

ІНФОРМАЦІЙНА ТЕХНОЛОГІЯ АНАЛІЗУ ТА ПРОГНОЗУВАННЯ PM_{2.5} В АТМОСФЕРНОМУ ПОВІТРІ МІСТА ВІННИЦІ НА ОСНОВІ МОДЕЛЕЙ PROPHET, LSTM ТА ARIMA

Вінницький національний технічний університет

Анотація

В роботі розроблено інформаційну технологію аналізу та прогнозування стану атмосферного повітря міста Вінниці на основі методів машинного навчання. Дослідження проводилось для показника PM_{2.5}. Проведено огляд існуючих методів та технологій для аналізу та прогнозування даних. Виконано прогнозування даних з використанням трьох моделей машинного навчання: Prophet, LSTM та ARIMA. Проведено оцінку результатів.

Ключові слова: інформаційна технологія, моніторинг, повітря, якість повітря, прогноз, датасет, модель..

Abstract

In the work, the information technology of analysis and forecasting of the state of atmospheric air in the city of Vinnytsia was developed based on machine learning methods. The study was conducted for the PM_{2.5} indicator. An overview of existing methods and technologies for data analysis and forecasting was conducted. Data forecasting was performed using three machine learning models: Prophet, LSTM, and ARIMA. The results were evaluated.

Keywords: information technology, monitoring, air, air quality, forecast, dataset, model.

Актуальність дослідження

Якість атмосферного повітря безпосередньо впливає на якість життя людини та має важливе значення для нормального функціонування її організму. На сучасному етапі відзначається стрімке збільшення впливу людської діяльності на оточуюче середовище, зокрема на атмосферний шар планети.

Розширення автотранспорту, розвиток промисловості, неконтрольована рубка лісів та використання технологій, що не відповідають екологічним стандартам, призвели до погіршення якості повітря у багатьох містах України та світу. У зв'язку з цим виникла проблема поліпшення та відновлення якості атмосферного повітря. Для вирішення цієї проблеми важливо вдосконалити системи державного моніторингу якості повітря та впровадити нові технології у цій сфері [1].

Система моніторингу якості повітря закладає основу для прогнозування та відстеження якості повітря, а більш точне прогнозування якості повітря стає можливим завдяки покращенню даних моніторингу. Прогнозування якості повітря є ефективним підходом до захисту здоров'я людини шляхом надання попередніх попереджень про шкідливе забруднення повітря [2-4].

Постановка задачі

Для розв'язання задачі прогнозування ціни на природній газ необхідно розв'язати такі задачі:

- Виконати збір даних;
- Провести огляд існуючих рішень;
- Провести розвідувальний аналіз даних;
- Виконати побудову моделей машинного навчання;
- Провести прогнозування та оцінити результати.

В роботі пропонується дослідити такі моделі машинного навчання:

1. Prophet;
2. LSTM;
3. ARIMA.

Отримані результати

Під час виконання дослідження використовувалась система Kaggle. В даній системі створено новий датасет та підключенно його (рис.1).

	stations_id	stations_name	Lat	Long	stations_time	stations_offset	stations_params_id	stations_params_key	stations_params_name	stations_param:
0	650	дім	49.343484	28.72983	2023-09-18 01:00:53	0	A2	PM10	PM10	Пил 10 мкм
1	650	дім	49.343484	28.72983	2023-09-18 01:00:53	0	A3	PM2.5	PM2.5	Пил 2.5 мкм
2	650	дім	49.343484	28.72983	2023-09-18 01:00:53	0	C1	VOC (H ₂ CO)	VOC (H ₂ CO)	ЛОС (Формаль,
3	650	дім	49.343484	28.72983	2023-09-18 01:00:53	0	E1	Temperature	Temperature	Температура
4	650	дім	49.343484	28.72983	2023-09-18 01:00:53	0	E2	Humidity	Humidity	Вологість
...
148484	90	vinnytsia-90	49.217734	28.449795	2023-09-17T21:35:00Z	0	A3	PM2.5	PM2.5	Пил 2.5 мкм
148485	90	vinnytsia-90	49.217734	28.449795	2023-09-17T21:35:00Z	0	A2	PM100	PM100	Пил 10 мкм
148486	90	vinnytsia-90	49.217734	28.449795	2023-09-17T21:35:00Z	0	E1	Temperature	Temperature	Температура
148487	90	vinnytsia-90	49.217734	28.449795	2023-09-17T21:35:00Z	0	E2	Humidity	Humidity	Вологість
148488	90	vinnytsia-90	49.217734	28.449795	2023-09-17T21:35:00Z	0	E3	Pressure	Pressure	Атмосферний т

148489 rows × 1001 columns

Рисунок 1 – Фрагмент датасету

Даний датасет містить наступні колонки:

- Station_id. Ідентифікатор станції, що виконувала виміри;
- Station_name. Назва станції;
- Lat. Координати розташування станції (ширина);
- Long. Координати розташування станції (довжина);
- Station_time. Час в який станція зробила вимірювання;
- Station_offset.
- Station_params_id. Ідентифікатор параметру, що вимірювався;
- Station_params_key. Ключ параметра, що вимірювався (назва показника PM10, PM2.5 тощо);
- Station_params_name. Ключ параметра, що вимірювався (назва показника PM10, PM2.5 тощо);
- station_params_localName. Ключ параметра, що вимірювався (назва показника PM10, PM2.5 тощо)українською;
- Station_params_unit. Одиниця вимірювання;
- Station_params_localUnit. Одиниця вимірювання українською;
- Station_params_valu. Вимірне значення;
- Station_params_cr.
- Station_params_time. Час коли було здійснено вимірювання.

Проведено візуалізацію середніх показників за кожною станцією погодинно та за одну добу. Це дасть змогу краще зрозуміти динаміку даних у часі та дозволить краще налаштувати моделі для прогнозування (рис.2-3). Також визначено аномальні значення в датасеті. Це необхідно для того, щоб

знайти дані які вибиваються з загального контексту, завдяки чому можна буде краще натренувати модель, що дозволить отримати кращі прогнозовані значення (рис.4).

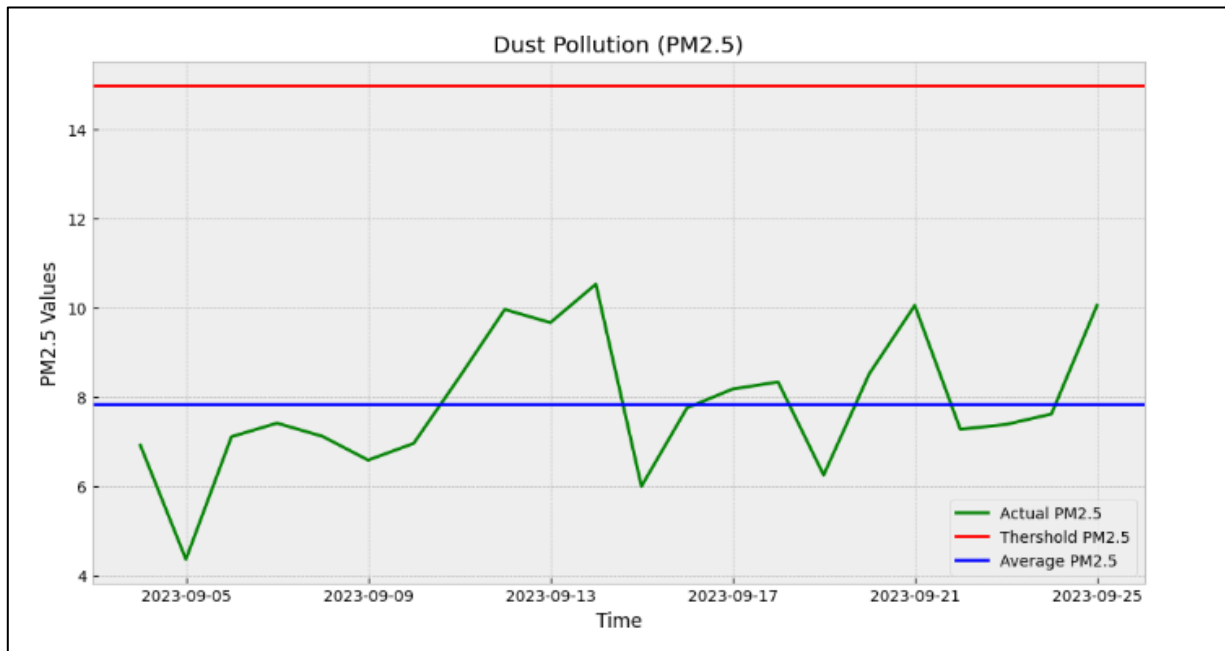


Рисунок 2 – Графік динаміки середньодобових значень частинок PM2.5

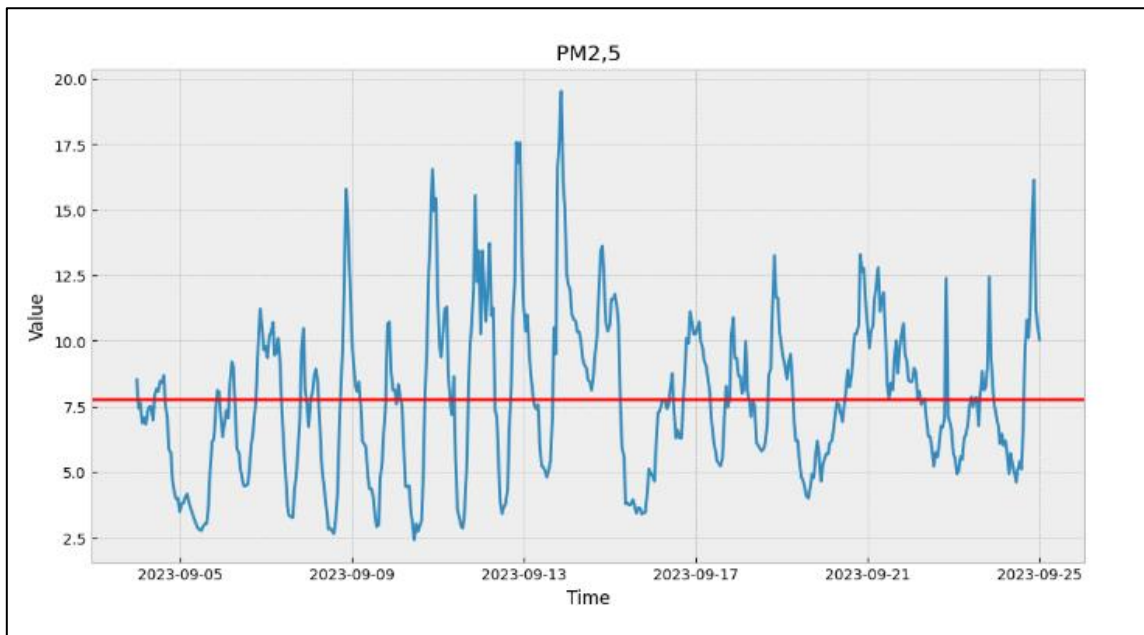


Рисунок 3 – Графік динаміки погодинних значень частинок PM2.5

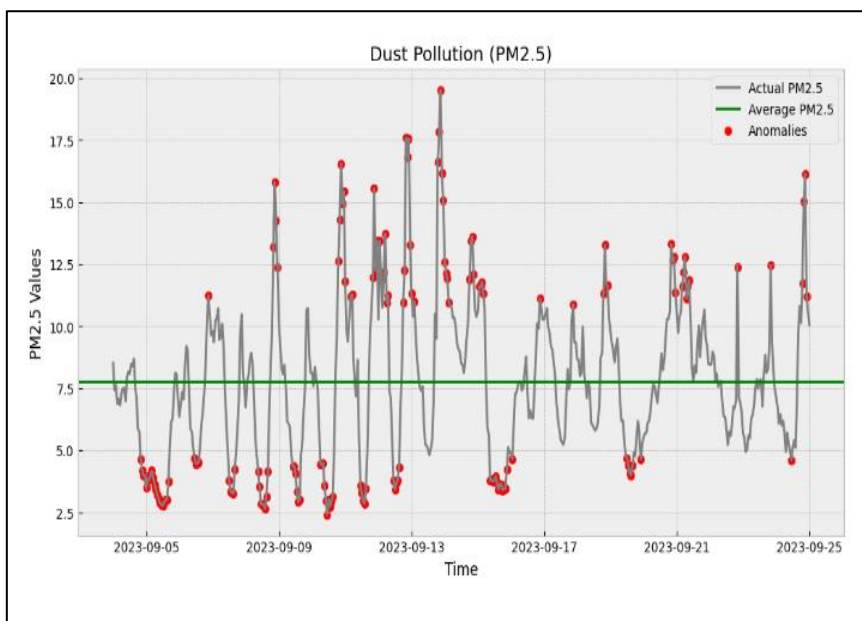


Рисунок 4 – Графік аномальних значень

На рисунку 5 наведено результати прогнозування моделей за наступними метриками: r^2_score , $rmse$ та mae .

	name	r2_score_train	r2_score_test	rmse_train	rmse_test	mae_train	mae_test
0	Prophet	0.91	0.73	0.93	1.69	0.71	1.47
1	LSTM	0.99	0.89	0.426	1.502	0.295	1.015
2	ARIMA	0.85	0.82	1.17	1.38	0.78	0.91

Рисунок 5 – Результати роботи моделей

Графік прогнозу наведено на рисунку 6.

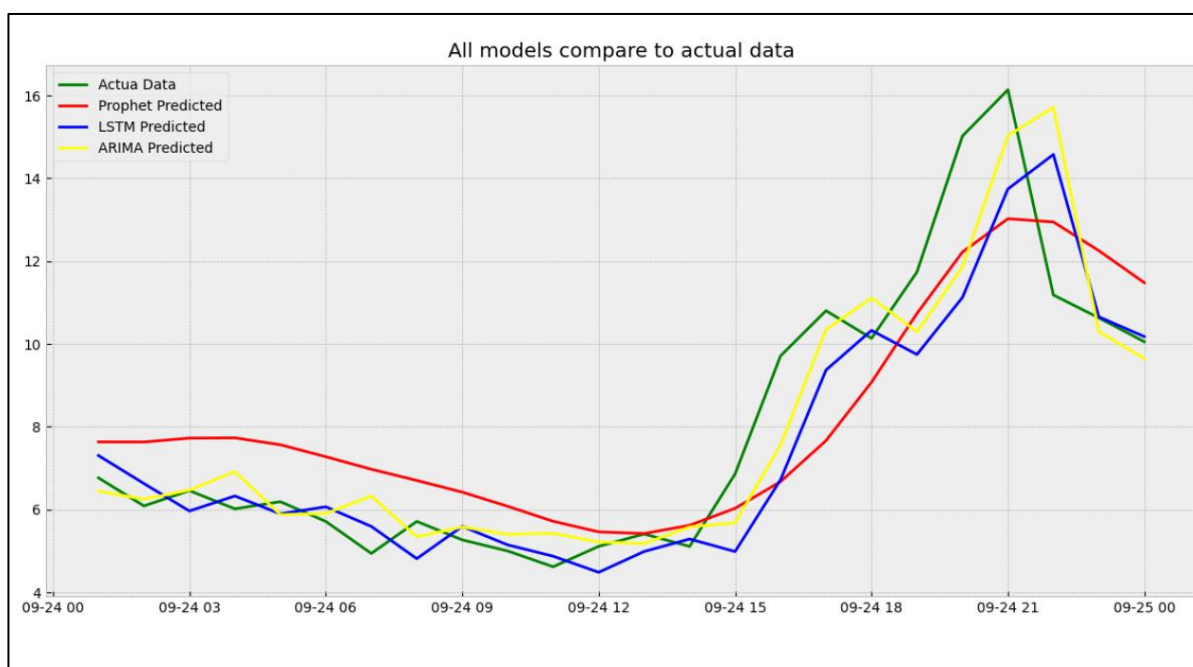


Рисунок 6 – Графік прогнозу

Висновки

В результаті дослідження створено інформаційну технологію аналізу та прогнозування якості атмосферного повітря міста Вінниці. Проведено огляд існуючих інформаційних технологій. Здійснено розвідувальний аналіз даних та побудовано три моделі машинного навчання: Prophet, LSTM, ARIMA. Найкращий результат прогнозування продемонструвала модель LSTM, яка за метрикою $r2_score$ показала результат 0,89 для тестового датасету. Прогнозування проводилось для показника PM2.5.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Вінницький обласний центр контролю та профілактики хвороб МОЗ України. URL: http://cgz.vn.ua/problematika-gromadskogo-zdorovya/problematika-gromadskogo-zdorovya_455.html (дата звернення: 01.12.2023).
2. В. Б. Мокін, Д. Ю. Дзюняк, К. О. Бондалетов, і В. В. Олійник, «МЕТОД І ТЕХНОЛОГІЯ МОНІТОРИНГУ СТАНУ АТМОСФЕРНОГО ПОВІТРЯ ЗА ДОПОМОГОЮ УНІВЕРСАЛЬНОЇ ІНФОРМАЦІЙНО-ВИМІРЮВАЛЬНОЇ СИСТЕМИ З ВИКОРИСТАННЯМ МОБІЛЬНИХ ПРИСТРОЇВ», *НаукПраці ВНТУ*, вип. 4, Січ 2016.
3. PM2.5 forecasting for an urban area based on deep learning and decomposition method / N. Zaini et al. *Scientific Reports*. 2022. Vol. 12, no. 1. URL: <https://doi.org/10.1038/s41598-022-21769-1> (date of access: 01.12.2023).
4. Artificial intelligence technologies for forecasting air pollution and human health: a narrative review / S. Subramaniam et al. *Sustainability*. 2022. Vol. 14, no. 16. P. 9951. URL: <https://doi.org/10.3390/su14169951> (date of access: 15.10.2023).

Пінчук Владислав Павлович – студент групи 2ІСТ-22м, Факультет інтелектуальних інформаційних технологій та автоматизації, Вінницький національний технічний університет, Вінниця, e-mail: vlpin4uk228@gmail.com

Жуков Сергій Олександрович – к.т.н., доцент кафедри системного аналізу та інформаційних технологій, Вінницький національний технічний університет, Вінниця, e-mail: sazhukov@gmail.com

Pinchuk Vladislav P. - student of group 2IST-22m, Faculty of Intelligent Information Technologies and Automation, Vinnytsia National Technical University, Vinnytsia, e-mail: vlpin4uk228@gmail.com

Zhukov Sergey O. – Ph.D., associate professor of the System Analysis and Information Technologies Department, Vinnytsia National Technical University, Vinnytsia, e-mail: sazhukov@gmail.com