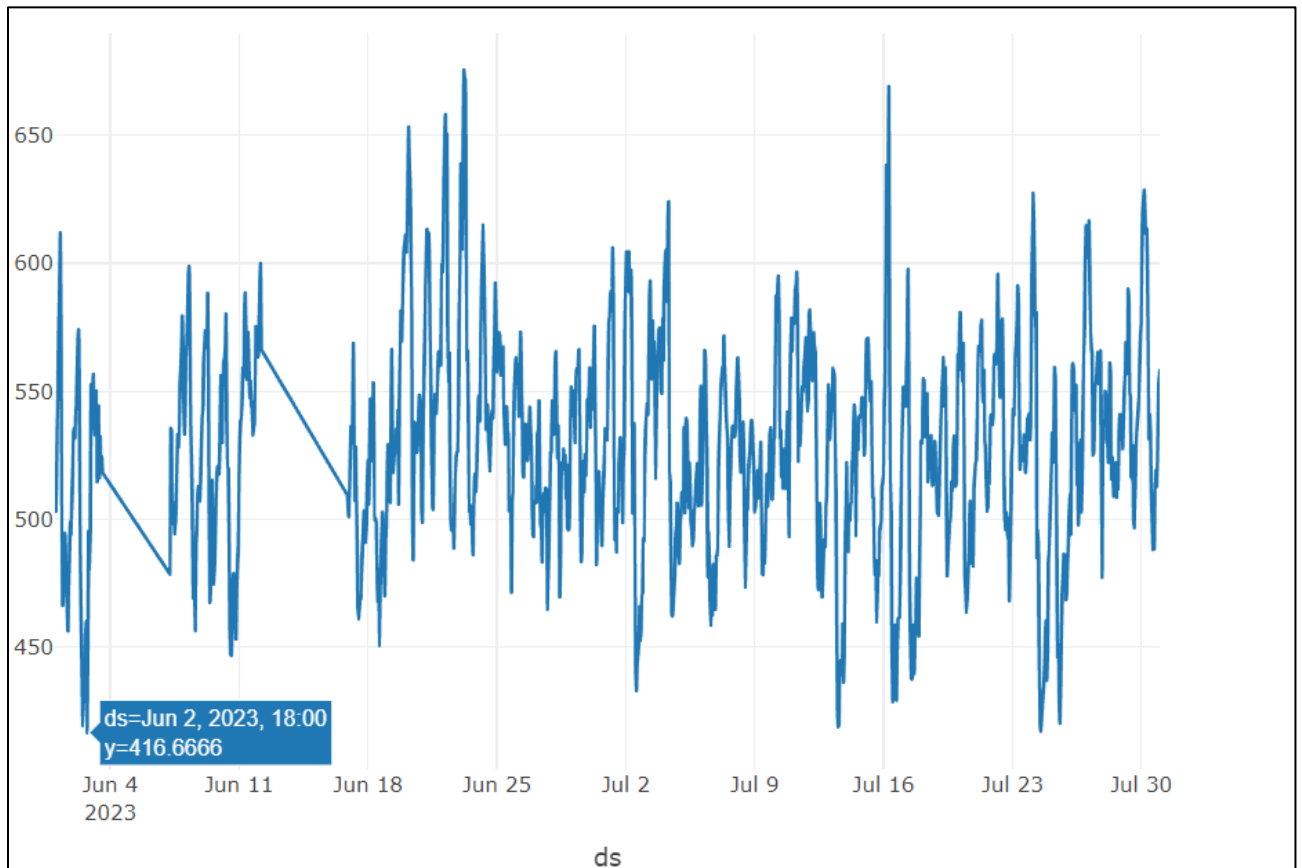


Рисунок 2.4 – Аномалія 2.06.23р. зниження вмісту CO2 в атмосфері

Наступна аномалія сталася в період з 4.06.2023 – 7.06.2023 і пов’язана або з відключенням електроенергії через масові обстріли російської федерації або через нестабільне інтернет підключення по тій самій причині. Також дана аномалія могла виникнути через збій самої системи моніторингу якості атмосферного повітря або некоректно передані дані внаслідок різноманітних причин (рис. 2.5).

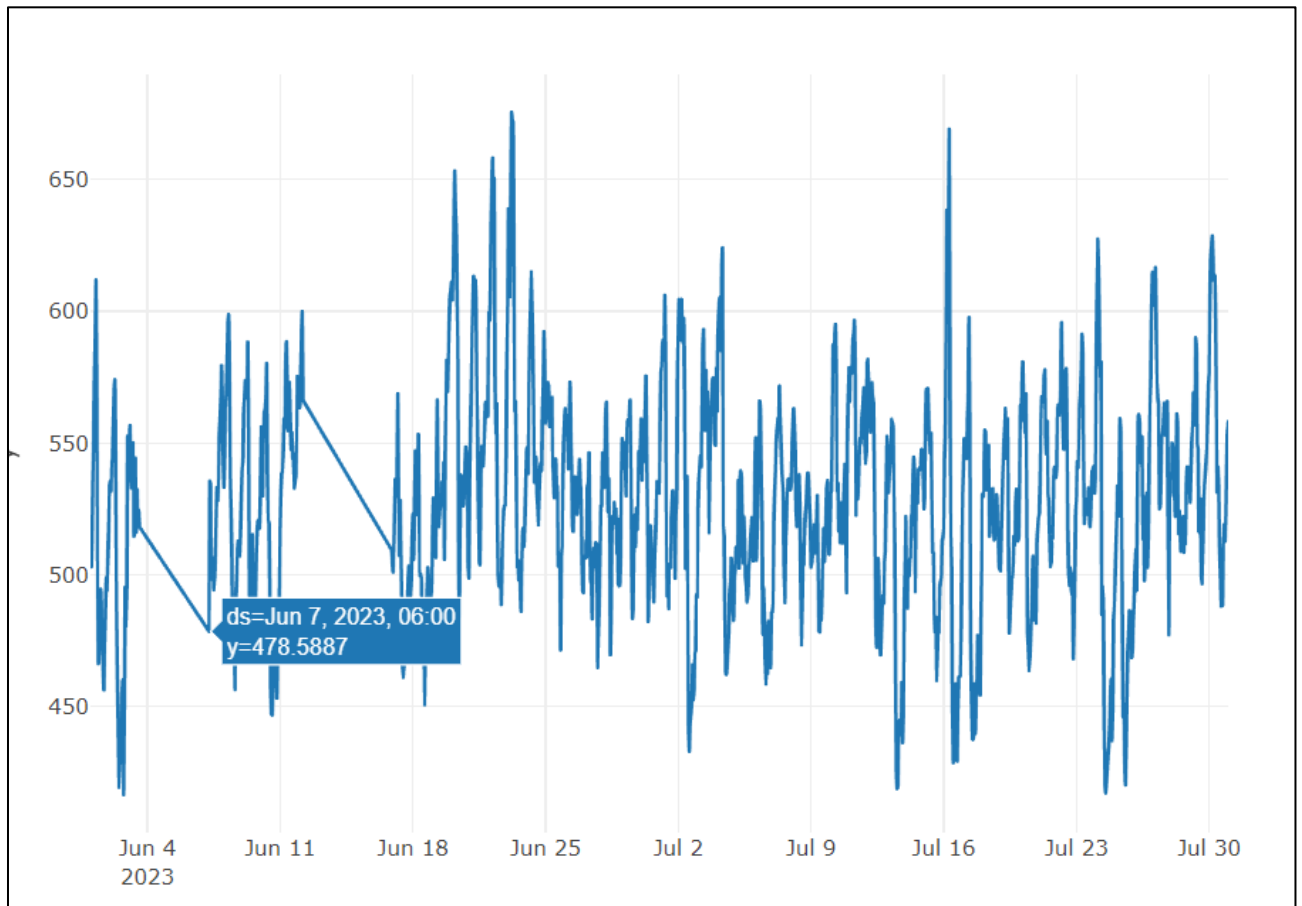


Рисунок 2.5 – Аномалія 4.06.2023 – 7.06.2023

Наступна аномалія знову виникла через нестабільний інтернет зв'язок, вимкнення електроенергії внаслідок обстрілів, некоректно передані дані чи збій роботи самої системи моніторингу якості повітря (рис. 2.6).

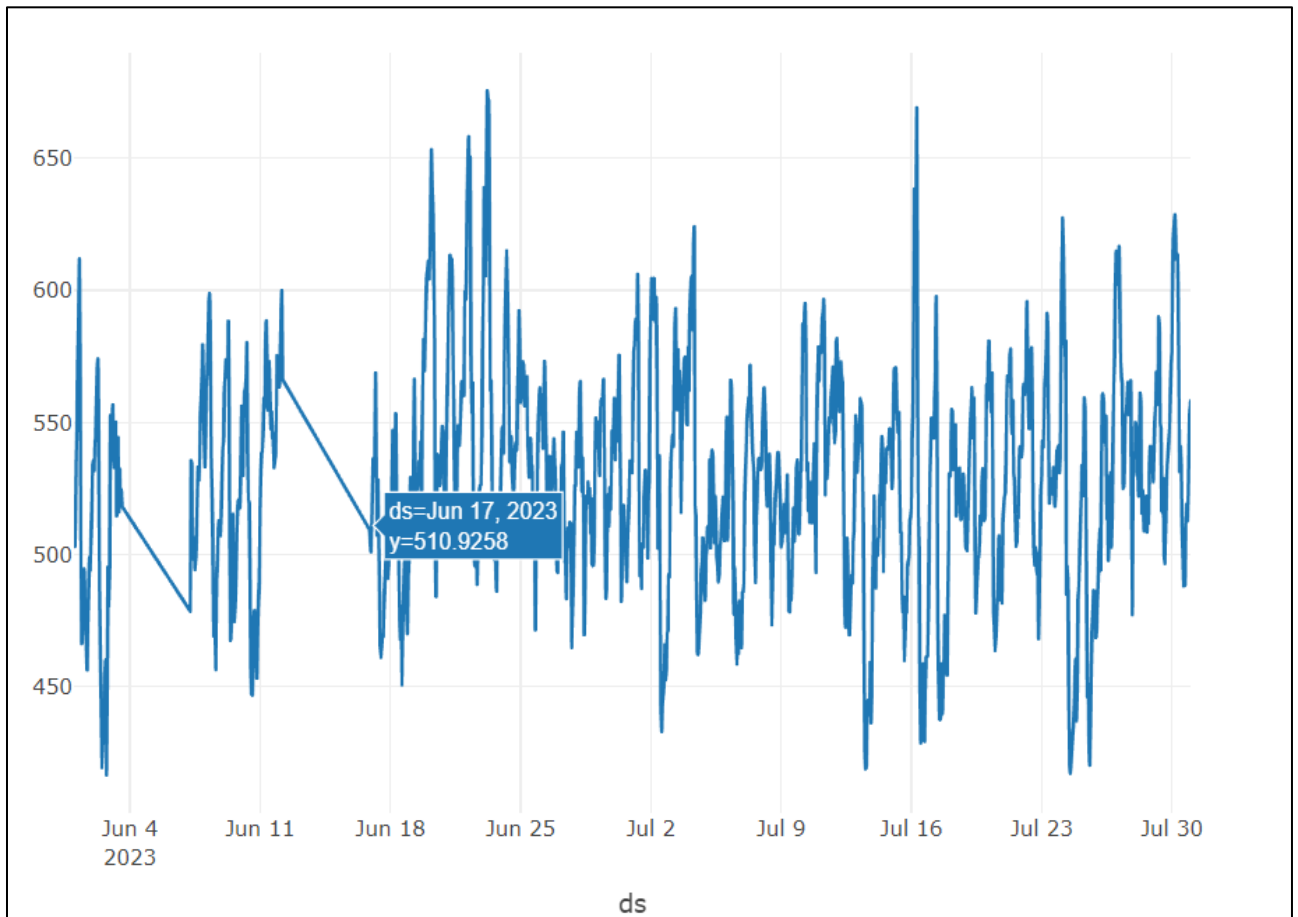


Рисунок 2.6 – Аномалія 13.06.23 – 17.06.23р.

Наступна аномалія могла виникнути також через певні причини. Збільшення вмісту вуглекислого газу в атмосферному повітрі можуть бути пов'язані з підсиленою роботою забруднюючих підприємств, що знаходяться поблизу, також через великий автомобільний трафік чи погодні умови, що цьому сприяли (рис. 2.7).



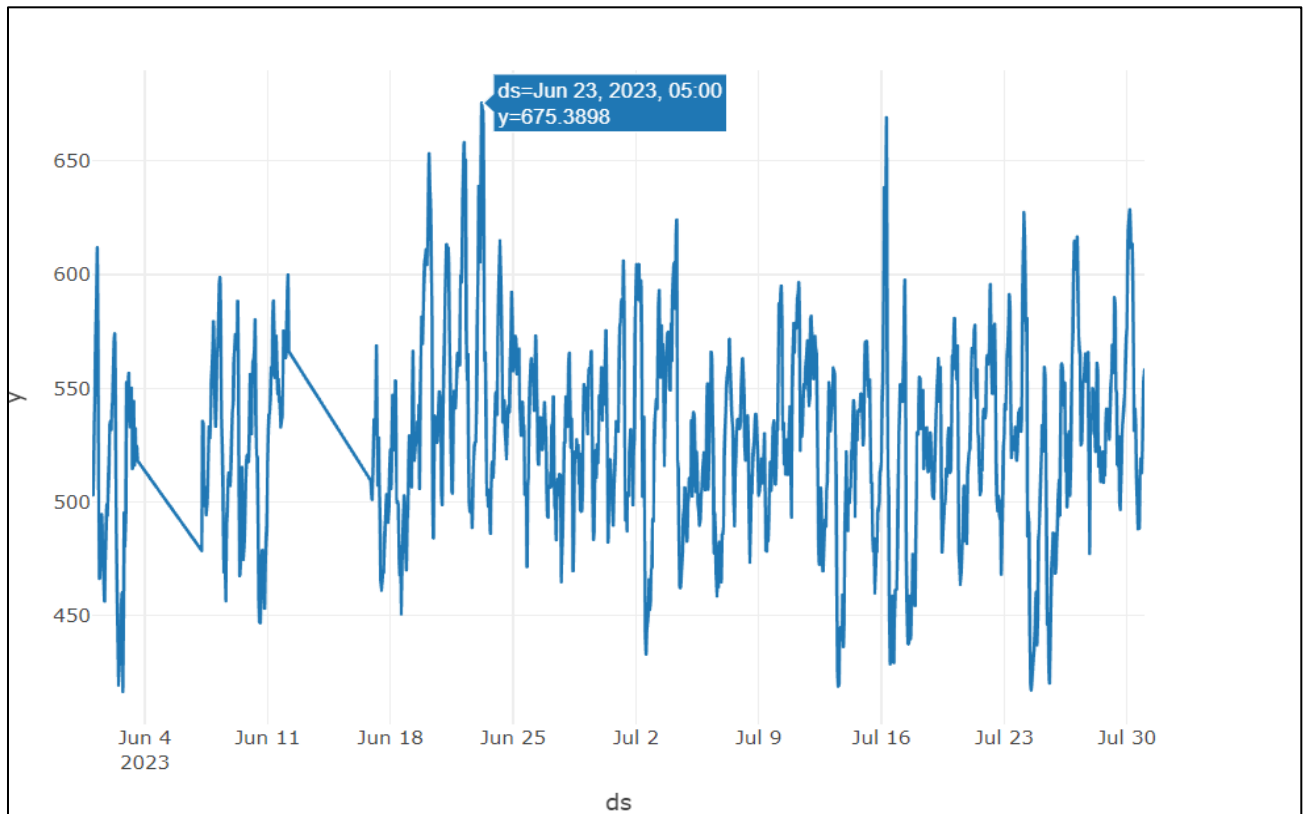


Рисунок 2.7 – Аномалія 23.06.2023 різке збільшення вуглекислого газу

Наступна аномалія різкого зниження вмісту вуглекислого газу в атмосферному повітрі відбулась 2.07.2023 це була неділя і така аномалія пов'язана з зменшеним потоком автомобільного трафіку чи вихідним днем на забруднюючих підприємствах (рис. 2.8).

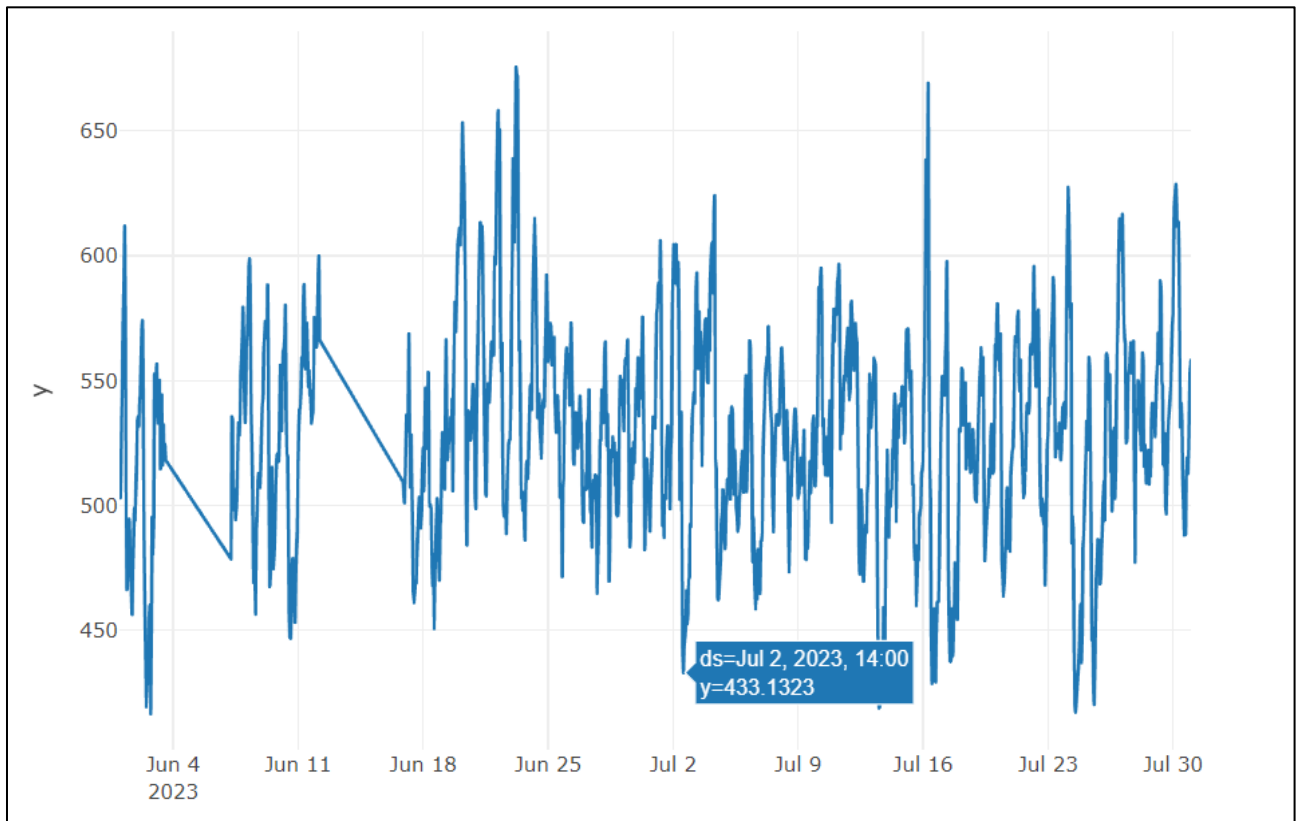


Рисунок 2.8 – Аномалія 2.07.2023 різке зниження вмісту вуглекислого газу

Наступна аномалія зниження вмісту вуглекислого газу відбулася 13.07.2023 це був четвер і пов'язана з зменшенням автомобільного трафіку, зменшенням викидів забруднюючих підприємств, погодні умови чи можливо збій роботи системи моніторингу якості атмосферного повітря чи некоректність введених даних (рис. 2.9).

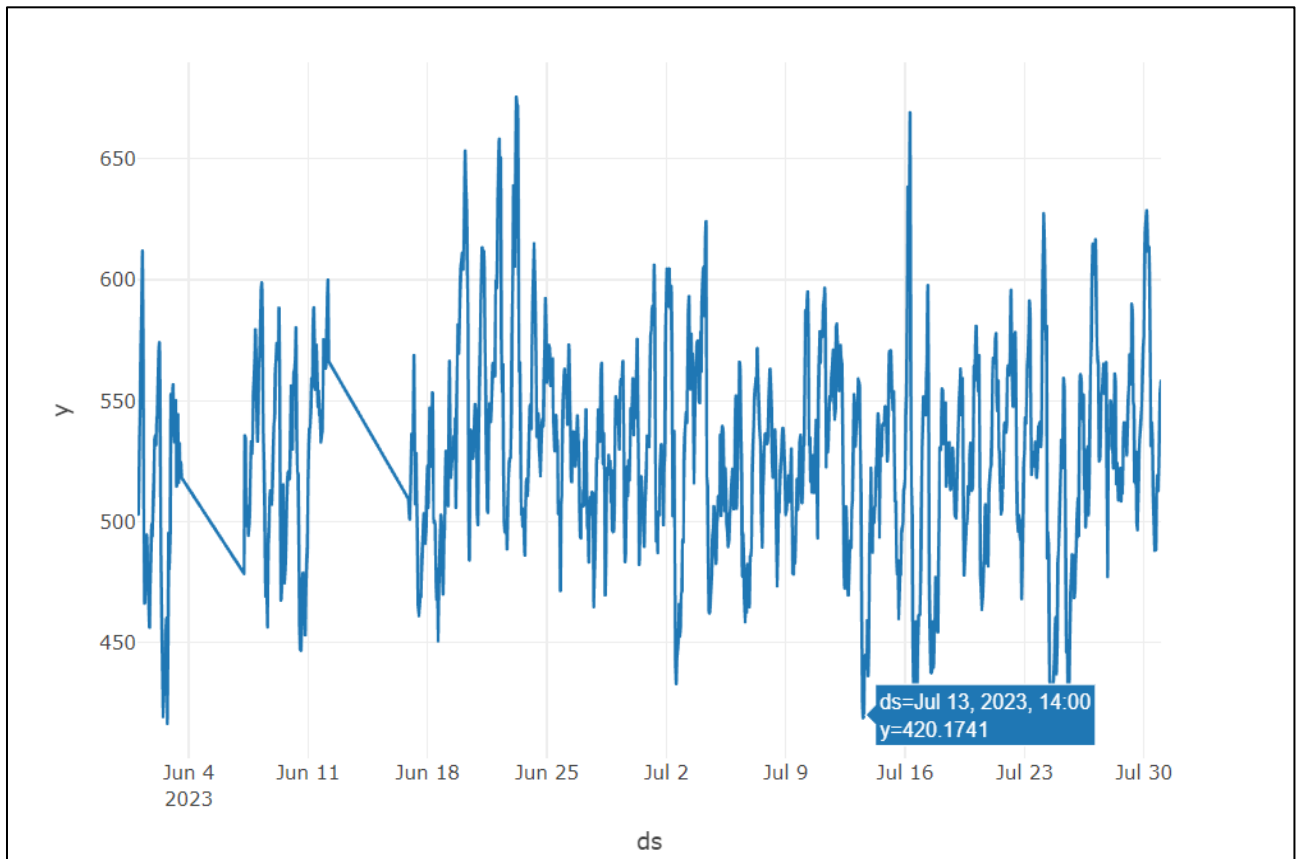


Рисунок 2.9 – Аномалія 13.07.2023 різкого зниження вмісту вуглекислого газу в атмосфері

Настпуна аномалія могла виникнути внаслідок збільшення викидів забруднюючих підприємств, що знаходяться поблизу, збільшення руху автомобільного трафіку або некоректна робота системи моніторингу (рис. 2.10).

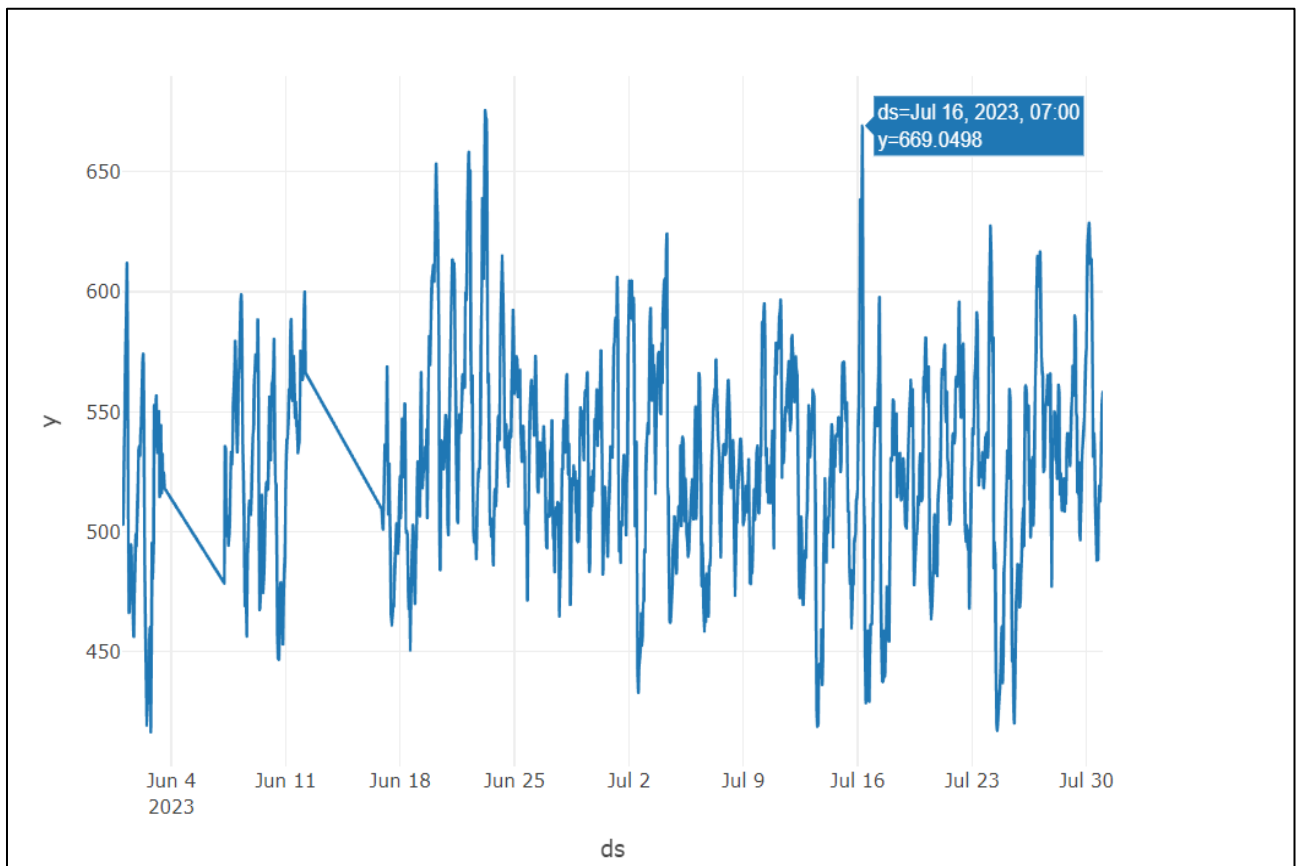


Рисунок 2.10 – Аномалія 16.07.2023 07:00 збільшення вмісту вуглекислого газу в атмосферному повітрі

Оскільки того самого дня о 7:00 ранку спостерігалось різке збільшення вмісту вуглекислого газу, то можна припустити, що дана аномалія також пов'язана із проблемами роботи системи, погодними умовами чи різким зменшенням викидів забруднюючих джерел (рис. 2.11).

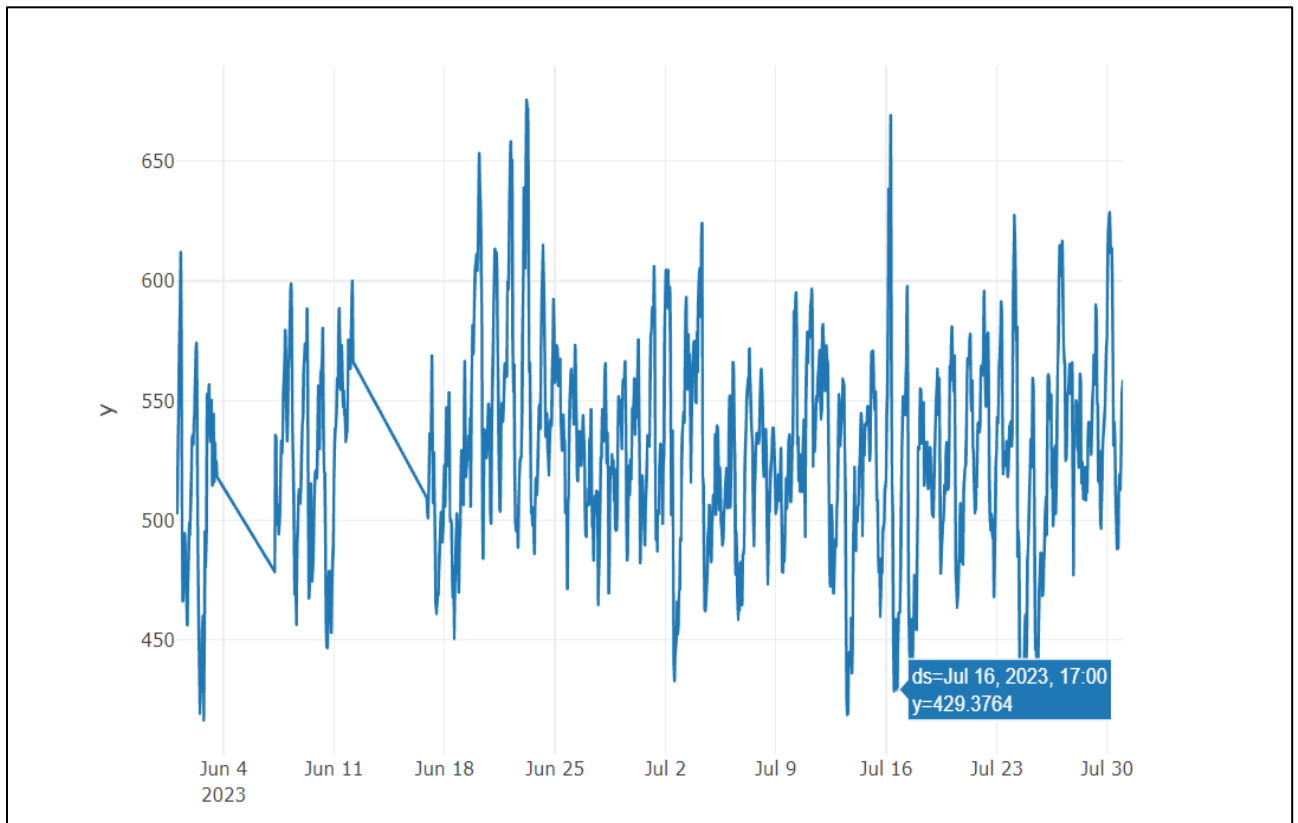


Рисунок 2.11 – Аномалія 16.07.2023 17:00 різке зменшення вмісту вуглекислого газу в атмосфері

Наступна аномалія припадає на понеділок, що є початком робочого тижня і скоріше за все підвищена активність забруднюючих джерел: заводів, підприємств, тощо, а також підвищена температура могли вплинути на підвищення рівня вуглекислого газу в атмосферному повітрі (рис. 2.12).

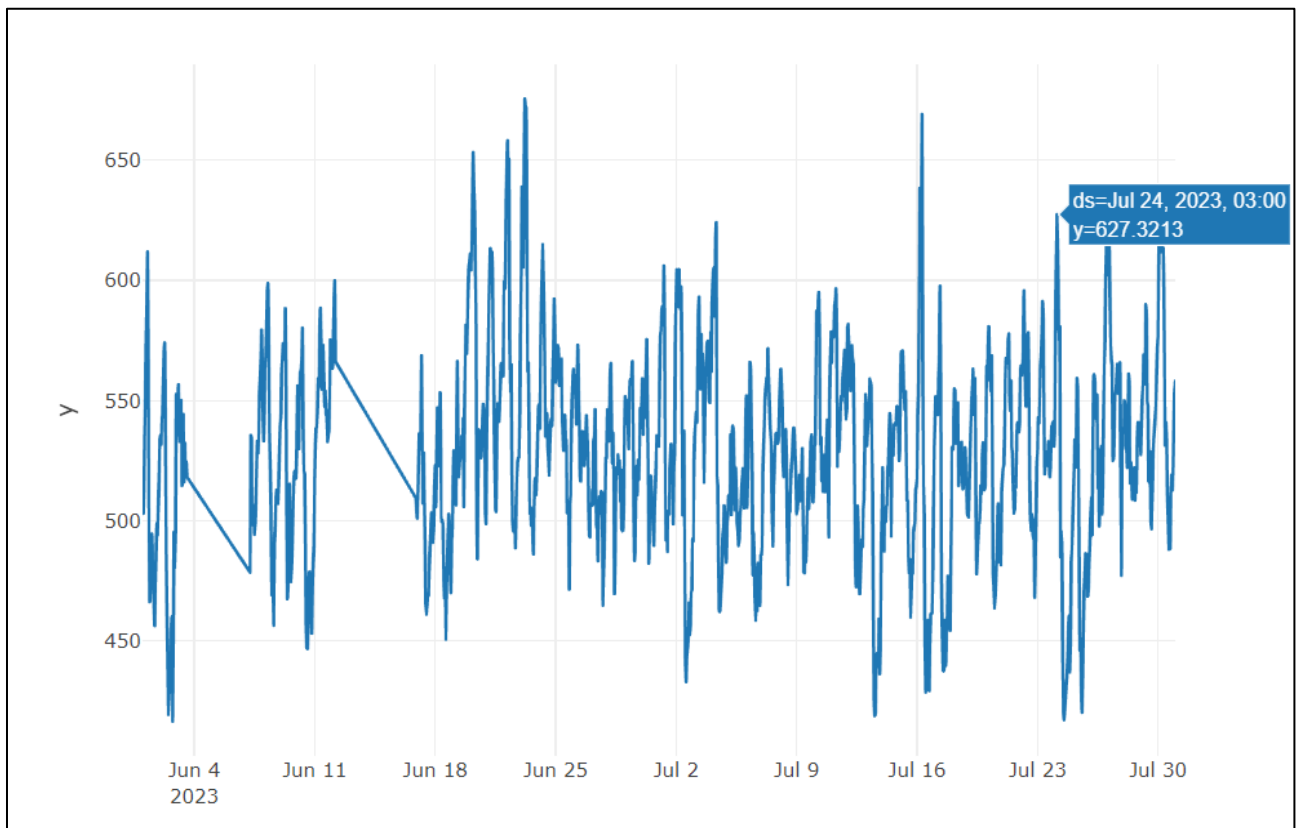


Рисунок 2.12 – Аномалія 24.07.2023 03:00 різке збільшення вмісту вуглекислого газу в атмосфері

Наступна аномалія була зафіксована в той день, що й минула, але уже зі зменшенням вмісту вуглекислого газу в атмосферному повітрі, скоріше за все, вона виникла через неправильну фіксацію даних або через зміни погодних умов, можливо в той день був дощ чи інше погодне явище (рис. 2.13).

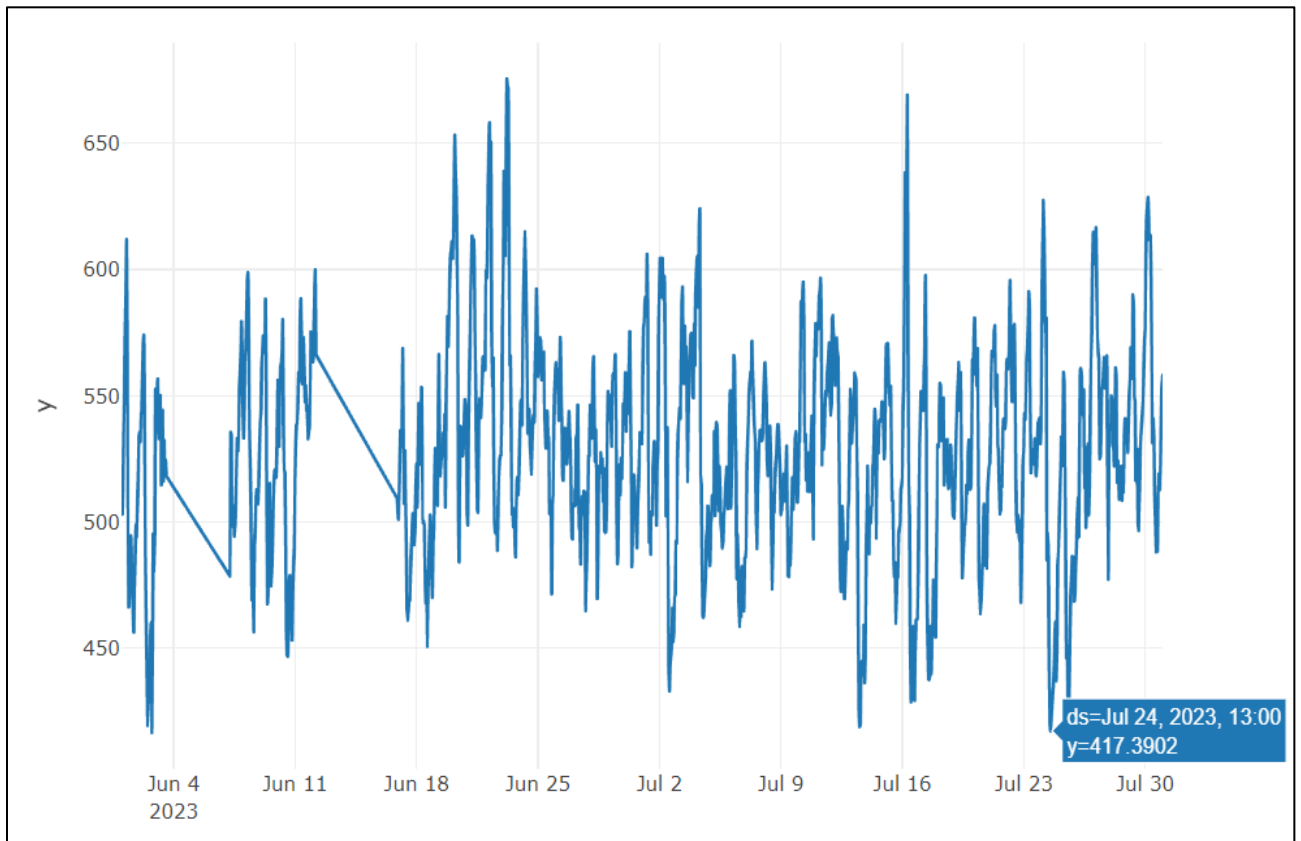


Рисунок 2.13 – Аномалія 24.07.2023 13:00 різке зменшення вмісту вуглекислого газу в атмосферному повітрі

Оскільки зменшений вміст вуглекислого газу у повітрі відбувається другий день підряд, то можна зробити висновок, що дана аномалія виникла по тим самим причинам, що й минула, а саме: певні погодні умови, некоректна фіксація даних або зменшена кількість викидів забруднюючих джерел біля станції (рис. 2.14).



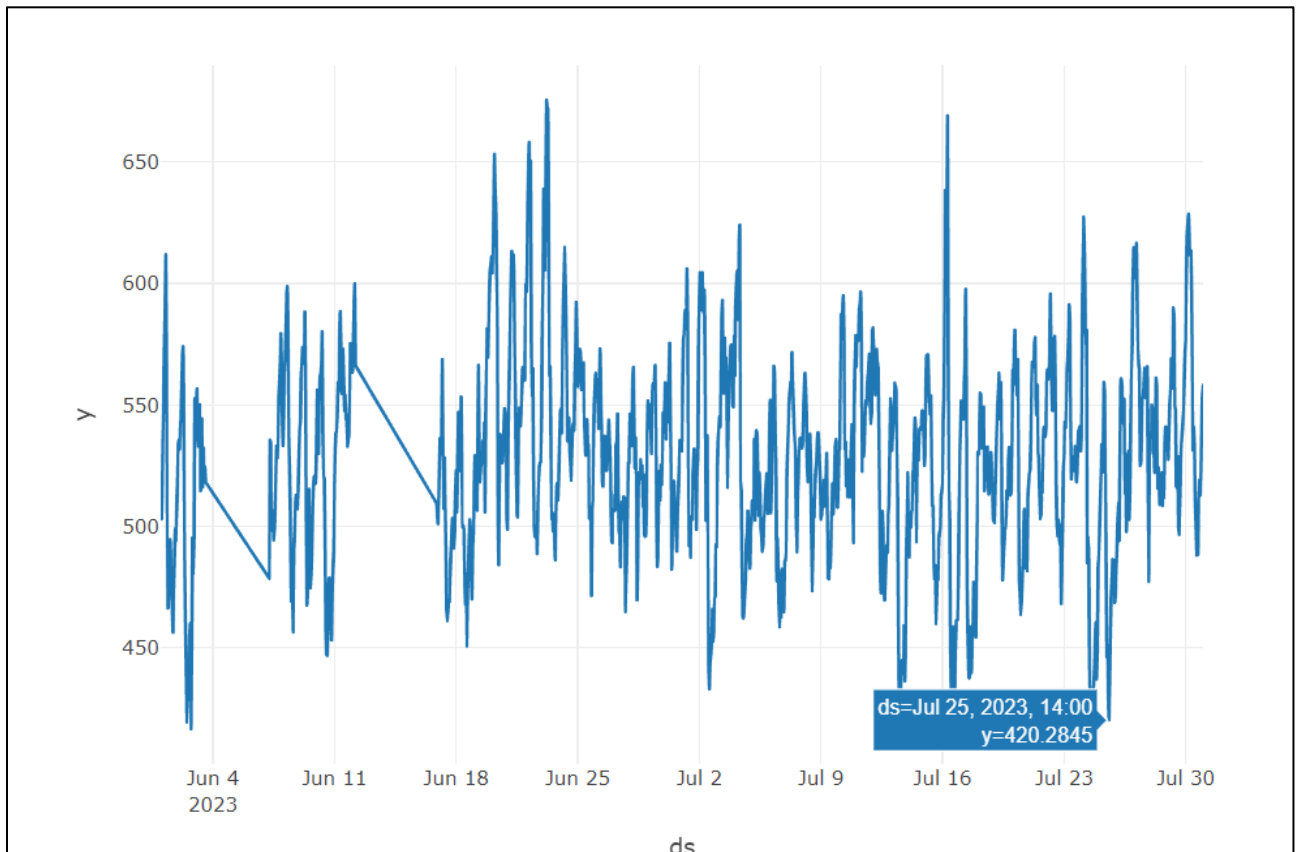


Рисунок 2.14 – Аномалія 25.07.2023 14:00 різке зменшення вмісту вуглекислого газу в атмосферному повітрі

Наступна аномалія могла виникнути через низку причин. Збільшення викидів забруднюючих речовин різними джерелами, погодні умови, що викликали дану аномалію, неправильне передавання даних чи їх фіксація станцією моніторингу атмосферного повітря (рис. 2.15).

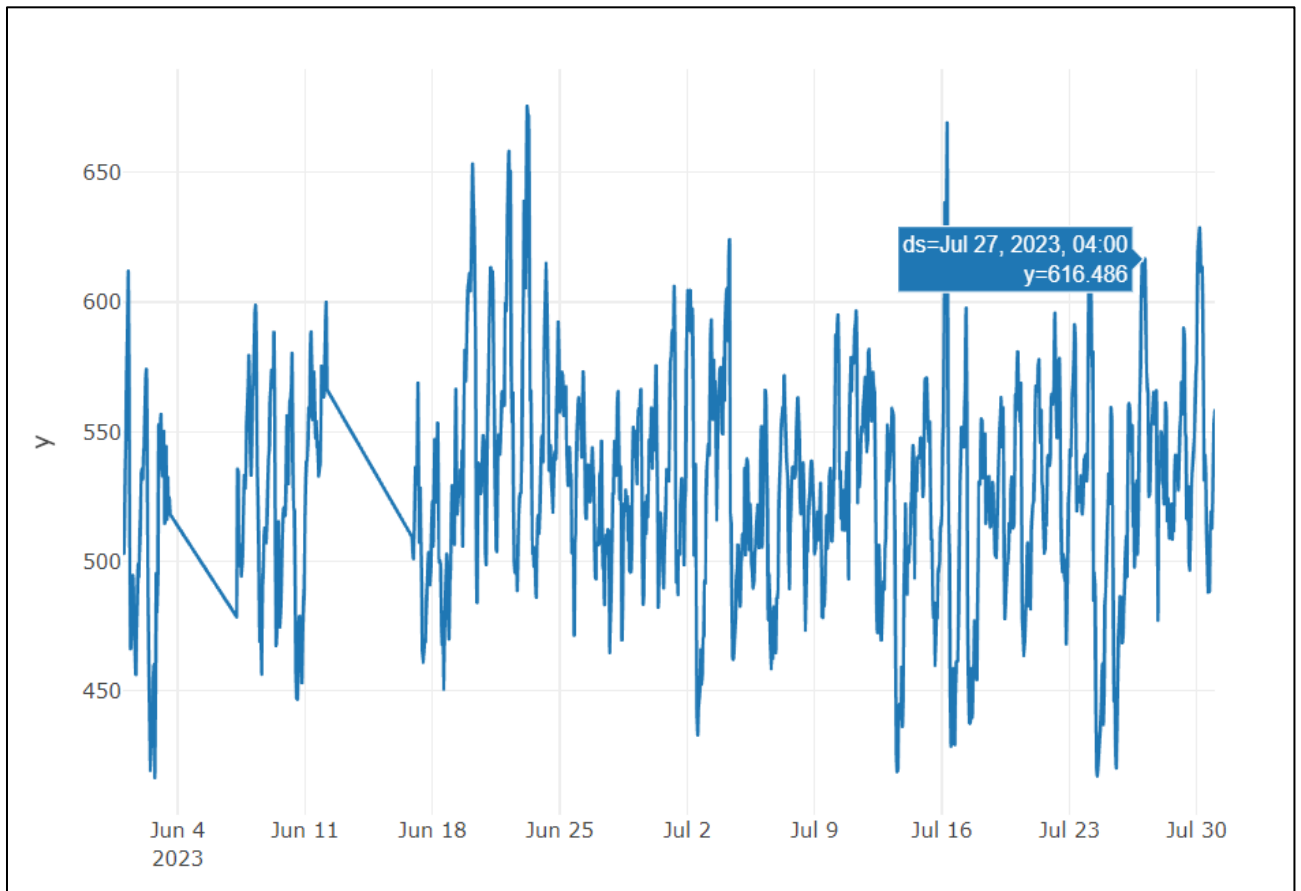


Рисунок 2.15 – Аномалія 27.07.2023 04:00 різке збільшення вмісту вуглекислого газу в атмосферному повітрі

Природу даної аномалії можна пояснити збільшенням викидів забруднюючих речовин, автомобільним трафіком чи високою температурою, що і спричинила високий вміст вуглекислого газу в атмосферному повітрі чи можливо неправильною фіксацією даних і їх передавання (рис. 2.16).

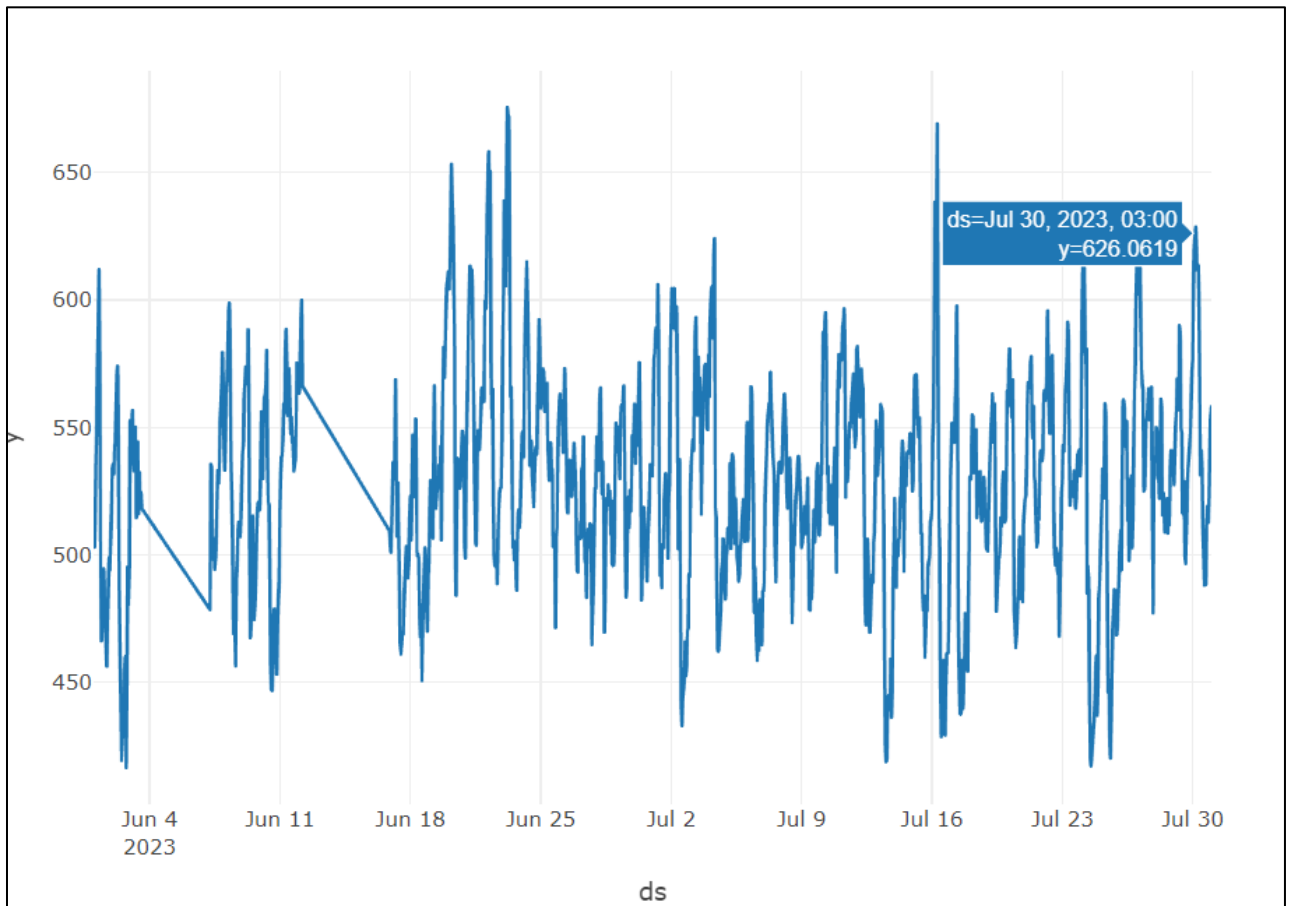


Рисунок 2.16 – Аномалія 30.07.2023 03:00 різке збільшення вмісту вуглекислого газу в атмосферному повітрі

На рисунку 2.17 наведено сформований датасет із таких дат аномальних змін, який варто використати під час передбачення вмісту вуглекислого газу.

|    | ds         | lower_window | upper_window | prior_scale | holiday         |
|----|------------|--------------|--------------|-------------|-----------------|
| 0  | 2023-06-02 | 0            | 0            | 10          | anomalous_dates |
| 1  | 2023-06-03 | 0            | 0            | 10          | anomalous_dates |
| 2  | 2023-06-07 | 0            | 0            | 10          | anomalous_dates |
| 3  | 2023-06-20 | 0            | 0            | 10          | anomalous_dates |
| 4  | 2023-06-22 | 0            | 0            | 10          | anomalous_dates |
| 5  | 2023-06-23 | 0            | 0            | 10          | anomalous_dates |
| 6  | 2023-07-02 | 0            | 0            | 10          | anomalous_dates |
| 7  | 2023-07-04 | 0            | 0            | 10          | anomalous_dates |
| 8  | 2023-07-13 | 0            | 0            | 10          | anomalous_dates |
| 9  | 2023-07-16 | 0            | 0            | 10          | anomalous_dates |
| 10 | 2023-07-17 | 0            | 0            | 10          | anomalous_dates |
| 11 | 2023-07-24 | 0            | 0            | 10          | anomalous_dates |
| 12 | 2023-07-25 | 0            | 0            | 10          | anomalous_dates |
| 13 | 2023-07-27 | 0            | 0            | 10          | anomalous_dates |
| 14 | 2023-07-30 | 0            | 0            | 10          | anomalous_dates |

Рисунок 2.17 – Датасет знайдених аномалій

Дані аномалії можуть завадити точності передбачення чи призвести до неправильних висновків, що може потім призвести до помилкових рішень, що в свою чергу може нанести великі збитки, особливо в критичних галузях, наприклад, в медицині.

## 2.2 Побудова моделей часових рядів з використанням Facebook Prophet

Модель FB Prophet є ефективною тільки для самих значень ряду, а не для його різниці, які – надто зашумлені.

Побудовані тренувальний, валідаційний та тестовий датасети з усіма ознаками вмісту вуглекислого газу наведено на рисунку 2.18.

```
Origin dataset has 720 rows and 2 features
Get training dataset with 710 rows
Get validation dataset with 5 rows
Get test dataset with 5 rows
```

Рисунок 2.18 – Розмір побудованих тренувального, валідаційного та тестового датасетів з усіма ознаками вмісту вуглекислого газу

Prophet - це процедура передбачення даних часових рядів на основі адитивної моделі, в якій нелінійні тренди підганяються під річну, тижневу та денну сезонність, а також під святкові ефекти. Найкраще працює з часовими рядами, які мають сильні сезонні ефекти та кілька сезонів історичних даних. Prophet стійкий до пропущених даних і зсувів у тренді.

Використаємо перебір різних варіантів параметрів моделі. Результат ідентифікації оптимальних параметрів та структури моделі подано на рисунку 2.19.

```
model.add_seasonality(name='seasonality', period=period_days,
                      fourier_order=fourier_order_seasonality,
                      mode = 'multiplicative', prior_scale = 0.5)
```

Рисунок 2.19 – Ідентифіковані оптимальні параметри та структура моделі для передбачення вмісту вуглекислого газу в атмосферному повітрі технологією Facebook Prophet

На рисунку 2.20 наведено приклад результату передбачення валідаційних даних за цією моделлю.

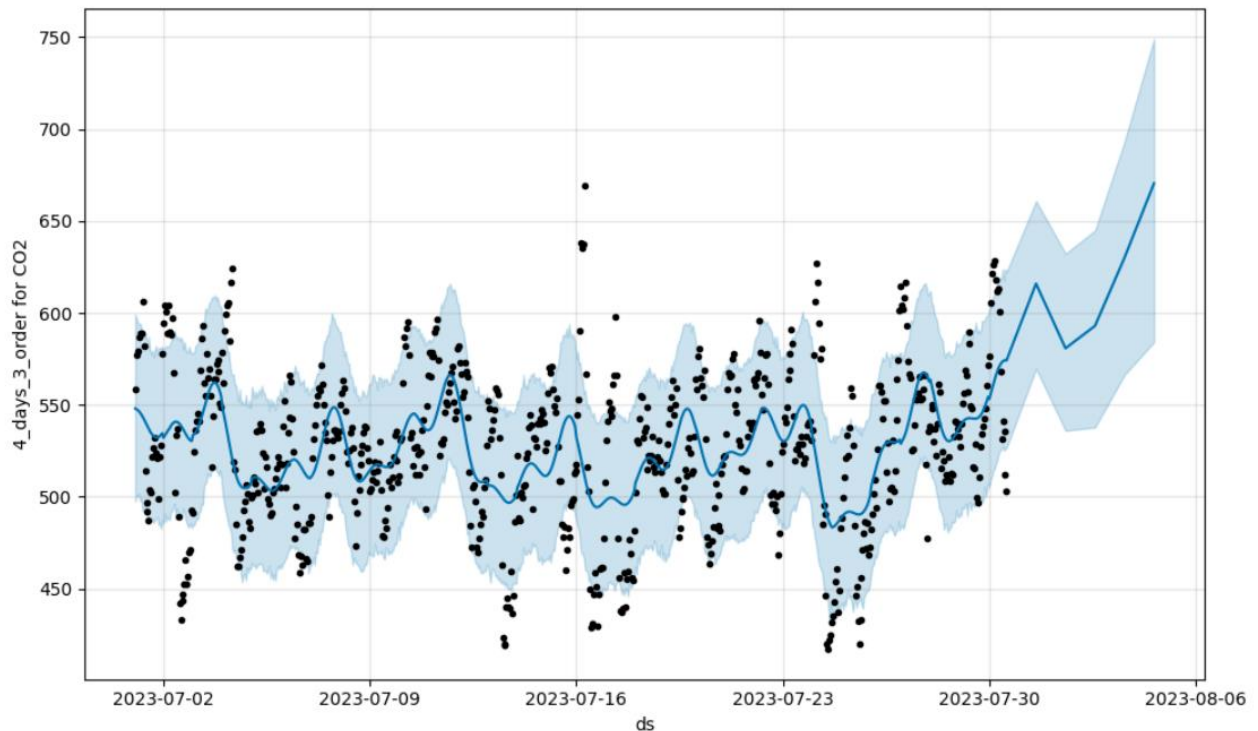


Рисунок 2.20 – Результат передбачення вмісту вуглекислого газу в атмосферному повітрі за валідаційними даними на основі моделі Facebook Prophet

Як видно з даного графіку чорні точки це реальні дані. Синя лінія це прогнозовані значення моделі, синя область навколо лінії вказує на ступінь невизначеності прогнозу, адже чим ширша область, тим більша невизначеність. На графіку видні циклічні коливання, а також ми можемо спостерігати як в кінці прогнозується збільшений вміст вуглекислого газу, що може бути наслідком довгострокового тренду, який модель розраховувала на основі історичних даних. Загалом синя лінія добре співпадає з чорними точками, що свідчить про її якісне моделювання.

## 2.3 Побудова інших моделей машинного навчання

На рисунку 2.21 наведено тренувальний, валідаційний та тестовий датасети з усіма ознаками вмісту вуглекислого газу.

```
Origin dataset has 715 rows and 79 features
Get training dataset with 705 rows
Get validation dataset with 5 rows
Get test dataset with 5 rows
```

Рисунок 2.21 – Тренувальний, валідаційний та тестовий датасети з усіма ознаками вмісту вуглекислого газу

На рисунку 2.22 зазначено вибір всіх моделей для передбачення вмісту вуглекислого газу.

```
tuning model 'Linear Regression'
best parameters: {'fit_intercept': True}

tuning model 'KNeighbors Regressor'
best parameters: {'leaf_size': 10, 'n_neighbors': 30}

tuning model 'Support Vector Machines'
best parameters: {'C': 4.0, 'kernel': 'rbf', 'tol': 0.001}

tuning model 'Linear SVR'
best parameters: {'C': 1.0}

tuning model 'Random Forest Regressor'
best parameters: {'max_depth': 3, 'max_features': 'auto', 'min_samples_leaf': 10, 'min_samples_split': 40, 'n_estimators': 80}

tuning model 'Bagging Regressor'
best parameters: {'max_features': 0.05, 'n_estimators': 6, 'warm_start': False}

tuning model 'XGB Regressor'
best parameters: {'learning_rate': 0.01, 'max_depth': 3, 'n_estimators': 90}

tuning model 'MLP Regressor'
best parameters: {'hidden_layer_sizes': 2, 'learning_rate': 'adaptive', 'learning_rate_init': 0.001, 'max_iter': 1000, 'solver': 'sgd'}
```

Рисунок 2.22 – Всі моделі для передбачення вмісту вуглекислого газу

## 2.4 Вибір оптимальної моделі

Після обчислення усіх моделей на валідаційних даних здійснюється порівняння прогнозів з відомими цільовими значеннями «CO<sub>2</sub>» у валідаційному датасеті за 1-ю метрикою: MAPE. Середня абсолютна відсоткова помилка (MAPE), також відома як середнє абсолютне відсоткове відхилення (MAPD), є мірою точності передбачення методу передбачення



в статистиці. Зазвичай він виражає точність у вигляді співвідношення, яке визначається формулою (2.1):

$$MAPE = 100 \frac{1}{n} \sum_{t=1}^n \left| \frac{A_t - F_t}{A_t} \right| \quad (2.1)$$

де  $A_t$  – фактичне значення, а  $F_t$  – прогнозоване значення. Їх різницю ділять на фактичне значення  $A_t$ . Абсолютне значення цього співвідношення підсумовується для кожного прогнозованого моменту часу та ділиться на кількість підігнаних точок  $n$ .

Результат наведено на рисунку 2.23.

|    | name_model                | type_data | mape      |
|----|---------------------------|-----------|-----------|
| 12 | Linear SVR                | valid     | 1.93953   |
| 13 | Random Forest Regressor   | valid     | 1.991942  |
| 15 | XGB Regressor             | valid     | 2.025112  |
| 8  | ARIMA_auto                | valid     | 2.039046  |
| 10 | KNeighbors Regressor      | valid     | 2.18653   |
| 9  | Linear Regression         | valid     | 2.498551  |
| 16 | MLP Regressor             | valid     | 2.814234  |
| 11 | Support Vector Machines   | valid     | 2.996162  |
| 14 | Bagging Regressor         | valid     | 3.003043  |
| 5  | Prophet_30_days_12_order  | valid     | 4.728317  |
| 7  | Prophet_365_days_12_order | valid     | 12.123947 |
| 1  | Prophet_4_days_12_order   | valid     | 17.116291 |
| 0  | Prophet_4_days_3_order    | valid     | 21.692297 |
| 6  | Prophet_365_days_3_order  | valid     | 25.851544 |
| 3  | Prophet_7_days_12_order   | valid     | 28.985652 |
| 2  | Prophet_7_days_3_order    | valid     | 32.211638 |
| 4  | Prophet_30_days_3_order   | valid     | 43.620066 |

Рисунок 2.23 – Похибки усіх моделей, побудованих для валідаційних даних передбачення вмісту вуглекислого газу в атмосферному повітрі

З рисунку видно, що найкраща модель за метрикою MAPE є Linear SVR з похибкою в 1,93953%.

## 2.5 Висновки

У другому розділі було проведено розвідувальний аналіз даних (EDA), проведено аналіз на стаціонарність та позначенні аномалії даних ряду та описані можливі чинники виникнення даних аномалій. Було побудовано модель використовуючи технологію FB Prophet, а також ряд інших моделей для визначення оптимальної з них. За результатами тренування було обрано модель Linear SVR за метрикою MAPE з абсолютною похибкою в 1,93953%.

### **3 РОЗРОБЛЕННЯ ІНФОРМАЦІЙНОЇ ТЕХНОЛОГІЇ ПЕРЕДБАЧЕННЯ ВМІСТУ ДІОКСИДУ ВУГЛЕЦЮ У ПОВІТРІ**

#### **3.1 Розроблення інформаційної технології передбачення вмісту вуглекислого газу**

Для розв'язання поставленої задачі для тестових даних пропонується такий алгоритм (рис. 3.1):

1. Аналіз даних;
2. Формулювання цільової ознаки;
3. Формування тренувального, валідаційного та тестових датасетів для різних типів моделей.
4. Вибір і тренування моделей на тренувальному датасеті (train). Передбачення за кожною ідентифікованою моделлю валідаційних даних (valid).
5. Вибір оптимальної моделі.
6. Тренування оптимальної моделі за розширеним датасетом (train+valid). Передбачення за ідентифікованою моделлю тестових даних (test).
7. Оцінювання точності прогнозу тестових даних.

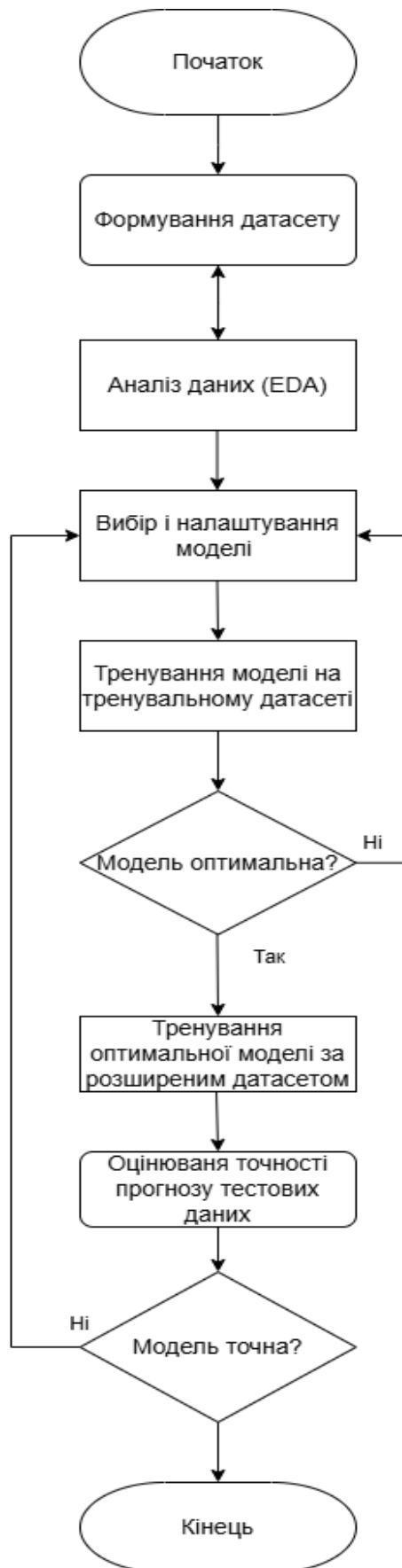


Рисунок 3.1 – Блок-схема алгоритму інформаційної технології передбачення вмісту вуглекислого газу в атмосферному повітрі

3.2 Передбачення тестових даних

Вибрана у розділі 2 оптимальна модель була заново ідентифікована на вже на розширеному датасеті «train+valid», після чого було зроблено передбачення вже тестових даних та порівняння з реальними даними (рис. 3.2, рис. 3.3).

|    | name_model | type_data | mape     |
|----|------------|-----------|----------|
| 12 | Linear SVR | valid     | 1.927118 |

Optimal model by metrics "mape" is "Linear SVR" with type "Linear SVR" parameters {'C': 1.0}

Рисунок 3.2 – Похибки і оптимальні параметри моделі Linear SVR, оптимальної за метрикою MAPE

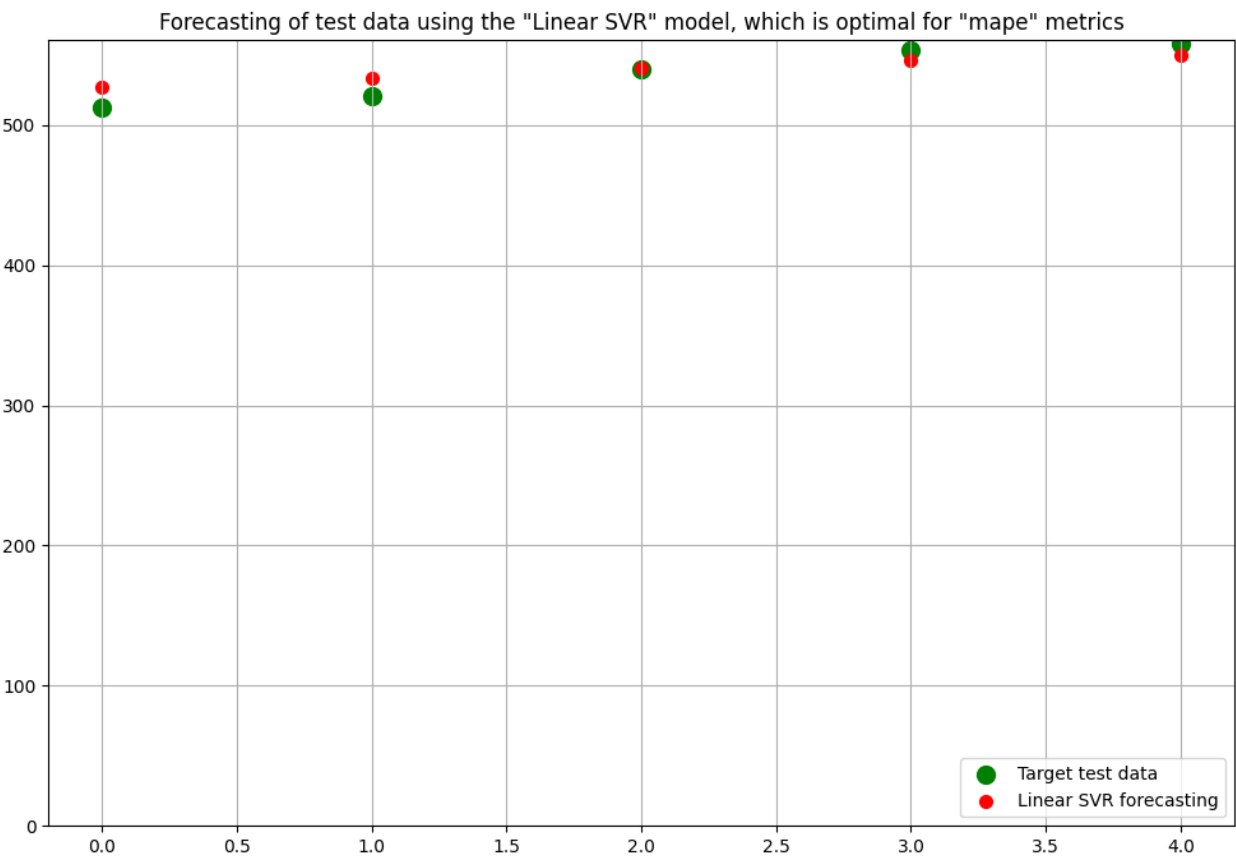


Рисунок 3.3 – Реальні тестові дані вмісту вуглекислого газу та їх прогноз за моделлю Linear SVR, оптимальної за метрикою MAPE

Як бачимо з рисунку 3.3 зелені точки є реальними тестовими даними, в той час червоні точки – прогнозованими даними, вони достатньо близько розташовуються одна до одної, що свідчить про високий відсоток передбачення.

### **3.3 Висновки**

У третьому розділі було розроблено інформаційну технологію аналізу та передбачення вмісту вуглекислого газу, представлено алгоритм дій для вирішення поставленої задачі. Даний алгоритм складався з формування датасету, потім визначення задачі, яку потрібно вирішити, вибір моделі і проведення її налаштування, потім тренування моделі на тренувальному датасеті, після цього визначаємо, яка із вибраних моделей була оптимальна, потім проводилось тренування на розширеному датасеті, що включав в себе і тренувальний і валідаційний датасети, в кінці ми визначаємо чи дана модель добре підходить для вирішення поставленої задачі чи ні. Також було проведено аналіз результатів моделі та порівняно результати тестування з реальними результатами. Звідси слідує, що дана модель добре підходить для вирішення поставленої задачі та має похибку в 1,927118%.

## 4 ЕКОНОМІЧНА ЧАСТИНА

Науково-технічна розробка має право на існування та впровадження, якщо вона відповідає вимогам часу, як в напрямку науково-технічного прогресу та і в плані економіки. Тому для науково-дослідної роботи необхідно оцінювати економічну ефективність результатів виконаної роботи.

### 4.1 Проведення комерційного та технологічного аудиту науково-технічної розробки

Метою проведення комерційного і технологічного аудиту дослідження за темою «Інформаційна технологія аналізу та передбачення діоксиду вуглецю за даними громадського моніторингу атмосферного повітря у м. Вінниці» є оцінювання науково-технічного рівня та рівня комерційного потенціалу розробки, створеної в результаті науково-технічної діяльності.

Оцінювання науково-технічного рівня розробки та її комерційного потенціалу рекомендується здійснювати із застосуванням 5-ти бальної системи оцінювання за 12-ма критеріями [25].

Таблиця 4.1 – Результати оцінювання науково-технічного рівня і комерційного потенціалу розробки експертами

| Критерії                                     | Експерт (ПІБ, посада) |   |   |
|----------------------------------------------|-----------------------|---|---|
|                                              | 1                     | 2 | 3 |
|                                              | Бали:                 |   |   |
| 1. Технічна здійсненність концепції          | 4                     | 5 | 4 |
| 2. Ринкові переваги (наявність аналогів)     | 4                     | 3 | 3 |
| 3. Ринкові переваги (ціна продукту)          | 3                     | 3 | 3 |
| 4. Ринкові переваги (технічні властивості)   | 3                     | 3 | 3 |
| 5. Ринкові переваги (експлуатаційні витрати) | 2                     | 2 | 3 |
| 6. Ринкові перспективи (розмір ринку)        | 3                     | 3 | 3 |

Продовження таблиці 4.1

| Критерії                                                    | Експерт (ПІБ, посада) |    |    |
|-------------------------------------------------------------|-----------------------|----|----|
|                                                             | 1                     | 2  | 3  |
|                                                             | Бали                  |    |    |
| 7. Ринкові перспективи (конкуренція)                        | 2                     | 2  | 2  |
| 8. Практична здійсненність (наявність фахівців)             | 4                     | 4  | 4  |
| 9. Практична здійсненність (наявність фінансів)             | 2                     | 3  | 2  |
| 10. Практична здійсненність (необхідність нових матеріалів) | 3                     | 4  | 4  |
| 11. Практична здійсненність (термін реалізації)             | 3                     | 4  | 4  |
| 12. Практична здійсненність (розроблення документів)        | 4                     | 4  | 3  |
| Сума балів                                                  | 37                    | 40 | 38 |
| Середньоарифметична сума балів $CB_c$                       | 38,3                  |    |    |

За результатами розрахунків, наведених в таблиці 4.1, зробимо висновок щодо науково-технічного рівня і рівня комерційного потенціалу розробки. При цьому використаємо рекомендації, наведені в методичних вказівках до виконання економічної частини магістерських кваліфікаційних робіт [25].

Згідно проведених досліджень рівень комерційного потенціалу розробки за темою «Інформаційна технологія аналізу та передбачення діоксиду вуглецю за даними громадського моніторингу атмосферного повітря у м. Вінниці» становить 38,3 бала, що, відповідно до методичних вказівок до виконання економічної частини магістерських кваліфікаційних робіт [25], свідчить про комерційну важливість проведення даних досліджень (рівень комерційного потенціалу розробки вище середнього).

## 4.2 Розрахунок узагальненого коефіцієнта якості розробки

Узагальнений коефіцієнт якості ( $B_n$ ) для нового технічного рішення розрахуємо за формулою [26]:



$$B_n = \sum_{i=1}^k \alpha_i \cdot \beta_i, \quad (4.1)$$

де  $k$  – кількість найбільш важливих технічних показників, які впливають на якість нового технічного рішення;

$\alpha_i$  – коефіцієнт, який враховує питому вагу  $i$ -го технічного показника в загальній якості розробки. Коефіцієнт  $\alpha_i$  визначається експертним шляхом і при

цьому має виконуватись умова  $\sum_{i=1}^k \alpha_i = 1$ ;

$\beta_i$  – відносне значення  $i$ -го технічного показника якості нової розробки.

Результати порівняння зведемо до таблиці 4.2.

Таблиця 4.2 – Порівняння основних параметрів розробки та аналога.

| Показники<br>(параметри)             | Одиниця<br>вимірю-<br>вання | Аналог | Проектований<br>продукт | Відношення<br>параметрів<br>нової<br>розробки до<br>аналога | Питома вага<br>показника |
|--------------------------------------|-----------------------------|--------|-------------------------|-------------------------------------------------------------|--------------------------|
| Швидкість<br>обчислення              | с                           | 600    | 300                     | 2                                                           | 0,3                      |
| Похибка<br>обчислення                | %                           | 1,92   | 3                       | 0,64                                                        | 0,2                      |
| Кількість<br>графіків                | шт                          | 10     | 20                      | 2                                                           | 0,15                     |
| Споживання<br>оперативної<br>пам'яті | гб                          | 5      | 2                       | 2,5                                                         | 0,25                     |
| Кількість<br>моделей<br>передбачення | шт                          | 3      | 8                       | 2,67                                                        | 0,1                      |

Узагальнений коефіцієнт якості ( $B_n$ ) для нового технічного рішення складе:

$$B_n = \sum_{i=1}^k \alpha_i \cdot \beta_i = 2 \cdot 0,3 + 0,64 \cdot 0,2 + 2 \cdot 0,15 + 2,5 \cdot 0,25 + 2,67 \cdot 0,1 = 1,92.$$

Отже за технічними параметрами, згідно узагальненого коефіцієнту якості розробки, науково-технічна розробка переважає існуючі аналоги приблизно в 1,92 рази.

### 4.3 Розрахунок витрат на проведення науково-дослідної роботи

Витрати, пов'язані з проведенням науково-дослідної роботи на тему «Інформаційна технологія аналізу та передбачення діоксиду вуглецю за даними громадського моніторингу атмосферного повітря у м. Вінниці», під час планування, обліку і калькулювання собівартості науково-дослідної роботи групуємо за відповідними статтями.

#### 4.3.1 Витрати на оплату праці

Основна заробітна плата дослідників

Витрати на основну заробітну плату дослідників ( $Z_o$ ) розраховуємо у відповідності до посадових окладів працівників, за формулою [25]:

$$Z_o = \sum_{i=1}^k \frac{M_{ni} \cdot t_i}{T_p}, \quad (4.2)$$

де  $k$  – кількість посад дослідників залучених до процесу досліджень;

$M_{ni}$  – місячний посадовий оклад конкретного дослідника, грн;

$t_i$  – число днів роботи конкретного дослідника, дн.;

$T_p$  – середнє число робочих днів в місяці,  $T_p=21$  дні.

$$Z_o = 18300,00 \cdot 32 / 21 = 27885,71 \text{ грн.}$$

Проведені розрахунки зведемо до таблиці 4.3.

Таблиця 4.3 – Витрати на заробітну плату дослідників

| Найменування посади                    | Місячний посадовий оклад, грн | Оплата за робочий день, грн | Число днів роботи | Витрати на заробітну плату, грн |
|----------------------------------------|-------------------------------|-----------------------------|-------------------|---------------------------------|
| Керівник науково-дослідної роботи      | 18300,00                      | 871,43                      | 32                | 27885,71                        |
| Інженер-розробник інформаційних систем | 17500,00                      | 833,33                      | 28                | 23333,33                        |
| Консультант (експерт-еколог)           | 18000,00                      | 857,14                      | 5                 | 4285,71                         |
| Фахівець 1-ї категорії                 | 8500,00                       | 404,76                      | 21                | 8500,00                         |
| Всього                                 |                               |                             |                   | 64004,76                        |

#### Основна заробітна плата робітників

Витрати на основну заробітну плату робітників ( $Z_p$ ) за відповідними найменуваннями робіт НДР на тему «Інформаційна технологія аналізу та передбачення діоксиду вуглецю за даними громадського моніторингу атмосферного повітря у м. Вінниці» розраховуємо за формулою 4.3:

$$Z_p = \sum_{i=1}^n C_i \cdot t_i, \quad (4.3)$$

де  $C_i$  – погодинна тарифна ставка робітника відповідного розряду, за виконану відповідну роботу, грн/год;

$t_i$  – час роботи робітника при виконанні визначеної роботи, год.

Погодинну тарифну ставку робітника відповідного розряду  $C_i$  можна визначити за формулою 4.4:

$$C_i = \frac{M_M \cdot K_i \cdot K_c}{T_p \cdot t_{zm}}, \quad (4.4)$$

де  $M_M$  – розмір мінімальної місячної заробітної плати, прийmemo  $M_M=8000,00$  грн;

$K_i$  – коефіцієнт міжкваліфікаційного співвідношення (табл. Б.2, додаток Б) [26];

$K_c$  – мінімальний коефіцієнт співвідношень місячних тарифних ставок;

$T_p$  – середнє число робочих днів в місяці, приблизно  $T_p = 21$  дн;

$t_{зм}$  – тривалість зміни, год.

$$C_l = 8000,00 \cdot 1,10 \cdot 1,15 / (21 \cdot 8) = 60,24 \text{ грн.}$$

$$Z_{pl} = 60,24 \cdot 7,20 = 433,71 \text{ грн.}$$

Таблиця 4.4 – Величина витрат на основну заробітну плату робітників

| Найменування робіт                                                    | Тривалість роботи, год | Розряд роботи | Тарифний коефіцієнт | Погодинна тарифна ставка, грн | Величина оплати на робітника грн |
|-----------------------------------------------------------------------|------------------------|---------------|---------------------|-------------------------------|----------------------------------|
| Підготовка робочого місця розробника інформаційної системи            | 7,20                   | 2             | 1,10                | 60,24                         | 433,71                           |
| Інсталяція програмного забезпечення середовища розробки і моделювання | 7,30                   | 3             | 1,35                | 73,93                         | 539,68                           |
| Компіляція програмних блоків моделювання інформаційної системи        | 5,80                   | 4             | 1,50                | 82,14                         | 476,43                           |
| Підготовка локального серверного обладнання для проведення досліджень | 6,00                   | 5             | 1,70                | 93,10                         | 558,57                           |
| Налагодження програмних блоків збору інформації                       | 6,00                   | 5             | 1,70                | 93,10                         | 558,57                           |
| Тестування системи                                                    | 8,00                   | 2             | 1,10                | 60,24                         | 481,90                           |
| Всього                                                                |                        |               |                     |                               | 3048,87                          |

Додаткова заробітна плата дослідників та робітників

Додаткову заробітну плату розраховуємо як 10 ... 12% від суми основної заробітної плати дослідників та робітників за формулою 4.5:

$$Z_{\text{доп}} = (Z_o + Z_p) \cdot \frac{H_{\text{доп}}}{100\%}, \quad (4.5)$$

де  $H_{\text{доп}}$  – норма нарахування додаткової заробітної плати. Прийmemo 10%.

$$Z_{\text{доп}} = (64004,76 + 3048,87) \cdot 10 / 100\% = 6705,36 \text{ грн}$$

.

#### 4.3.2 Відрахування на соціальні заходи

Нарахування на заробітну плату дослідників та робітників розраховуємо як 22% від суми основної та додаткової заробітної плати дослідників і робітників за формулою 4.6:

$$Z_n = (Z_o + Z_p + Z_{\text{доп}}) \cdot \frac{H_n}{100\%} \quad (4.6)$$

де  $H_n$  – норма нарахування на заробітну плату. Приймаємо 22%.

$$Z_n = (64004,76 + 3048,87 + 6705,36) \cdot 22 / 100\% = 16226,98 \text{ грн.}$$

#### 4.3.3 Сировина та матеріали

Витрати на матеріали ( $M$ ), у вартісному вираженні розраховуються окремо по кожному виду матеріалів за формулою 4.7:

$$M = \sum_{j=1}^n H_j \cdot C_j \cdot K_j - \sum_{j=1}^n B_j \cdot C_{\text{в}j}, \quad (4.7)$$

де  $H_j$  – норма витрат матеріалу  $j$ -го найменування, кг;

$n$  – кількість видів матеріалів;

$C_j$  – вартість матеріалу  $j$ -го найменування, грн/кг;

$K_j$  – коефіцієнт транспортних витрат, ( $K_j = 1,1 \dots 1,15$ );

$B_j$  – маса відходів  $j$ -го найменування, кг;

$C_{\text{в}j}$  – вартість відходів  $j$ -го найменування, грн/кг.

$$M_1 = 4,0 \cdot 205,00 \cdot 1,1 - 0 \cdot 0 = 902,00 \text{ грн.}$$

Проведені розрахунки зведемо до таблиці 4.5.

Таблиця 4.5 – Витрати на матеріали

| Найменування матеріалу, марка, тип, сорт          | Ціна за 1 кг, грн | Норма витрат, кг | Величина відходів, кг | Ціна відходів, грн/кг | Вартість витраченого матеріалу, грн |
|---------------------------------------------------|-------------------|------------------|-----------------------|-----------------------|-------------------------------------|
| Офісний папір SOFT Ultra Plus                     | 205,00            | 4,0              | 0                     | 0                     | 902,00                              |
| Папір для записів Parers SOFT Light A5            | 130,00            | 5,0              | 0                     | 0                     | 715,00                              |
| Органайзер офісний OFFICE SOFT                    | 185,00            | 2,0              | 0                     | 0                     | 407,00                              |
| Канцелярське приладдя (набір офісного працівника) | 126,00            | 3,0              | 0                     | 0                     | 415,80                              |
| Картридж для принтера Erixon EZ2500               | 720,00            | 1,0              | 0                     | 0                     | 792,00                              |
| Диск оптичний NewOice CD-RW                       | 27,00             | 3,0              | 0                     | 0                     | 89,10                               |
| Flesh-пам'ять Kingston 64 GB                      | 269,00            | 1,0              | 0                     | 0                     | 295,90                              |
| Тека для паперів BOSS BOX                         | 65,00             | 5,0              | 0                     | 0                     | 357,50                              |
| Інші матеріали                                    | 195,00            | 1,0              | 0                     | 0                     | 214,50                              |
| Всього                                            |                   |                  |                       |                       | 4188,80                             |

#### 4.3.4 Розрахунок витрат на комплектуючі

Витрати на комплектуючі ( $K_e$ ), які використовують при проведенні НДР на тему «Інформаційна технологія аналізу та передбачення діоксиду вуглецю за даними громадського моніторингу атмосферного повітря у м. Вінниці», розраховуємо, згідно з їхньою номенклатурою, за формулою 4.8:

$$K_e = \sum_{j=1}^n H_j \cdot C_j \cdot K_j \quad (4.8)$$

де  $H_j$  – кількість комплектуючих  $j$ -го виду, шт.;

$C_j$  – покупна ціна комплектуючих  $j$ -го виду, грн;

$K_j$  – коефіцієнт транспортних витрат, ( $K_j = 1,1 \dots 1,15$ ).

$$K_e = 1 \cdot 3820,00 \cdot 1,1 = 4202,00 \text{ грн.}$$

Проведені розрахунки зведемо до таблиці 4.6.

Таблиця 4.6 – Витрати на комплектуючі

| Найменування комплектуючих                                                                   | Кількість, шт. | Ціна за штуку, грн | Сума, грн |
|----------------------------------------------------------------------------------------------|----------------|--------------------|-----------|
| RouterBOARD TP-Link230                                                                       | 1              | 3820,00            | 4202,00   |
| Диск жорсткий зовнішній<br>Transcend StoreJet 1TB USB<br>3.1 Iron Gray Slim<br>(TS1TSJ25M3S) | 1              | 2699,00            | 2968,90   |
| Всього                                                                                       |                |                    | 7170,90   |

#### 4.3.5 Спецустаткування для наукових (експериментальних) робіт

Балансову вартість спецустаткування розраховуємо за формулою 4.9:

$$B_{\text{спец}} = \sum_{i=1}^k C_i \cdot C_{\text{нр.і}} \cdot K_i, \quad (4.9)$$

де  $C_i$  – ціна придбання одиниці спецустаткування даного виду, марки, грн;

$C_{np.i}$  – кількість одиниць устаткування відповідного найменування, які придбані для проведення досліджень, шт.;

$K_i$  – коефіцієнт, що враховує доставку, монтаж, налагодження устаткування тощо, ( $K_i = 1,10 \dots 1,12$ );

$k$  – кількість найменувань устаткування.

$$B_{спец} = 10120,00 \cdot 1 \cdot 1,1 = 11132,00 \text{ грн.}$$

Отримані результати зведемо до таблиці 4.7:

Таблиця 4.7 – Витрати на придбання спецустаткування по кожному виду

| Найменування устаткування                                                                                               | Кількість, шт | Ціна за<br>одиницю, грн | Вартість, грн |
|-------------------------------------------------------------------------------------------------------------------------|---------------|-------------------------|---------------|
| Маршрутизатор VPN TP-Link ER7212PC                                                                                      | 1             | 10120,00                | 11132,00      |
| Серверне обладнання (локальне) на базі QUBE QB Ryzen 5 7600X RTX 4070 Ti 12GB W1 3210 (Ryzen57600XRTX4070Ti12GB W13210) | 1             | 65999,00                | 72598,90      |
| Імітатор станції досліджень моніторингу діоксиду вуглецю                                                                | 1             | 26999,00                | 29698,90      |
| Всього                                                                                                                  |               |                         | 113429,80     |

#### 4.3.6 Програмне забезпечення для наукових (експериментальних) робіт

Балансову вартість програмного забезпечення розраховуємо за формулою 4.10:

$$B_{прог} = \sum_{i=1}^k C_{прог} \cdot C_{npz.i} \cdot K_i, \quad (4.10)$$



де  $C_{inprg}$  – ціна придбання одиниці програмного засобу даного виду, грн;

$C_{prg.i}$  – кількість одиниць програмного забезпечення відповідного найменування, які придбані для проведення досліджень, шт.;

$K_i$  – коефіцієнт, що враховує інсталяцію, налагодження програмного засобу тощо, ( $K_i = 1, 10 \dots 1, 12$ );

$k$  – кількість найменувань програмних засобів.

$$B_{prg} = 7499,00 \cdot 1 \cdot 1,1 = 8248,90 \text{ грн.}$$

Отримані результати зведемо до таблиці 4.8:

Таблиця 4.8 – Витрати на придбання програмних засобів по кожному виду

| Найменування програмного засобу                         | Кількість, шт | Ціна за одиницю, грн | Вартість, грн |
|---------------------------------------------------------|---------------|----------------------|---------------|
| Мова програмування Python та її бібліотеки,             | 1             | 7499,00              | 8248,90       |
| Панель керування cPanel з підтримкою PHP,               | 1             | 6820,00              | 7502,00       |
| Система керування реляційними базами даних MySQL        | 1             | 8459,00              | 9304,90       |
| Доступ до мережі Internet (високошвидкісний) грн/місяць | 2             | 350,00               | 770,00        |
| Всього                                                  |               |                      | 25825,80      |

#### 4.3.7 Амортизація обладнання, програмних засобів та приміщень

В спрощеному вигляді амортизаційні відрахування по кожному виду обладнання, приміщень та програмному забезпеченню тощо, розраховуємо з використанням прямолінійного методу амортизації за формулою 4.11:

$$A_{obl} = \frac{C_{\text{б}}}{T_{\text{г}}} \cdot \frac{t_{\text{вик}}}{12}, \quad (4.11)$$

де  $C_0$  – балансова вартість обладнання, програмних засобів, приміщень тощо, які використовувались для проведення досліджень, грн;

$t_{вик}$  – термін використання обладнання, програмних засобів, приміщень під час досліджень, місяців;

$T_с$  – строк корисного використання обладнання, програмних засобів, приміщень тощо, років.

$$A_{обл} = (25400,00 \cdot 1) / (3 \cdot 12) = 705,56 \text{ грн.}$$

Проведені розрахунки зведемо до таблиці 4.9.

Таблиця 4.9 – Амортизаційні відрахування по кожному виду обладнання

| Найменування обладнання                                                 | Балансова вартість, грн | Строк корисного використання, років | Термін використання обладнання, місяців | Амортизаційні відрахування, грн |
|-------------------------------------------------------------------------|-------------------------|-------------------------------------|-----------------------------------------|---------------------------------|
| Обчислювальний комплекс та комп'ютеризована система проектування        | 25400,00                | 3                                   | 1                                       | 705,56                          |
| Обчислювальний комплекс Master-133                                      | 11200,00                | 4                                   | 1                                       | 233,33                          |
| Обчислювальний комплекс Master-800                                      | 15600,00                | 5                                   | 1                                       | 260,00                          |
| Блок інтерфейсів ОКР-5                                                  | 12400,00                | 5                                   | 1                                       | 206,67                          |
| Вимірювальний комплекс метрологічної системи оцінки газового середовища | 19800,00                | 4                                   | 1                                       | 412,50                          |
| Монтажно-модульне обладнання                                            | 13400,00                | 5                                   | 1                                       | 223,33                          |
| Дослідницька лабораторія                                                | 340000,00               | 25                                  | 1                                       | 1133,33                         |
| Прикладний пакет Microsoft Office                                       | 6700,00                 | 3                                   | 1                                       | 186,11                          |
| ОС Windows 11                                                           | 7400,00                 | 3                                   | 1                                       | 205,56                          |
| Всього                                                                  |                         |                                     |                                         | 3566,39                         |

### 4.3.8 Паливо та енергія для науково-виробничих цілей

Витрати на силову електроенергію ( $B_e$ ) розраховуємо за формулою 4.12:

$$B_e = \sum_{i=1}^n \frac{W_{yi} \cdot t_i \cdot C_e \cdot K_{eni}}{\eta_i}, \quad (4.12)$$

де  $W_{yi}$  – встановлена потужність обладнання на визначеному етапі розробки, кВт;

$t_i$  – тривалість роботи обладнання на етапі дослідження, год;

$C_e$  – вартість 1 кВт-години електроенергії, грн; прийmemo  $C_e = 10,98$  грн;

$K_{eni}$  – коефіцієнт, що враховує використання потужності,  $K_{eni} < 1$ ;

$\eta_i$  – коефіцієнт корисної дії обладнання,  $\eta_i < 1$ .

$$B_e = 0,28 \cdot 240,0 \cdot 10,98 \cdot 0,95 / 0,97 = 737,86 \text{ грн.}$$

Проведені розрахунки зведемо до таблиці 4.10.

Таблиця 4.10 – Витрати на електроенергію

| Найменування обладнання                                                                                                 | Встановлена потужність, кВт | Тривалість роботи, год | Сума, грн |
|-------------------------------------------------------------------------------------------------------------------------|-----------------------------|------------------------|-----------|
| Обчислювальний комплекс та комп'ютеризована система проектування                                                        | 0,28                        | 240,0                  | 737,86    |
| Обчислювальний комплекс Master-133                                                                                      | 0,32                        | 200,0                  | 702,72    |
| Обчислювальний комплекс Master-800                                                                                      | 0,32                        | 200,0                  | 702,72    |
| Блок інтерфейсів ОКР-5                                                                                                  | 0,02                        | 180,0                  | 39,53     |
| Вимірювальний комплекс метрологічної системи оцінки газового середовища                                                 | 0,46                        | 180,0                  | 909,14    |
| Монтажно-модульне обладнання                                                                                            | 0,10                        | 15,0                   | 16,47     |
| Маршрутизатор VPN TP-Link ER7212PC                                                                                      | 0,03                        | 240,0                  | 79,06     |
| Серверне обладнання (локальне) на базі QUBE QB Ryzen 5 7600X RTX 4070 Ti 12GB W1 3210 (Ryzen57600XRTX4070Ti12GB W13210) | 0,34                        | 240,0                  | 895,97    |
| Імітатор станції досліджень моніторингу діоксиду вуглецю                                                                | 0,12                        | 180,0                  | 237,17    |
| Всього                                                                                                                  |                             |                        | 4320,63   |

### 4.3.9 Службові відрядження

Витрати за статтею «Службові відрядження» розраховуємо як 20...25% від суми основної заробітної плати дослідників та робітників за формулою 4.13:

$$B_{cv} = (Z_o + Z_p) \cdot \frac{H_{cv}}{100\%}, \quad (4.13)$$

де  $H_{cv}$  – норма нарахування за статтею «Службові відрядження», прийmemo  $H_{cv} = 20\%$ .

$$B_{cv} = (64004,76 + 3048,87) \cdot 20 / 100\% = 13410,73 \text{ грн.}$$

### 4.3.10 Витрати на роботи, які виконують сторонні підприємства, установи і організації

Витрати розраховуємо як 30...45% від суми основної заробітної плати дослідників та робітників за формулою 4.14:

$$B_{cn} = (Z_o + Z_p) \cdot \frac{H_{cn}}{100\%}, \quad (4.14)$$

де  $H_{cn}$  – норма нарахування за статтею «Витрати на роботи, які виконують сторонні підприємства, установи і організації», прийmemo  $H_{cn} = 30\%$ .

$$B_{cn} = (64004,76 + 3048,87) \cdot 30 / 100\% = 20116,09 \text{ грн.}$$

### 4.3.11 Інші витрати

Витрати за статтею «Інші витрати» розраховуємо як 50...100% від суми основної заробітної плати дослідників та робітників за формулою 4.15:

$$I_{\varepsilon} = (Z_o + Z_p) \cdot \frac{H_{i\varepsilon}}{100\%}, \quad (4.15)$$

де  $H_{i\varepsilon}$  – норма нарахування за статтею «Інші витрати», прийmemo  $H_{i\varepsilon} = 50\%$ .

$$I_{\varepsilon} = (64004,76 + 3048,87) \cdot 50 / 100\% = 33526,82 \text{ грн.}$$

#### 4.3.12 Накладні (загальновиробничі) витрати

Витрати за статтею «Накладні (загальновиробничі) витрати» розраховуємо як 100...150% від суми основної заробітної плати дослідників та робітників за формулою 4.16:

$$B_{нзв} = (Z_o + Z_p) \cdot \frac{H_{нзв}}{100\%}, \quad (4.16)$$

де  $H_{нзв}$  – норма нарахування за статтею «Накладні (загальновиробничі) витрати», прийmemo  $H_{нзв} = 100\%$ .

$$B_{нзв} = (64004,76 + 3048,87) \cdot 100 / 100\% = 67053,63 \text{ грн.}$$

Витрати на проведення науково-дослідної роботи на тему «Інформаційна технологія аналізу та передбачення діоксиду вуглецю за даними громадського моніторингу атмосферного повітря у м. Вінниці» розраховуємо як суму всіх попередніх статей витрат за формулою:

$$B_{заг} = Z_o + Z_p + Z_{доо} + Z_n + M + K_e + B_{спец} + B_{прз} + A_{обл} + B_e + B_{св} + B_{сп} + I_e + B_{нзв}. \quad (4.17)$$

$$B_{заг} = 64004,76 + 3048,87 + 6705,36 + 16226,98 + 4188,80 + 7170,90 + 113429,80 + 25825,80 + 3566,39 + 4320,63 + 13410,73 + 20116,09 + 33526,82 + 67053,63 = 382595,55 \text{ грн.}$$

Загальні витрати  $ЗВ$  на завершення науково-дослідної (науково-технічної) роботи та оформлення її результатів розраховується за формулою 4.18:

$$ЗВ = \frac{B_{заг}}{\eta}, \quad (4.18)$$

де  $\eta$  - коефіцієнт, який характеризує етап (стадію) виконання науково-дослідної роботи, прийmemo  $\eta = 0,9$ .

$$ЗВ = 382595,55 / 0,9 = 425106,17 \text{ грн.}$$

#### 4.4 Розрахунок економічної ефективності науково-технічної розробки при її можливій комерціалізації потенційним інвестором

Результати дослідження проведені за темою «Інформаційна технологія аналізу та передбачення діоксиду вуглецю за даними громадського моніторингу атмосферного повітря у м. Вінниці» передбачають комерціалізацію протягом 4-х років реалізації на ринку.

За комерційним напрямом дослідження відносяться до ситуації: 1 – «Розроблення інформаційної системи (web-сайт, консолідований ресурс тощо) на основі нових алгоритмів, програмних або технічних засобів» [25]. В цьому випадку основу майбутнього економічного ефекту будуть формувати:

$\Delta N$  – збільшення кількості споживачів яким надається відповідна інформаційна послуга у періоди часу, що аналізуються;

| Показник                              | 1-й рік | 2-й рік | 3-й рік | 4-й рік |
|---------------------------------------|---------|---------|---------|---------|
| Збільшення кількості споживачів, осіб | 1100    | 1200    | 1800    | 1500    |

$N$  – кількість споживачів яким надавалась відповідна інформаційна послуга у році до впровадження результатів нової науково-технічної розробки, прийmemo 210000 осіб;

$C_o$  – вартість послуги у році до впровадження інформаційної системи, прийmemo 50,00 грн;

$\pm \Delta C_o$  – зміна вартості послуги від впровадження результатів, прийmemo 16,00 грн.

Можливе збільшення чистого прибутку у потенційного інвестора  $\Delta \Pi_i$  для кожного із 4-х років, протягом яких очікується отримання позитивних результатів від можливого впровадження та комерціалізації науково-технічної розробки, розраховуємо за формулою [25]:

$$\Delta \Pi_i = (\pm \Delta C_o \cdot N + C_o \cdot \Delta N)_i \cdot \lambda \cdot \rho \cdot \left(1 - \frac{\rho}{100}\right), \quad (4.19)$$

де  $\lambda$  – коефіцієнт, який враховує сплату потенційним інвестором податку на додану вартість. У 2024 році ставка податку на додану вартість складає 20%, а коефіцієнт  $\lambda = 0,8333$ ;

$\rho$  – коефіцієнт, який враховує рентабельність інноваційного продукту).  
Прийmemo  $\rho = 42\%$ ;

$\vartheta$  – ставка податку на прибуток, який має сплачувати потенційний інвестор, у 2024 році  $\vartheta = 18\%$ ;

Збільшення чистого прибутку 1-го року:

$$\Delta\Pi_1 = (16,00 \cdot 210000,00 + 66,00 \cdot 1100) \cdot 0,83 \cdot 0,42 \cdot (1 - 0,18/100\%) = 981215,58 \text{ грн.}$$

Збільшення чистого прибутку 2-го року:

$$\Delta\Pi_2 = (16,00 \cdot 210000,00 + 66,00 \cdot 2300) \cdot 0,83 \cdot 0,42 \cdot (1 - 0,18/100\%) = 1003855,05 \text{ грн.}$$

Збільшення чистого прибутку 3-го року:

$$\Delta\Pi_3 = (16,00 \cdot 210000,00 + 66,00 \cdot 4100) \cdot 0,83 \cdot 0,42 \cdot (1 - 0,18/100\%) = 1037814,27 \text{ грн.}$$

Збільшення чистого прибутку 4-го року:

$$\Delta\Pi_4 = (16,00 \cdot 210000,00 + 66,00 \cdot 5600) \cdot 0,83 \cdot 0,42 \cdot (1 - 0,18/100\%) = 1066113,62 \text{ грн.}$$

Приведена вартість збільшення всіх чистих прибутків  $\Pi\Pi$ , що їх може отримати потенційний інвестор від можливого впровадження та комерціалізації науково-технічної розробки:

$$\Pi\Pi = \sum_{i=1}^T \frac{\Delta\Pi_i}{(1 + \tau)^i}, \quad (4.20)$$

де  $\Delta\Pi_i$  – збільшення чистого прибутку у кожному з років, протягом яких виявляються результати впровадження науково-технічної розробки, грн;

$T$  – період часу, протягом якого очікується отримання позитивних результатів від впровадження та комерціалізації науково-технічної розробки, роки;

$\tau$  – ставка дисконтування, за яку можна взяти щорічний прогнозований рівень інфляції в країні,  $\tau = 0,15$ ;

$t$  – період часу (в роках) від моменту початку впровадження науково-технічної розробки до моменту отримання потенційним інвестором додаткових чистих прибутків у цьому році.

$$\begin{aligned} III = & 981215,58/(1+0,15)^1 + 1003855,05/(1+0,15)^2 + 1037814,27/(1+0,15)^3 + \\ & + 1066113,62/(1+0,15)^4 = 853230,93 + 759058,64 + 682379,73 + 609553,92 = 2904223,23 \\ & \text{грн.} \end{aligned}$$

Величина початкових інвестицій  $PV$ , які потенційний інвестор має вкласти для впровадження і комерціалізації науково-технічної розробки:

$$PV = k_{инв} \cdot 3B, \quad (4.21)$$

де  $k_{инв}$  – коефіцієнт, що враховує витрати інвестора на впровадження науково-технічної розробки та її комерціалізацію, приймаємо  $k_{инв} = 2$ ;

$3B$  – загальні витрати на проведення науково-технічної розробки та оформлення її результатів, приймаємо 425106,17 грн.

$$PV = k_{инв} \cdot 3B = 2 \cdot 425106,17 = 850212,34 \text{ грн.}$$

Абсолютний економічний ефект  $E_{абс}$  для потенційного інвестора від можливого впровадження та комерціалізації науково-технічної розробки становитиме:

$$E_{абс} = III - PV \quad (4.22)$$

де  $III$  – приведена вартість зростання всіх чистих прибутків від можливого впровадження та комерціалізації науково-технічної розробки, 2904223,23 грн;

$PV$  – теперішня вартість початкових інвестицій, 850212,34 грн.

$$E_{абс} = III - PV = 2904223,23 - 850212,34 = 2054010,89 \text{ грн.}$$

Внутрішня економічна дохідність інвестицій  $E_g$ , які можуть бути вкладені потенційним інвестором у впровадження та комерціалізацію науково-технічної розробки:



$$E_{\epsilon} = \sqrt[T_{жс}]{1 + \frac{E_{абс}}{PV}} - 1, \quad (4.23)$$

де  $E_{абс}$  – абсолютний економічний ефект вкладених інвестицій, 2054010,89 грн;

$PV$  – теперішня вартість початкових інвестицій, 850212,34 грн;

$T_{жс}$  – життєвий цикл науково-технічної розробки, тобто час від початку її розробки до закінчення отримування позитивних результатів від її впровадження, 4 роки.

$$E_{\epsilon} = \sqrt[T_{жс}]{1 + \frac{E_{абс}}{PV}} - 1 = (1 + 2054010,89/850212,34)^{1/4} = 0,36.$$

Мінімальна внутрішня економічна дохідність вкладених інвестицій  $\tau_{мін}$ :

$$\tau_{мін} = d + f, \quad (4.24)$$

де  $d$  – середньозважена ставка за депозитними операціями в комерційних банках; в 2024 році в Україні  $d = 0,09$ ;

$f$  – показник, що характеризує ризикованість вкладення інвестицій, приймемо 0,15.

$\tau_{мін} = 0,09 + 0,15 = 0,24 < 0,36$  свідчить про те, що внутрішня економічна дохідність інвестицій  $E_{\epsilon}$ , вища мінімальної внутрішньої дохідності. Тобто інвестувати в науково-дослідну роботу за темою «Інформаційна технологія аналізу та передбачення діоксиду вуглецю за даними громадського моніторингу атмосферного повітря у м. Вінниці» доцільно.

Період окупності інвестицій  $T_{ок}$  які можуть бути вкладені потенційним інвестором у впровадження та комерціалізацію науково-технічної розробки:

$$T_{ок} = \frac{1}{E_{\epsilon}}, \quad (4.25)$$

де  $E_e$  – внутрішня економічна дохідність вкладених інвестицій.

$$T_{ок} = 1 / 0,36 = 2,78 \text{ р.}$$

$T_{ок} < 3$ -х років, що свідчить про комерційну привабливість науково-технічної розробки і може спонукати потенційного інвестора профінансувати впровадження даної розробки та виведення її на ринок.

#### 4.5 Висновки

Згідно проведених досліджень рівень комерційного потенціалу розробки за темою «Інформаційна технологія аналізу та передбачення діоксиду вуглецю за даними громадського моніторингу атмосферного повітря у м. Вінниці» становить 38,3 бала, що, свідчить про комерційну важливість проведення даних досліджень (рівень комерційного потенціалу розробки вище середнього).

При оцінюванні за технічними параметрами, згідно узагальненого коефіцієнту якості розробки, науково-технічна розробка переважає існуючі аналоги приблизно в 1,92 рази.

Також термін окупності становить 2,78 р., що менше 3-х років, що свідчить про комерційну привабливість науково-технічної розробки і може спонукати потенційного інвестора профінансувати впровадження даної розробки та виведення її на ринок.

Отже можна зробити висновок про доцільність проведення науково-дослідної роботи за темою «Інформаційна технологія аналізу та передбачення діоксиду вуглецю за даними громадського моніторингу атмосферного повітря у м. Вінниці».

## ВИСНОВКИ

Магістерська кваліфікаційна робота присвячена розробці інформаційної технології технологія аналізу та передбачення діоксиду вуглецю за даними громадського моніторингу атмосферного повітря у м. Вінниці.

У першому розділі охарактеризовано причини та наслідки забруднення та зростання діоксиду вуглецю в атмосферному повітрі. Розглянуто засоби та методи моніторингу атмосферного повітря і боротьба з наслідками забруднення в м. Вінниця. Здійснено вибір технологій та методів для розв'язування поставленої задачі. Також було описано датасет і охарактеризовано моделі машинного навчання для вирішення поставленої задачі.

У другому розділі проведено аналіз платформи вирішення поставленої задачі Kaggle. Здійснено розвідувальний аналіз даних, за допомогою якого було очищено датасет від різних аномалій, також було охарактеризовано частину аномалій та причини їх виникнення. Створено та налаштовано різні типи моделей машинного навчання, найкращими за метрикою MAPE виявились Linear SVR – 1, 1,93953%, Random Forest Regressor – 1,991942%, XGB Regressor – 2,025112%, ARIMA\_auto – 2,039046%, метод k-найближчих сусідів – 2,18653%. Найкращою моделлю було вибрано Linear SVR за найменшою відносною похибкою.

У третьому розділі було розроблено інформаційну технологію передбачення вмісту вуглекислого газу в атмосфері, побудовано блок-схему алгоритму вирішення заданої задачі. Проведено тренування моделі Linear SVR з відносною похибкою в 1,927118% на розширеному датасеті, що включав в себе тренувальні та валідаційні дані. Порівняно результати передбачення з реальними даними та зображення порівняння за допомогою графіку, що свідчить про гарні результати в передбаченні вмісту вуглекислого газу.

У четвертому розділі проведено оцінку комерційного потенціалу інформаційної технології аналізу та передбачення діоксиду вуглецю за даними громадського моніторингу атмосферного повітря у м. Вінниці. Прогнозування

витрат на виконання науково-дослідної роботи по кожній з статей витрат складе 425106,17 грн

Вкладені інвестиції в даний проект окупляться через 2,7 років, приведена вартість всіх чистих прибутків, що їх отримає підприємство від реалізації результатів наукової розробки склала 2054010,89 грн.

В результаті виконаної магістерської роботи розроблено інформаційну технологію передбачення вмісту діоксиду вуглецю в атмосферному повітрі регіону за рахунок оптимізації інтелектуальних моделей методами машинного навчання, що дозволяє підвищити точність передбачення його концентрації в атмосферному повітрі м. Вінниці.

## СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Губар О.М., Мокін В. Б., Розроблення інформаційної технології аналізу та передбачення діоксиду вуглецю за даними громадського моніторингу атмосферного повітря у м. Вінниці. Всеукраїнська науково-практична інтернет-конференція «Молодь в науці: дослідження, проблеми, перспективи» (Вінниця, 2024-2025 рр.). URL: <https://conferences.vntu.edu.ua/index.php/>. (Дата звернення: 10.11.2024).
2. Joseph Black. URL: [https://en.wikipedia.org/wiki/Joseph\\_Black](https://en.wikipedia.org/wiki/Joseph_Black) (дата звернення: 18.10.2024).
3. Donovan A. L. Antoine Lavoisier - Oxygen, Combustion, Chemistry | Britannica. *Encyclopedia Britannica*. URL: <https://www.britannica.com/biography/Antoine-Lavoisier/Oxygen-theory-of-combustion> (дата звернення: 19.10.2024).
4. Hot topic: Discovering the greenhouse effect | Science Museum. *Science Museum*. URL: <https://www.sciencemuseum.org.uk/objects-and-stories/our-environment/hot-topic-discovering-greenhouse-effect> (дата звернення: 20.10.2024).
5. Future Calculations. *Science History Institute*. URL: <https://www.sciencehistory.org/stories/magazine/future-calculations/> (дата звернення: 21.10.2024).
6. Guy Stewart Callendar - a scientist ahead of his time. *Weather and Climate*. URL: <https://blogs.reading.ac.uk/weather-and-climate-at-reading/2014/guy-stewart-callendar-1898-1964-a-scientist-ahead-of-his-time/> (дата звернення: 21.10.2024).
7. Monroe R. The History of the Keeling Curve. URL: <https://keelingcurve.ucsd.edu/2013/04/03/the-history-of-the-keeling-curve/> (дата звернення: 22.10.2024).
8. Frontiers | Next-Generation Digital Ecosystem for Climate Data Mining and Knowledge Discovery: A Review of Digital Data Collection Technologies / A. Hsu et al. *Frontiers*. URL: <https://www.frontiersin.org/journals/big-data/articles/10.3389/fdata.2020.00029/full> (дата звернення: 23.10.2024).

9. NASA. Why Earth Is Getting Warmer: The Greenhouse Effect Explained. *SkyTechDaily*. URL: <https://scitechdaily.com/why-earth-is-getting-warmer-the-greenhouse-effect-explained/> (дата звернення: 24.10.2024).

10. Charting the Exceptional, Unexpected Heat of 2023 and 2024. *NASA Earth Observatory* - *Home*. URL: <https://earthobservatory.nasa.gov/images/153588/charting-the-exceptional-unexpected-heat-of-2023-and-2024> (дата звернення: 24.10.2024).

11. Climate change threatens one in six species with extinction, study finds - Carbon Brief. *Carbon Brief*. URL: <https://www.carbonbrief.org/climate-change-threatens-one-in-six-species-with-extinction-study-finds/> (дата звернення: 25.10.2024).

12. Quantifying health and aviation hazards from Icelandic volcanic eruptions to inform government policy - Centre for the Observation and Modelling of Earthquakes, Volcanoes and Tectonics. *Centre for the Observation and Modelling of Earthquakes, Volcanoes and Tectonics*. URL: <https://comet.nerc.ac.uk/research-impact/quantifying-health-and-aviation-hazards-from-icelandic-volcanic-eruptions-to-inform-government-policy/> (дата звернення: 25.10.2024).

13. Saharan dust leaves air across southern Europe 'like custard powder'. *The Independent*. URL: <https://www.independent.co.uk/news/world/europe/sahara-dust-cloud-greece-italy-cyprus-malta-visibility-b2520357.html> (дата звернення: 26.10.2024).

14. Smog in Kyiv. *New Voice*. URL: <https://english.nv.ua/nation/smog-in-kyiv-rescuers-fight-fires-as-city-advises-residents-to-limit-outdoor-activity-50447958.html> (дата звернення: 27.10.2024).

15. NASA Scientific Visualization Studio | Methane Emissions from Wetlands. *NASA Scientific Visualization Studio*. URL: <https://svs.gsfc.nasa.gov/5054> (дата звернення: 28.10.2024).

16. Air Pollution from Agriculture. *Home - Defra, UK*. URL: [https://uk-air.defra.gov.uk/assets/documents/reports/aqeg/2800829\\_Agricultural\\_emissions\\_vfinal2.pdf](https://uk-air.defra.gov.uk/assets/documents/reports/aqeg/2800829_Agricultural_emissions_vfinal2.pdf) (дата звернення: 29.10.2024).

17. Sageguardian the ozone layer. *IPCC*.  
URL: [https://www.ipcc.ch/site/assets/uploads/2018/03/sroc\\_full-1.pdf](https://www.ipcc.ch/site/assets/uploads/2018/03/sroc_full-1.pdf) (дата звернення: 29.10.2024).
18. Програма державного моніторингу у галузі охорони атмосферного повітря агломерації «Вінниця» на 2021 - 2025 роки. *Вінницька міська рада*.  
URL: <https://www.vmr.gov.ua/media/Сервіси/Ecology/Програма%20остаточний%20варіант%2026.11.21.pdf> (дата звернення: 30.10.2024).
19. Звіт про виконання у 2022 році заходів. *Вінницька міська рада*.  
URL: <https://www.vmr.gov.ua/media/Сервіси/Ecology/Звіт%20за%202022%20рік.pdf> (дата звернення: 31.10.2024).
20. Air Quality Monitoring from EcoCity. *Kaggle*.  
URL: <https://www.kaggle.com/datasets/vbmokin/air-quality-monitoring-from-ecocity> (дата звернення: 02.11.2024).
21. Air Quality City - Mapping. *Kaggle*.  
URL: <https://www.kaggle.com/code/vbmokin/air-quality-city-mapping> (дата звернення: 03.11.2024).
22. Мокін В. Б., Дратований М. В. Наука про дані: машинне навчання та інтелектуальний аналіз даних. *VNTU*.  
URL: <https://iq.vntu.edu.ua/repository/getfile.php/8163.pdf> (дата звернення: 04.11.2024).
23. Forecasting and mitigation of global environmental carbon dioxide emission using machine learning techniques / Н. Brat et al.  
URL: <https://www.sciencedirect.com/science/article/pii/S2772782323000037> (дата звернення: 05.11.2024).
24. *Kaggle*. URL: <https://www.kaggle.com/> (дата звернення: 01.11.2024).
25. Козловський В. О., Лесько О. Й., Кавецький В. В. Методичні вказівки до виконання економічної частини магістерських кваліфікаційних робіт. Вінниця : ВНТУ, 2021. 42 с.
26. Кавецький В. В., Козловський В. О., Причепа І. В. Економічне обґрунтування інноваційних рішень: практикум. Вінниця : ВНТУ, 2016. 113 с.

Додаток А  
Технічне завдання  
Міністерство освіти і науки України  
Вінницький національний технічний університет  
Факультет інтелектуальних інформаційних технологій та автоматизації

ЗАТВЕРДЖУЮ

Завідувач кафедри САІТ

\_\_\_\_\_ д.т.н., проф. Віталій МОКІН

«\_\_» \_\_\_\_\_ 2024 р.

ТЕХНІЧНЕ ЗАВДАННЯ

на магістерську кваліфікаційну роботу

«Інформаційна технологія аналізу та передбачення діоксиду вуглецю за даними  
громадського моніторингу атмосферного повітря у м. Вінниці»

08-34.МКР.002.02.000.ТЗ

Керівник: д.т.н., проф. каф. САІТ

\_\_\_\_\_ Олександр МОКІН

«\_\_» \_\_\_\_\_ 2024 р.

Розробив: студент гр. ІІСТ-23м

\_\_\_\_\_ Олексій ГУБАР

«\_\_» \_\_\_\_\_ 2024 р.



## 1. Підстава для проведення робіт

Підставою для виконання роботи є наказ № \_\_ по ВНТУ від «\_\_» \_\_\_\_\_ 2024 р., та індивідуальне завдання на МКР, затверджене протоколом № \_\_ засідання кафедри САІТ від «\_\_» \_\_\_\_\_ 2024 р.

## 2. Джерела розробки:

– Наука про дані: машинне навчання та інтелектуальний аналіз даних : електронний навчальний посібник комбінованого (локального та мережевого) використання [Електронний ресурс] / В. Б. Мокін, М. В. Дратований – Вінниця : ВНТУ, 2024. – 258 с. – Режим доступу: <https://docs.vntu.edu.ua/card.php?id=8163>

– A Comprehensive Guide to Machine Learning [Електронний ресурс] / S. Nasiriany et al. 2018. 176 с. – Режим доступу: <https://github.com/juanmartinsantos/books/blob/main/docs/Machine%20Learning/A%20Comprehensive%20Guide%20to%20Machine%20Learning.pdf>

3. Мета і призначення роботи: є підвищення точності передбачення концентрації діоксиду вуглецю в атмосферному повітрі м. Вінниці шляхом розроблення інформаційної технології

## 4. Вихідні дані для проведення робіт:

– датасет Kaggle «Air Quality Monitoring from EcoCity» (<https://www.kaggle.com/datasets/vbmokin/air-quality-monitoring-from-ecocity>).

## 5. Методи дослідження:

1. Розвідувальний аналіз;
2. Методи машинного навчання.

## 6. Етапи роботи і терміни їх виконання:

- загальна характеристика об'єкту досліджень ..... – \_\_\_\_;
- побудова моделей та вибір оптимальної для розв'язання поставленої задачі ..... – \_\_\_\_;
- розроблення інформаційної технології передбачення вмісту діоксиду вуглецю у повітрі ..... – \_\_\_\_;
- економічна частина..... – \_\_\_\_;
- оформлення МКР..... – \_\_\_\_.

## 7. Очікувані результати та порядок реалізації:

Інформаційна технологія аналізу та передбачення діоксиду вуглецю за даними громадського моніторингу атмосферного повітря у м. Вінниці, результатом якої є підвищення точності передбачення за рахунок використання методів машинного навчання.

## 8. Вимоги до розробленої документації

Пояснювальна записка оформлена у відповідності до вимог «Методичних вказівок до виконання магістерських кваліфікаційних робіт для студентів спеціальності 126 «Інформаційні системи та технології» (освітня програма «Інформаційні технології аналізу даних та зображень»)

9. Порядок приймання роботи

Публічний захист ..... «\_\_» \_\_\_\_\_ 2024 р.  
Початок розробки..... «\_\_» \_\_\_\_\_ 2024 р.  
Граничні терміни виконання МКР ..... «\_\_» \_\_\_\_\_ 2024 р.  
Розробив студент групи ІСТ-23м \_\_\_\_\_ Олексій ГУБАР

## Додаток Б

Протокол перевірки кваліфікаційної роботи на наявність текстових запозичень

Назва роботи: «Інформаційна технологія аналізу та передбачення діоксиду вуглецю за даними громадського моніторингу атмосферного повітря у м. Вінниці»

Тип роботи: магістерська кваліфікаційна робота

Підрозділ: кафедра САІТ

Керівник: Олександр МОКІН, д.т.н., проф. каф. САІТ

## Показники звіту подібності Turnitin

Оригінальність 94 %

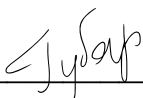
Загальна схожість 6 %

Аналіз звіту подібності (відмітити потрібне):

- Запозичення, виявлені у роботі, оформлені коректно і не містять ознак плагіату.
  - Виявлені у роботі запозичення не мають ознак плагіату, але їх надмірна кількість викликає сумніви щодо цінності роботи і відсутності самостійності її автора. Роботу направити на доопрацювання.
  - Виявлені у роботі запозичення є недобросовісними і мають ознаки плагіату та/або в ній містяться навмисні спотворення тексту, що вказують на спроби приховування недобросовісних запозичень.

Заявляю, що ознайомлений з повним звітом подібності, який був згенерований системою Turnitin щодо роботи.

Автор роботи

  
(підпис)

Олексій ГУБАР

Опис прийнятого рішення

Робота допускається до захисту

Особа, відповідальна за перевірку

  
(підпис)

Сергій ЖУКОВ

## Додаток В

## Лістинг програми

```
# Import libraries
import random
import os
import numpy as np
import pandas as pd
import requests
# Date
import datetime as dt
from datetime import date, timedelta, datetime
# EDA
import matplotlib.pyplot as plt
from matplotlib.pylab import rcParams
import plotly.express as px
import plotly.graph_objects as go
from plotly.offline import init_notebook_mode
init_notebook_mode(connected=True)
# FE
from tsfresh import extract_features, select_features, extract_relevant_features
from tsfresh.utilities.dataframe_functions import impute
from sklearn.inspection import permutation_importance
import eli5
from eli5.sklearn import PermutationImportance
import shap
# Time Series - EDA and Modelling
import statsmodels.api as sm
from statsmodels.graphics.tsaplots import plot_acf, plot_pacf
from statsmodels.tsa.stattools import adfuller
from statsmodels.tsa.seasonal import seasonal_decompose
from statsmodels.tsa.arima_model import ARIMA
# Metrics
from sklearn.metrics import r2_score
from sklearn.metrics import mean_squared_error, mean_absolute_percentage_error
# Modeling and preprocessing
from sklearn.preprocessing import StandardScaler, MinMaxScaler
from sklearn.model_selection import train_test_split, GridSearchCV
from sklearn.linear_model import LinearRegression
from sklearn.svm import SVR, LinearSVR
from sklearn.neighbors import KNeighborsRegressor
from sklearn.ensemble import RandomForestRegressor
from sklearn.ensemble import BaggingRegressor, AdaBoostRegressor
from sklearn.neural_network import MLPRegressor
from prophet import Prophet
```

```

import xgboost as xgb
from xgboost import XGBRegressor
import lightgbm as lgb
from lightgbm import LGBMRegressor
import warnings
warnings.filterwarnings("ignore")
# What EDA & FE techniques use?
is_anomalies = True # or False - Take into account anomalies or no?
# What type of model to use?
is_Prophet = True # or False - Facebook Prophet
is_ARIMA = True # or False - ARIMA and AutoARIMA
is_other_ML = True # or False - multi-factors models: trees, neural networks, etc.
#is_other_ML = False # or False - multi-factors models: trees, neural networks, etc.
# Automatic building ARIMA for Time Series
if is_ARIMA:
    import pmdarima as pm
# Set random state
def fix_all_seeds(seed):
    np.random.seed(seed)
    random.seed(seed)
    os.environ['PYTHONHASHSEED'] = str(seed)
random_state = 42
fix_all_seeds(random_state)
# Set main parameters
series_name = 'CO2'
target = 'y'
forecasting_days = 5
# Set time interval of data
date_start = dt.datetime(2023, 7, 1)
date_end = dt.datetime(2023, 7, 31)
print(f"Time interval: from {date_start} to {date_end}")
def get_data(name_param, station_id):
    # Get data for name_param from station_id
    station = pd.read_csv(f'/kaggle/input/air-quality-monitoring-from-ecocity/ecocity_hourly_{station_id}_2024.csv', sep=',', header=0)
    df = station.pivot_table(index=['id_station', 'datetime'], columns='indicator_name',
    values='value', aggfunc='first')
    df.reset_index(inplace=True)
    df['datetime'] = pd.to_datetime(df['datetime'])
    df.columns.name = None
    df = df[df['datetime'].between('2023-07-01', '2023-07-31')]
    df2 = df[['datetime', name_param]].dropna()
    df2 = df2.sort_values('datetime').reset_index(drop=True)
    df2.columns = ['ds', 'y']

```

```
    return df2
name_param = 'CO2'
id_station = 1315
df = get_data(name_param, id_station)
df
```

## Додаток Г

**ІЛЮСТРАТИВНА ЧАСТИНА**

ІНФОРМАЦІЙНА ТЕХНОЛОГІЯ АНАЛІЗУ ТА ПЕРЕДБАЧЕННЯ  
ДІОКСИДУ ВУГЛЕЦЮ ЗА ДАНИМИ ГРОМАДСЬКОГО  
МОНІТОРИНГУ АТМОСФЕРНОГО ПОВІТРЯ У М. ВІННИЦІ

Нормоконтроль: к.т.н., доцент



Сергій ЖУКОВ

«\_\_\_» \_\_\_\_\_ 2024 р.

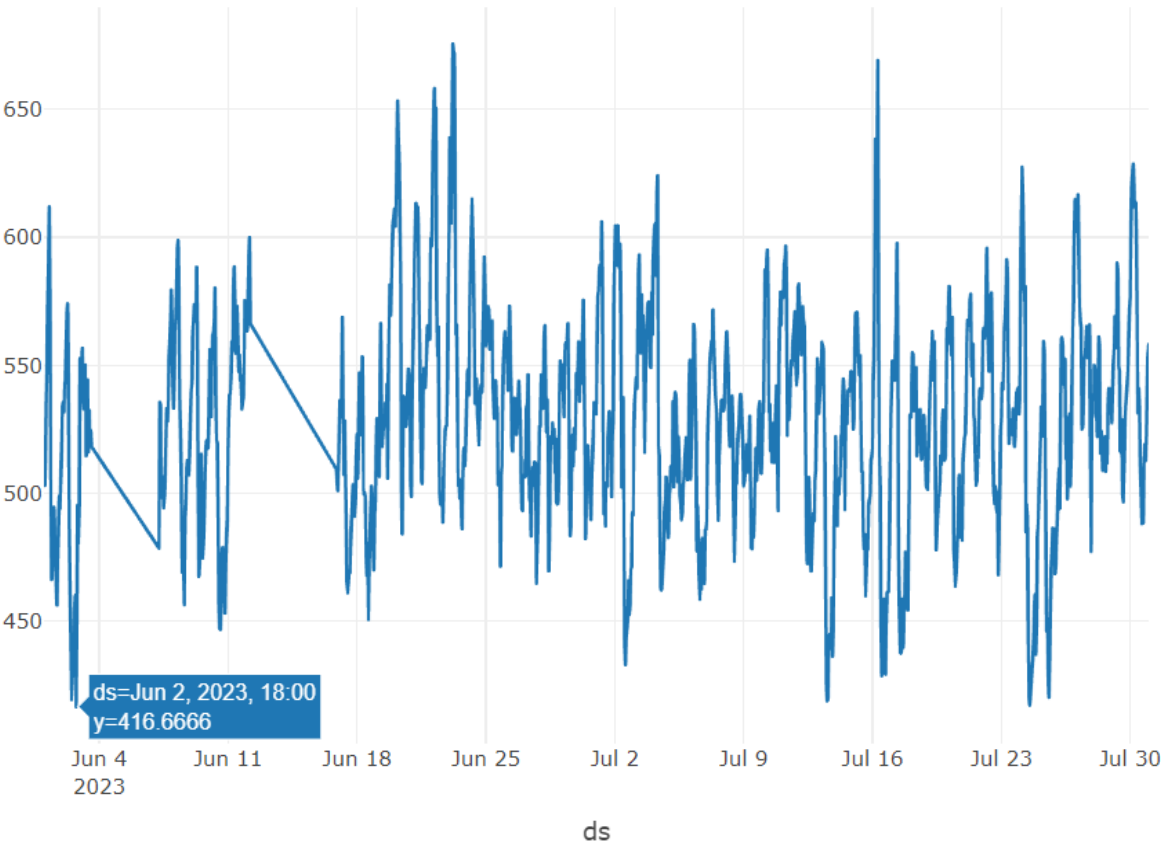


Рисунок Г.1 – Зображення аномалії

|    | ds         | lower_window | upper_window | prior_scale | holiday         |
|----|------------|--------------|--------------|-------------|-----------------|
| 0  | 2023-06-02 | 0            | 0            | 10          | anomalous_dates |
| 1  | 2023-06-03 | 0            | 0            | 10          | anomalous_dates |
| 2  | 2023-06-07 | 0            | 0            | 10          | anomalous_dates |
| 3  | 2023-06-20 | 0            | 0            | 10          | anomalous_dates |
| 4  | 2023-06-22 | 0            | 0            | 10          | anomalous_dates |
| 5  | 2023-06-23 | 0            | 0            | 10          | anomalous_dates |
| 6  | 2023-07-02 | 0            | 0            | 10          | anomalous_dates |
| 7  | 2023-07-04 | 0            | 0            | 10          | anomalous_dates |
| 8  | 2023-07-13 | 0            | 0            | 10          | anomalous_dates |
| 9  | 2023-07-16 | 0            | 0            | 10          | anomalous_dates |
| 10 | 2023-07-17 | 0            | 0            | 10          | anomalous_dates |
| 11 | 2023-07-24 | 0            | 0            | 10          | anomalous_dates |
| 12 | 2023-07-25 | 0            | 0            | 10          | anomalous_dates |
| 13 | 2023-07-27 | 0            | 0            | 10          | anomalous_dates |
| 14 | 2023-07-30 | 0            | 0            | 10          | anomalous_dates |

Рисунок Г.2 – Зображення датасету аномалій



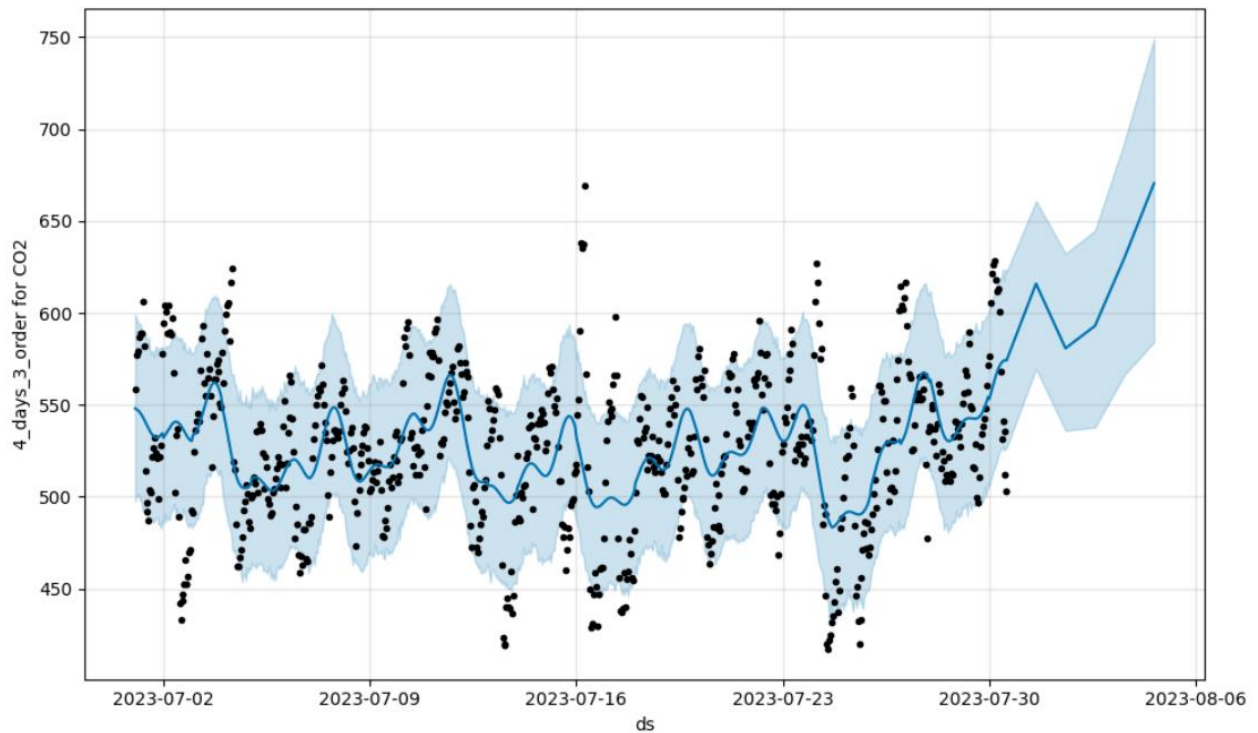


Рисунок Г.3 – Результат передбачення вмісту вуглекислого газу в атмосферному повітрі за валідаційними даними на основі моделі FB Prophet

```

Tuning model 'Linear Regression'
Best parameters: {'fit_intercept': True}

Tuning model 'KNeighbors Regressor'
Best parameters: {'leaf_size': 10, 'n_neighbors': 30}

Tuning model 'Support Vector Machines'
Best parameters: {'C': 4.0, 'kernel': 'rbf', 'tol': 0.001}

Tuning model 'Linear SVR'
Best parameters: {'C': 1.0}

Tuning model 'Random Forest Regressor'
Best parameters: {'max_depth': 3, 'max_features': 'auto', 'min_samples_leaf': 10, 'min_samples_split': 40, 'n_estimators': 80}

Tuning model 'Bagging Regressor'
Best parameters: {'max_features': 0.05, 'n_estimators': 6, 'warm_start': False}

Tuning model 'XGB Regressor'
Best parameters: {'learning_rate': 0.01, 'max_depth': 3, 'n_estimators': 90}

Tuning model 'MLP Regressor'
Best parameters: {'hidden_layer_sizes': 2, 'learning_rate': 'adaptive', 'learning_rate_init': 0.001, 'max_iter': 1000, 'solver': 'sgd'}

```

Рисунок Г.4 – Побудовані моделі перебачення

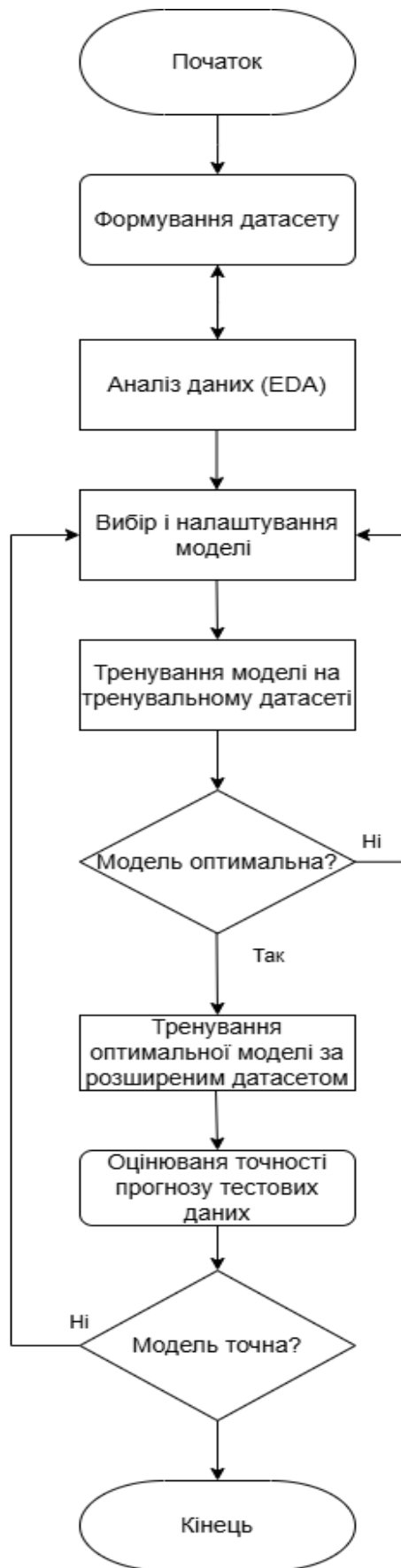


Рисунок Г.5 – Зображення алгоритму виконання задачі

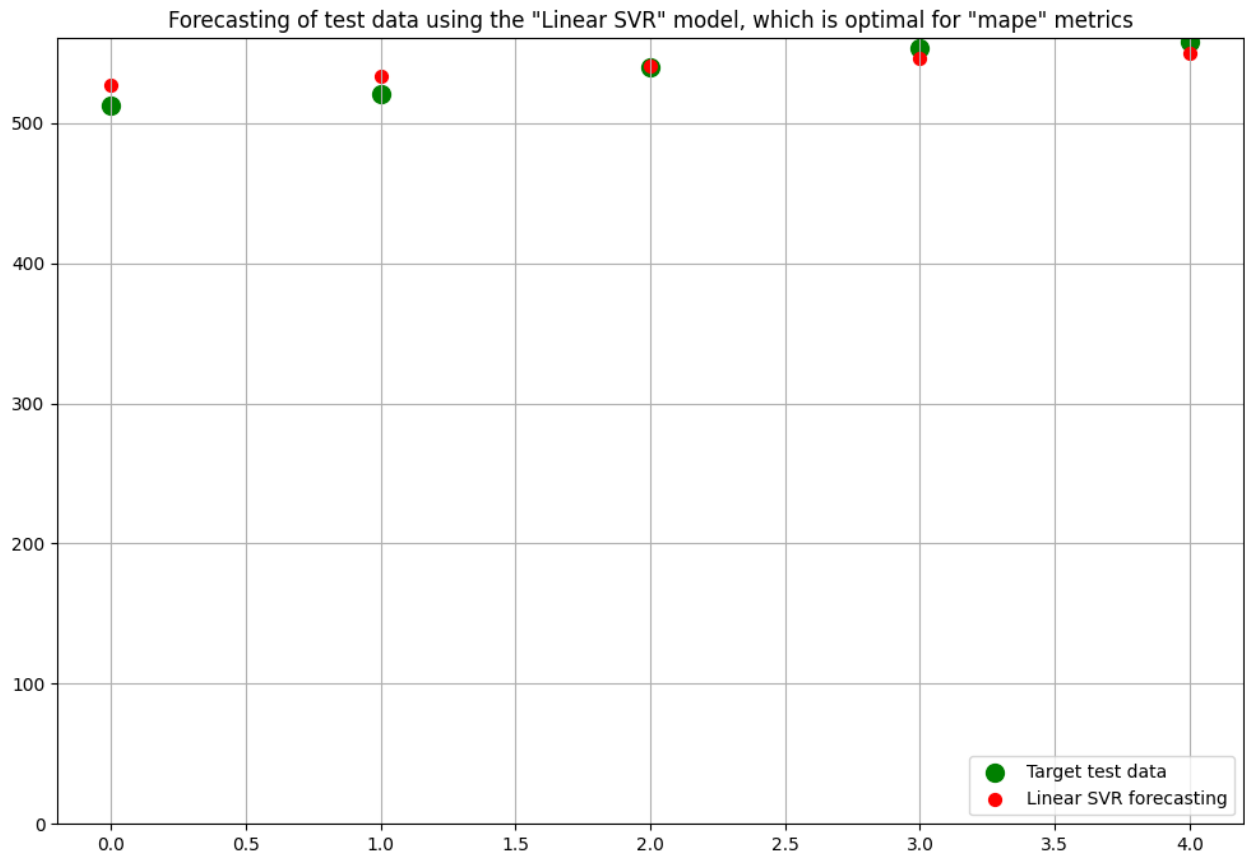


Рисунок Г.6 – Реальні тестові дані вмісту вуглекислого газу та їх прогноз за моделлю Linear SVR, оптимальної за метрикою MAPE