

Ломазов П.К., аспірант спеціальності 183 Технології захисту навколишнього середовища

Наукові керівники: Павличенко А.В., д.т.н., професор кафедри екології та технологій захисту навколишнього середовища; Бучавий Ю.В., к.б.н., доцент кафедри екології та технологій захисту навколишнього середовища

(Національний технічний університет «Дніпровська політехніка», м. Дніпро, Україна).

## СУЧАСНІ ПІДХОДИ ДО ПРОГНОЗУВАННЯ ЯКОСТІ ПОВІТРЯ ЗА ДОПОМОГОЮ МАШИННОГО НАВЧАННЯ

Сучасні методи прогнозування базуються на аналізі даних та алгоритмах машинного навчання, зокрема Random Forest і LSTM. Вони враховують зв'язок між джерелами забруднення, атмосферними умовами та їхньою динамікою. Аналіз температури, вологості та вітру дозволяє прогнозувати рівень забруднення з високою точністю, що важливо для екологічного моніторингу та попередження населення.

Для ефективного прогнозування якості повітря необхідний збір точних та різноманітних даних. Головними джерелами інформації є:

Станції моніторингу – стаціонарні та мобільні екологічні пости, оснащені сенсорами для вимірювання концентрацій NO<sub>2</sub>, CO, SO<sub>2</sub>, O<sub>3</sub>, PM10 та PM2.5. Вони надають локальні, високоточні дані, проте їхня кількість обмежена, що ускладнює створення повної картини забруднення.

Супутникові знімки – дистанційне зондування атмосфери за допомогою супутників NASA, ESA та інших платформ. Вони дозволяють отримувати глобальні дані про пил, озон, парникові гази, доповнюючи інформацію зі станцій у віддалених регіонах.

Громадські датчики – компактні пристрої, розміщені у містах та приватних об'єктах, які передають дані через інтернет. Вони забезпечують широку мережу моніторингу, але потребують калібрування для підвищення точності [1].

Для ефективного використання даних в алгоритмах машинного навчання необхідно провести їх ретельну попередню обробку, оскільки первісні дані часто містять шум, відсутні значення та інші проблеми, які можуть значно вплинути на точність прогнозів (табл. 1) [2].

Таблиця 1

### Види процесів

Назва процесу	Опис	Методи обробки
Обробка шумів	Усунення або згладжування випадкових похибок і нерелевантної інформації («шуму») в даних.	Фільтрація даних (напр., медіанний фільтр); згладжування сигналу; видалення викидів.
Відсутні значення	Робота з відсутніми (пропущеними) даними, що можуть спотворювати результати аналізу або призводити до некоректної роботи алгоритмів.	Видалення записів з пропусками; заповнення пропущених значень середнім, медіанним чи модою; використання методів імпутації (напр., прогнозування відсутніх значень).
Нормалізація даних	Приведення різнорідних даних до спільного масштабу для коректного аналізу.	Масштабування до діапазону [0,1] (мінімаксна нормалізація); стандартизація (Z-нормалізація).

Для прогнозування забруднення повітря використовуються алгоритми машинного навчання, що забезпечують високу точність завдяки обробці великих обсягів даних. Найпоширеніші підходи включають Random Forest, LSTM, Gradient Boosting Machines (GBM), Support Vector Machines (SVM) і Multilayer Perceptron (MLP) [3].

Random Forest (RF) є ансамблевим методом на основі дерев рішень, що підвищує точність прогнозів завдяки беггінгу. Його перевагою є стійкість до переобчислення, проте він не підходить для часових залежностей і може бути чутливим до розмірності даних.

LSTM – це рекурентна нейронна мережа, що добре моделює часові ряди, зокрема динамічні зміни забруднення. Її недоліками є висока обчислювальна складність і потреба у великому обсязі даних.

GBM (XGBoost, LightGBM, CatBoost) є ефективними для великих наборів даних, оскільки навчаються поетапно, мінімізуючи помилки. Вони забезпечують високу точність, проте можуть бути чутливими до шумів і вимагати значних обчислювальних ресурсів.

SVM застосовується для класифікації та регресії, особливо на малих наборах даних, однак для великих обсягів інформації має обмежену ефективність через обчислювальні витрати та чутливість до вибору ядра.

MLP – базова нейромережа, здатна моделювати нелінійні залежності. Вона менш ресурсомістка, ніж LSTM, але поступається їй у прогнозуванні часових рядів. Для складних задач MLP потребує більшої глибини, що ускладнює її навчання.

Вибір алгоритму для прогнозування забруднення повітря залежить від типу даних. Random Forest і GBM ефективні для статичних даних, тоді як LSTM оптимальна для часових рядів. У разі обмежених ресурсів використовують SVM або MLP. Поєднання GBM для вибору ознак і LSTM для часових залежностей дає найкращі результати.

Так, комбінація Random Forest і LSTM дозволяє об'єднати їхні сильні сторони: RF виділяє ключові ознаки, а LSTM враховує часові залежності, що покращує прогнози. Такий підхід допомагає аналізувати як короткострокові, так і довготривалі зміни забруднення. Моделі прогнозування забруднення можуть інтегруватися у міське управління. Вони дозволяють проводити оперативний моніторинг, створювати системи раннього попередження та оптимізувати транспортні потоки для зниження викидів. Наприклад, прогноз концентрації PM<sub>2.5</sub> допомагає інформувати громадськість і адаптувати міську політику [3].

Таким чином, результати дослідження підтверджують, що алгоритми машинного навчання є потужним інструментом для прогнозування рівня забруднення повітря. Використання моделей, таких як Random Forest, LSTM, GBM, SVM та MLP, дозволяє аналізувати великі масиви даних і виявляти складні взаємозв'язки між параметрами. Random Forest і GBM найкраще підходять для роботи зі статичними наборами даних, тоді як LSTM ефективно моделює часові ряди. Комбінування цих методів дозволяє покращити точність прогнозів завдяки використанню їхніх унікальних переваг.

#### **Список використаних джерел:**

1. Вишневський А. В., Потапова Н. А. Методи аналізу даних в інформаційній системі прогнозування показників забруднення повітря // Прикладні аспекти сучасних міждисциплінарних досліджень. – 2024. – С. 156–157.

2. Ameer S., Ali Shah M., Khan A. Comparative analysis of machine learning techniques for predicting air quality in smart cities // *IEEE Access*. – 2019. – Т. 7. – С. 1–10. – DOI: [10.1109/RMKMATE59243.2023.10369180](https://doi.org/10.1109/RMKMATE59243.2023.10369180).

3. Ломазов П. К., Павличенко А. В., Бучавий Ю. В. Методи прогнозування забруднення атмосферного повітря на основі машинного навчання // *Збірник наукових праць НГУ*. – 2025. – №78.