

УДК 004.415.2.043

І.І. Бородій, К.І. Поліщук

(Тернопільський національний технічний університет імені Івана Пулюя, Україна)

**ВИБІР ДЖЕРЕЛ ДАНИХ ТА ТЕХНІЧНИХ ЗАСОБІВ ДЛЯ ПРОЄКТУВАННЯ
ПРОГРАМНОЇ СИСТЕМИ ПРОГНОЗУВАННЯ СТАНУ ЯКОСТІ
АТМОСФЕРНОГО ПОВІТРЯ**

I.I. Borodii, K.I. Polishchuk

**SELECTION OF DATA SOURCES AND TECHNICAL TOOLS FOR DESIGNING A
SOFTWARE SYSTEM FOR FORECASTING THE STATE OF ATMOSPHERIC AIR
QUALITY**

Забруднення атмосферного повітря є глобальною проблемою, яка має серйозні наслідки для екології, здоров'я населення та економіки. Важливість прогнозування якості атмосферного повітря зумовлена необхідністю прийняття обґрунтованих рішень у галузі екологічної політики, планування інфраструктури та зменшення шкоди для здоров'я, що робить дану тематику актуальною на сьогодні.

У зв'язку з цим виникає потреба у створенні програмної системи, яка б дозволила прогнозувати стан якості повітря. Перед початком розробки програмної системи необхідно здійснити етап проєктування [8]. Ключовими складовими процесу проєктування таких систем є вибір джерел даних та технічних засобів для реалізації такої системи.

Вибір джерел даних суттєво впливає на аналіз, моделювання та прогнозування, при цьому. У рамках дослідження використовуються сучасні технології збору даних із відкритих джерел, які забезпечують глобальне охоплення й доступ до інформації в реальному часі. Усі джерела даних відбиралися з урахуванням їхньої доступності через API, точності та надійності. Для створення багатофакторної моделі, яка адекватно відображає поточний стан атмосфери та прогнозує його зміну, визначено такі метрики як температура повітря, рівень забруднюючих речовин, активність пожеж і промислових підприємств, що відповідає сучасним підходам у виборі даних і технологій для екологічного моделювання [6].

Для моніторингу концентрації забруднюючих речовин використовується World Air Quality Index API [7], який надає інформацію про стан повітря у режимі реального часу по більшості країн світу. Температурні показники отримуються через GEM API від Open-Meteo [4], який забезпечує глобальний доступ до високоточних погодних даних. Для відстеження пожеж надано перевагу NASA FIRMS API [3], який надає актуальні дані про активні точки займання, включаючи координати, час виявлення та інтенсивність пожежі. Для збору даних про промислові об'єкти зупинено вибір на OpenStreetMap (OSM) через Overpass API [5], що надає можливість ідентифікувати місцезнаходження великих промислових зон і підприємств, аналізувати їхній просторовий вплив на якість повітря.

Вибір технічних засобів для реалізації програмної системи прогнозування стану якості атмосферного повітря має забезпечити виконання ключових операцій, таких як збір, обробка, завантаження та прогнозування надвеликих масивів даних з використанням найбільш оптимізованих підходів.

Для оркестрації та автоматизації процесів обрано інструмент Apache Airflow. У системі прогнозування його призначенням є створення орієнтованих ациклічних графів

(DAG), кожен з яких виконує завдання пов'язані зі збором даних, їх обробкою та завантаженням в централізоване сховище даних системи [1].

Сховище даних програмної системи організоване у форматі збереження даних Apache Iceberg. Цей інструмент використовується для зберігання даних у колонковому форматі. Особливістю цього формату є підтримка версіонування, що дозволяє уникати конфліктів під час читання чи запису даних [2].

Інструмент PySpark використовується як інструмент збору, обробки та завантаження даних. PySpark забезпечує паралельну обробку на кластерах, що пришвидшує роботу з великими обсягами інформації у розподіленому середовищі. Завдяки своїй масштабованій архітектурі PySpark є ідеальним інструментом для великих проєктів, де потрібно обробляти петабайти інформації з високою швидкістю та надійністю.

Платформа Docker дозволяє виконати контейнеризацію компонентів системи. Контейнеризація дозволяє значно спростити процеси розробки та розгортання програмного забезпечення. У системі прогнозування стану атмосферного повітря Docker застосовується як засіб контейнеризації для кожного компонента системи, що дозволяє ізолювати такі ключові компоненти, як Apache Airflow, PySpark і Apache Iceberg. Це підтримує їхню стабільність та незалежність від апаратного чи програмного забезпечення сервера.

Запропоновано поєднати класичні методи машинного навчання, такі як Random Forest та XGBoost, а також сучасні нейронні мережі для виконання процесу прогнозування стану якості атмосферного повітря. Метод Random Forest дозволяє обробляти складні нелінійні залежності та є стійким до шумів у даних. XGBoost, завдяки своїй швидкості та точності, добре підходить для роботи з великими наборами даних і враховує взаємодії між змінними. Після прогнозування результати інтегруються з платформою Power BI, яка дозволяє візуалізувати дані у вигляді інтерактивних графіків, таблиць та мап.

Література

1. Apache Airflow [Electronic resource] / Apache Airflow – Access mode: <https://airflow.apache.org/>
2. Apache Iceberg [Electronic resource] / dremio – Access mode: <https://www.dremio.com/resources/guides/apache-iceberg/>
3. Fire Information for Resource Management System [Electronic resource] / Nasa – Access mode: <https://firms.modaps.eosdis.nasa.gov/>
4. open-meteo.com [Electronic resource] / Open Meteo – Access mode: <https://open-meteo.com/>
5. overpass-turbo.eu [Electronic resource] / Overpass-turbo.eu – Access mode: <https://overpass-turbo.eu/>
6. Palamar A., Karpinski M., Palamar M., Osukhivska H., Mytnyk M. Remote Air Pollution Monitoring System Based on Internet of Things. CEUR Workshop Proceedings, 2nd International Workshop on Information Technologies: Theoretical and Applied Problems, Ternopil, Ukraine, November 22–24, 2022. Vol. 3309. P. 194-204.
7. Wqi.info [Electronic resource] / World Air Quality Index – Access mode: <https://waqi.info/>
8. І.І. Бородій. Вибір технології проєктування програмної системи формування агрегованих надвеликих масивів даних. Збірник тез доповідей XII Міжнародної науково-практичної конференції молодих учених та студентів «Актуальні задачі сучасних технологій», 6-7 грудня 2023 року. Тернопіль: ФОП Паляниця В. А., 2023. С. 368-369.