

*Вишневський А. В., здобувач вищої освіти,
Потапова Н. А., канд. екон. наук, доцент,
доцент кафедри інформаційних технологій,
Донецький національний університет імені Василя Стуса*

МЕТОДИ АНАЛІЗУ ДАНИХ В ІНФОРМАЦІЙНІЙ СИСТЕМІ ПРОГНОЗУВАННЯ ПОКАЗНИКІВ ЗАБРУДНЕННЯ ПОВІТРЯ

Анотація. В роботі висвітлюються основні підходи до аналізу даних під час прогнозування метеорологічних та екологічних показників забруднення повітря. Аналізуються недоліки та переваги підходів на основі використання штучного інтелекту, математичних та статистичних моделей.

Ключові слова: аналіз даних, прогнозування, математичні моделі, статистичні моделі, штучний інтелект.

Вступ. Сучасні технології, як-от машинне навчання та нейронні мережі, відкривають нові можливості для покращення точності й оперативності прогнозів забруднення повітря. Використання цих технологій дає змогу обробляти великі обсяги даних і швидко адаптуватися до змін умов. Існують різні методи і техніки для прогнозування забруднення повітря, зокрема статистичні моделі, математичне моделювання та штучний інтелект. Основними характеристиками цих методів є оцінка таких показників: точність, вартість впровадження, здатність до інтеграції з іншими системами моніторингу.

Одним із підходів проведення аналізу даних забруднення повітря є використання машинного навчання, що має перевагу в оцінці точності та зниженню помилок у короткострокових прогнозах.

Актуальність. Забруднення повітря визначається як наявність однієї чи декількох речовин в атмосфері, які шкідливі для здоров'я людини, рослин або тварин, які можуть завдати шкоди матеріалам, знизити видимість або викликати неприємний запах. Забруднення повітря можна класифікувати так [1]:

- Природні джерела. Вони включають вулкани, які викидають сірководень, попіл, і металеві оксиди; лісові пожежі, які викидають дим і вуглекислий газ; пил з пустель, який може переноситися на великі відстані.

- Антропогенні джерела. Вони включають викиди від автомобілів, які включають вуглекислий газ, оксиди азоту, вуглеводні; промислові викиди, що складаються з сірководню, аміаку, хлору і важких металів, а також теплові електростанції, які викидають сірку, діоксид і золу.

Природні та антропогенні джерела є основними джерелами викидів забруднюючих речовин у повітря, тому одним з основних завдань дослідження є вивчення джерел забруднення, їх характеристик і методів вимірювання та контролю концентрацій забруднювачів. Важливим аспектом є вплив метеорологічних умов на розповсюдження та концентрацію забруднювачів у повітрі. Основними підходами в аналізі даних забруднення повітря є використання методів статистичного, математичного моделювання та моделей штучного інтелекту [2].

Основний текст. *Математичні моделі* формалізують процес забруднення повітря шляхом використання рівнянь опису процесів на основі детермінованих фізичних зв'язків, що утворюються в екологічній системі. Основними перевагами таких моделей є точність під час відтворення фізичних процесів із достатньо деталізованою інформацією про умови використання. Побудова моделей на фізичних законах робить їх зрозумілими для фахівців у цій галузі. Недоліком такого підходу є високий рівень вимог до точності вхідних даних, особливо параметрів забруднення, метеорологічних та географічних даних. Вагомими є обчислювальні витрати для вирішення складних рівнянь.

Підхід аналізу на основі використання статистичних моделей має велику гнучкість, яка дає змогу легко адаптувати модель під різні типи даних і умови. Типові статистичні обчислення потребують менше ресурсів, порівняно з детерміністичними моделями. Проте основним недоліком є залежність від даних, що обумовлено якістю історичних даних. Також існує ризик виявлення статистичних зв'язків без розуміння причинно-наслідкових категорій.

Моделі на основі штучного інтелекту мають високу точність і можуть ефективно обробляти великі набори даних з виявленням складних патернів, які бувають невидимими для традиційних методів. Моделі є адаптивними, можуть навчатися і пристосовуватись до нових даних у реальному часі. Проте отримані розв'язки можуть мати складність в інтерпретації результатів, і виникає постійна потреба у великих обсягах даних для ефективного навчання.

Використання цих підходів пов'язане з необхідністю запровадження системного підходу в отриманні та обробці даних про забруднення, що можливо досягти за допомогою розробки інформаційної системи прогнозування показників забруднення. Інформаційна система складається з таких ключових компонентів: збору даних за допомогою AWS Lambda-функцій, зберігання даних у S3-бакетах, обробки та аналізу даних за допомогою Amazon SageMaker, а також розгортання прогнозної моделі через API Gateway.

Висновки. Розробка системи прогнозування забруднення повітря, заснованої на аналізі метеорологічних і екологічних даних, дає змогу підвищити ефективність екологічного моніторингу та сприяє своєчасному виявленню екологічних загроз. Для оцінки ефективності системи доцільно використовувати метрики, як от середня абсолютна помилка (MAE), корінь середньоквадратичної помилки (RMSE) та коефіцієнт детермінації (R^2).

Список використаних джерел

1. Запорожець А. О. Аналіз засобів моніторингу забруднення повітря навколишнього середовища. *Наукоємні технології*. 2017. № 3. URL: <https://jrnل.nau.edu.ua/index.php/SBT/article/view/11844> (дата звернення: 22.10.2024).
2. Савенець М. В. Інтегрований погляд на сучасний стан забруднення атмосферного повітря в Україні. *Вісник НАН України*. 2023. № 9. С. 80–86. DOI: 10.15407/visn2023.09.080.
3. Моніторинг забруднення атмосферного повітря. *Офіційний сайт Держекоінспекції України*. URL: <https://dei.gov.ua> (дата звернення: 22.10.2024).