

# ІНФОРМАЦІЙНА ТЕХНОЛОГІЯ АНАЛІЗУ ТА ПЕРЕДБАЧЕННЯ ДІОКСИДУ ВУГЛЕЦЮ ЗА ДАНИМИ ГРОМАДСЬКОГО МОНІТОРИНГУ АТМОСФЕРНОГО ПОВІТРЯ У М. ВІННИЦІ

Вінницький національний технічний університет

## Анотація

Проведено розвідувальний аналіз Kaggle-додатку „Air Quality Monitoring from EcoCity” з даними вмісту різних забруднюючих речовин у повітрі. Для вирішення поставленої задачі побудовано інтелектуальні моделі передбачення даних Linear SVR, XGB Regressor та RandomForestRegressor. Оптимальною є Linear SVR. Розроблено інформаційну технологію передбачення вмісту вуглекислого газу за рахунок удосконалення методів машинного навчання та розвідувального аналізу, що дозволяє підвищити точність цього передбачення у м. Вінниці.

**Ключові слова:** моніторинг якості атмосферного повітря, машинне навчання, передбачення, м. Вінниця.

## Abstracts

An exploratory analysis of the Kaggle dataset “Air Quality Monitoring from EcoCity” with data on the content of various pollutants in the air is carried out. To solve this problem, we built intelligent data prediction models Linear SVR, XGB Regressor and RandomForestRegressor. The optimal one is Linear SVR. An information technology for predicting the carbon dioxide content has been developed by improving machine learning and intelligence analysis methods, which allows to increase the accuracy of this prediction in Vinnytsia city.

**Keywords:** air quality monitoring, machine learning, prediction, Vinnytsia city.

## Вступ

Зростання рівня CO<sub>2</sub> є однією з головних причин глобального потепління. Передбачення концентрації вуглекислого газу допомагає оцінити темпи підвищення температури, екстремальних погодних явищ і довгострокові зміни клімату. Адже високий вміст CO<sub>2</sub> негативно впливає на екосистеми, зокрема спричиняє окислення океанів, що шкодить морській фауні та флорі, а також впливає на продуктивність сільського господарства. призводить до економічних втрат через погіршення умов ведення сільського господарства, пошкодження інфраструктури внаслідок екстремальних погодних умов, а також підвищення витрат на охорону здоров'я через зростання частоти природних катастроф.

Аби уникнути майбутніх наслідків, треба вміти передбачати вміст вуглекислого газу в майбутньому за для розробки нових методів та практик для боротьби з підвищеним вмістом CO<sub>2</sub> та оцінювання їх ефективності.

Метою даного дослідження є підвищення точності передбачення вмісту вуглекислого газу в атмосферному повітрі використовуючи методи машинного навчання.

## Розвідувальний аналіз даних

Використовуючи датасет «Air Quality Monitoring from EcoCity» на онлайн-платформі Kaggle (<https://www.kaggle.com/datasets/vbmokin/air-quality-monitoring-from-ecocity>). Було проведено розвідувальний аналіз, в результаті якого було знайдено та вилучено аномалії даних, що могли б вплинути на кінцевий результат прогнозування. На рисунку 1 зображено таку аномалію даних.

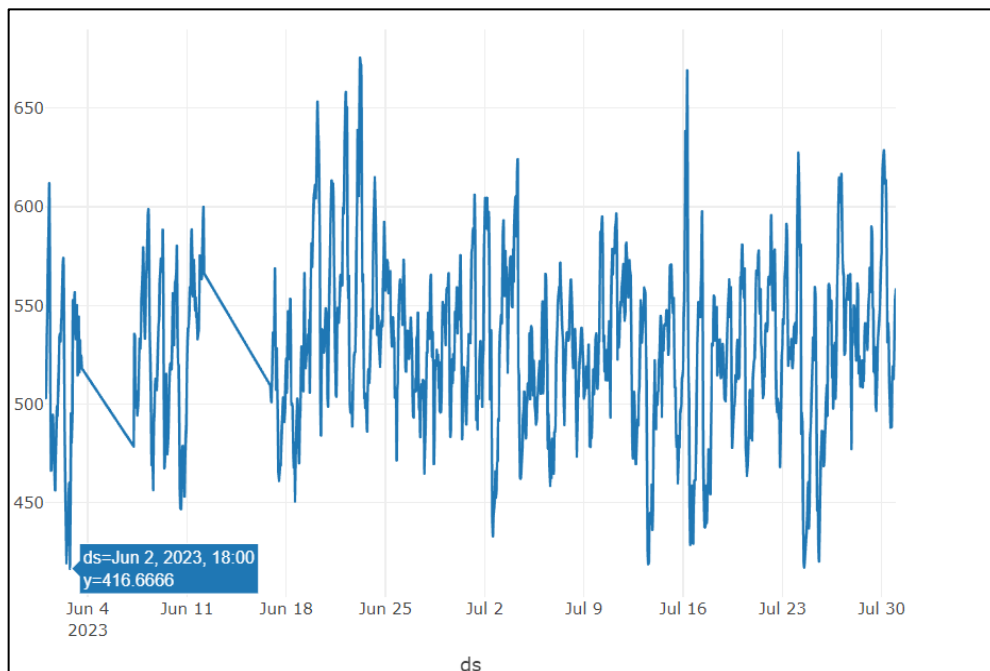


Рисунок 1 – Аномалія 2.06.23р. зниження вмісту CO<sub>2</sub> в атмосфері

Дана аномалія могла виникнути з безлічі причин, серед яких короткий робочий день підприємств поблизу або зменшений потік автотранспорту, оскільки це було в п'ятницю ввечері або через певні погодні умови, що посприяли зменшенню вмісту вуглекислого газу в атмосферному повітрі.

Після урахування всіх аномалій було побудовано біля 2-х десятків інтелектуальних моделей з різними параметрами з використанням методів машинного навчання [1, 2], задля виявлення оптимальної з них, що підходить під дане дослідження. На рисунку 2 зображено всі побудовані моделі з метрикою, визначеною на валідаційних даних.

	name_model	type_data	mape
12	Linear SVR	valid	1.93953
13	Random Forest Regressor	valid	1.991942
15	XGB Regressor	valid	2.025112
8	ARIMA_auto	valid	2.039046
10	KNeighbors Regressor	valid	2.18653
9	Linear Regression	valid	2.498551
16	MLP Regressor	valid	2.814234
11	Support Vector Machines	valid	2.996162
14	Bagging Regressor	valid	3.003043
5	Prophet_30_days_12_order	valid	4.728317
7	Prophet_365_days_12_order	valid	12.123947
1	Prophet_4_days_12_order	valid	17.116291
0	Prophet_4_days_3_order	valid	21.692297
6	Prophet_365_days_3_order	valid	25.851544
3	Prophet_7_days_12_order	valid	28.985652
2	Prophet_7_days_3_order	valid	32.211638
4	Prophet_30_days_3_order	valid	43.620066

Рисунок 2 – Побудовані моделі передбачення

Як видно з рисунку 2, найкращою моделлю є Linear SVR з метрикою MAPE (середня абсолютна похибка) що складає 1,93953%.

### Результати дослідження

Взявши за основу Linear SVR далі було проведено навчання з використанням розширеного датасету, що включав в себе тренувальні та валідаційні дані. В цей раз похибка склала лише 1,927118% (рис. 3).

name_model	type_data	mape
12	Linear SVR	valid 1.927118

Рисунок 3 – Середня абсолютна похибка на розширеному датасеті

Далі було порівняно результати передбачення моделі з реальними даними (рис. 4).

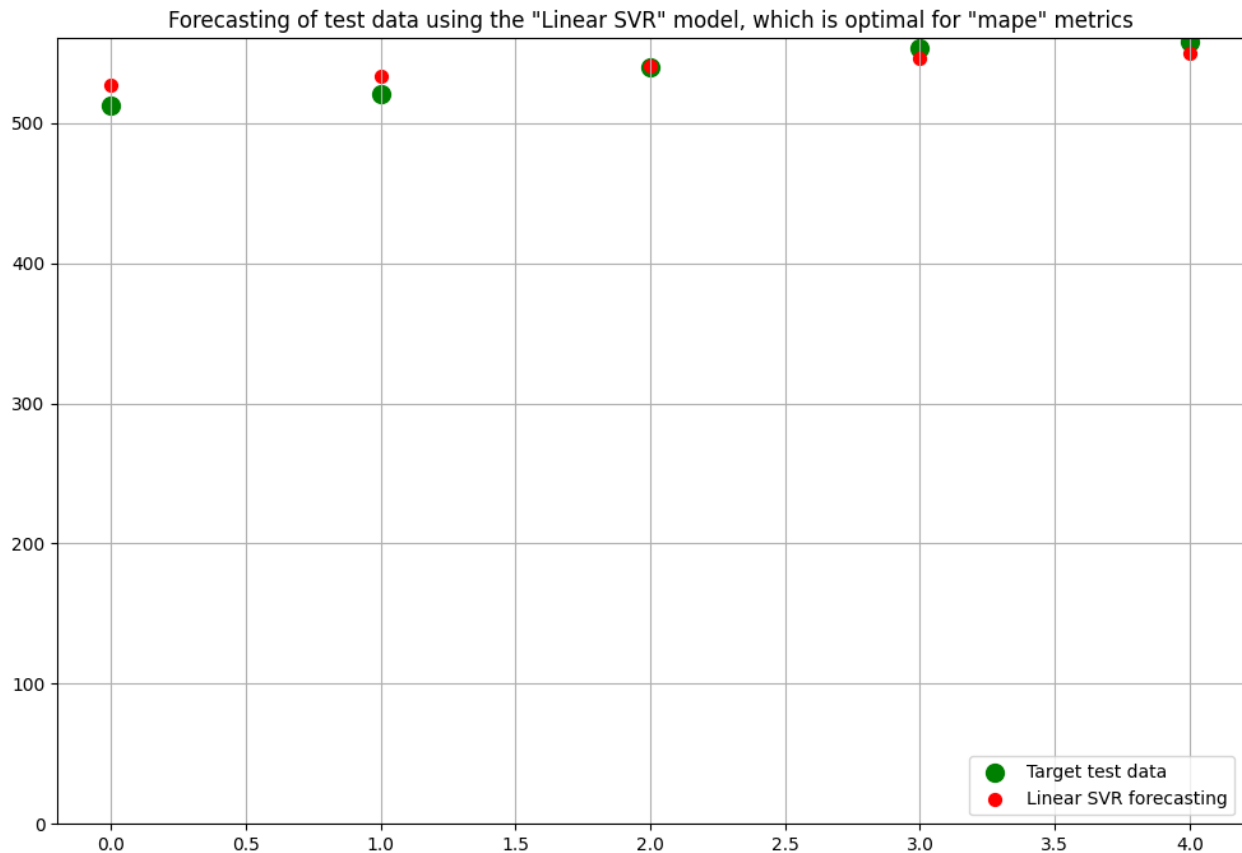


Рисунок 4 – Результати порівняння реальних даних з даними передбачення

На даному графіці зображено червоними крапками результати передбачення, а зеленими крапками – реальні дані. Оскільки червоні точки знаходяться дуже близько до зелених це свідчить про гарне навчання та здатність моделі виконувати поставлену задачу.

### Висновки

В даному дослідженні виконано побудову та налаштування різних моделей машинного навчання. Проведено розвідувальний аналіз датасету, з якого були вилучені аномалії даних. Було виконано навчання на валідаційних даних за для визначення оптимальної моделі. Нею виявилась модель Linear SVR з метрикою MAPE і складала 1,93953%. Потім було використано розширений датасет, що включав в себе валідаційні та тренувальні дані. Що дало вже 1,927118% похибки. Також було проаналізовано результати передбачення та порівняно з реальними результатами. Отже, побудована інтелектуальна модель добре підходить для виконання поставленої задачі та має перспективи в практичному використанні.

## СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. A Comprehensive Guide to Machine Learning [Електронний ресурс] / S. Nasiriany et al. 2018. 176 с. – Режим доступу:

<https://github.com/juanmartinsantos/books/blob/main/docs/Machine%20Learning/A%20Comprehensive%20Guide%20to%20Machine%20Learning.pdf>

2. Наука про дані: машинне навчання та інтелектуальний аналіз даних : електронний навчальний посібник комбінованого (локального та мережевого) використання [Електронний ресурс] / В. Б. Мокін, М. В. Дратованій – Вінниця : ВНТУ, 2024. – 258 с. – Режим доступу: <https://docs.vntu.edu.ua/card.php?id=8163>

**Губар Олексій Михайлович** — студент групи ІІСТ-23м, Вінницький національний технічний університет, Вінниця, [olexiihubar@gmail.com](mailto:olexiihubar@gmail.com)

**Мокін Віталій Борисович** – д-р техн. наук, проф., завідувач кафедри системного аналізу та інформаційних технологій, Вінницький національний технічний університет, Вінниця, e-mail: [ybmokin@vntu.edu.ua](mailto:ybmokin@vntu.edu.ua)

**Hubar Oleksii M.**– student of group IIST-23m, Vinnytsia National Technical University, Vinnytsia, [olexiihubar@gmail.com](mailto:olexiihubar@gmail.com)

**Mokin Vitalii B.** – Dr. Tech. Sciences, Prof., Head of the Department of System Analysis and Information Technologies, Vinnytsia National Technical University, Vinnytsia, e-mail: [ybmokin@vntu.edu.ua](mailto:ybmokin@vntu.edu.ua)