

РОЗРОБЛЕННЯ ІНФОРМАЦІЙНОЇ ТЕХНОЛОГІЇ АНАЛІЗУ ТА ПРОГНОЗУВАННЯ ПОШИРЕННЯ ПИЛУ САХАРИ У ПОВІТРІ УКРАЇНИ ЗА ДАНИМИ ГРОМАДСЬКОГО МОНІТОРИНГУ НА ОСНОВІ PROPHET

Вінницький національний технічний університет

Анотація

Проведено дослідження проблеми прогнозування поширення пилу Сахари в атмосферному повітрі України на основі даних громадського моніторингу EcoCity. Розроблено інформаційну технологію, що включає етапи збору, очищення, об'єднання та аналізу даних, а також алгоритми прогнозування концентрацій PM1 за допомогою моделей Prophet та ARIMA.

Ключові слова: якість повітря, PM₁, машинне навчання, Prophet, ARIMA, прогнозування, пил Сахари, інформаційна технологія.

Abstract

The study investigates the problem of forecasting Saharan dust spread in the air of Ukraine based on public air-quality monitoring data from EcoCity. An information technology was developed, including data acquisition, cleaning, preprocessing, analysis, and forecasting using Prophet and ARIMA models.

Keywords: air quality, PM₁, machine learning, Prophet, ARIMA, forecasting, Saharan dust.

Вступ

Якість атмосферного повітря в Україні суттєво змінюється під впливом як локальних джерел забруднення, так і глобальних природних процесів. Одним із найбільш помітних та небезпечних явищ останніх років стало перенесення пилу Сахари, що періодично спричиняє підвищення концентрацій дрібнодисперсних частинок і погіршення видимості [1, 2]. Події 2021 та 2024 років підтвердили необхідність створення технологій, здатних прогнозувати такі ситуації [2].

Використання даних системи EcoCity створює можливість аналізувати фактичні вимірювання PM₁ та застосовувати сучасні інструменти моделювання для отримання короткострокових прогнозів, які можуть бути корисними для екологічних служб, науковців та населення [3]. Саме тому було поставлено задачу розробити інформаційну технологію, що забезпечить точне прогнозування концентрацій PM₁ у періоди перенесення пилу Сахари.

Методологія дослідження

У дослідженні використано більш ніж дев'ятнадцять тисяч записів EcoCity, що охоплюють дані з декількох станцій спостереження, включаючи їхні координати, часові мітки, індикатор PM₁ та допоміжні атрибути [4]. Дані були попередньо очищені від дублікатів і пропусків, приведені до єдиного формату, просторово узгоджені та об'єднані у єдиний датасет. У процесі розвідувального аналізу виявлено, що значення PM₁ мають асиметричний розподіл із наявністю значної кількості викидів, що потребувало уваги в подальших етапах моделювання.

Фрагмент підготовлених даних наведено на рисунку 1.

[27]:

	id_station	location	date	time_point	Column6	indicator	dimension	value	address
0	1315	49.23327 28.409161	2024-03-25	1	18	PM1.0	ug/m3	26.8422	Vinnytsia, str. Khmelnytsky shose, 95, VNTU, b...
1	1315	49.23327 28.409161	2024-03-25	2	20	PM1.0	ug/m3	25.4215	Vinnytsia, str. Khmelnytsky shose, 95, VNTU, b...
2	1315	49.23327 28.409161	2024-03-25	3	18	PM1.0	ug/m3	24.4250	Vinnytsia, str. Khmelnytsky shose, 95, VNTU, b...
3	1315	49.23327 28.409161	2024-03-25	4	20	PM1.0	ug/m3	22.8410	Vinnytsia, str. Khmelnytsky shose, 95, VNTU, b...
4	1315	49.23327 28.409161	2024-03-25	5	20	PM1.0	ug/m3	22.6225	Vinnytsia, str. Khmelnytsky shose, 95, VNTU, b...
...	...	...	...	...	...	...	...	...	...
19642	1876	49.2048561 28.5288355	2024-05-10	68	20	PM1.0	ug/m3	6.3255	Prov. Bohdana Khmel'nyts'koho, Vinnyts'ki Khutoru
19643	1876	49.2048561 28.5288355	2024-05-10	69	20	PM1.0	ug/m3	5.8295	Prov. Bohdana Khmel'nyts'koho, Vinnyts'ki Khutoru
19644	1876	49.2048561 28.5288355	2024-05-10	70	20	PM1.0	ug/m3	3.7260	Prov. Bohdana Khmel'nyts'koho, Vinnyts'ki Khutoru
19645	1876	49.2048561 28.5288355	2024-05-10	71	20	PM1.0	ug/m3	3.9925	Prov. Bohdana Khmel'nyts'koho, Vinnyts'ki Khutoru
19646	1876	49.2048561 28.5288355	2024-05-10	72	20	PM1.0	ug/m3	4.0595	Prov. Bohdana Khmel'nyts'koho, Vinnyts'ki Khutoru

19647 rows × 9 columns

Рисунок 1 - Фрагмент датасету EcoCity

З метою отримання найбільш точних прогнозів було розроблено алгоритм інформаційної технології, який охоплює всі основні етапи моделювання: визначення цільової ознаки, аналіз закономірностей даних, виявлення аномальних значень, формування вибірок, навчання та тюнінг моделей, а також оцінку їхньої точності [5]. Структура алгоритму наведена на рисунку 2.

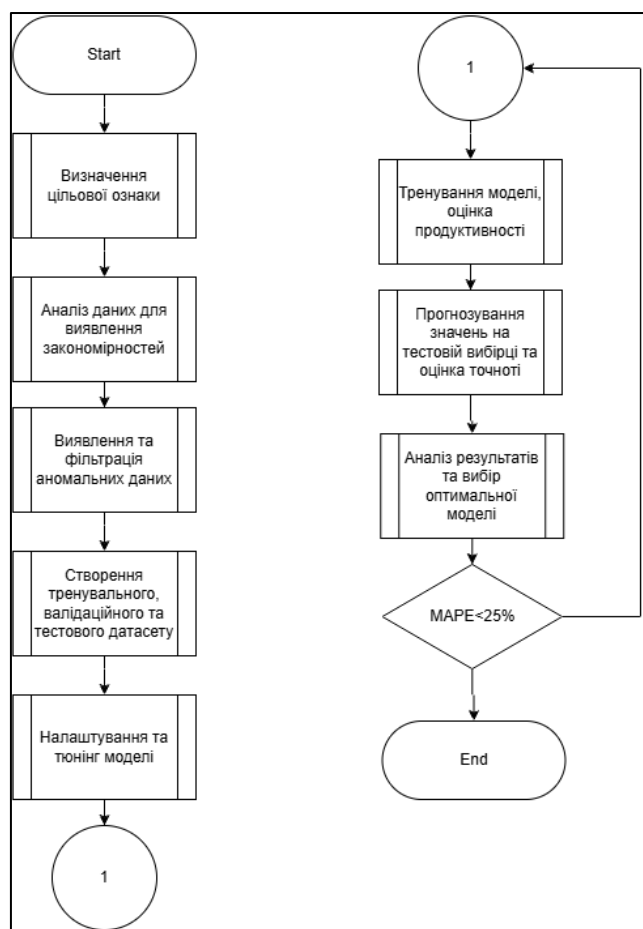


Рисунок 2 - Алгоритм інформаційної технології прогнозування поширення пилу Сахари

У моделюванні використано алгоритми ARIMA та Prophet. Для ARIMA проводився підбір порядків моделі, а для Prophet - оптимізація ключових гіперпараметрів, зокрема довжини сезонних періодів, порядку Фур'є-розкладу, параметрів зміни тренду (changepoint range та changepoint prior scale), типу сезонності (additive/multiplicative) та налаштувань holiday-ефектів, що дозволило сформувати тринадцять комбінацій, придатних для порівняння за метриками RMSE та MAPE [6, 7].

Результати дослідження

Побудовані моделі показали різний рівень точності, що дало можливість виділити оптимальну конфігурацію. Найкращий результат продемонструвала модель « Prophet_6_period_21_order », яка досягла значення MAPE = 7,38 % і продемонструвала високу відповідність прогнозних значень фактичним даним на тестовому наборі [8]. Це вказує на її здатність коректно відтворювати короткострокову динаміку концентрацій PM₁ під час перенесення пилу.

Графік порівняння фактичних тестових та прогнозованих значень наведено на рисунку 3.

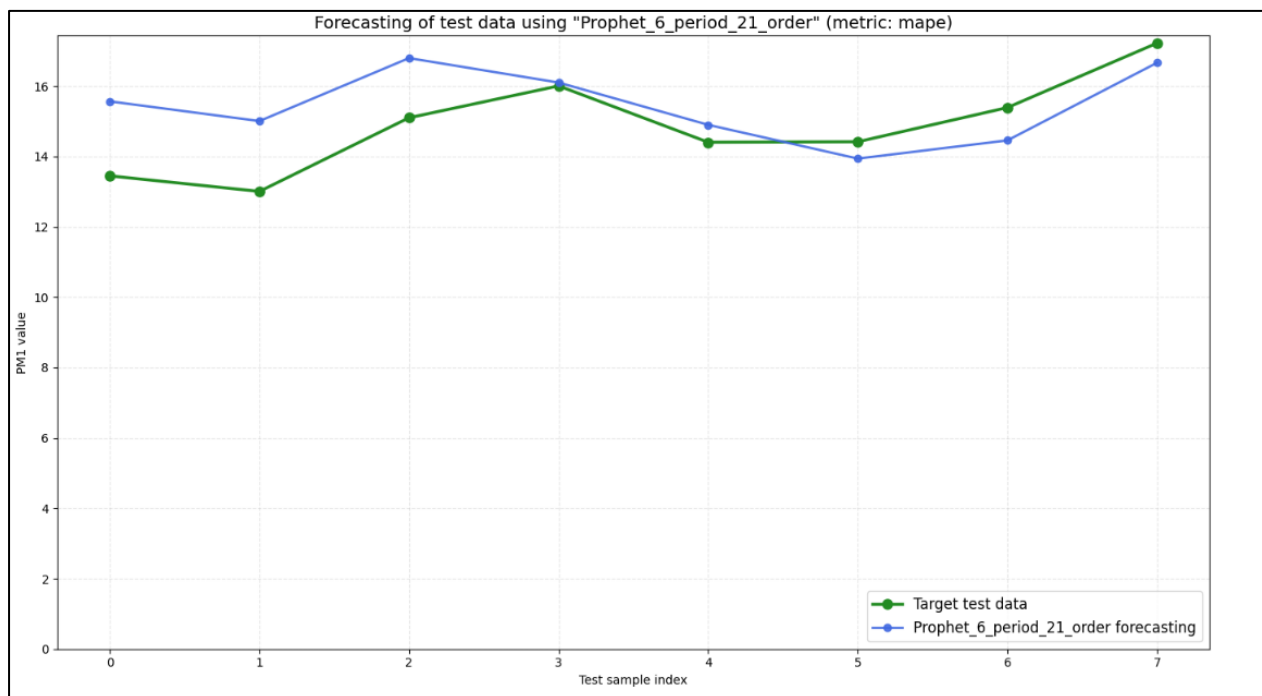


Рисунок 3 - Прогноз концентрацій PM₁ (модель Prophet_6_period_21_order)

Отримані результати свідчать, що модель Prophet є більш ефективною, порівняно з ARIMA, для часових рядів такого типу, оскільки здатна враховувати як короткострокові коливання, так і сезонні компоненти, властиві екологічним даним [8]. Підвищити точність прогнозування можна, лише застосувавши більш складні моделі гетероскедастичних часових рядів чи їх ансамблі, як нами було зроблено у роботі [9].

Висновки

У ході роботи створено інформаційну технологію, що поєднує етапи обробки, аналізу та моделювання даних громадського моніторингу EcoCity. Розроблена система забезпечує доволі точне прогнозування концентрацій PM₁ у періоди перенесення пилу Сахари та може використовуватися

для завчасного інформування населення, екологічних служб і органів місцевого самоврядування. Технологія має потенціал для інтеграції в реальні системи моніторингу та подальшого розширення шляхом включення додаткових моделей, параметрів забруднення чи географічних регіонів.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Sharp increase in Saharan dust intrusions over the western Euro-Mediterranean in February–March 2020–2022 [Електронний ресурс]. Режим доступу: <https://acp.copernicus.org/articles/24/4083/2024/>
2. World Health Organization. Ambient air pollution: Health impacts. WHO, 2023 [Електронний ресурс]. Режим доступу: [https://www.who.int/news-room/fact-sheets/detail/ambient-\(outdoor\)-air-quality-and-health](https://www.who.int/news-room/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health)
3. Карта моніторингу якості повітря EcoCity. [Електронний ресурс]. Режим доступу: <https://reborn.eco-city.org.ua/>
4. Air Quality Monitoring from EcoCity [Електронний ресурс]. Режим доступу: <https://www.kaggle.com/datasets/vbmokin/air-quality-monitoring-from-ecocity>
5. Мокін В. Б., Дратований М. В. Наука про дані: машинне навчання та інтелектуальний аналіз даних : електронний навчальний посібник комбінованого (локального та мережевого) використання [Електронний ресурс]. Вінниця : ВНТУ, 2024, 258 с. Режим доступу: <https://docs.vntu.edu.ua/card.php?id=8163>
6. FBProphet [Електронний ресурс]. Режим доступу: <https://pypi.org/project/fbprophet>
7. Hyndman R. J., Athanasopoulos G. Forecasting: principles and practice. 2nd ed. Oxford : Oxford University Press, 2014..
8. Sahara Influence forecasting and EDAv2 [Електронний ресурс]. Режим доступу: <https://www.kaggle.com/code/tarasskrynnik/sahara-influence-forecasting-and-edav2>
9. Копняк В. Є., Мокін В. Б., Жуков С. О., Варчук І. В., Скринник Т. В., Метод бустингу гетероскедастичних моделей для прогнозування концентрацій пилу Сахари в атмосферному повітрі України, Наукові праці ВНТУ [Електронний ресурс]. Вип. 2, Лип 2024. Режим доступу: doi <https://doi.org/10.31649/2307-5376-2024-2-28-38>

Скринник Тарас Васильович – студент групи 2ІСТ-24м, факультет інтелектуальних інформаційних технологій та автоматизації, Вінницький національний технічний університет, Вінниця, e-mail: tarasskrynnik1@gmail.com

Науковий керівник: **Мокін Віталій Борисович** – д-р техн. наук, проф., завідувач кафедри системного аналізу та інформаційних технологій, Вінницький національний технічний університет, Вінниця, e-mail: vbmokin@vntu.edu.ua

Skrynnik Taras V. – student of group 2IST-24m, faculty of Intellectual Information Technologies and Automation, Vinnytsia National Technical University, Vinnytsia, e-mail: tarasskrynnik1@gmail.com

Scientific supervisor: **Mokin Vitaliy B.** – Doctor of Technical Sciences, Professor, Head of the Department of System Analysis and Information Technologies, Vinnytsia National Technical University, Vinnytsia, e-mail: vbmokin@vntu.edu.ua