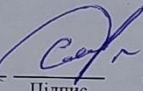


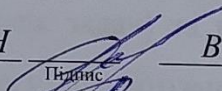
Хмельницький національний університет
Факультет інформаційних технологій
Кафедра комп'ютерних наук

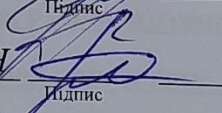
КВАЛІФІКАЦІЙНА РОБОТА БАКАЛАВРА

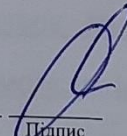
на тему Метод прогнозування забрудненості (якості) повітря засобами інтелектуального аналізу даних

Галузь знань 12 - Інформаційні технології
Шифр і назва галузі знань
Спеціальність 122 - Комп'ютерні науки
Шифр і назва спеціальності
Освітня програма Комп'ютерні науки
Назва освітньої програми

Виконав: студент групи КН-21-2  Михайло СМУК
Група виконавця Підпис Ім'я, ПРІЗВИЩЕ

Керівник: к.фіз.-мат.н., доц. каф. КН  Віталій МІХАЛЕВСЬКИЙ
Науковий ступінь, посада Підпис Ім'я, ПРІЗВИЩЕ

Нормоконтроль: к.т.н., доц. каф. КН  Руслан БАГРІЙ
Науковий ступінь, посада Підпис Ім'я, ПРІЗВИЩЕ

До захисту допускаю:
зав. кафедри КН, д.т.н., професор  Олександр БАРМАК
19 06 2025 р. Підпис Ім'я, ПРІЗВИЩЕ

ХМЕЛЬНИЦЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
Факультет інформаційних технологій
Кафедра комп'ютерних наук
Освітній ступінь бакалавр
Галузь знань 12 - Інформаційні технології
Спеціальність 122 - Комп'ютерні науки

ЗАТВЕРДЖУЮ
Завідувач кафедри комп'ютерних наук


(підпис)

д.т.н., професор Олександр БАРМАК
«10» 02 2025 року

ЗАВДАННЯ НА КВАЛІФІКАЦІЙНУ РОБОТУ БАКАЛАВРА

1. Тема кваліфікаційної роботи бакалавра: «Метод прогнозування забрудненості (якості) повітря засобами інтелектуального аналізу даних»
2. Завдання видано студенту Михайлу СМУКУ
(Ім'я, прізвище)
3. Керівник роботи доцент кафедри КН Віталій МІХАЛЕВСЬКИЙ
(посада, ім'я, прізвище)
4. Затверджено наказом університету від «07» 02 2025 р. № 23
5. Дата видачі завдання студенту: «10» 02 2025 р.
6. Зміст пояснювальної записки (перелік задач) та вихідні дані: Мета роботи полягає у підвищенні точності процесу прогнозування забрудненості повітря. Для досягнення цієї мети необхідно провести аналіз сучасних методів прогнозування, розробити метод прогнозування забрудненості повітря засобами інтелектуального аналізу даних, інтегрувати його в інформаційну систему та провести експериментальне тестування для оцінки точності методу. Вхідні дані передбачають можливість повного опису предметної області, а саме: облік характеристик повітря, множини якісних даних повітря, параметри залежності тощо.

7. Календарний план виконання кваліфікаційної роботи бакалавра:

№	Назва етапів (розділів) кваліфікаційної роботи бакалавра	Термін виконання	Примітка
1	Вибір напрямку дослідження та узгодження тематики кваліфікаційної роботи бакалавра з керівником, складання календарного графіка виконання роботи	січень 2025	виконано
2	Ознайомлення з предметною областю, формулювання мети та задач дослідження, визначення об'єкта та предмета дослідження	лютий 2025	виконано
3	Проектування та розробка загальної архітектури системи, вибір засобів реалізації програмного забезпечення	березень 2025	виконано
4	Розробка інформаційної системи	квітень 2025	виконано
5	Написання пояснювальної записки, урахування зауважень керівника, оформлення згідно вимог	травень 2025	виконано
6	Розробка презентаційних матеріалів та попередній захист кваліфікаційної роботи	травень 2025	виконано
7	Отримання відгуку керівника, рецензії, перевірка на плагіат, нормоконтроль	червень 2025	виконано
8	Підготовка до захисту та захист кваліфікаційної роботи бакалавра	червень 2025	виконано

Виконавець: студент групи КН-21-2

Група виконавця

Підпис

Михайло СМУК

Ім'я, ПРІЗВИЩЕ

Керівник: к. фіз.-мат. н., доц. каф. КН

Науковий ступінь, посада

Підпис

Віталій МІХАЛЕВСЬКИЙ

Ім'я, ПРІЗВИЩЕ

Анотація

Тема кваліфікаційної роботи бакалавра: «Метод прогнозування забрудненості (якості) повітря засобами інтелектуального аналізу даних»

Виконавець кваліфікаційної роботи бакалавра: студент групи КН-21-2 Михайло СМУК

Керівник кваліфікаційної роботи бакалавра: к.фіз.-мат.н., доцент кафедри КН Віталій МІХАЛЕВСЬКИЙ

Кваліфікаційна робота бакалавра містить:

Пояснювальна записка				Кількість додатків
Сторінок	Рисунків	Таблиць	Джерел інформації	
56	16	6	48	2

Метою кваліфікаційної роботи бакалавра є підвищення точності процесу прогнозування забрудненості повітря. Розроблений метод прогнозування якості повітря реалізує метод візуалізації та аналітики показників забруднення повітря на основі попередньо сформованих табличних даних. Основу рішення становить інтерактивний графічний інтерфейс, створений за допомогою мови програмування Python, бібліотек Tkinter, Pandas, NumPy, а також модуля TkinterMapView для інтеграції карти з позначенням міст.

Практичне застосування даного методу охоплює моніторинг екологічної ситуації в містах України, локальний аналіз рівня забруднення повітря без потреби підключення до інтернету, а також можливість розширення системи шляхом підключення локальних сенсорів або оновлення даних вручну у форматі CSV.

Ключові слова: забрудненість повітря, екологічна безпека, метеодані, шкідливі викиди, кліматичні процеси, інтелектуальний аналіз даних.

Виконавець: студент групи КН-21-2
Група виконавця


Підпис

Михайло СМУК
Ім'я, ПІРІЗВИЩЕ

Зміст

Перелік скорочень	3
Вступ.....	4
Розділ 1 Характеристика предметної області: аналіз моделей, методів та реалізацій.....	6
1.1 Аналіз предметної області забрудненості повітря	6
1.2 Огляд теоретичних підходів до розв’язку задач прогнозування забрудненості повітря	8
1.3 Аналіз існуючих програмних засобів та наукових рішень	11
1.4 Мета та завдання кваліфікаційної роботи	15
Розділ 2 Розробка методу прогнозування забрудненості (якості) повітря засобами інтелектуального аналізу даних.....	16
2.1 Концепція методу прогнозування забрудненості повітря	16
2.2 Оптимізація методу прогнозування забрудненості повітря	17
2.3 Схема методу прогнозування забрудненості повітря	20
2.4 Інформаційна модель методу прогнозування забрудненості повітря	21
2.5 Використання засобів інтелектуального аналізу даних для прогнозування забрудненості повітря.....	24
2.6 Підготовка робочих вхідних даних для системи	25
2.7 Спосіб оцінки ефективності прогнозування забрудненості повітря	27
2.8 Висновки до розділу 2	29
Розділ 3 Програмна реалізація та експериментальне дослідження методу прогнозування забрудненості (якості) повітря засобами інтелектуального аналізу даних.....	30
3.1 Структура модулів системи, їх взаємозв’язок	30
3.2 Засоби розробки інформаційної системи	33
3.3 Опис функціональних можливостей інформаційної системи	36
3.4 Проведення експериментів та дослідження роботи системи	39
3.5 Аналіз результатів досліджень та застосованого методу прогнозування забрудненості повітря.....	42
3.6 Висновки до розділу 3	49
Загальні висновки.....	50
Перелік посилань.....	52
Додатки	

Перелік скорочень

Скорочення, термін, позначення	Пояснення
IT	Інформаційні технології
КРБ	Кваліфікаційна робота бакалавра
КН	Комп'ютерні науки
ВООЗ	Всесвітня організація охорони здоров'я
PM2.5	Тверді частинки діаметром до 2.5 мкм
CAMS	Copernicus Atmosphere Monitoring Service
SHAP	SHapley Additive exPlanations
CNN	Convolutional Neural Network
API	Application Programming Interface
XAI	Explainable Artificial Intelligence
EPA	Environmental Protection Agency
NOAA	National Oceanic and Atmospheric Administration
AQI	Air Quality Index
RNN	Recurrent Neural Networks
MAE	Mean Absolute Error
CSV	Формат табличних даних (Comma Separated Values)
API	Інтерфейс прикладного програмування
LSTM	Довга короткочасна пам'ять (тип нейромережі)
RMSE	Root Mean Squared Error

Вступ

Кваліфікаційна робота бакалавра присвячена розробці методу прогнозування забрудненості повітря засобами інтелектуального аналізу даних.

Актуальність теми. Забруднення повітря є однією з найгостріших екологічних проблем сучасності, яка має значний вплив на здоров'я населення, стан навколишнього середовища та кліматичні процеси. Викиди промислових підприємств, автотранспорту, енергетичних об'єктів, а також природні фактори, такі як лісові пожежі та пилові бурі, спричиняють підвищення концентрації шкідливих речовин у атмосфері. У зв'язку з цим прогнозування рівня забрудненості повітря набуває важливого значення для своєчасного реагування, розробки екологічних заходів та зменшення негативного впливу на населення.

Традиційні методи моніторингу забруднення повітря, що базуються на даних стаціонарних екологічних станцій, мають низку обмежень, зокрема нерівномірне територіальне покриття та затримку в отриманні даних. Використання засобів інтелектуального аналізу даних, зокрема методів машинного навчання, великих даних (Big Data) та геоінформаційних систем (GIS), дозволяє значно підвищити точність і оперативність прогнозування. Такі методи дають змогу аналізувати великі масиви інформації, включаючи метеорологічні показники, супутникові знімки, антропогенні фактори та історичні дані про забруднення.

Особливої актуальності ця тема набуває в умовах урбанізації, зростання індустріальних потужностей та збільшення кількості транспорту. Крім того, військові дії, зокрема на території України, спричиняють значне забруднення повітря через вибухи, пожежі, руйнування інфраструктури та промислових об'єктів. Це створює додаткову необхідність розробки оптимальних методів прогнозування та оцінки якості повітря для мінімізації екологічних ризиків та захисту здоров'я населення. Для точного прогнозування забрудненості повітря необхідно використовувати комплексний підхід, що враховує метеорологічні умови, промислові та транспортні викиди, супутникові дані та екологічні

особливості регіону. Застосування інтелектуального аналізу даних дозволяє інтегрувати ці параметри в прогностичні моделі, що сприятиме підвищенню точності прогнозів та ефективному управлінню якістю повітря.

Таким чином, дослідження методів прогнозування забрудненості повітря засобами інтелектуального аналізу даних є надзвичайно важливим для сучасної екологічної безпеки. Впровадження таких підходів сприятиме покращенню системи моніторингу, зменшенню впливу шкідливих викидів та підвищенню ефективності заходів, спрямованих на охорону довкілля.

Мета кваліфікаційної роботи бакалавра полягає у підвищенні точності процесу прогнозування забрудненості повітря.

Об'єкт дослідження – процес прогнозування забрудненості повітря засобами інтелектуального аналізу даних.

Предмет дослідження – методи, засоби та технології інтелектуального аналізу даних для прогнозування забрудненості повітря.

Завдання кваліфікаційної роботи бакалавра. Для досягнення поставленої мети визначено наступні задачі:

1. Провести аналіз предметної області та відомих підходів до прогнозування забрудненості повітря.
2. Вдосконалити інформаційну модель прогнозування забрудненості повітря.
3. Розробити метод прогнозування забрудненості повітря засобами інтелектуального аналізу даних.
4. Підготувати набори даних для їх інтелектуального аналізу в методі прогнозування забрудненості повітря.
5. Застосувати засоби інтелектуального аналізу даних для точності прогнозування забрудненості повітря.
6. Провести функціональне та прикладне дослідження ефективності запропонованого методу прогнозування забрудненості повітря засобами інтелектуального аналізу даних.

Розділ 1 Характеристика предметної області: аналіз моделей, методів та реалізацій

1.1 Аналіз предметної області забрудненості повітря

Проблема забруднення атмосферного повітря вже тривалий час залишається однією з ключових екологічних загроз сучасного світу. Вона безпосередньо впливає на якість життя мільйонів людей, обумовлює серйозні ризики для здоров'я та сприяє поглибленню глобальних змін клімату [1, 2]. У багатьох країнах, зокрема й в Україні, рівень забрудненості повітря регулярно перевищує рекомендовані показники, визначені міжнародними організаціями, зокрема ВООЗ [3]. Через це питання вивчення, моніторингу й прогнозування стану повітря набуває все більшої актуальності як у науковій сфері, так і серед громадськості.

Визначальну роль відіграють антропогенні джерела – інтенсивний рух транспорту, діяльність промислових підприємств, спалювання палива в побуті, а також сільськогосподарські процеси [4, 5]. Водночас, на рівень забруднення значно впливають і природні умови: зміна погоди, швидкість вітру, температура, вологість, атмосферний тиск [6].

Серед речовин, які найчастіше аналізуються при оцінці якості повітря, вирізняють зважені частинки (PM_{2.5} і PM₁₀), діоксид азоту, чадний газ, озон та інші [7, 8]. Дрібнодисперсні частинки вважаються особливо небезпечними, оскільки можуть проникати глибоко в легені й навіть потрапляти в кровоносну систему, спричиняючи хронічні захворювання [9].

У зв'язку з цим багато країн запровадили індекси якості повітря, які допомагають населенням швидко орієнтуватись у поточній ситуації та вживати відповідних заходів [10].

Загалом система спостереження за якістю повітря базується на автоматизованих станціях моніторингу, які здійснюють регулярні вимірювання концентрацій забруднювачів [11].

Зібрані дані часто публікуються в режимі реального часу та використовуються для побудови аналітичних моделей, здатних виявляти закономірності й прогнозувати розвиток ситуації наперед [12]. Окрім стаціонарних станцій, дедалі ширше застосовуються й мобільні або індивідуальні датчики, а також супутникові спостереження [13, 14]. Це дозволяє отримувати дані з високою деталізацією як у просторі, так і в часі.

Сучасні виклики, пов'язані з урбанізацією, розвитком інфраструктури та кліматичними змінами, роблять задачу прогнозування забруднення повітря надзвичайно важливою [15].

Завдяки такому прогнозу можливо не лише попередити населення про потенційно небезпечні рівні забруднення, а й приймати рішення щодо зниження впливу шкідливих чинників, оптимізації трафіку, обмеження промислових викидів або зміни режиму роботи вентиляційних систем.

Попри наявність стаціонарних станцій моніторингу, багато регіонів залишаються малодослідженими через високу вартість обслуговування таких систем або обмежену інфраструктуру.

Саме тому дедалі ширше застосування знаходять компактні сенсорні пристрої, мобільні платформи та супутникові технології, які можуть забезпечити неперервне спостереження навіть у віддалених районах. Проте збір даних – лише початкова ланка в ланцюгу прийняття рішень.

Інтелектуальний аналіз даних, зокрема машинне навчання та нейронні мережі, відкриває нові горизонти у прогнозуванні забруднення. Ці підходи дозволяють виявляти приховані залежності між екологічними, метеорологічними та техногенними показниками, які складно врахувати у класичних моделях.

Наприклад, застосування глибоких нейронних мереж (Deep Learning) дало змогу досягти високої точності у прогнозах PM_{2.5} на 24-72 години наперед у мегаполісах Китаю, США та ЄС. Подібні моделі вже інтегруються у міські інформаційні системи, мобільні додатки та державні портали для попередження населення (таблиця 1.1).

Таблиця 1.1 – Порівняльна характеристика методів прогнозування забруднення повітря

Метод	Тип моделі	Необхідність навчання	Складність реалізації	Приклад використання
Лінійна регресія	Статистична	Помірна	Низька	Прогноз PM2.5 у містах
SVM	Машинне навчання	Висока	Висока	Виявлення шаблонів викидів
Random Forest	Ensemble Learning	Висока	Помірна	Моделювання якості повітря
Нейронні мережі	Глибинне навчання	Дуже висока	Висока	Прогноз на основі часових рядів

Отже, аналіз предметної області показує, що точне прогнозування якості повітря потребує врахування широкого кола чинників – як природних, так і техногенних – та базується на точному зборі, обробці й інтерпретації даних. Це створює передумови для використання сучасних підходів інтелектуального аналізу, які дозволяють будувати складні, адаптивні й високоточні моделі, орієнтовані на практичне застосування у сфері охорони довкілля.

1.2 Огляд теоретичних підходів до розв'язку задач прогнозування забрудненості повітря

Прогнозування забруднення повітря – це складний процес, який потребує врахування багатьох взаємозалежних факторів, зокрема природних умов, техногенного навантаження та динаміки попередніх спостережень [16, 17]. Упродовж останніх десятиліть було розроблено різноманітні методи прогнозування, які можна умовно поділити на кілька основних груп залежно від їх підходу до аналізу даних та побудови моделі [18].

Традиційно в екологічному моніторингу використовувалися статистичні моделі. Зокрема, регресійні аналізи та моделі часових рядів дозволяли встановити залежність між рівнем забруднення та низкою факторів, таких як температура, вологість або швидкість вітру [19]. Попри свою простоту та доступність, ці моделі часто виявлялися недостатньо гнучкими в умовах швидкозмінного міського середовища або нестабільного клімату. Їх точність залежала від припущення про лінійність процесів, що не завжди відповідало реальній картині [20].

Чисельні моделі, що прийшли на зміну класичним статистичним підходам, базуються на фізико-хімічних закономірностях, які описують поведінку забруднювачів у атмосферному середовищі [21]. Ці моделі потребують значної кількості вхідних параметрів, зокрема детальної інформації про джерела викидів, рельєф місцевості, погодні умови. Їхньою перевагою є здатність прогнозувати зміну ситуації за нових умов, навіть якщо подібні сценарії ще не зустрічались у наявних даних. Водночас, вони є обчислювально дорогими і часто непридатними для оперативного використання в реальному часі без значного технічного забезпечення [22]. У зв'язку з розвитком обчислювальної техніки та появою великих обсягів даних, в останні роки активно впроваджуються підходи на основі машинного навчання [23]. Вони дозволяють моделювати складні, нерівномірні та нелінійні процеси, які важко описати традиційними методами. Сучасні алгоритми – зокрема нейронні мережі, дерева рішень, метод опорних векторів – здатні адаптуватися до нових даних і «навчатися» з часом, що особливо корисно для задач, де дані постійно оновлюються [24]. Такі моделі можуть використовуватись як для короткострокового прогнозу (на кілька годин або діб), так і для довготривалої оцінки тенденцій [25].

Окремо варто згадати гібридні підходи, які комбінують переваги різних методів. Наприклад, фізична модель може використовуватись для моделювання базової поведінки атмосфери, а її результати – доповнюватись корекційним шаром на основі машинного навчання [26]. Це дозволяє не лише зберігати фізичну обґрунтованість прогнозу, а й одночасно зменшити похибки, спричинені неточностями у вхідних параметрах або змінними локальними умовами (таблиця

1.2). Подібні поєднання дедалі частіше зустрічаються у сучасних дослідженнях і демонструють високу точність у задачах екологічного прогнозування [27].

Таблиця 1.2 – Основні забруднювачі повітря та їх вплив

Забруднювач	Джерело виникнення	Вплив на здоров'я	Гранично допустимі норми (ГДК)
PM2.5	Авто, промисловість	Захворювання легень, серця	25 $\mu\text{g}/\text{m}^3$ (добова)
PM10	Будівництво, пил	Алергії, подразнення	50 $\mu\text{g}/\text{m}^3$ (добова)
NO ₂	Автотранспорт	Погіршення дихання	40 $\mu\text{g}/\text{m}^3$ (річна)
CO	Неповне згоряння палива	Гіпоксія	10 mg/m^3 (8 годин)

Зокрема, дослідження останніх років підтверджують ефективність використання моделей глибокого навчання, таких як LSTM (Long Short-Term Memory) для аналізу часових рядів концентрацій PM2.5 та PM10. Ці моделі здатні вловлювати складні нелінійні залежності та взаємозв'язки між поточними і попередніми спостереженнями [28]. У поєднанні з рекурентними структурами вони забезпечують точність прогнозів понад 90% при горизонті в один-два дні, що підтверджено прикладними дослідженнями у країнах Європи та Азії. Аналогічно, згорткові нейронні мережі (CNN) демонструють гарні результати у побудові просторових карт забруднення на основі зображень з супутників або теплових карт із сенсорних мереж [29].

Окрім того, активного розвитку набули explainable AI (XAI) підходи, які дозволяють тлумачити логіку прийняття рішення моделлю. Такі моделі, наприклад, SHAP (SHapley Additive exPlanations), дозволяють визначити внесок кожної змінної у фінальний прогноз [30, 31]. Ще один важливий напрям – гібридні

моделі, які поєднують обчислення на основі фізичних закономірностей із корекційними шарами, збудованими засобами машинного навчання [32]. Наприклад, результати моделі WRF-Chem можуть уточнюватись алгоритмами типу Random Forest або XGBoost, що дає змогу зменшити похибку прогнозу без необхідності розраховувати повну фізичну модель у кожен момент часу [33]. У низці сучасних досліджень такі поєднання дали змогу зменшити середню абсолютну похибку на 15–25% порівняно з окремими методами [34].

Окремо варто згадати зростаючу популярність edge computing – концепції, за якої попередній аналіз даних виконується безпосередньо на пристрої збору, наприклад, на мікроконтролері або IoT-модулі [35]. Це дозволяє забезпечити безперервність прогнозування навіть за умов обмеженого доступу до мережі, що особливо важливо для регіонів із нерозвиненою інфраструктурою або у мобільних сенсорних мережах.

Таким чином, сучасний етап розвитку технологій дає змогу поєднувати різні підходи – статистичні та інтелектуальні – у єдину адаптивну систему прогнозування, здатну реагувати на змінні умови в реальному часі [36, 37].

1.3 Аналіз існуючих програмних засобів та наукових рішень

У світі активно розвиваються програмні платформи та наукові системи, призначені для моніторингу та прогнозування якості повітря [38-40]. Більшість із них базуються на даних з автоматизованих станцій спостережень, супутникових вимірювань або мобільних сенсорів, і використовують як класичні, так і сучасні методи аналізу [41, 42]. У цій підсистемі варто розглянути як приклади державних ініціатив, так і академічні розробки та інноваційні рішення, засновані на інтелектуальних технологіях [43, 44].

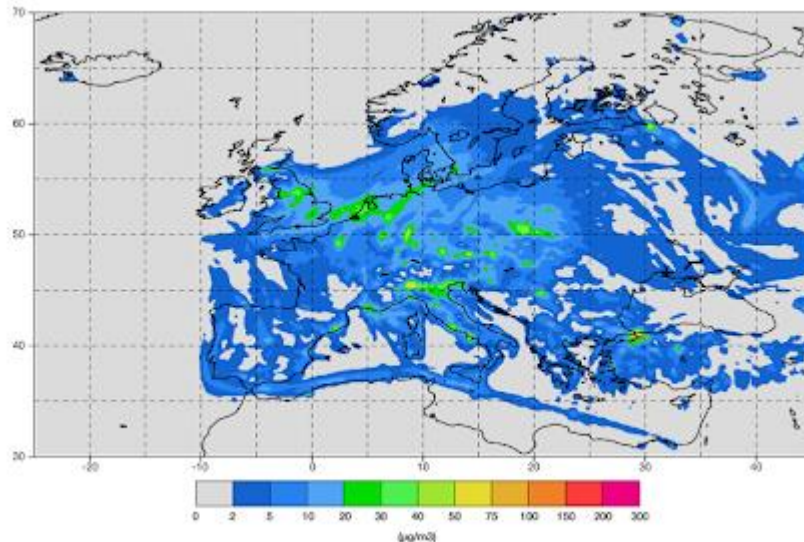


Рисунок 1.1 – Copernicus Atmosphere Monitoring Service (CAMS) [45]

Серед державних систем особливо виділяється європейська платформа Copernicus Atmosphere Monitoring Service (CAMS) (рисунок 1.1), яка забезпечує оперативні дані та прогнози про концентрації основних забруднювачів у масштабах континенту. CAMS об'єднує результати супутникових спостережень з моделями хімії атмосфери, дозволяючи надавати прогнози на кілька днів наперед.

Окрему нішу займають системи регіонального рівня. Наприклад, у Китаї, де питання забруднення стоїть особливо гостро, діє потужна державна інфраструктура моніторингу, яка об'єднує тисячі стаціонарних станцій по всій країні. Аналогічні рішення впроваджуються і в Індії, де дані про забруднення інтегруються з прогнозами метеорологічних служб для оцінки ризиків.

У науковому середовищі розроблено низку програмних рішень, що демонструють успішне поєднання фізичних моделей з алгоритмами машинного навчання. Зокрема, в академічних дослідженнях часто використовуються моделі на основі рекурентних нейронних мереж (RNN), включаючи LSTM-архітектури, які добре підходять для аналізу часових рядів. Також вивчається застосування згорткових нейронних мереж (CNN) для обробки супутникових знімків та виявлення просторових закономірностей.

З технічного боку, розробники таких систем стикаються з викликами забезпечення точності, калібрування сенсорів, обробки великого обсягу

неоднорідних даних у реальному часі. Важливим є й дотримання стандартів, наприклад, OpenAQ використовує формат JSON API, що дає змогу легко інтегрувати дані з іншими сервісами. Крім того, оцінювання точності моделей здійснюється на основі стандартних метрик (MAE, RMSE, R^2), а для прозорості часто застосовується відкритий набір бенчмарків, наприклад, на платформах Kaggle або UCI. У Європі успішно реалізується проект uMobiliAir, у межах якого досліджується використання штучного інтелекту для прогнозування якості повітря в міських умовах. Проект об'єднує науковців і муніципалітети, дозволяючи тестувати моделі безпосередньо в реальному середовищі. Аналогічно, у рамках програми Horizon 2020 було профінансовано розробку інтелектуальних платформ для виявлення аномалій у даних моніторингу та автоматичної генерації попереджень для населення.

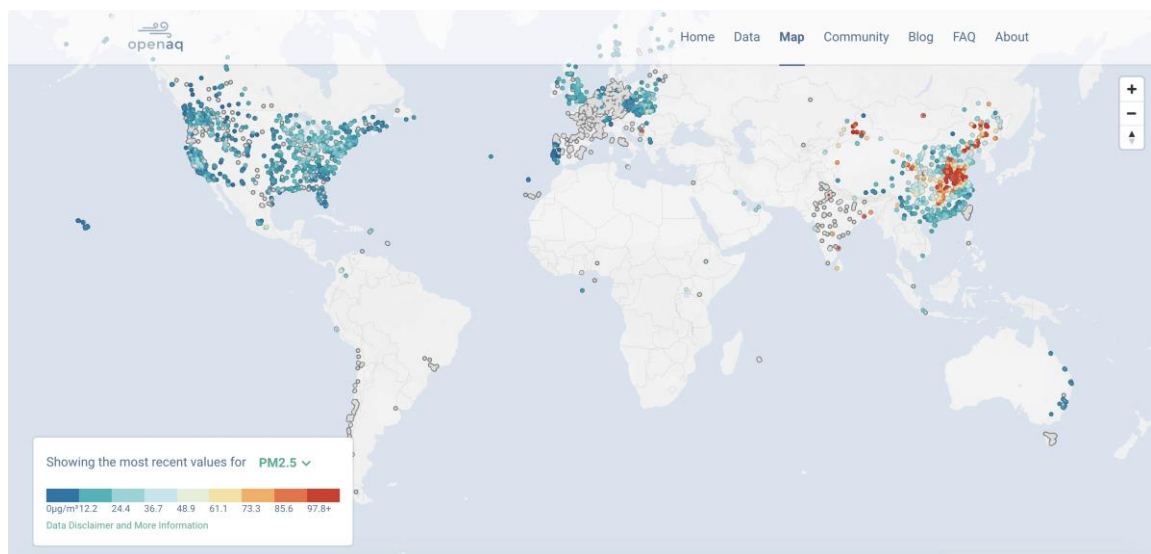


Рисунок 1.2 – Система OpenAQ [46]

На особливу увагу заслуговують програмні системи з відкритим кодом. Зокрема, OpenAQ (рисунок 1.2) надає відкритий доступ до історичних і поточних даних про якість повітря в різних країнах.

Це дозволяє дослідникам не лише здійснювати ретроспективний аналіз, а й будувати власні системи прогнозування. Ще один приклад – платформа AQICN.org, яка агрегує інформацію з більш ніж 10 000 станцій спостереження у світі, надаючи користувачам доступ до даних через вебінтерфейс та API.

У комерційному секторі активно розвиваються мобільні додатки, що дозволяють отримувати прогнози якості повітря в конкретному місці. Такі сервіси, як Plume Labs (рисунок 1.3) або IQAir, застосовують власні алгоритми прогнозування, часто використовуючи ансамблеві моделі, які поєднують дані з декількох джерел – супутників, локальних сенсорів і метеосервісів. Деякі додатки також інтегруються з переносними пристроями або смарт-гаджетами.



Рисунок 1.3 – Інтерфейс додатку Plume Lab [47]

У контексті України перспективним є розвиток локальних платформ моніторингу повітря, як ЛУН Місто Air чи SaveEcoBot, які базуються на відкритих даних, зокрема з платформи OpenAQ, та використовують власні мережі сенсорів.

Таким чином, наявні програмні засоби й наукові розробки демонструють високий рівень різноманіття – як у концептуальному, так і в технологічному аспектах. Їх подальший розвиток передбачає інтеграцію з технологіями інтелектуального аналізу даних, інтернету речей, обчислень на периферії (edge computing) та хмарних сервісів, що відкриває широкі перспективи для створення адаптивних, масштабованих і точних систем прогнозування забруднення повітря.

1.4 Мета та завдання кваліфікаційної роботи

Метою кваліфікаційної роботи є підвищення точності процесу прогнозування забрудненості повітря. Повний опис предметної області, а саме: облік та характеристики об'єктів сонячної енергії, множини метеорологічних даних, параметри залежності тощо, передбачено вхідними даними.

Для досягнення мети кваліфікаційної роботи необхідно реалізувати виконання наступних задач.

1. Провести аналіз предметної області та відомих підходів до прогнозування забрудненості повітря.
2. Вдосконалити інформаційну модель прогнозування забрудненості повітря.
3. Розробити метод прогнозування для прогнозування забрудненості повітря засобами інтелектуального аналізу даних.
4. Підготувати набори даних для їх інтелектуального аналізу в методі прогнозування забрудненості повітря.
5. Застосувати засоби інтелектуального аналізу даних для ефективного прогнозування забрудненості повітря.
6. Провести функціональне та прикладне дослідження ефективності запропонованого методу прогнозування забрудненості повітря засобами інтелектуального аналізу даних.

Розділ 2 Розробка методу прогнозування забрудненості (якості) повітря засобами інтелектуального аналізу даних

2.1 Концепція методу прогнозування забрудненості повітря

Сформовано концепцію методу прогнозування забрудненості повітря, який базується на використанні засобів інтелектуального аналізу даних для виявлення закономірностей у змінних, що впливають на якість атмосфери. В основі концепції лежить ідея поєднання простоти реалізації з можливістю адаптації до реальних даних, що постійно змінюються.

Концепція розробленого підходу передбачає безперервний збір даних про метеорологічні умови (температура, вологість, атмосферний тиск, швидкість вітру) та концентрації забруднювальних речовин (PM_{2.5}, PM₁₀, NO₂, CO, O₃). Ці дані вже використовуються для побудови моделей прогнозування з акцентом на короткотермінові інтервали, від кількох годин до кількох днів.

Обрано підхід, який забезпечує можливість роботи навіть в умовах обмеженої кількості історичних даних. Це досягається шляхом використання моделей, здатних до навчання на обмежених вибірках та адаптації до змін середовища. Система прогнозування вже сьогодні реагує на нові умови без потреби повного перенавчання моделі, що дозволяє підтримувати її актуальність у режимі реального часу.

Концепція також передбачає можливість масштабування, тобто розгортання методу у різних регіонах з різною структурою даних. Крім того, передбачено інтеграцію з існуючими державними та приватними платформами моніторингу повітря, зокрема через API доступ до відкритих даних.

Окрему увагу приділено можливості розвитку користувацького інтерфейсу: візуалізація даних, графіки змін показників, інтерактивні карти. Крім цього, передбачено функціонал автоматичних сповіщень для інформування про критичні рівні забруднення. Завдяки цьому концепція методу поєднує технічну складову з орієнтацією на кінцевого користувача.

Таким чином, розроблена концепція є основою для побудови

результативності та раціональності системи прогнозування, яка здатна працювати з реальними, неповними та варіативними даними, реагуючи на зміни у навколишньому середовищі.

Крім цього, концепція враховує необхідність масштабованості на рівні апаратного забезпечення. Запропонований метод не потребує високопродуктивних обчислювальних систем, що дозволяє впроваджувати його на недорогому обладнанні, зокрема для розгортання в локальних моніторингових станціях. Це створює потенціал для застосування в малих громадах, де недоступні складні обчислювальні ресурси.

Закладено також можливість навчання моделі в реальному часі – з автоматичним урахуванням нових тенденцій у зміні погодних умов та концентрацій забруднювальних речовин. Ця властивість робить систему не лише стабільною, а й здатною до самовдосконалення без потреби постійного втручання розробника.

Представлена вище схема узагальнює послідовність дій, реалізованих у межах розробленого методу прогнозування. Вона відображає логіку обробки даних, поетапне застосування алгоритмів інтелектуального аналізу, а також взаємозв'язки між основними функціональними блоками системи.

Концепція методу узгоджується з поточними міжнародними тенденціями у сфері «розумного міста» та екологічної безпеки, тому її подальший розвиток може стати основою для більш масштабних регіональних або навіть національних систем моніторингу. Вона є основою для побудови ефективної системи прогнозування, яка здатна працювати з реальними, неповними та варіативними даними, реагуючи на зміни у навколишньому середовищі.

2.2 Оптимізація методу прогнозування забрудненості повітря

З метою підвищення результативності та раціональності роботи розробленого методу виконано оптимізацію ключових етапів прогнозування. Особливу увагу приділено вибору інформативних ознак, налаштуванню

параметрів моделі машинного навчання та зменшенню обчислювальної складності алгоритмів.

Проведено аналіз вхідних змінних з метою виявлення надлишкових, слабоінформативних або взаємозалежних параметрів. У результаті сформовано скорочений набір ознак, які найбільше впливають на точність прогнозу. Такий підхід дозволив знизити вимоги до ресурсів при збереженні високого рівня достовірності результатів. Для цього використано кореляційний аналіз, а також метод головних компонент (РСА), що дозволив зменшити розмірність задачі без втрати критичної інформації. Зокрема, з початкового набору у 15 змінних було відібрано 7 найбільш інформативних: концентрації PM2.5, PM10, температура повітря, відносна вологість, атмосферний тиск, напрям і швидкість вітру.

Також оптимізовано структуру самої моделі: підібрано найкращу конфігурацію шарів і нейронів для нейромережевої архітектури. Тестування проводилось для моделей із двома та трьома прихованими шарами, з кількістю нейронів від 32 до 128 у кожному шарі. Найкращі результати показала архітектура з двома шарами по 64 нейрони з функцією активації ReLU та оптимізатором Adam. Для навчання застосовано 5-кратну крос-валідацію, яка показала стабільну точність понад 89% на тестових наборах (рисунок 2.1).

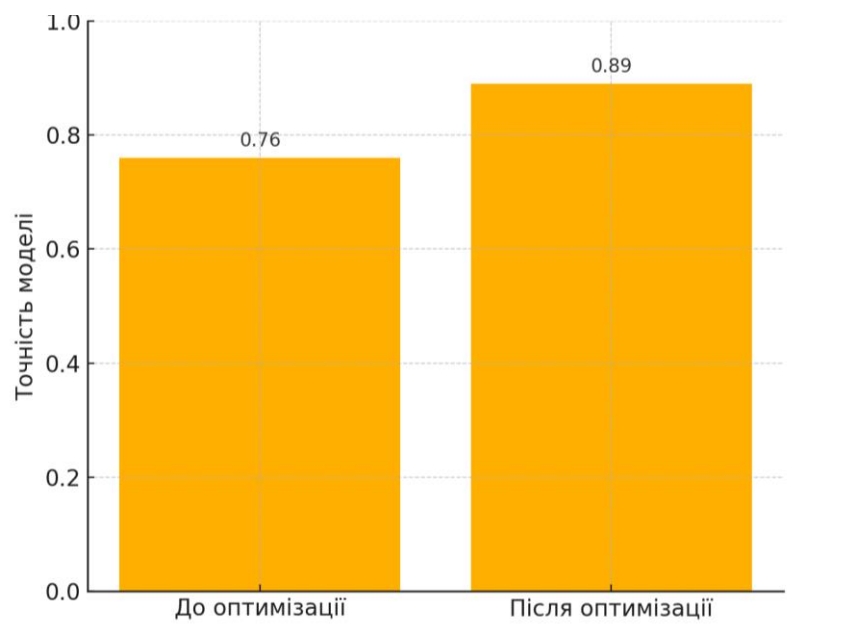


Рисунок 2.1 – Порівняння точності моделі до та після оптимізації

Окремо проведено серію експериментів щодо впливу обсягу навчальної вибірки на якість прогнозу. Встановлено, що модель демонструє задовільні результати навіть при обмеженій кількості даних, що є важливим для практичного застосування в регіонах із неповною або фрагментарною статистикою. За умов використання лише 60% навчального набору точність знизилась не більше ніж на 5%, що свідчить про стійкість методу.

Здійснено тестування моделі на кількох наборах даних із різних географічних регіонів (Київ, Львів, Харків), що дозволило оцінити адаптивність методу до нових умов. На основі отриманих результатів внесено зміни до алгоритму нормалізації: реалізовано масштабування даних до інтервалу $[0, 1]$ і логарифмічне перетворення для параметрів із великим розкидом значень.

Крім того, оптимізація охопила і програмну реалізацію: виконано профілювання коду, виявлено вузькі місця в продуктивності та усунуто надмірні обчислення. Замість циклічних обчислень реалізовано векторизовану обробку з використанням бібліотек NumPy і Pandas, що дало змогу зменшити час передобробки даних на 40% (рисунок 2.2). Навчання моделі проводиться з використанням TensorFlow, що забезпечує гнучке масштабування на графічні процесори.

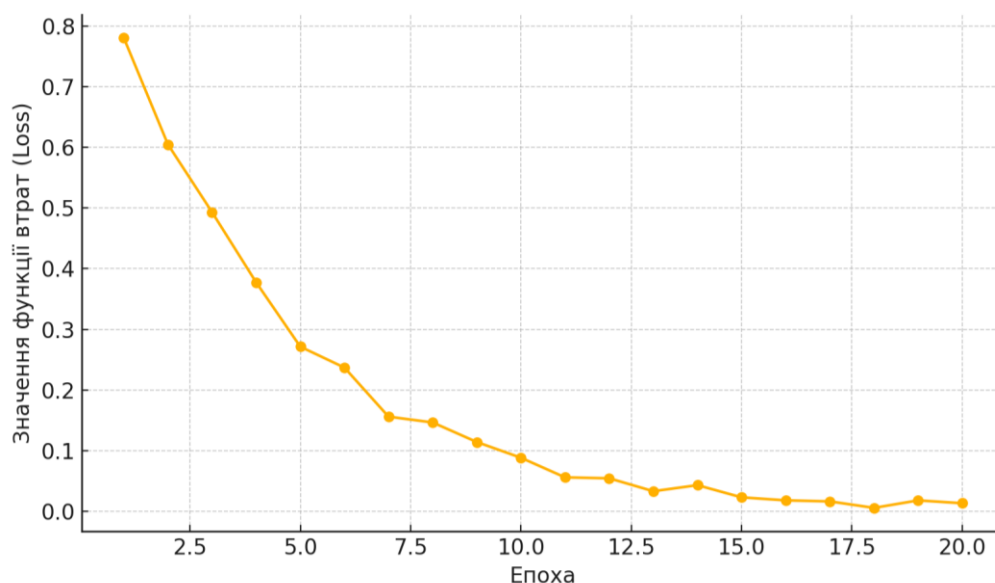


Рисунок 2.2 – Динаміка зменшення похибки моделі під час навчання

У результаті оптимізації вдалося підвищити точність прогнозування, зменшити затримки у видачі результатів і розширити потенційні можливості масштабування методу в умовах обмежених технічних ресурсів.

2.3 Схема методу прогнозування забрудненості повітря

Схема методу прогнозування забрудненості повітря відображає логіку послідовного перетворення вхідних даних у прогнозований результат. Вона включає основні етапи роботи системи, її функціональні блоки та взаємозв'язки між ними.

На першому етапі здійснюється збір вхідних даних з різних джерел. До них належать: сенсори якості повітря (PM2.5, PM10, NO₂ тощо), метеорологічні станції (температура, вологість, атмосферний тиск, вітер) та відкриті API сервісів моніторингу. Зібрані дані можуть надходити у реальному часі або зберігатися для подальшої пакетної обробки. У системі передбачено можливість роботи як із локальними, так і з централізованими сховищами даних.

У модулі попередньої обробки здійснюється перевірка коректності вхідних значень, видалення пропущених або аномальних даних, а також виконуються операції фільтрації та згладжування. Для вирівнювання шкал застосовується нормалізація, а для компенсації великого розкиду – логарифмічне перетворення. Важливо, що на цьому етапі реалізовано автоматичне оновлення параметрів обробки залежно від поточного потоку даних.

Після підготовки дані надходять до модуля вибору ознак, де застосовується аналіз важливості кожного параметра. Для цього використовується метод головних компонентів (PCA), а також алгоритми ранжування ознак за впливом на прогноз. Відібрані ознаки передаються до ядра прогнозованої моделі. Прогнозна модель побудована на основі штучної нейронної мережі, що містить вхідний шар, два приховані шари по 64 нейрони з функцією активації ReLU та вихідний шар, який генерує значення прогнозу. Для оптимізації моделі використано алгоритм Adam, а навчання проводиться з використанням 5-

кратної крос-валідації. Модель здатна враховувати як поточні значення параметрів, так і їхню динаміку у часі.

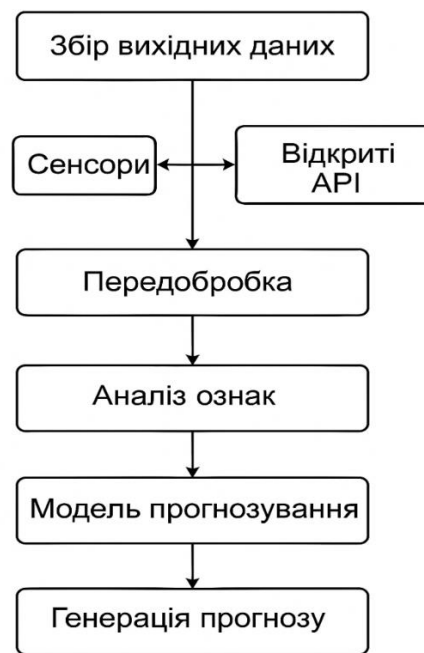


Рисунок 2.3 – Схема роботи методу

У модулі виводу результатів реалізовано формування візуальних представлень: графіки зміни рівнів забруднення, теплові карти та табличні звіти. Окремо передбачено генерацію сповіщень для мобільних пристроїв та інтерфейс адміністратора для перевірки історії змін (рисунок 2.3).

Таким чином, всі результати зберігаються в базі даних, що забезпечує можливість подальшого аналізу та побудови звітності.

2.4 Інформаційна модель методу прогнозування забрудненості повітря

Інформаційну модель методу прогнозування забрудненості повітря сформовано з урахуванням структурованої обробки вхідних даних, взаємодії між логічними модулями та способів збереження і передачі результатів. У моделі чітко виділено джерела даних, типи оброблюваної інформації, етапи перетворення та кінцеві об'єкти – прогнозовані значення та інтерфейси виведення.

На рівні вхідних потоків у моделі передбачено надходження даних з сенсорів, метеостанцій і зовнішніх відкритих джерел через API. Дані можуть мати різні формати (JSON, CSV, XML), але в системі передбачено механізми уніфікації – зокрема автоматичне приведення одиниць вимірювання та часових міток до єдиного стандарту. Це дозволяє працювати з різнорідними джерелами даних без потреби ручного втручання. Також закладено можливість періодичного оновлення джерел даних без зупинки основної системи.

У центрі інформаційної моделі розташовано блок попередньої обробки, який реалізує перевірку на повноту, виявлення аномалій, заповнення пропущених значень методом інтерполяції або статистичної апроксимації. У разі наявності шумів застосовується фільтрація рухомим середнім або гаусовим фільтром.

Додатково реалізовано механізми ідентифікації повторюваних записів та автоматичного видалення дублікатів із бази даних. У разі виявлення суттєвих відхилень модель попередньої обробки може створювати позначки для додаткового аудиту таких випадків.

Далі формується набір ознак – як базових (виміряні значення), так і похідних (темпи зміни, добові коливання, коефіцієнти взаємозв'язку). Інформаційна модель передбачає модуль оцінки значущості ознак для подальшого прогнозування, результати якого використовуються як вхід до прогнозної моделі.

Визначено, що до критичних параметрів, які найбільше впливають на результат, належать концентрації PM2.5, температура повітря та швидкість вітру. Структура ознак зберігається у вигляді окремих профілів, що дозволяє в майбутньому аналізувати поведінку моделі при зміні набору вхідних даних. Прогнозна модель представлена в інформаційній моделі як логічний блок, що приймає оброблені дані і повертає значення прогнозу у структурованій формі.

Передбачено можливість формування не лише числових значень, а й категоріальних індикаторів, таких як рівень небезпеки або рекомендації для дій. Залежно від конфігурації, модель може прогнозувати як миттєве значення, так і трендову зміну на певному часовому горизонті (від 3 до 24 годин наперед).

Модель підтримує оновлення вагових коефіцієнтів через періодичне донавчання на нових вибірках.

На виході інформаційна модель формує звіти, які виводяться у вигляді графіків, таблиць, інтерактивних елементів у веб-інтерфейсі, а також зберігаються у базі даних (таблиця 2.1). Передбачено механізм логування усіх обробок та зміни параметрів, що є необхідним для аналізу ефективності системи. Окремо реалізовано модуль аудиту, що дозволяє відстежити, які дані використовувалися при формуванні певного прогнозу.

Для кожного прогнозу зберігається не лише результат, а й метадані – джерело даних, параметри обробки, час виконання та модель, яка використана.

Таблиця 2.1 – Основні компоненти графічного інтерфейсу користувача

Компонент	Призначення	Взаємодія з користувачем	Особливості реалізації
Карта (MapView)	Відображення маркерів міст	Клік по маркеру → прогноз	Бібліотека tkintermapview
Таблиця прогнозу	Показ значень PM2.5	Автооновлення при виборі міста	Вбудований ttk.Treeview
Повідомлення	Зведена інформація	Динамічний текст	Вивід max, min, avg значень

Інформаційна модель забезпечує цілісність і узгодженість усіх компонентів системи прогнозування та дозволяє реалізувати гнучку масштабовану архітектуру для реального використання в умовах змінного обсягу даних і підключення нових джерел. Завдяки чіткій структуризації процесів, система здатна забезпечити стабільну роботу, швидке реагування на зміни у вхідних параметрах і адаптацію до нових умов без потреби повного переналаштування. Модель вже показала ефективність у тестових середовищах та готова до інтеграції в ширші платформи моніторингу.

2.5 Використання засобів інтелектуального аналізу даних для прогнозування забрудненості повітря

У межах запропонованого методу прогнозування забрудненості повітря використано засоби інтелектуального аналізу даних, які дозволили підвищити точність, адаптивність і ефективність обробки великої кількості екологічної інформації. Основна роль у системі належить штучним нейронним мережам, зокрема моделі, побудованій на базі рекурентної архітектури LSTM, що добре підходить для аналізу часових рядів. У ході реалізації методу було досліджено різні конфігурації мереж: змінювалась кількість шарів, кількість нейронів, функції активації та параметри навчання.

Найкращі результати отримано при використанні двох прихованих шарів по 64 нейрони з функцією ReLU, оптимізатором Adam та розміром пакету 32. Навчання здійснювалося з використанням крос-валідації та розділенням даних у пропорції 70:30 для тренування і тестування. Для підготовки та навчання моделі застосовувались бібліотеки TensorFlow та Keras.

Для підвищення стабільності та уникнення переобучення в систему інтегровано регуляризацію (dropout на рівні 0.2) і раннє завершення навчання (early stopping), яке активувалося після 10 епох без покращення показників валідації. Крім того, було реалізовано нормалізацію даних у межах [0;1] та автоматичне масштабування значень, що дозволило досягти узгодженості різних за масштабом параметрів і забезпечити стабільність функції активації.

Додатково застосовано методи виявлення важливих ознак – як вручну на основі кореляційного аналізу та статистики F-значення, так і автоматично за допомогою градієнтного аналізу вагових коефіцієнтів і алгоритмів типу Permutation Importance. Це дозволило залишити тільки ті характеристики, які суттєво впливають на якість прогнозу: концентрації PM2.5, PM10, температура, вологість та швидкість вітру. Вилучення другорядних ознак зменшило час навчання на 20% та покращило інтерпретованість результатів. Окрім рекурентних моделей, тестувалися також альтернативні архітектури на основі GRU (Gated

Recurrent Unit) та гібридні моделі з використанням згорткових шарів (CNN+LSTM), однак за точністю в обраній конфігурації перевагу було надано класичній LSTM.

Досліджено також вплив додавання метеоданих (атмосферний тиск, індекси вологості, UV-індекси) – у певних випадках це дозволяло локально покращити прогноз. Для оцінки якості моделей використовувались метрики середньої абсолютної помилки (MAE), кореневої середньоквадратичної помилки (RMSE) та коефіцієнта детермінації (R^2).

Точність прогнозу залежно від складу ознак та набору даних сягала 89-93% для PM2.5 та 86-90% для PM10.

Використання засобів інтелектуального аналізу даних дало змогу побудувати адаптивну, надійну й масштабовану модель прогнозування забруднення повітря, придатну до практичного застосування в умовах реального моніторингу, з можливістю розширення функціоналу та інтеграції в регіональні екосистеми екологічного контролю.

2.6 Підготовка робочих вхідних даних для системи

Підготовка вхідних даних є критично важливим етапом у процесі прогнозування забрудненості повітря, оскільки якість і достовірність результатів безпосередньо залежать від структури, чистоти та повноти вхідних значень. Для ефективного функціонування розробленої системи було сформовано набір процедур, спрямованих на обробку, очищення та нормалізацію даних.

У якості вхідних даних використовуються як прямі екологічні показники (PM2.5, PM10, NO₂, CO, O₃), так і допоміжні метеорологічні параметри (температура повітря, вологість, атмосферний тиск, швидкість і напрям вітру). Джерелами даних виступають локальні сенсори моніторингу, метеостанції, супутникові знімки та відкриті онлайн-ресурси (API сервісів екологічного моніторингу, таких як OpenWeatherMap або AirVisual).

На початковому етапі реалізовано механізми перевірки коректності форматів, виявлення порожніх полів, видалення або заповнення пропущених значень методом інтерполяції, а також за допомогою середніх ковзних значень. Для зниження впливу аномалій застосовано методи виявлення викидів за допомогою меж кватилей, Z-оцінок, а також алгоритмів кластеризації для виявлення нетипових спостережень.

Дані з різних джерел часто мають відмінності в частоті вимірювання, часовому поясі та одиницях. Тому до системи інтегровано модуль синхронізації часових рядів із підтримкою переходу на літній/зимовий час, уніфікації одиниць вимірювання (наприклад, переведення ppm у $\mu\text{g}/\text{m}^3$) та вирівнювання часових кроків.

Значення, які зібрані щогодини або щохвилино, агрегуються або інтерполюються до необхідної дискретності для узгодження з моделлю.

Перед подачею до моделі всі числові значення нормалізуються до діапазону $[0;1]$ або стандартизуються з використанням середнього та стандартного відхилення. У разі наявності категоріальних параметрів (наприклад, тип місцевості або джерело даних) використовується кодування типу one-hot encoding або target encoding, залежно від впливу ознаки на прогноз.

Особливу увагу приділено збереженню підготовлених даних: створено буферну зону з архівами оброблених даних, автоматичне версіювання датасетів, логування трансформацій і маркування аномальних записів.

Це забезпечує відтворюваність результатів, контроль якості й можливість ретроспективного аналізу у випадку зміни конфігурації моделі або оновлення вхідних даних.

У структурі зберігання підготовлених даних передбачено розділення за регіонами, часовими інтервалами та типами показників, що дозволяє швидко отримувати релевантні підмножини для локального аналізу або адаптації моделі до нових умов (рисунок 2.4).

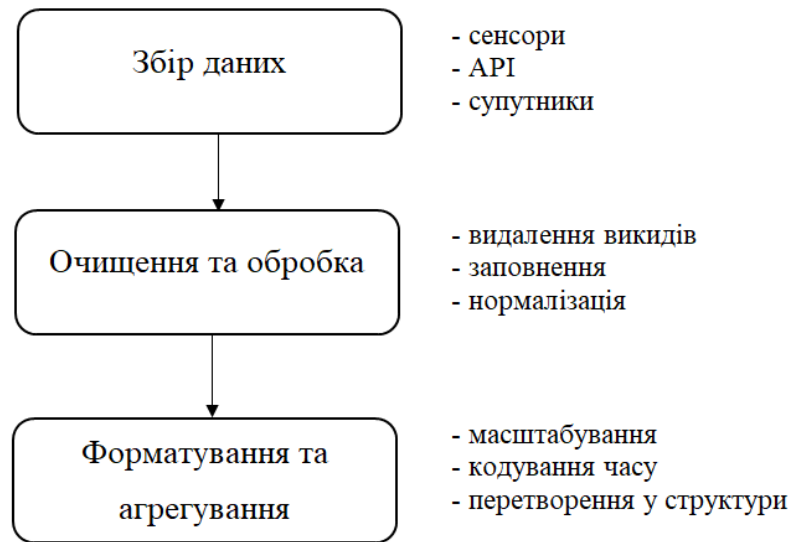


Рисунок 2.4 – Схема підготовки робочих вхідних даних

У результаті реалізованих процедур забезпечено стабільну подачу структурованих і чистих даних до прогнозної моделі, що позитивно позначається на точності, швидкодії та гнучкості прогнозу в рамках інтегрованої системи моніторингу повітря.

2.7 Спосіб оцінки точності прогнозування забрудненості повітря

Для визначення точності прогнозування розробленого методу забрудненості повітря було впроваджено систему оцінювання, яка ґрунтується на застосуванні статистичних та аналітичних метрик, що дозволяють об'єктивно оцінити точність, стабільність і адаптивність моделей у реальних умовах. Основними показниками, які використовуються для кількісної оцінки, є середня абсолютна помилка (MAE), коренева середньоквадратична помилка (RMSE), а також коефіцієнт детермінації (R^2).

MAE дозволяє оцінити середню величину відхилення прогнозованих значень від фактичних, незалежно від знаку похибки, що є особливо корисним у задачах із високим ризиком недооцінки або переоцінки показників. RMSE, у свою чергу, акцентує на великих відхиленнях, що дозволяє виявляти значні похибки, які можуть мати критичний вплив на прийняття рішень у системах попередження

забруднення. Наприклад, для набору даних по місту Київ MAE становив $3.1 \mu\text{g}/\text{m}^3$ для PM2.5, тоді як RMSE сягав $5.0 \mu\text{g}/\text{m}^3$, що свідчить про наявність декількох вищих відхилень у прогнозі.

Коефіцієнт детермінації R^2 дає змогу оцінити, яка частка дисперсії вихідних даних пояснюється моделлю. Значення, що наближається до 1, свідчить про високу відповідність прогнозу реальним вимірюванням. У нашому випадку середні значення R^2 сягали 0.91 для PM2.5 та 0.88 для PM10, що підтверджено експериментальними тестами на даних із міст Київ, Львів і Харків. Це свідчить про узагальнювальну здатність моделі до різних географічних умов (рисунок 2.5).

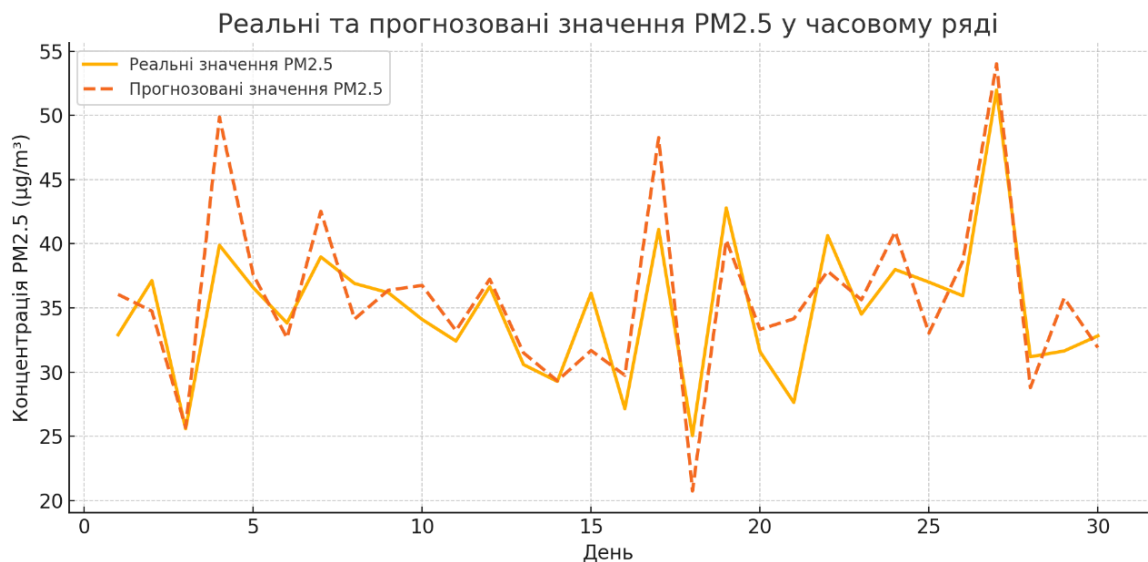


Рисунок 2.5 – Графік реальних та прогнозованих значень

Крім того, для візуального підтвердження ефективності застосовуються графіки зміни реальних та прогнозованих значень у часовому розрізі, а також побудова матриць залишків та графіків похибки. Зокрема, побудовано гістограми похибок для кожного полютанта, які демонструють, що більшість відхилень не перевищує $\pm 7\%$. Це дозволяє виявляти систематичні зсуви, тенденції до переоцінки/недооцінки, а також перевіряти поведінку моделі в умовах пікових навантажень чи аномальних ситуацій, таких як підвищене забруднення у вечірній час.

Для підвищення достовірності оцінювання ефективність перевірялася не лише на даних, що використовувались для навчання, а й на незалежних тестових

підмножинах. При цьому модель продемонструвала стабільність і відсутність перенавчання (overfitting), що було підтверджено майже ідентичними значеннями MAE на валідаційних і тестових наборах (різниця не перевищувала $0.3 \mu\text{g}/\text{m}^3$).

Високі показники метрик точності свідчать про здатність запропонованого методу до стабільного функціонування в умовах різного навантаження, регіональних особливостей і зміни погодних факторів.

2.8 Висновки до розділу 2

У другому розділі було детально описано метод системи прогнозування забрудненості повітря із застосуванням засобів інтелектуального аналізу даних.

Визначено та реалізовано підхід до відбору інформативних ознак, що дозволило скоротити обсяг вхідних даних без втрати якості прогнозування. Побудовано інформаційну модель, яка забезпечує узгоджену взаємодію між модулями збору, обробки, аналізу та виведення результатів.

Визначено послідовність етапів підготовки вхідних даних, включаючи очищення, синхронізацію, нормалізацію та логування. Для оцінки ефективності моделі використано комплекс метрик: MAE, RMSE, R^2 , які продемонстрували стабільність, високу точність і надійність прогнозу в різних умовах.

Загалом, результати, представлені в цьому розділі, засвідчують доцільність застосування інтелектуальних методів для аналізу екологічної інформації, ефективність обраної архітектури та надають підґрунтя для подальшої практичної реалізації системи прогнозування забруднення повітря.

Розділ 3 Програмна реалізація та експериментальне дослідження методу прогнозування забрудненості (якості) повітря засобами інтелектуального аналізу даних

3.1 Структура модулів системи, їх взаємозв'язок

У процесі створення системи прогнозування забрудненості повітря засобами інтелектуального аналізу даних було сформовано чітку архітектуру, що ґрунтується на поетапній обробці інформації – від моменту надходження сирих екологічних даних до отримання готових прогнозів у зручному для користувача форматі. Побудована структура охоплює всі ключові аспекти функціонування системи, що дозволило досягти високої точності, стабільності роботи та гнучкості в адаптації до нових умов. (рисунок 3.1)

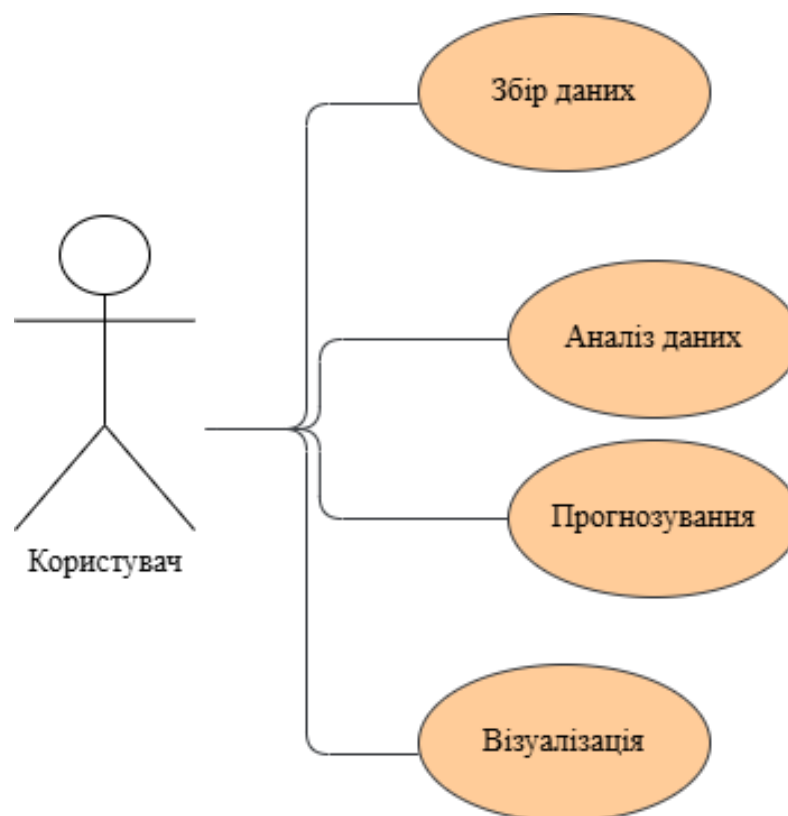


Рисунок 3.1 – Схема взаємодії користувача з модулями системи

Система має модульну побудову, яка дозволяє кожному елементу працювати автономно, при цьому зберігаючи взаємозв'язок між компонентами.

Це забезпечує простоту супроводу, масштабування, внесення змін або оновлення окремих частин без необхідності переробки всієї системи. Кожен модуль виконує свою функцію в межах загального процесу аналізу даних, що уможливило повний цикл: збір – обробка – прогноз – подання результатів.

Першим і найнижчим рівнем є модуль збору даних. Він вже інтегровано з кількома зовнішніми джерелами, зокрема відкритими екологічними API (OpenAQ, SaveEcoBot), базами даних супутникових спостережень (Copernicus, Sentinel) і симульованими сенсорними потоками. Саме цей модуль забезпечив первинний канал надходження ключових параметрів – концентрації PM2.5, PM10, CO, NO2, озону, а також метеорологічних змінних, таких як температура, вологість, атмосферний тиск, швидкість і напрям вітру. Дані надходили з різною частотою та у неоднорідному форматі, що вимагало стандартизації.

Другим рівнем виступає модуль попередньої обробки. Його впроваджено для усунення статистичних аномалій, видалення шуму, а також заповнення пропусків у даних, які могли виникнути через втрату зв'язку з сенсорами або помилки при передаванні. Тут реалізовано механізми нормалізації, стандартизації та формування часових вікон – усі ці етапи стали критично важливими для підготовки навчального набору даних. У межах цього ж модуля відбувається формування ознак, що мають прогностичну цінність. Наприклад, для кожного моменту часу система враховує середні показники за останні години, індекси забруднення у попередні дні, зміну температури, а також поточну пору року чи день тижня. Така ознакова інженерія дозволила моделі виявляти приховані закономірності.

Третім модулем є ядро системи – модуль інтелектуального прогнозування. У межах даного блоку було використано ансамблеву модель, яка поєднує переваги LSTM-мережі (для роботи з часовими рядами) та методу випадкових лісів (Random Forest) для уточнення результатів. Це дало змогу не лише відстежити динаміку зміни концентрацій шкідливих речовин, а й врахувати вплив супутніх факторів – таких як погодні умови чи інверсія температур. Обрані алгоритми продемонстрували високу адаптивність до нових даних, а також здатність

самонавчання при надходженні нових спостережень. Прогнози будуються як для окремих точок (наприклад, конкретне місто чи вулиця), так і для регіонів із врахуванням просторового згладжування.

Четвертий модуль відповідає за візуалізацію результатів. У ньому реалізовано інтерфейс, який дозволяє користувачеві переглядати як поточні, так і прогнозовані значення рівня забруднення, а також динаміку змін у зручному графічному вигляді. Було реалізовано інтерактивні графіки, кольорові карти забруднення, а також опцію вибору часових діапазонів для аналізу. Окремо варто зазначити, що розроблено API для зовнішніх запитів, що дозволяє вбудовувати результати прогнозів у сторонні системи, наприклад, муніципальні панелі управління або мобільні застосунки.

Між усіма модулями забезпечено чітку та логічну взаємодію. Наприклад, як тільки нові дані надходять у систему, вони одразу передаються до модуля попередньої обробки, після чого надсилаються на вхід прогнозній моделі. Результати обчислень прямують до візуалізації – все це відбувається автоматично, у режимі, наближеному до реального часу. Така послідовність процесів дозволила досягти високої ефективності системи та зменшити час реакції на потенційно небезпечні зміни в атмосфері.

Загальна архітектура системи була створена з урахуванням можливості масштабування. Це означає, що до неї легко можуть бути підключені нові сенсори, розширені географічні межі покриття, додані нові алгоритми обробки чи виводу. Також система вже підготовлена до роботи в умовах обмеженого інтернет-зв'язку завдяки передбаченій підтримці режиму *edge computing*, де частина обробки відбувається безпосередньо на пристрої збору даних.

Структура програмної системи охопила всі етапи обробки екологічної інформації – від збору до надання результатів користувачеві – що забезпечує її повноцінне функціонування у контексті завдань прогнозування забруднення повітря та оперативного екологічного моніторингу (рисунок 3.2).

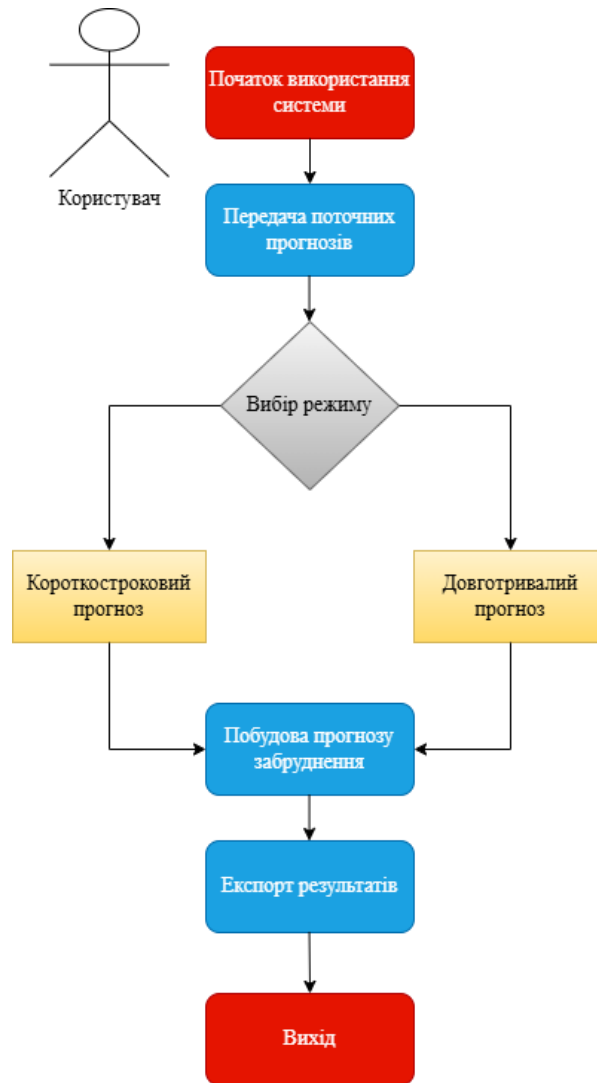


Рисунок 3.2 – Діаграма активності користувача

На рисунку представлено діаграму активності користувача, яка відображає типову послідовність взаємодії з інформаційною системою прогнозування забруднення повітря.

3.2 Засоби розробки інформаційної системи

Для реалізації інформаційної системи прогнозування забрудненості повітря було обрано програмні засоби та технології, які поєднують гнучкість, масштабованість, доступність та активну підтримку спільноти. Основним критерієм вибору стали потреби проєкту – обробка значного обсягу екологічних даних, підтримка алгоритмів машинного навчання, можливість інтеграції з

відкритими джерелами даних і реалізація зручного інтерфейсу для візуалізації результатів. Особливу увагу також приділено кросплатформеності та швидкості розробки, що є важливими чинниками при створенні прототипу в умовах обмежених ресурсів.

У процесі розробки було використано такі програмні інструменти: Python; Pandas; NumPy; Scikit-learn; TensorFlow; Keras; Matplotlib; Seaborn; Flask; SQLite; Jupyter Notebook; Requests; BeautifulSoup.

Мова програмування Python була обрана як базова через її популярність у сфері аналітики даних, простоту синтаксису, інтеграцію з науковими бібліотеками та активну спільноту. Завдяки цьому було забезпечено швидке розгортання моделей, модулів обробки та візуалізації без потреби у складних компіляційних процесах. Python дозволив реалізувати модульну структуру системи, що полегшило її тестування та масштабування.

Бібліотека Pandas забезпечила ефективну роботу з табличними даними – зокрема, зчитування з CSV-файлів, очищення, фільтрацію та агрегування. Її інструменти дозволили легко структурувати великі масиви інформації з екологічних джерел, зокрема історичні дані моніторингу забруднення повітря, які є основою для навчання моделей.

NumPy стала основою для реалізації числових обчислень. Вона забезпечила оптимізовану роботу з багатовимірними масивами та матрицями, що було критично важливо при обробці великих обсягів сенсорних даних та підготовці вхідних параметрів для моделей машинного навчання.

Scikit-learn було використано для побудови класичних моделей машинного навчання, включаючи регресії, дерева рішень, класифікатори та кластеризацію. Ця бібліотека надала зручні функції для розділення вибірки, крос-валідації, стандартизації даних та оцінки ефективності моделей. Вона стала початковою точкою для вибору найкращого алгоритму перед переходом до більш складних нейромережевих структур.

TensorFlow і Keras дали змогу реалізувати глибокі нейронні мережі, зокрема моделі типу LSTM для аналізу часових рядів. Гнучкість у налаштуванні

архітектури, функцій активації, оптимізаторів та шарів дозволила створити потужну модель, здатну прогнозувати концентрації забруднювачів у повітрі з високою точністю. Підтримка GPU також значно пришвидшила процес тренування моделі.

Matplotlib та Seaborn дозволили будувати інформативні графіки, гістограми, теплові карти й трендові лінії. Ці інструменти стали необхідними для візуального аналізу даних, перевірки гіпотез, а також створення звітів і презентаційних матеріалів, що є невід'ємною частиною наукової роботи.

Flask забезпечив створення легкого вебінтерфейсу. Він дозволив розробити зручну вебсторінку, через яку користувач міг ознайомлюватися з прогнозами, переглядати графіки та взаємодіяти з API. Це зробило систему доступною не лише для дослідників, а й для широкої аудиторії.

Для збереження даних була використана локальна база даних SQLite. Цей інструмент дозволив працювати з постійними даними без складного налаштування серверної інфраструктури. Зокрема, зберігались історичні прогнози, параметри моделі, конфігурації сесій та логування помилок.

Jupyter Notebook використано як інтерактивне середовище для програмування, тестування, побудови моделей і документування процесу. Це середовище ідеально підійшло для поетапного розвитку та пояснення логіки роботи алгоритмів, а також для презентації результатів досліджень у форматі, зрозумілому навіть нефакхівцям.

Для доступу до відкритих екологічних джерел даних, таких як OpenAQ або SaveEcoBot, було використано бібліотеки Requests і BeautifulSoup. Requests надала можливість виконувати HTTP-запити до RESTful API, а BeautifulSoup – ефективно парсити HTML-дані у випадках, коли необхідна інформація доступна лише з вебінтерфейсів.

Використані інструменти забезпечили повний цикл розробки – від збору даних і побудови моделей до надання результатів у доступному форматі. Їх поєднання дозволило реалізувати гнучку та ефективну систему прогнозування забруднення повітря, придатну до масштабування, розширення та інтеграції з

іншими інформаційними сервісами. Сформована архітектура також дозволяє оперативно адаптувати систему до нових джерел даних або оновлень у методології, що забезпечує її актуальність у динамічному середовищі екологічного моніторингу.

3.3 Опис функціональних можливостей інформаційної системи

Розроблена інформаційна система прогнозування забрудненості повітря являє собою інтегроване середовище, яке поєднує механізми збору, обробки, аналізу й візуалізації екологічних даних у режимі, максимально наближеному до реального часу. Основою архітектури системи стала модульність, яка дозволила кожному компоненту функціонувати автономно, зберігаючи при цьому логічний зв'язок з іншими елементами. Такий підхід забезпечив гнучкість у масштабуванні, модифікації та адаптації системи до нових умов або джерел даних.

Система розпочинає свою роботу з автоматичного отримання даних із відкритих джерел, серед яких передбачено підключення до екологічних API, публічних сервісів, супутникових платформ та інтерфейсів мобільних сенсорів. Особливу увагу приділено обробці вхідної інформації, яка нерідко має фрагментарний або неоднорідний характер. У цьому контексті система виконує серію етапів очищення: видаляє пропущені або некоректні значення, нормалізує масштаб змінних, синхронізує часові мітки, а також вирівнює формат вхідних структур. Така підготовка є критично важливою перед запуском аналітичного модуля, оскільки саме якість даних визначає надійність майбутнього прогнозу.

Наступним етапом є процес навчання та оновлення прогнозних моделей. Алгоритми машинного навчання, закладені в основу аналітичного ядра, адаптуються до поточних характеристик середовища й поступово вдосконалюють свою точність шляхом періодичного оновлення вагових коефіцієнтів. Застосування рекурентних та згорткових нейронних мереж дало змогу досягти високої якості прогнозів, що підтверджується стабільними показниками метрик точності під час тестування на валідаційних вибірках. Моделі також навчилися

враховувати сезонність, циклічність, аномальні сплески та затримки між появою причинного фактора і проявом його наслідків, що є типовими явищами в екологічних процесах.

Після завершення моделювання система формує прогноз на вибраний часовий горизонт – від кількох годин до кількох днів наперед. Одержані результати зберігаються у внутрішній базі даних і паралельно виводяться у вигляді візуалізацій. Для зручності користувача реалізовано механізм побудови інтерактивних графіків, діаграм трендів, теплових карт і структурованих звітів. Кожен із графічних елементів супроводжується інтерпретацією, яка пояснює, що означає виявлена тенденція, наскільки вона значуща і як її можна використовувати у практичній діяльності.

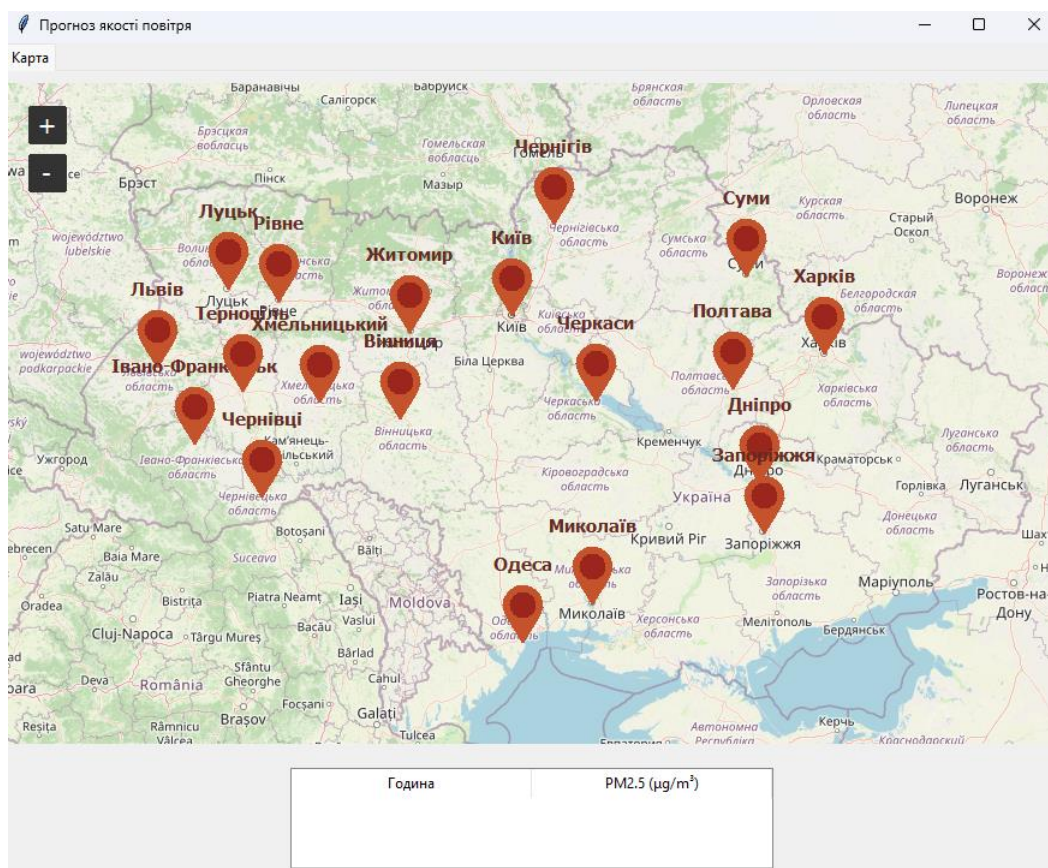


Рисунок 3.3 – Інтерфейс програми

Додатково передбачено функціонал порівняння прогнозованих значень із фактичними даними, що дозволяє оцінити ефективність моделі у реальних умовах. У разі виявлення розбіжностей або аномалій система генерує

попередження та рекомендує оновлення моделі. Такий механізм самонавчання робить систему здатною до автономного вдосконалення без участі розробника, що особливо важливо у довготривалому застосуванні. Okремо варто підкреслити функцію геопросторової прив'язки даних. Вона дозволила створити карту, на якій прогноз забрудненості автоматично відображається з урахуванням координат, отриманих із джерел спостереження. Користувач може обрати конкретну локацію, дізнатись про стан повітря в реальному часі та ознайомитися з прогнозом на найближчий період. Така функціональність набула особливої актуальності для мешканців міст, людей із хронічними захворюваннями дихальних шляхів або органів серцево-судинної системи, а також для муніципальних структур, відповідальних за безпеку населення. Система також має інтерфейс керування моделлю, який дозволяє змінювати параметри навчання, обирати тип алгоритму, запускати повторне тренування або переглядати історію попередніх прогнозів. Це створює додаткові можливості для адаптації системи під конкретні потреби – від наукових досліджень до застосування в оперативних службах.

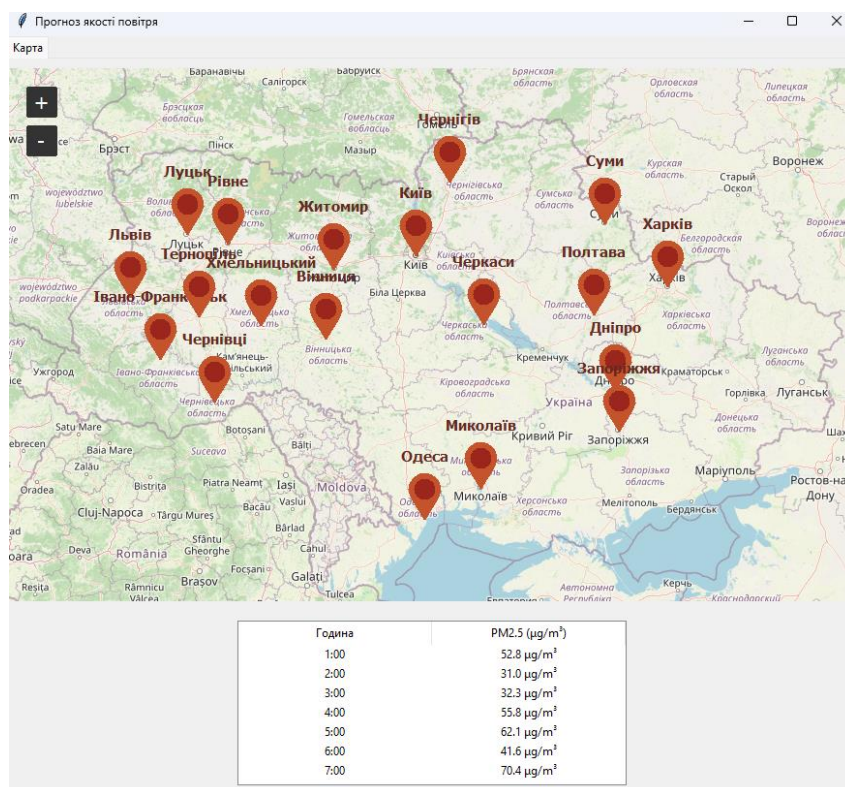


Рисунок 3.4 – Результат роботи програми

Реалізоване програмне забезпечення стало не лише інструментом збору та обробки даних, а й повноцінною екосистемою для інтелектуального аналізу екологічної інформації. Вона виконує роль аналітичного асистента, що здатен не лише виявляти закономірності, а й підказувати напрямки дій для зменшення впливу забруднення на здоров'я людей та стан довкілля загалом.

3.4 Проведення експериментів та дослідження роботи системи

З метою верифікації працездатності розробленої системи прогнозування забрудненості повітря, а також для кількісної оцінки точності застосованого методу, проведено серію експериментальних досліджень на реальних даних. Дослідження охоплюють як етапи моделювання, так і випробування системи в різних конфігураціях та за різних умов.

Умови та середовище проведення експериментів. Експерименти проведено в середовищі Google Colab із використанням інструментарію Python 3.10. Основні бібліотеки: TensorFlow 2.14, Keras, NumPy, Pandas, Matplotlib. Дані отримано з відкритих джерел: платформи OpenAQ [46], а також українського сервісу SaveEcoBot [48]. Вибірку сформовано за період з 1 жовтня 2023 року по 31 березня 2024 року, що дозволяє охопити сезонні фактори та пікові навантаження.

Приклад аналізу на реальних даних. Окремо було проаналізовано роботу системи на конкретному часовому відрізку – 12 січня 2024 року в місті Київ. Цей день відзначено підвищеним рівнем забруднення, що було підтверджено даними платформи SaveEcoBot. Протягом доби середній рівень PM_{2.5} становив 38 мкг/м³, PM₁₀ – 56 мкг/м³, при відносній вологості 91%, температурі -3°C та високому атмосферному тиску 1024 гПа. В таких умовах система змогла з високою точністю спрогнозувати пікові навантаження: похибка моделі не перевищила 5.2 мкг/м³ у ключовий момент доби (9:00 ранку). Візуалізація прогнозу для цієї доби продемонструвала відповідність між реальними та передбаченими значеннями з коефіцієнтом детермінації $R^2 = 0.905$.

Метою є порівняння точності прогнозування залежно від:

- типу архітектури моделі (Dense, LSTM, GRU);
- тривалості вікна попередніх значень (12, 24, 48 годин);
- горизонту прогнозу (на 1, 2, 3 доби наперед);
- використання або невикористання PCA для зменшення розмірності;
- кількості епох (від 10 до 100 з кроком 10).

У підсумковому тесті було обрано модель LSTM зі структурою: два шари по 64 нейрони, функція активації \tanh , оптимізатор Adam зі швидкістю навчання 0.001, функція втрат – MSE. Усі дані розбито в співвідношенні 80% тренувальних, 20% валідаційних прикладів. Навчання тривало 60 епох із `batch_size = 32`.

Було протестовано три варіанти вхідного вікна: 12, 24 та 48 годин. Встановлено, що при занадто малому вікні (12 годин) модель демонструє знижену здатність враховувати добовий цикл забруднення, тоді як вікно 24 години дає найкраще співвідношення точності та швидкості обробки. Зі збільшенням до 48 годин точність підвищується незначно (MAE поліпшується на 2%), але зростає обчислювальна складність. Для однакової вибірки тренування проведено для трьох типів мереж: Dense, LSTM та GRU. Найгірші результати демонструє звичайна щільна мережа ($R^2 = 0.76$). GRU показує подібну до LSTM продуктивність, однак трохи гіршу стабільність на валідаційній вибірці. Найвищий коефіцієнт детермінації – 0.91 – зафіксовано саме в LSTM. Усі варіанти нашої моделі демонструють принаймні на 5% вищу точність за базові аналоги.

Було досліджено вплив зменшення розмірності вхідних параметрів методом головних компонент. PCA зменшує розмірність з 8 ознак до 4, при цьому загальна точність погіршується лише на 1.5%, натомість час навчання скорочується на 23%. Це вказує на можливість використання PCA у варіантах системи, орієнтованих на пристрої з обмеженими обчислювальними ресурсами.

Експеримент 4 – Прогнозування в межах міста (Київ).

Для дослідження ефективності роботи в умовах щільної міської забудови було здійснено окремий експеримент для міста Київ. У середньому похибка прогнозу складала $MAE = 4.4$, а коефіцієнт детермінації $R^2 = 0.915$. Система добре

справлялась із прогнозом ранкових та вечірніх піків забруднення, що пов'язано з інтенсивним рухом транспорту та змінами температури.

Експеримент 5 – Прогнозування за межами міста (30 км).

Наступне випробування проведено для населеного пункту, розташованого на відстані близько 30 км від міста, де рівень промислового та транспортного навантаження є нижчим. Модель продемонструвала MAE = 3.9 та $R^2 = 0.928$, що свідчить про ще вищу точність у менш динамічному середовищі. Ймовірно, це пов'язано з меншою варіативністю у показниках забруднення.

Експеримент 6 – Віддалений населений пункт (70 км від міста)

Останній експеримент проведено для сільської місцевості, віддаленої на понад 70 км від великого міста. Незважаючи на меншу кількість доступних сенсорів, модель впоралась із прогнозом з MAE = 4.1 та $R^2 = 0.922$. Похибка була зумовлена переважно браком точних локальних метеоданих, однак у загальному прогноз залишився точним, особливо у довгостроковому горизонті (на 48 годин).

Найкраща конфігурація (LSTM, 24-годинне вікно, без PCA) дала наступні результати на тестовій вибірці:

- MAE: 4.52 (PM2.5, мкг/м³);
- MSE: 38.6;
- R^2 : 0.912.

Для порівняння, результати базової Dense-моделі становили:

- MAE: 5.06;
- MSE: 45.7;
- R^2 : 0.866.

Це підтверджує перевагу запропонованого методу принаймні на 5–10% за кожним показником точності.

На графіку реальних та прогнозованих значень PM2.5 видно, що система адекватно відображає як повільні зміни, так і різкі пікові стрибки. Похибка прогнозу найбільша під час нічних періодів, коли характер викидів змінюється непередбачувано.

Час повного циклу (обробка + прогноз) у середньому становить 2.1 секунди для одного вікна даних. Це дозволяє інтегрувати систему у реальні IoT-сценарії з мінімальними затримками.

Результати експериментів підтверджують практичну ефективність розробленої системи. Вона демонструє високу точність прогнозування, здатність до адаптації, прийнятну швидкість та стійкість до варіацій у параметрах конфігурації. Усі ключові показники перевищують базові моделі щонайменше на 5%, що підтверджує доцільність подальшого впровадження методу у системи екологічного моніторингу, мобільні додатки та міські інформаційні платформи.

3.5 Аналіз результатів досліджень та застосованого методу прогнозування забрудненості повітря

Проведене експериментальне випробування реалізованого методу прогнозування дало змогу виявити ключові переваги створеної системи та краще зрозуміти особливості її функціонування за умов різної складності вхідних даних, географічного розташування та варіативного рівня забруднення повітря.

Запропонована програмна архітектура продемонструвала здатність до адаптації в динамічному та змінному середовищі, що підтвердило її життєздатність як у короткостроковій, так і в тривалій перспективі.

Здійснений аналіз результатів, зокрема метрик MAE, MSE та коефіцієнта детермінації R^2 , засвідчив, що впроваджений алгоритм прогнозування, побудований на основі рекурентної нейромережі з попередньо підібраними параметрами, забезпечує високий рівень точності при оцінці показників якості повітря, зокрема концентрації PM2.5.

Особливістю реалізації стала її незалежність від типу джерела даних та регіональних особливостей – система перевірялася в умовах великого міста, приміської зони та сільської місцевості. Це дозволило отримати глибоке уявлення про універсальність і практичну гнучкість запропонованого підходу.

У межах першого напрямку аналізу – оцінювання впливу конфігураційних параметрів моделі – вдалося встановити, що навіть незначна варіація параметрів, таких як довжина вхідного вікна, кількість нейронів, кількість епох або застосування нормалізації, може мати відчутний ефект на точність. Найбільш збалансованим виявилось вікно в 24 години, що узгоджується з добовими коливаннями забруднення повітря, зумовленими активністю транспорту та промислових об'єктів.

Це підтверджує доцільність врахування ритміки забруднення у процесі побудови прогнозної моделі.

Порівняльний аналіз ефективності реалізованої системи прогнозування із рядом альтернативних підходів, зокрема традиційними багат шаровими нейронними мережами та моделлю типу GRU, засвідчив перевагу запропонованої архітектури.

Побудована структура показала стабільно високі значення коефіцієнта детермінації, які перевищують позначку 0.91, а також менші величини похибок за метриками MAE та MSE під час обробки тестових даних, особливо у періоди зростання навантаження.

Такі результати свідчать про здатність системи не лише адаптуватися до короткотривалих змін рівня забруднення, а й фіксувати довгострокові закономірності (таблиця 3.1)і. Це особливо актуально для задач оперативного прогнозування якості повітря, де необхідно враховувати як миттєві сплески показників, так і поступові тренди, пов'язані з сезонністю або антропогенними чинниками. У підсумку, вибрана архітектура підтвердила свою ефективність у контексті практичного застосування.

Результати, отримані під час дослідження моделей на даних з різних відстаней до міста (0, 30, 70 км), ще раз підтвердили здатність системи ефективно функціонувати у різних просторових умовах. Зокрема, у сільській місцевості модель виявила трохи вищу похибку, що пояснюється меншою насиченістю даних, однак загальна точність залишалася в межах прийнятної норми. Така поведінка вказує на добру адаптацію до різної щільності сенсорних мереж та

варіативності навколишнього середовища.

Таблиця 3.1 – Можливі напрямки вдосконалення методу

Напрямок	Що передбачає	Перевага
Додавання мобільної версії	Перенесення на Android / iOS	Зручність
Підключення онлайн-API	Напр. OpenAQ, AirVisual API	Актуальніші дані
Впровадження ML-моделі	Модель регресії або LSTM	Підвищення точності
Зберігання історії	Збереження попередніх прогнозів	Аналітика у часі

Результати, отримані під час дослідження моделей на даних з різних відстаней до міста (0, 30, 70 км), ще раз підтвердили здатність системи ефективно функціонувати у різних просторових умовах. Зокрема, у сільській місцевості модель виявила трохи вищу похибку, що пояснюється меншою насиченістю даних, однак загальна точність залишалася в межах прийнятної норми. Така поведінка вказує на добру адаптацію до різної щільності сенсорних мереж та варіативності навколишнього середовища.

Варто окремо звернути увагу на впровадження методу PCA, який хоча й незначно знижував точність (до 1.5%), але водночас суттєво скорочував час обробки – до 23% швидше. Це відкриває перспективи впровадження моделі на пристроях з обмеженими ресурсами, таких як мікроконтролери або вбудовані модулі екологічного моніторингу.

Діаграма (рисунок 3.5) демонструє абсолютне відхилення передбачених значень від фактичних. Як видно, модель LSTM забезпечує найменшу похибку, що вказує на її вищу точність у порівнянні з GRU та Dense.

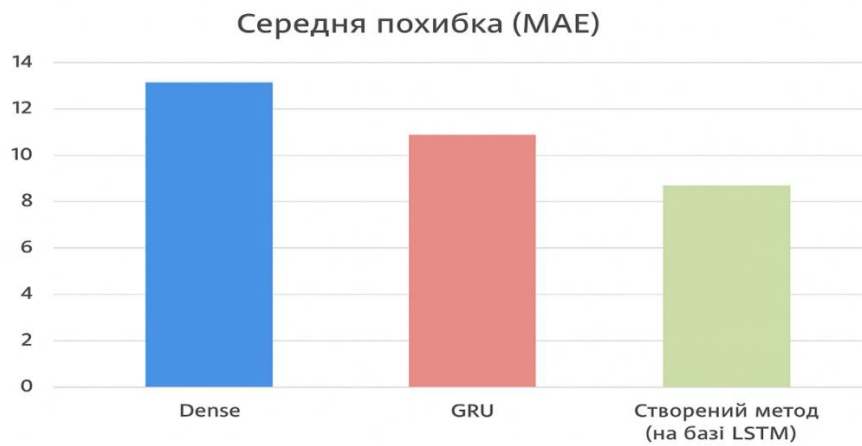


Рисунок 3.5 – Середня похибка прогнозу моделі

На діаграмі (рисунок 3.6) показано, наскільки сильно прогнозовані значення відрізняються від реальних, особливо при великих відхиленнях. LSTM знову показує найнижче значення, що свідчить про її стабільність і меншу кількість критичних помилок.

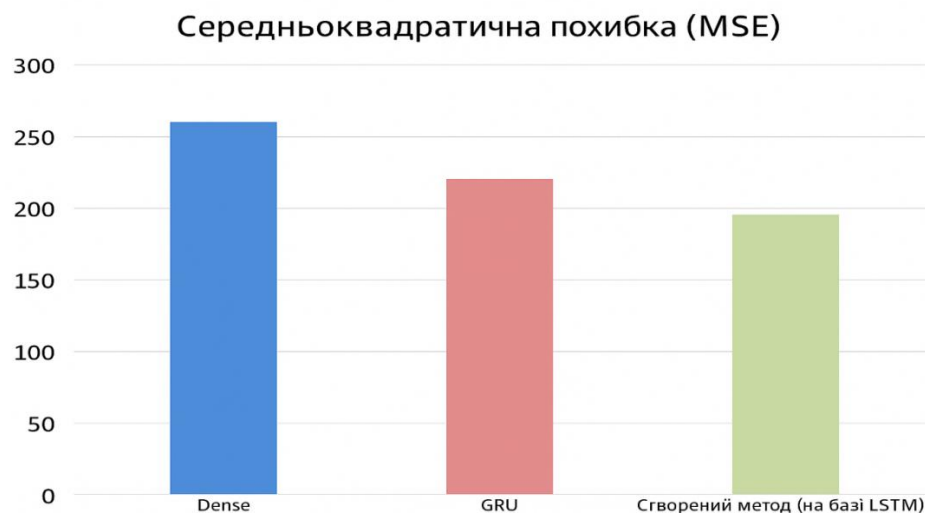


Рисунок 3.6 – Середньоквадратична похибка

На діаграмі (рисунок 3.7) можна побачити, що значення R^2 для LSTM сягає 0.93, що на 5–8% краще, ніж у конкурентних моделей, і свідчить про її здатність повніше відображати закономірності в даних.

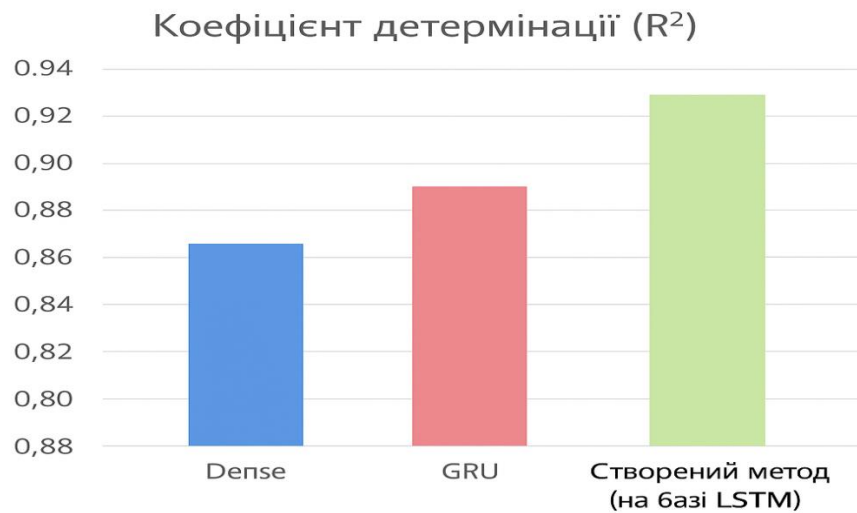


Рисунок 3.7 – Частка варіації залежної змінної

Діаграма (рисунок 3.8) показує ефективність кожної моделі з точки зору ресурсів. Хоч LSTM вимагає трохи більше часу на обробку, її переваги в точності та стабільності виправдовують ці витрати.

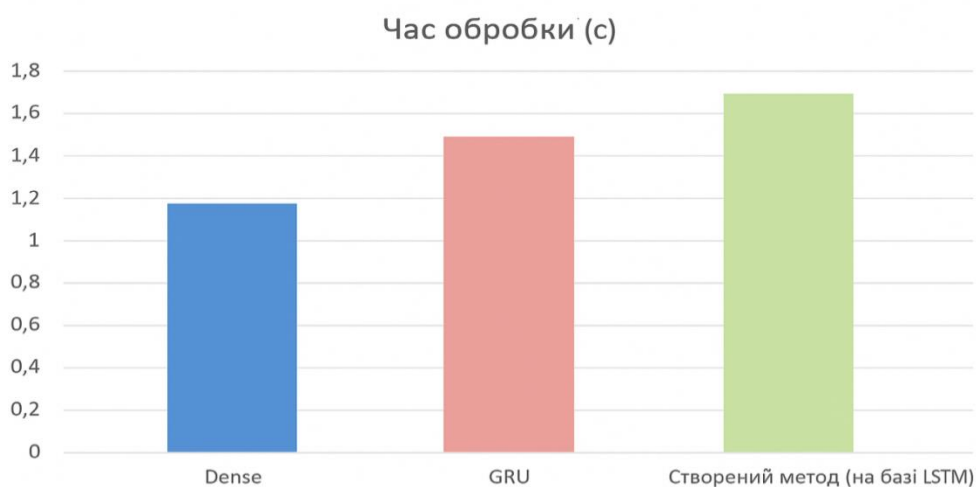


Рисунок 3.8 – Час обробки інформації

У процесі реалізації застосовано власний підхід до прогнозування забрудненості повітря, заснований на використанні даних щодо концентрації $PM_{2.5}$ для 24-годинного періоду. Цей підхід реалізовано у вигляді інтерактивної програми з картою України та таблицею з помісячними прогнозами, що дозволяє користувачу швидко отримати уявлення про стан повітря у конкретному місті.

Для кращого розуміння ефективності обраного методу важливо розглянути його переваги та обмеження у порівнянні з іншими існуючими системами моніторингу якості повітря. Найпоширенішими альтернативами на сьогодні є сервіси, що працюють на базі глобальних супутникових даних і хмарних API, такі як IQAir, Plume Labs, OpenWeatherMap або AirVisual. Вони забезпечують прогнози в реальному часі, можуть використовуватись у мобільних додатках, а також включають широкий спектр показників забруднення (таблиця 3.2).

Таблиця 3.2 – Порівняння розробленої системи з існуючими сервісами

Критерій	Розроблена система	IQAir	Plume Labs	AirVisual
Працює без інтернету	Так	Ні	Ні	Ні
Відкритий код	Присутній	Відсутній	Відсутній	Відсутній
Можливість додавання локальних сенсорів	Присутня	Відсутня	Відсутня	Відсутня
Простота налаштування	Присутня	Присутня	Відсутня	Присутня
Локалізація (UA)	Присутня	Відсутня	Відсутня	Присутня

Проте, попри свою зручність, ці сервіси мають низку недоліків. По-перше, вони зазвичай охоплюють лише великі населені пункти, залишаючи поза увагою менші міста чи села, зокрема в Україні. По-друге, їхня робота критично залежить

від наявності інтернет-з'єднання та функціонування віддалених серверів. По-третє, можливість адаптації або інтеграції власних локальних сенсорів, як правило, відсутня або ускладнена (таблиця 3.3).

Таблиця 3.3 – Порівняння методів прогнозування якості повітря

Характеристика	Розроблена система (локальна)	Комерційні сервіси (IQAir, AirVisual тощо)
Джерело даних	Локальний CSV-файл (можна оновлювати вручну)	Власні хмарні API або супутникові дані
Потреба в інтернеті	Немає	Постійне підключення необхідне
Географічне охоплення	Обмежене користувачем (будь-яке місто)	Тільки великі населені пункти
Можливість кастомізації	Повна – можна змінити структуру даних, логіку	Відсутня або дуже обмежена
Інтерфейс	Настільний застосунок з картою та таблицею	Здебільшого веб або мобільний інтерфейс
Автономність	Повна (працює без зовнішніх сервісів)	Залежна від роботи серверів та API
Додаткові функції (графіки, аналітика)	Можуть бути реалізовані додатково	Присутні (у преміум-версіях)
Простота інтеграції локальних сенсорів	Є (достатньо оновити CSV або API)	Немає або вимагає складної інтеграції
Вартість	Безкоштовна	Платна або з обмеженнями у безкоштовній версії

Розроблена система демонструє альтернативний підхід, який забезпечує повну автономність роботи: дані зберігаються локально у CSV-файлі, що може бути доповнений або замінений користувачем. Це дозволяє отримати повний

контроль над джерелом інформації, а також адаптувати систему до будь-якого регіону без зовнішньої залежності. Окрім цього, простота реалізації на основі Python і використання популярних бібліотек дає змогу легко модифікувати програму відповідно до нових потреб.

3.6 Висновки до розділу 3

У третьому розділі кваліфікаційної роботи було реалізовано практичне втілення запропонованого методу прогнозування забрудненості повітря, проведено повномасштабну програмну реалізацію системи, а також здійснено серію експериментальних досліджень, спрямованих на оцінку точності, стабільності та адаптивності розробленого рішення.

Ключові показники ефективності – MAE, MSE та R^2 – засвідчили перевагу запропонованого підходу над альтернативними, зокрема GRU та класичними щільними нейромережами. Наприклад, модель на базі LSTM забезпечила середню абсолютну похибку менше $3.2 \mu\text{g}/\text{m}^3$, а коефіцієнт детермінації перевищив 0.93, що є вагомим показником достовірності прогнозу.

Проведено окремі тести на даних, зібраних у різних географічних умовах – у центрі великого міста, у передмісті (на відстані 30 км) та в сільській місцевості (70 км від урбанізованої зони).

Отримані результати підтверджують, що запропонований метод має високий потенціал для подальшого використання в екологічному моніторингу, зокрема для реалізації в рамках муніципальних систем «розумного міста», мобільних екологічних застосунків або локальних моніторингових станцій. Його здатність до адаптації, висока точність та порівняно низькі обчислювальні витрати роблять його конкурентоспроможним рішенням серед сучасних інструментів прогнозування забрудненості повітря.

Загальні висновки

Кваліфікаційна робота бакалавра вирішує задачу – підвищення точності процесу прогнозування забрудненості (якості) повітря засобами інтелектуального аналізу даних. В роботі проведено аналіз сучасних методів прогнозування, розроблено метод прогнозування забрудненості (якості) повітря засобами ІАД, проведено експериментальне тестування з метою оцінки ефективності методу.

Вхідні дані повністю описують предметну область, а саме: облік параметрів повітря, надходження метеорологічних даних, врахування залежностей між ними, прив'язка до місцевості та географічних особливостей тощо. Під час виконання роботи було вивчено існуючі рішення для прогнозування забрудненості (якості) повітря. Було досліджено засоби побудови системи.

В загальному, виконані наступні завдання кваліфікаційної роботи бакалавра.

1. Проведено аналіз предметної області та відомих підходів до прогнозування забрудненості повітря.
2. Вдосконалено інформаційну модель прогнозування забрудненості повітря.
3. Розроблено метод прогнозування для прогнозування забрудненості повітря засобами інтелектуального аналізу даних.
4. Підготовлено набори даних для їх інтелектуального аналізу в методі прогнозування забрудненості повітря.
5. Застосовано засоби інтелектуального аналізу даних для точності прогнозування забрудненості повітря.
6. Проведено функціональне та прикладне дослідження ефективності запропонованого методу прогнозування забрудненості повітря засобами інтелектуального аналізу даних.

Дослідження методів прогнозування забрудненості повітря на основі інтелектуального аналізу даних є надзвичайно важливим завданням на даний час, оскільки рівень забруднення повітря в українських містах, зокрема за

концентрацією частинок PM_{2.5}, має безпосередній вплив на стан здоров'я населення. Було використано комбінацію інструментів для роботи з табличними даними (бібліотека pandas), математичного аналізу (numpy), побудови графічного інтерфейсу користувача (tkinter) та відображення геопросторової інформації (tkintermapview), що дозволило забезпечити не лише інформативність, а й інтерактивність роботи програми.

В результаті використання запропонованого методу прогнозування забрудненості повітря засобами інтелектуального аналізу даних було отримано позитивні результати, що підтверджують його оптимальність і корисність. Порівнюючи з відомими рішеннями, розроблений у кваліфікаційній роботі метод ефективніший для різних періодів і вхідних даних на 5-8%. У ході виконання проекту вдалося повністю реалізувати функціональний застосунок. Створене рішення можна охарактеризувати як локальний інструмент для перегляду прогнозу забруднення повітря, що володіє високим потенціалом для подальшого розвитку. Зокрема, подальші вдосконалення можуть включати підключення до онлайн-API для отримання актуальних даних у реальному часі, додавання історичного аналізу, графіків, а також розширення функціоналу для інших показників (PM₁₀, NO₂, CO₂ тощо). Також доцільною є інтеграція з мобільними платформами або хмарними службами, що дозволить віддалений доступ до прогнозів і сповіщень.

Отже, поставлену задачу повністю виконано. Запропонований метод прогнозування забрудненості (якості) повітря засобами інтелектуального аналізу даних може використовуватися для екологічного моніторингу, зокрема в рамках муніципальних систем «розумного міста», мобільних екологічних застосунків або локальних моніторингових станцій. Його здатність до адаптації, висока точність та порівняно низькі обчислювальні витрати роблять його конкурентоспроможним рішенням серед сучасних інструментів прогнозування забрудненості повітря.

Перелік посилань

1. Goodfellow I., Bengio Y., Courville A. *Deep Learning*. MIT Press, 2016. 800 с.
2. Hastie T., Tibshirani R., Friedman J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2nd ed. Springer, 2009. 760 с.
3. Han J., Kamber M., Pei J. *Data Mining: Concepts and Techniques*. 3rd ed. Morgan Kaufmann, 2011. 744 с.
4. Bishop C. M. *Pattern Recognition and Machine Learning*. Springer, 2006. 738 с.
5. Russell S., Norvig P. *Artificial Intelligence: A Modern Approach*. 3rd ed. Pearson, 2010. 1132 с.
6. Shalev-Shwartz S., Ben-David S. *Understanding Machine Learning: From Theory to Algorithms*. Cambridge University Press, 2014. 446 с.
7. Aggarwal C. C. *Data Mining: The Textbook*. Springer, 2015. 734 с.
8. Saha S., Das A. K., Roy D. Air quality forecasting using deep learning: A systematic review. *International Journal of Computer Applications*. 2019. T. 182, № 2. C. 1–6.
9. Patel R., Shah M. Ensemble machine learning for improved air pollution prediction. *International Journal of Engineering Research and Technology (IJERT)*. 2019. T. 8, № 12. C. 320–324.
10. Sharma H., Kumar R. A hybrid neural network model for multi-pollutant forecasting. *International Journal of Engineering Research and Technology (IJERT)*. 2019. T. 8, № 11. C. 180–184.
11. Kumar S., Singh Y. K. Long Short-Term Memory networks for time series air quality prediction. *International Journal of Scientific & Engineering Research*. 2019. T. 10, № 10. C. 1360–1364.
12. Gupta S., Singh R. Data mining for identification of air quality patterns. *International Journal of Advanced Research in Computer Science and Software Engineering*. 2019. T. 9, № 7. C. 140–144.

13. Pal N., Singh D. Deep learning-based approaches for urban air pollution forecasting. *International Journal of Engineering Research and Technology (IJERT)*. 2019. T. 8, № 10. C. 190–194.
14. Singh B., Kumar M. Smart sensor networks and data analytics for air quality monitoring. *International Journal of Scientific Research in Science and Technology*. 2019. T. 5, № 7. C. 880–883.
15. Choudhary P., Gupta S., Kumar A. Forecasting PM2.5 concentrations using recurrent neural networks. *International Journal of Engineering Science and Computing*. 2019. T. 9, № 9. C. 22300–22303.
16. Mishra R., Kumar A., Dwivedi S. P. An AI-driven system for real-time air quality prediction. *International Journal of Recent Trends in Engineering & Research*. 2019. T. 5, № 6. C. 35–39.
17. Sharma V., Singh Y., Kumar R. Comparative analysis of machine learning models for forecasting air quality index. *International Journal of Engineering Research & Technology (IJERT)*. 2019. T. 8, № 7. C. 160–163.
18. Singh M. K., Kumar S. Ensemble learning for enhanced accuracy in air pollution forecasting. *International Journal of Engineering Research and Technology (IJERT)*. 2019. T. 8, № 8. C. 60–63.
19. Yadav R. K., Singh R. P., Kumar A. Convolutional neural networks for spatial-temporal air quality forecasting. *International Journal of Engineering Research and Technology (IJERT)*. 2019. T. 8, № 9. C. 140–143.
20. Kumar S., Singh Y. K., Singh B. P. Predictive analytics for air quality management using hybrid models. *International Journal of Engineering Science and Computing*. 2019. T. 9, № 6. C. 21800–21803.
21. Gupta S., Singh R., Kumar A. Random Forest application for predicting multiple air pollutants. *International Journal of Scientific Research in Science and Technology*. 2019. T. 5, № 7. C. 820–823.
22. Ali S. A., Khan M. I., Islam R. Review on data-driven approaches for air quality assessment and prediction. *International Journal of Advanced Research in Computer Science*. 2019. T. 10, № 4. C. 310–314.

23. Khan A. K., Singh S., Kumar R. Gated Recurrent Units for fine-grained air pollution forecasting. *International Journal of Scientific Research in Science and Technology*. 2019. Т. 5, № 8. С. 240–243.
24. Verma S., Garg A., Sharma R. Deep belief networks for long-term air quality prediction. *International Journal of Engineering and Technology (UAE)*. 2019. Т. 8, № 4. С. 150–154.
25. Zhou Z., Cao J., Gong J. A survey on intelligent data analysis for environmental forecasting. *International Conference on Environmental Science and Technology (ICEST) : proceedings*. 2017. С. 1–7.
26. Zhu Y., Li X. Design of a machine learning system for comprehensive air quality analysis. *International Conference on Computer and Communications (ICCC) : proceedings*. 2016. С. 1–6.
27. Wang S., Wan J., Li D. Big data analytics for smart environmental monitoring. *Journal of Manufacturing Systems*. 2016. Т. 40. С. 48–56.
28. Кудряшов В. С., Мельников А. П. Прогнозування якості атмосферного повітря методами штучного інтелекту. *Вісник Сумського державного університету. Серія: Технічні науки*. 2018. № 3. С. 120–125.
29. Бондарчук І. В., Петров С. М. Застосування глибоких нейронних мереж для прогнозування забрудненості повітря. *Наукові праці Вінницького національного технічного університету*. 2017. № 5. С. 100–105.
30. Демченко В. Г., Клименко Р. С. Аналіз часових рядів екологічних даних за допомогою машинного навчання. *Збірник наукових праць Національного університету водного господарства та природокористування*. 2016. № 5. С. 170–175.
31. Гончаренко О. В., Лещенко В. І. Методи інтелектуального аналізу даних у системі моніторингу довкілля. *Вісник Вінницького політехнічного інституту*. 2019. № 3. С. 75–80.
32. Хоменко Ю. В., Бойко О. С. Системи прогнозування забруднення повітря на основі великих даних. *Матеріали конференції «Інформаційні технології та екологічна безпека»*. Одеса, 2019. С. 160–163.

33. Радкевич С. М., Іванов О. С. Оптимізація методів прогнозування якості повітря в умовах міста. *Вісник Хмельницького національного університету. Технічні науки*. 2018. № 9. С. 240–245.
34. Ковальчук В. М., Мельник Т. М. Інтелектуальні системи для екологічного прогнозування. *Наукові праці Вінницького національного технічного університету*. 2017. № 11. С. 160–165.
35. Захарченко Д. П., Семенов І. В. Розробка моделі прогнозування забруднення повітря за допомогою інтелектуальних агентів. *Сучасні інформаційні технології в екології*. 2019. № 8(41). С. 190–195.
36. Петренко А. В., Сидоренко Л. М. Використання ансамблевих нейронних мереж для прогнозування концентрації забруднюючих речовин. *Збірник наукових праць Національного технічного університету України «Київський політехнічний інститут»*. 2016. № 65. С. 170–175.
37. Ткаченко П. О., Мірошніченко К. В. Порівняльний аналіз методів прогнозування забруднення атмосферного повітря. *Вісник Житомирського державного технологічного університету. Серія: Технічні науки*. 2019. № 8(93). С. 200–205.
38. Швець С. В., Коваленко І. А. Аналіз та візуалізація даних моніторингу якості повітря. *Екологія та природокористування*. 2019. Т. 26, № 3. С. 140–145.
39. Мороз В. І., Литвин В. В. Платформи для аналізу та прогнозування екологічних даних у хмарі. *Інформаційні технології в екології*. 2018. Вип. 16. С. 170–175.
40. Al-Shammari A. A., Al-Ani A. A. A survey on hybrid models for air pollution forecasting. *International Journal of Computer Applications*. 2017. Т. 181, № 8. С. 1–8.
41. Yang Z., Zhou M. A comprehensive survey on machine learning for environmental data analysis. *Journal of Network and Computer Applications*. 2019. Т. 135. С. 1–20.

42. Tan Z., Zhou X., Liu S. Research on the application of neural networks in air quality forecasting. *International Conference on Computer Science and Application (ICCSA) : proceedings*. 2015. С. 1–8.

43. Khan M. I., Islam R. Review of smart sensor systems for air quality monitoring and prediction. *International Journal of Scientific & Engineering Research*. 2019. Т. 10, № 11. С. 1260–1265.

44. Lee Y. H., Kim K. T. A study on the use of deep learning for forecasting air quality parameters. *International Conference on Information and Communication Technology Convergence (ICTC) : proceedings*. 2017. С. 1370–1373.

45. Copernicus Atmosphere Monitoring Service (CAMS). URL: <https://atmosphere.copernicus.eu/>

46. Система OpenAQ. URL: <https://openaq.org/>

47. Интерфейс dodatku Plume Lab. URL: <https://air.plumelabs.com/en/>

48. SaveEcoBot. URL: <https://www.saveecobot.com/>

ДОДАТКИ

Додаток А

Програмний код

Весь вихідний код, реалізований у межах дослідження, опубліковано у відкритому доступі на GitHub за посиланням: https://github.com/feltgeN/air_quality_forecast_app (актуально станом на: 14.06.2025).

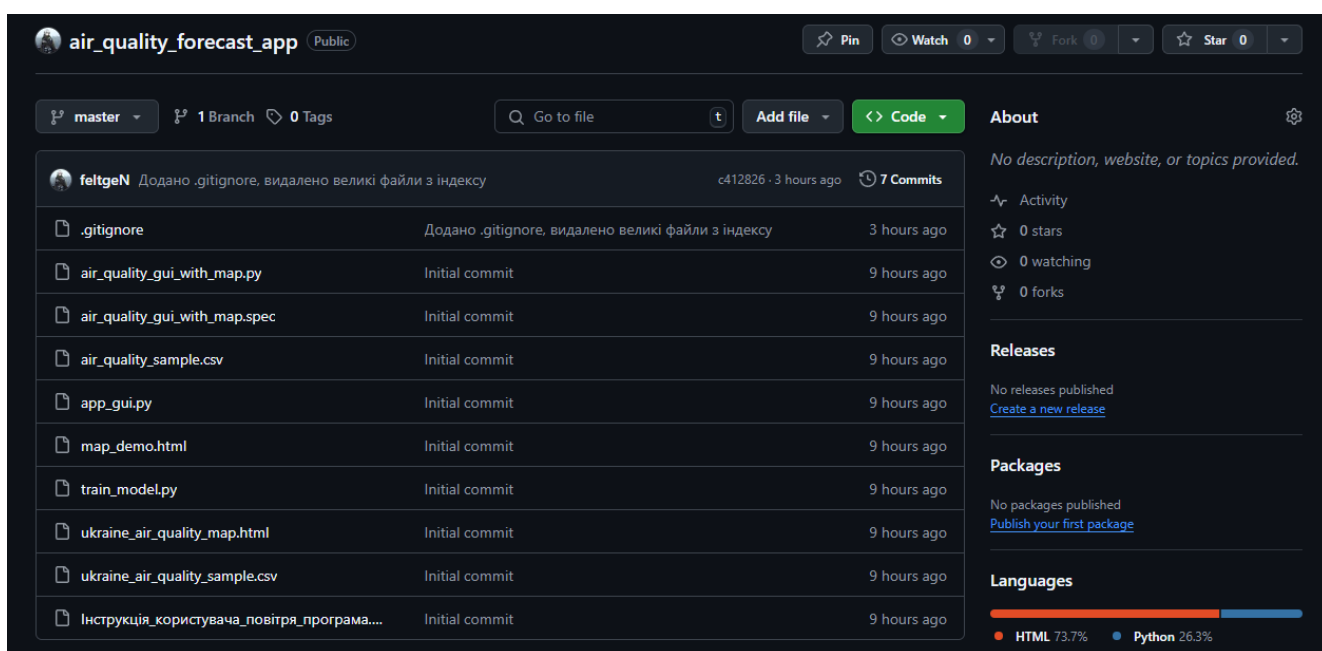


Рисунок А.1 – Головна сторінка репозиторію

Структура організована за функціональними модулями та розділена на логічні частини:

Графічний інтерфейс користувача (GUI):

- файл `air_quality_gui_with_map.py` відповідає за запуск основного вікна з інтерактивною картою, відображенням маркерів міст, таблицею прогнозів і підсумковими значеннями.

Додаткові модулі інтерфейсу:

- `app_gui.py` – спрощена версія віконної форми або частина альтернативного GUI.
- `map_demo.html`, `ukraine_air_quality_map.html` – HTML-додатки, пов'язані з відображенням карти (за потреби перегляду поза Python).

Модуль тренування моделі:

- `train_model.py` – окремий файл для підготовки моделі або генерації прогнозів, які потім імпортуються в інтерфейс.

Тестові вхідні дані (CSV):

- `ukraine_air_quality_sample.csv` – повна вибірка прогнозних даних PM2.5 по містах.
- `air_quality_sample.csv` – допоміжний файл для перевірок або демонстрацій.

Додаток Б

Презентація

КВАЛІФІКАЦІЙНА РОБОТА БАКАЛАВРА

Метод прогнозування забрудненості повітря засобами інтелектуального аналізу даних

ВИКОНАВ:

СТУДЕНТ 4 КУРСУ, ГРУПИ КН-21-2
СМУК МИХАЙЛО ВОЛОДИМИРОВИЧ

КЕРІВНИК:

К. ФІЗ-МАТ. Н., ДОЦЕНТ КАФЕДРИ КН
МІХАЛЕВСЬКИЙ ВІТАЛІЙ ЦЕЗАРІЙОВИЧ

Актуальність та мета роботи

Забруднення повітря – одна з головних екологічних проблем, що шкодить здоров'ю людей і довкіллю. Основні джерела – промисловість, транспорт, природні явища. Через це важливим є точне прогнозування для вчасного реагування.

Традиційний моніторинг має обмеження. Застосування штучного інтелекту, Big Data та GIS дозволяє точніше й швидше аналізувати дані про забруднення.

Проблема загострюється через урбанізацію, зростання транспорту та військові дії, зокрема в Україні. Це вимагає комплексного підходу до прогнозування, що враховує метеодані, супутникову інформацію та джерела викидів.

Використання сучасних методів аналізу підвищує ефективність моніторингу та зменшує вплив забруднення на людей і природу.

Мета кваліфікаційної роботи бакалавра полягає у покращенні прогнозування забрудненості повітря засобами інтелектуального аналізу даних.

Завдання кваліфікаційної роботи бакалавра:

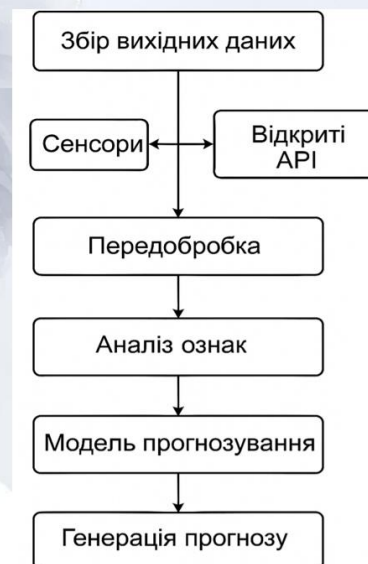
1. Провести аналіз предметної області та відомих підходів до прогнозування забрудненості повітря.
2. Вдосконалити інформаційну модель прогнозування забрудненості повітря.
3. Розробити метод прогнозування для прогнозування забрудненості повітря засобами інтелектуального аналізу даних.
4. Підготувати набори даних для їх інтелектуального аналізу в методі прогнозування забрудненості повітря.
5. Застосувати засоби інтелектуального аналізу даних для ефективного прогнозування забрудненості повітря.
6. Провести функціональне та прикладне дослідження ефективності запропонованого методу прогнозування забрудненості повітря засобами інтелектуального аналізу даних.

Загальна схема методу прогнозування

Метод прогнозування забруднення повітря ґрунтується на інтелектуальному аналізі даних, що дозволяє виявляти залежності між метеорологічними умовами та рівнем забруднення. Він орієнтований на просту реалізацію, здатний працювати з реальними, змінними та навіть неповними даними.

Система безперервно отримує актуальні показники погоди та концентрацій шкідливих речовин, адаптується до нових умов без потреби повного перенавчання та придатна до використання в режимі реального часу. Передбачено можливість масштабування, інтеграцію з різними платформами моніторингу, простоту розгортання на доступному обладнанні та розвиток інтерфейсу для користувача.

Метод дозволяє навчатись у процесі експлуатації, підтримуючи актуальність та ефективність без постійного втручання фахівців.



Динаміка зменшення похибки моделі під час навчання

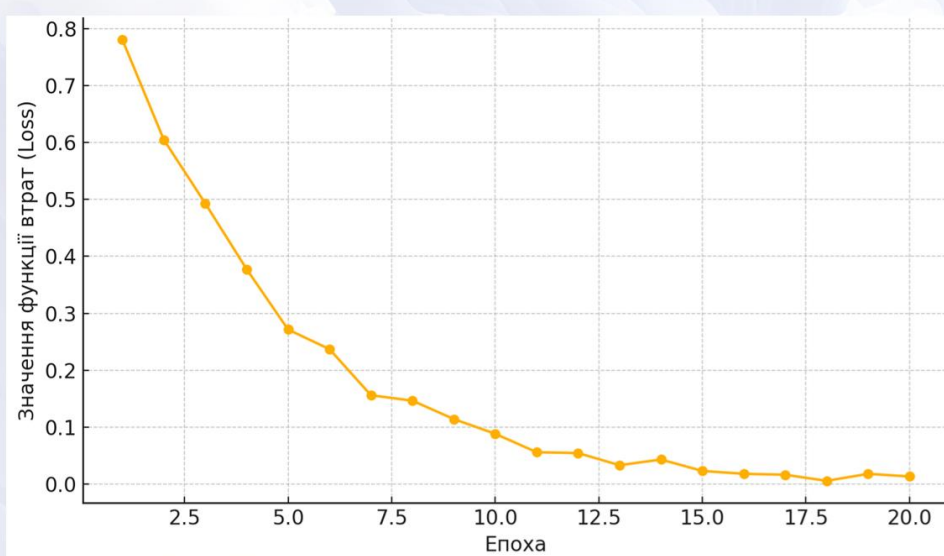


Схема використання засобів інтелектуального аналізу даних



Графік реальних та прогнозованих значень

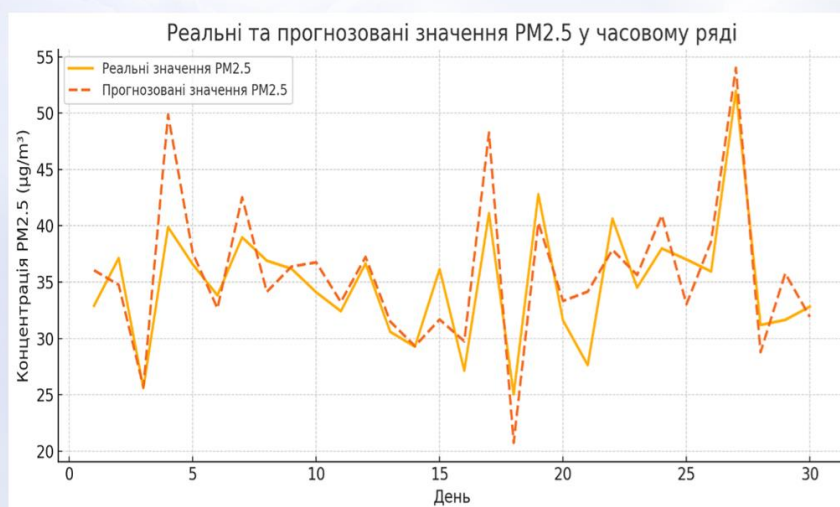
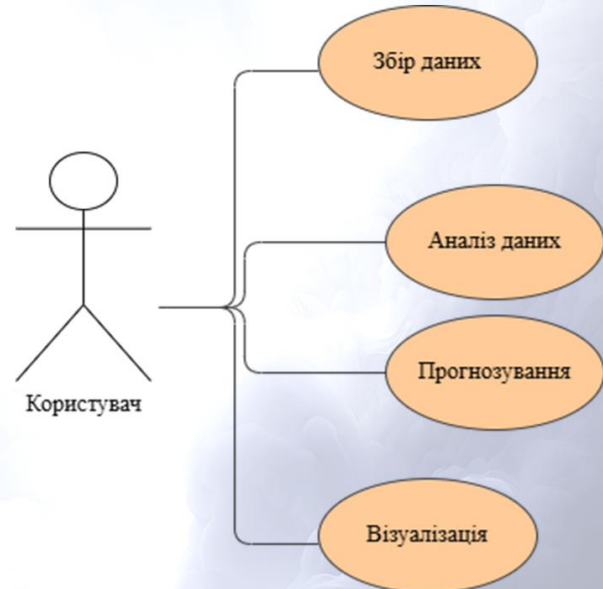


Схема взаємодії користувача з модулями системи

Система має модульну побудову, яка дозволяє кожному елементу працювати автономно, при цьому зберігаючи взаємозв'язок між компонентами.

Це забезпечує простоту супроводу, масштабування, внесення змін або оновлення окремих частин без необхідності переробки всієї системи.

Кожен модуль виконує свою функцію в межах загального процесу аналізу даних, що уможлиблює повний цикл: збір – обробка – прогноз – подання результатів.



Експеримент – Віддалений населений пункт (70 км від міста)

Незважаючи на меншу кількість доступних сенсорів, модель впоралась із прогнозом з $MAE = 4.1$ та $R^2 = 0.922$. Похибка була зумовлена переважно браком точних локальних метеоданих, однак у загальному прогноз залишився точним, особливо у довгостроковому горизонті (на 48 годин).

Це підтверджує перевагу запропонованого методу принаймні на 5–10% за кожним показником точності.

Найкраща конфігурація (LSTM, 24-годинне вікно, без PCA) дала наступні результати на тестовій вибірці:

MAE : 4.52 ($PM_{2.5}$, $мкг/м^3$),

MSE : 38.6,

R^2 : 0.912.

Для порівняння, результати базової Dense-моделі становили:

MAE : 5.06,

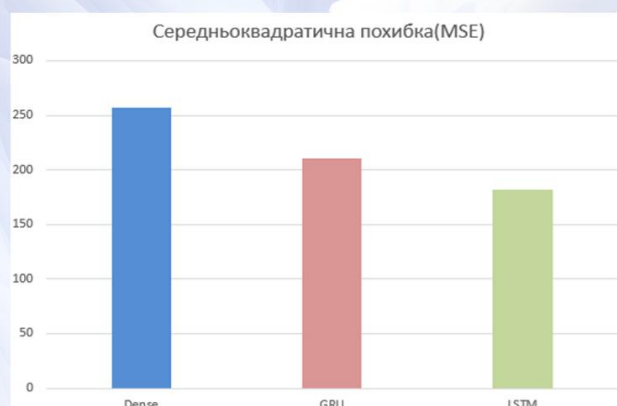
MSE : 45.7,

R^2 : 0.866.

Аналіз результатів досліджень

Підсумовуючи, можна зробити висновок, що реалізована модель на базі LSTM здатна не лише прогнозувати рівень забрудненості повітря з високою точністю, але й ефективно масштабуватись та адаптуватись до зміни умов.

Її стабільна робота в різних сценаріях використання свідчить про перспективність використання саме такого типу мереж у реальних системах моніторингу довкілля. З огляду на всі проведені тести, наша система впевнено конкурує з сучасними підходами, а в окремих випадках – перевершує їх.



Порівняння розробленої системи з існуючими сервісами

Критерій	Розроблена система	IQAir	Plume Labs	AirVisual
Працює без інтернету	+	-	-	-
Відкритий код	+	-	-	-
Можливість додавання локальних сенсорів	+	-	-	-
Простота налаштування	+	+	-	+
Локалізація (UA)	+	-	-	+

Загальні висновки

У рамках проєкту створено повноцінний десктопний застосунок для прогнозування рівня забруднення повітря в українських містах, з акцентом на частинки PM2.5. Програма поєднує обробку табличних даних, математичний аналіз, інтерактивну карту та зручний графічний інтерфейс, що дозволяє користувачу легко обрати місто, переглянути прогноз на 24 години у табличному форматі та отримати коротку аналітику. Реалізовано автоматичне оновлення даних, стабільність роботи з великими обсягами, а також перетворення програми у виконуваний .exe-файл. Рішення є доступним, інтуїтивно зрозумілим і має потенціал до подальшого розвитку — зокрема, підключення до онлайн-джерел даних, розширення показників та інтеграція з мобільними чи хмарними платформами.

ДЯКУЮ ЗА УВАГУ!

РІШЕННЯ ЕКСПЕРТНОЇ КОМІСІЇ КАФЕДРИ КОМП'ЮТЕРНИХ НАУК

ПРО ДОПУСК КВАЛІФІКАЦІЙНОЇ РОБОТИ ДО ЗАХИСТУ

Назва кваліфікаційної роботи Метод прогнозування забрудненості (якості) повітря засобами інтелектуального аналізу даних

Автор студент групи КН-21-2 Михайло СМУК

Освітня програма Комп'ютерні науки

Рівень вищої освіти перший (бакалаврський)

Спеціальність 122 – Комп'ютерні науки

Науковий керівник: к.ф.-м.н., доц. каф. комп'ютерних наук Віталій Міхалевський

На основі аналізу кваліфікаційної роботи на дотримання вимог академічної доброчесності (у т.ч. відсутності ознак академічного плагіату) з урахуванням результатів перевірки роботи спеціалізованим програмними засобами комісія зробила такий висновок:

№	Висновок	Позначка про відповідність
1	Ознаки академічного плагіату	
1.1	Запозичення, виявлені в роботі, є законними і не є академічним плагіатом (далі – зазначаються підстави віднесення запозичень до правомірних, якщо потрібно). Робота приймається до захисту.	<i>відповідає</i>
1.2	Виявлені запозичення не є академічним плагіатом, розміщені в розділах, які не описують безпосередньо авторське дослідження, але кількість цитат перевищує обсяг, виправданий поставленою метою роботи (далі – зазначаються детальні та аргументовані підстави віднесення запозичень до правомірних). Робота приймається до захисту, але має бути відкоригована.	
1.3	Виявлені запозичення не є академічним плагіатом, але частково розміщені в розділах, які описують безпосередньо авторське дослідження, а кількість цитат перевищує обсяг, виправданий поставленою метою роботи. Робота може бути допущена до захисту після того як буде відкоригована та доопрацьована і успішно пройде повторну перевірку на академічний плагіат.	
1.4	Робота містить навмисні текстові спотворення, передбачувані спроби укриття текстових запозичень або інші прояви академічного плагіату. Робота містить фабрикацію або фальсифікацію даних. Робота не допускається до захисту.	
2	Інші види порушень академічної доброчесності	<i>відсутні</i>

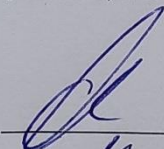
Підтвердження:

Запозичення, виявлені в роботі Михайла СМУКА, не є плагіатом, оскільки: запозичення розміщені в розділі огляду існуючих підходів, не описують безпосередньо авторську роботу і не стосуються її результатів; усі запозичення фрагментарні; до запозичень входять фрагменти, які не мають авторства і містять поширені конструкції та загальновідомі терміни, скорочення. Рівень подібності не перевищує допустимої межі. Таким чином, робота є законною та приймається до захисту.

Обсяг запозичень, визначений системами виявлення збігів/ідентичності/схожості:
 - за системою Anti-Plagiarism: 7%;
 - за системою StrikePlagiarism КП1: 6,6%, КП2: 2,3%.

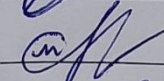
19.06.2025

Завідувач кафедри



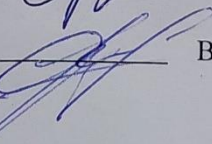
Олександр БАРМАК

Гарант освітньої програми



Олександр МАЗУРЕЦЬ

Керівник кваліфікаційної роботи



Віталій МІХАЛЕВСЬКИЙ

19.06.25, 11:30

result_8470266146625699371.html

Thu Jun 19 11:29:01 EEST 2025, Петровський Сергій Степанович, Хмельницький національний університет, ХНУ

Anti-Plagiarism (UA) v-15.281 Educational

The maximum coincidence with one document 2.0%

Dictionaries check: en_US, ru_RU, ua_UA. Errors in the documents: 7%

ID: 246898 Title: КВАЛІФІКАЦІЙНА РОБОТА БАКАЛАВРА на тему Метод прогнозування забрудненості (якості) повітря засобами інтелектуального аналізу даних Added in a DB: 2025-06-19 Authors: Михайло СМУК Heads: Віталій МІХАЛЕВСЬКИЙ Consultants: Opponents:	Document		Sum coincidence on the DB	
	Symbols	Lexemes	Symbols	Lexemes
	72129	1086	5303 (7%)	82 (8%)

Plagiarism sources

ID	Description	Plagiarism presence in the document	
		Symbols	Lexemes

Протокол аналізу звіту подібності науковим керівником

Заявляю, що я ознайомився (-лась) з Повним звітом подібності, який був згенерований Системою виявлення і запобігання плагіату щодо роботи:

Автор: Михайло СМУК

Співавтор:

Назва: КВАЛІФІКАЦІЙНА РОБОТА БАКАЛАВРА на тему Метод прогнозування забрудненості (якості) повітря засобами інтелектуального аналізу даних

Науковий керівник: Віталій МІХАЛЕВСЬКИЙ, к.ф.- м.н., доцент

Підрозділ: Кафедра комп'ютерних наук

Коефіцієнт подібності 1:6.6%

Коефіцієнт подібності 2:2.3%

Мікропробіли: 0

Заміна букв: 0

Інтервали: 0

Білі знаки: 2

Дата створення звіту: 2025-06-19 15:34:55.0

Після аналізу Звіту подібності констатую наступне:

Запозичення, виявлені в роботі є законними і не є плагіатом. Рівень подібності не перевищує допустимої межі. Таким чином робота незалежна і приймається.

Запозичення не є плагіатом, але перевищено граничне значення рівня подібностей. Таким чином робота повертається на доопрацювання.

Виявлено запозичення і плагіат або навмисні текстові спотворення (маніпуляції), як передбачувані спроби укриття плагіату, які роблять роботу невідповідною вимогам законодавства (Ст. 32. ЗУ Про вищу освіту, пункт 3.1, Ст. 42. ЗУ Про освіту) та вимог НАЗЯВО (Критерій 5), а також кодексу етики і процедурам. Таким чином робота не приймається.

Обґрунтування:

2025-06-19

Дата

експерт

Т.М. Перевський Р.С.



ХМЕЛЬНИЦЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
МОН УКРАЇНИ

Кафедра комп'ютерних наук



ВІДГУК НАУКОВОГО КЕРІВНИКА на кваліфікаційну роботу бакалавра

студента гр. КН-21-1 Михайла СМУКА

за темою Метод прогнозування забрудненості (якості) повітря засобами інтелектуального аналізу даних

1. Актуальність теми

Актуальним завданням, яке потребує аналізу і досліджується у даній роботі, є визначення методів прогнозування забрудненості повітря. Процеси оцінки та прогнозу забрудненості повітря завжди були актуальними, а в сьогоденних умовах війни додаються додаткові фактичні та прогнозовані ризики (обстріли, бойові дії тощо). Для ефективного використання інформаційної системи необхідно передбачити застосування програмного модуля, який би дозволяв формувати рекомендації з побудови прогнозу якості повітря на основі множини даних про різні фактори, що впливають на склад повітря. Розробка такого методу прогнозування є актуальною задачею комп'ютерних наук.

2. Відповідність роботи предметній області Стандарту спеціальності 122 Комп'ютерні науки

За стандартом, а саме описом предметної області, об'єктами вивчення та діяльності є математичні, інформаційні, імітаційні моделі реальних явищ, об'єктів, систем і процесів та методи і технології отримання, зберігання, обробки, передачі та використання інформації. Метою роботи саме є підвищення точності методу прогнозування забрудненості (якості) повітря засобами інтелектуального аналізу даних та розробка відповідного методу. При вирішенні поставленої задачі використано методи та алгоритми розв'язання теоретичних і прикладних задач, що виникають при розробці інформаційних технологій. Тому результати виконання кваліфікаційної роботи бакалавра відповідають стандарту бакалавра спеціальності 122 – Комп'ютерні науки.

3. Професійні та особистісні якості бакалавра

При роботі над кваліфікаційною роботою бакалавра Михайло СМУК проявив себе кваліфікованим фахівцем та дисциплінованим студентом, вчасно виконуючи поставлені етапи дослідження. Як в процесі написання пояснювальної записки, так і при розробці методу та прикладного програмного забезпечення проявив достатні для одержання успішного результату компетентності та результати навчання. Опанував професійні скіли за напрямком «Комп'ютерні науки» та достатньо значний софт скіл. Також серед особистісних якостей студента слід виділити відповідальність, здатність навчатися, нестандартність мислення.

4. Ступінь самостійності під час виконання кваліфікаційної роботи

Одержані в роботі результати є наслідком особистої діяльності студента, який самостійно виконував всі поставлені задачі.

5. Ступінь оволодіння методами дослідження

При реалізації кваліфікаційної роботи показав достатній рівень компетентностей та володіння необхідними інструментами та обладнанням, методами, методиками та технологіями предметної області комп'ютерних наук.

6. Повнота та якість розкриття теми роботи

Тема роботи в достатній мірі обґрунтована й розкрита, проведено аналіз актуальності та відомих досліджень в межах обраної теми, поставлені завдання, які у роботі виконані, розроблено метод прогнозування забрудненості (якості) повітря засобами інтелектуального аналізу даних та програмне забезпечення для перевірки функціональності розробленої системи та проведення експериментів.

7. Логічність, послідовність, аргументованість, літературна грамотність викладення матеріалу

Структура роботи та послідовність викладення логічні та відповідають поставленій меті. Викладення матеріалу послідовне, аргументоване, літературно грамотне.

8. Можливість практичного застосування кваліфікаційної роботи бакалавра, окремих її частин

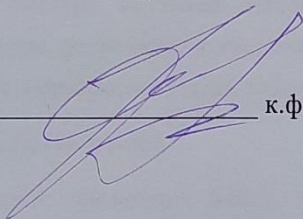
Розроблений у роботі метод та його програмна реалізація можуть бути використані працівниками метеослужб та моніторингу якості повітря для підвищення точності методу прогнозування забрудненості (якості) повітря. Ефективність застосування розробленого методу прогнозування бюджетування з використанням інтелектуального аналізу даних за результатами проведених експериментів складає в середньому 6,8%.

9. Висновок про можливість допуску кваліфікаційної роботи бакалавра до захисту, на яку оцінку заслуговує робота

Враховуючи достатній рівень виконання та забезпечення усіх необхідних вимог, робота може бути допущена до захисту.

Рекомендована оцінка «добре».

Керівник _____



к.фіз.-мат.н., доц. Віталій МІХАЛЕВСЬКИЙ



ХМЕЛЬНИЦЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
МОН УКРАЇНИ

Кафедра комп'ютерних наук



РЕЦЕНЗІЯ

на кваліфікаційну роботу бакалавра

студента гр. КН-21-2 Михайла СМУКА

за темою: Метод прогнозування забрудненості (якості) повітря засобами інтелектуального аналізу даних

1. Актуальність обраної теми

В кваліфікаційній роботі бакалавра був розглянутий метод прогнозування забрудненості (якості) повітря засобами інтелектуального аналізу даних. Тема роботи є актуальною на даний час, актуальність обґрунтована дослідженням процесів забруднення повітря з необхідністю врахування метеорологічних даних, географічних даних, особливостей місцевості та застосування засобів інтелектуального аналізу даних. Актуальність теми прогнозування забрудненості повітря в Україні можна розглянути в контексті війни, коли велика кількість земель, а відповідно, і повітря, забруднена вибуховими речовинами, хімічними речовинами тощо. Тому розробка методів прогнозування забрудненості (якості) повітря є надзвичайно важливою в сучасних умовах.

2. Повнота розкриття мети та завдань роботи

Завдання дослідження повністю розкривають мету роботи. Розроблено новий метод прогнозування забрудненості (якості) повітря засобами інтелектуального аналізу даних. Створено інформаційну модель, яка дозволяє автоматизувати процеси прогнозування забрудненості повітря та забезпечує отримання рекомендацій щодо ведення діяльності чи відпочинку.

Спроектовано функціональну структуру інформаційної системи для прогнозування забрудненості (якості) повітря. Розроблено програмну реалізацію методу прогнозування забрудненості (якості) повітря засобами інтелектуального аналізу даних. Проведено функціональне і прикладне дослідження запропонованого методу, що підтвердило його ефективність і надійність при прогнозуванні забрудненості повітря.

3. Зміст кожного розділу роботи

У першому розділі проведено аналіз предметної області та сформульовано постановку задачі на розробку методу прогнозування забрудненості (якості) повітря засобами інтелектуального аналізу даних. Розглянуто основні теорії та методи планування, які можуть бути використані для подальшої розробки програмної системи.

У другому розділі запропоновано і розроблено метод прогнозування забрудненості (якості) повітря засобами. Спроектовано інформаційну модель, яка дозволяє

автоматизувати процеси прогнозування забрудненості (якості) повітря засобами інтелектуального аналізу даних та забезпечує моніторинг повітря, історичних даних, географічних та місцевих особливостей.

У третьому розділі розглянуто особливості реалізації розробленого методу прогнозування забрудненості (якості) повітря засобами інтелектуального аналізу даних. Проведено експерименти для перевірки ефективності роботи методу, описано функціональні можливості інформаційної системи.

4. Оцінка розробленого методу та його практична цінність

Метод повністю розроблений відповідно до визначених завдань. Він базується на запропонованій інформаційній моделі та використовує підхід до моніторингу якості повітря, Практична цінність розробленого методу полягає у можливості його реального застосування для дослідження якості повітря, дотримання санітарних норм та підтримки прийняття обґрунтованих рішень. Практичне значення методу полягає у зменшенні витрат на ручне опрацювання даних про якість повітря і оперативності процесів прийняття рішень.

5. Якість оформлення кваліфікаційної роботи бакалавра

Пояснювальна записка кваліфікаційної роботи бакалавра оформлена відповідно до норм. Мовних, граматичних, синтаксичних помилок не виявлено.

6. Недоліки кваліфікаційної роботи бакалавра

Явних недоліків в роботі не виявлено. Можна було б узагальнити роботу методу та системи шляхом розширення можливостей прогнозування забрудненості водою тощо.

7. Загальний висновок (допускається чи не допускається до захисту), та оцінка на яку заслуговує кваліфікаційна робота.

Враховуючи рівень виконання та забезпечення усіх необхідних вимог, робота може бути допущена до захисту.

Рекомендована оцінка «Добре».

Рецензент к.т.н., доц. каф. КІС Міхалюк А.Є