

# ІНФОРМАЦІЙНА СИСТЕМА АНАЛІЗУ ТА ПРОГНОЗУВАННЯ ЯКОСТІ ВОДИ В РІЧЦІ ПІВДЕННИЙ БУГ

Вінницький національний технічний університет

## Анотація

*Створено систему для аналізу та прогнозування даних в річці Південний Буг з використанням технологій аналізу та прогнозування даних, геоінформаційних технологій, баз даних, технологій веб-розробки, а також було проведено її успішне випробування за даними моніторингу річки Південний Буг за 2000-2023 роки.*

**Ключові слова:** веб-додаток, інформаційна система, ГІС, бази даних, аналіз даних, прогнозування.

## Abstract

*A system for analyzing and forecasting data in the Southern Bug River was created using data analysis and forecasting technologies, geographic information technologies, databases, web development technologies, and its successful testing was carried out using data from the Southern Bug River monitoring for 2000-2023*

**.Keywords:** web application, information system, GIS, databases, data analysis, forecasting.

## Актуальність дослідження

Проблема забруднення води на сьогодні є однією з найбільш актуальних екологічних проблем, рівноцінною з забрудненням повітря та вимиранням тварин. Збільшення кількості шкідливих речовин прямо впливає на екосистему всього водного басейну, спричиняючи серйозні наслідки. Зокрема, підвищення концентрації токсичних речовин у поверхневих водах негативно впливає на екосистему, призводячи до масового вимирання живих організмів в забрудненому середовищі [1-5]. Сучасні джерела забруднення вод включають несанкціоновані викиди сміття у річки, скиди від підприємств і інші джерела [1].

На сьогоднішній день особливо актуальною є задача створення інформаційної системи, призначеної для вивчення динаміки зміни концентрації шкідливих речовин, а також аналізу та прогнозування загальної якості води.

## Створення системи аналізу та прогнозування якості води у річці Південний Буг

На основі аналізу зібраних вхідних даних про вміст шкідливих речовин в поверхневих водах Південного Бугу за останніх більш ніж 20 років, за основу яких був взятий датасет під назвою «River Water Quality EDA and Forecasting» [2], а також дані Державного агентства водних ресурсів України [3], створено структуру інформаційної системи аналізу та прогнозування якості води Південного Бугу. Візуалізація структури приведена на рисунку 1.

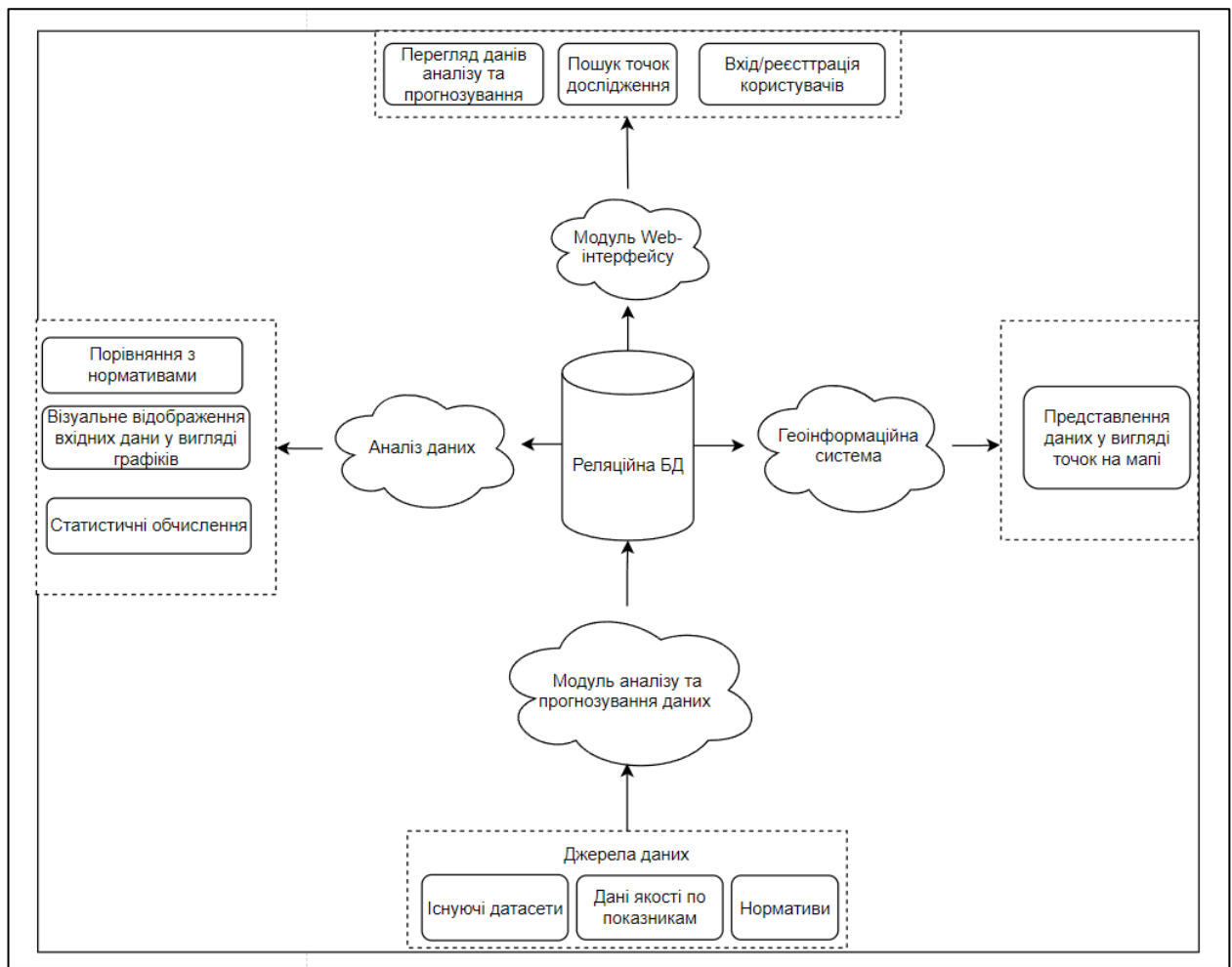


Рис. 1. Архітектура системи аналізу та прогнозування якості води

Одним з основних модулів є модуль аналізу та прогнозування даних, даний модуль виконується на платформі Kaggle, де здійснюються основні обрахунки а їх результати імпортуються в спроектовану веб-систему. Дане рішення зумовлене тим що дана платформа включає в себе всі необхідні функції та бібліотеки для проведення аналізу та прогнозування. На рисунках 2 та 3 зображено елементи коду імпорту та обробки вхідних даних для проектування системи.

```

# Initialize an empty list to store DataFrames
dfs = []

# Specify the columns you want to keep
selected_columns = ['Post_ID', 'Post_Name', 'Controle_Date', 'BSK5', 'Zavisli', 'Kisen', 'Sulfat', 'Hlorid', 'Amoniy', 'Nitrat', '

for csv_file in csv_files:
    # Assuming the CSV files have a header row
    current_data = pd.read_csv(csv_file, sep=';', header=0, usecols=selected_columns)

    # Filter rows where 'Riverbas_Name' is equal to 'Південний Буг'
    current_data = current_data[current_data['Post_Name'].str.startswith('п. Південний Буг')]

    dfs.append(current_data)

```

Рисунок 2 – Імпорт даних моніторингу Державного агентства водних ресурсів України

```

other_data = pd.read_csv('../input/wq-southern-bug-river-01052021/PB_All_2000_2021.csv', sep=';', header=0)

# Rename columns in combined_data
combined_data.rename(columns={
    'Post_ID': 'id',
    'Controle_Date': 'date',
    'Amoniy': 'NH4',
    'BSK5': 'BSK5',
    'Zavisli': 'Suspended',
    'Kisen': 'O2',
    'Nitrat': 'N03',
    'Fosfat': 'P04',
    'Hlorid': 'CL',
    'Nitrit': 'N02',
    'Sulfat': 'S04'
}, inplace=True)

```

Рисунок 3 – Імпорт даних датасету «River Water Quality EDA and Forecasting» та обробка даних Держводагентства

На основі імпортованих даних було проведено розвідувальний аналіз, а також побудовано моделі для прогнозування, проведено навчання моделей та виведення результатів на тестових, тренувальних та валідних даних, результати аналізу та прогнозування було експортовано з нотбука Kaggle та імпортовано в спроектовану систему. Результати тестування системи з графіком вмісту кисню в воді на станції м. Ладизин зображено на рисунку 4.

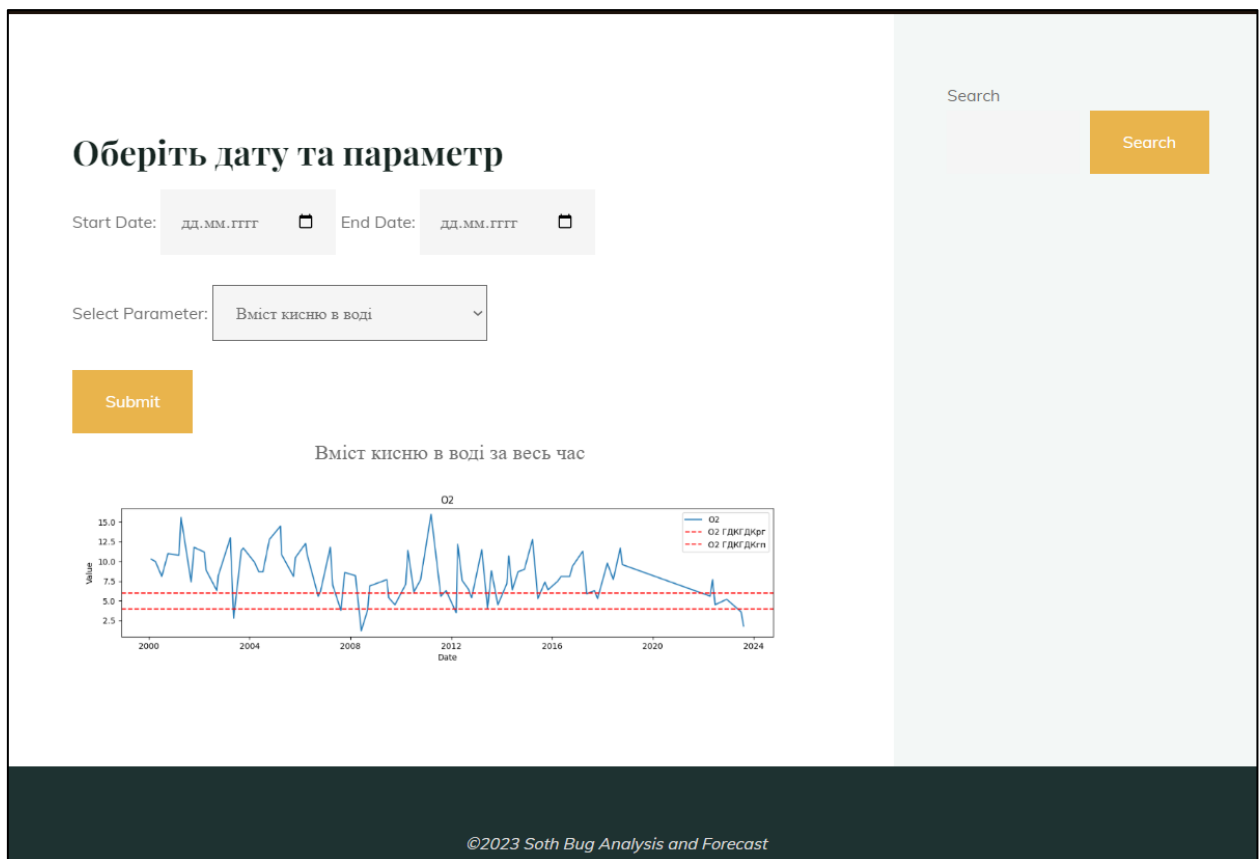


Рис. 4 – Зображення вмісту кисню на станції м. Ладизин в спроектованій системі.

Також, важливим блоком даної системи є геоінформаційні технології, що представляють собою мапу з точками дослідження для покращеного орієнтації в розробленій веб-системі – кожна точка на мапі

представляє собою посилання на сторінку кожної станції, в якій реалізовано подальші функції відображення даних. Приклад мапи та код реалізації розробленої мапи зображено на рисунках 5 та 6.

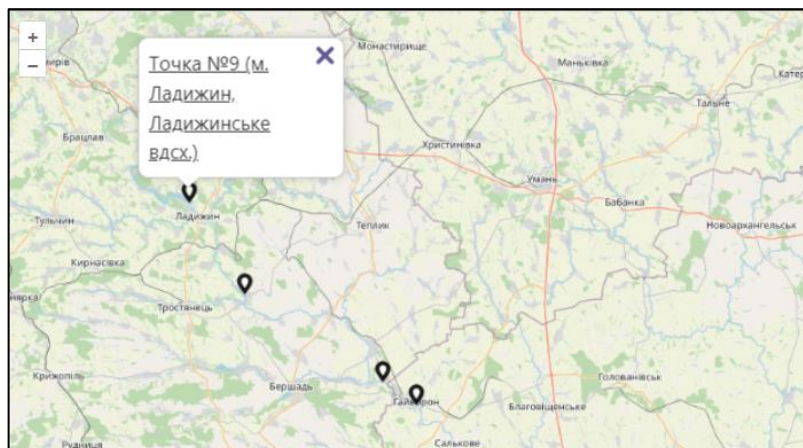


Рисунок 5 – Додана в систему мапа з точками станцій дослідження

```
[osm_map_v3 map_center="43.7586,6.9242" zoom="11.0" width="100%" height="450" tagged_type="page" marker_name="mic_black_pinother_02.png" map_border="thin solid grey" tagged_param="cluster" tagged_color="black"]
```

Рисунок 6 – Код реалізації мапи в системі

Заключним етапом розроблення системи є тестування модуля прогнозування даних. В даному блоці користувачу надано вибір моделей а також станцій досліджень, ключової ознаки та факторів для моделювання. Приклад сторінки вибору зображено на рисунку 7.

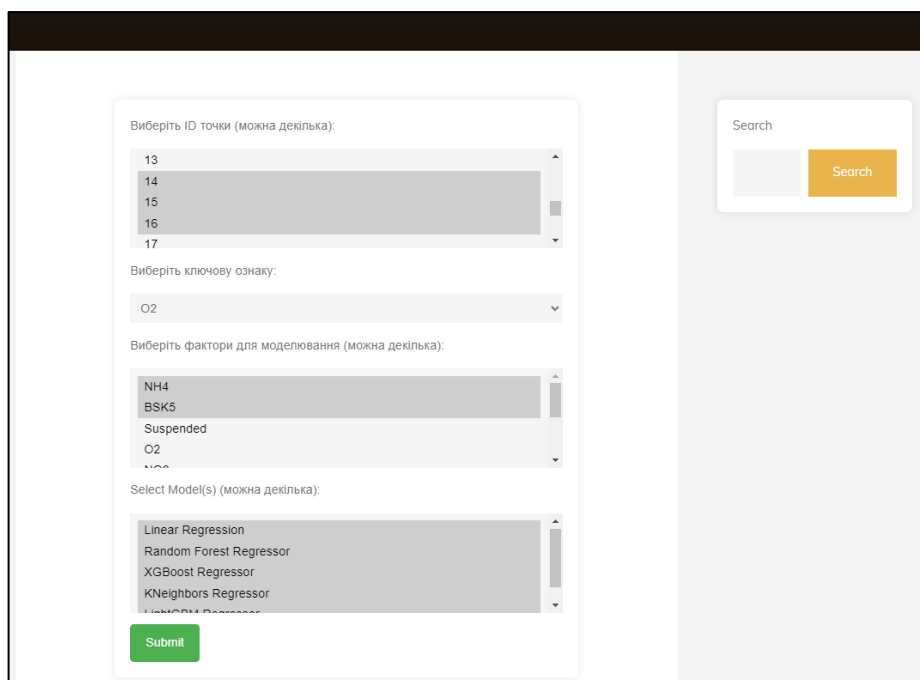


Рисунок 7 – Вибір параметрів для відображення результатів прогнозування

В результаті користувач отримує оцінку  $r2\_score$  для кожної моделі та обирає модель що показала найкращі результати. Результати відображення точності моделей а також вибір найкращої з них зображено на рисунках 8 та 9.

Точність моделей для обраних моделей			
model	Модель	train_score	valid_score
1	Random Forest Regressor	0.7	0.63
0	Linear Regression	0.66	0.58
2	XGBoost Regressor	0.99	0.70
4	LightGBM Regressor	0.66	0.58
3	KNeighbors Regressor	1.00	0.59

Рисунок 9 – Точність кожної обраної моделі

Найкраща модель		
Модель	train_score	valid_score
Random Forest Regressor	0.7	0.63

Рисунок 10 – Вибір найкращої моделі для обраних даних

На основі моделі що була визнаною найкращою для обраних даних, будуються графіки що наочно показують відхилення від тестових, валідних та тренувальних даних. Дані графіки зображено на рисунках 11 – 13.

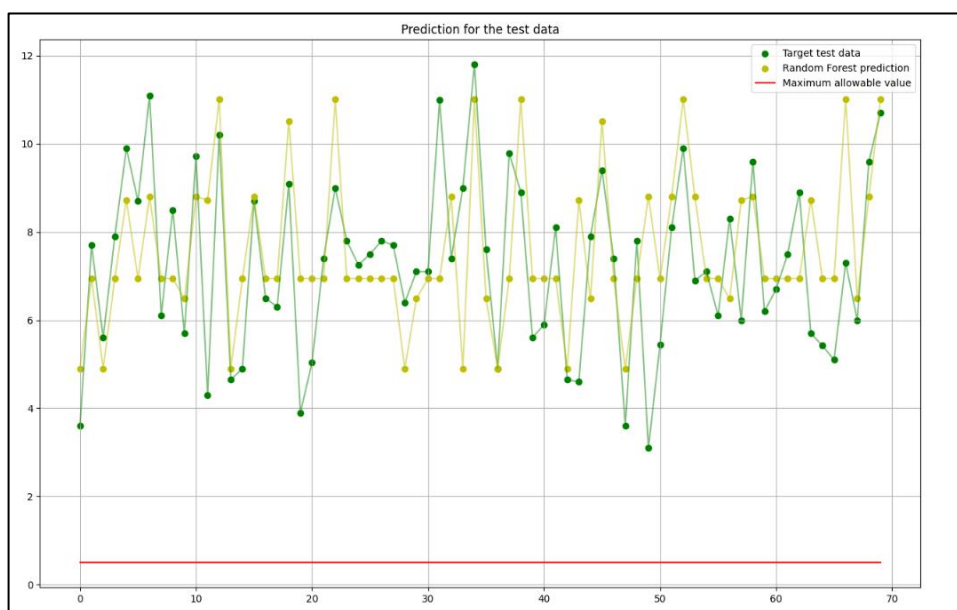


Рисунок 11 – Графік прогнозування моделі RandomForestRegressor для тестових даних

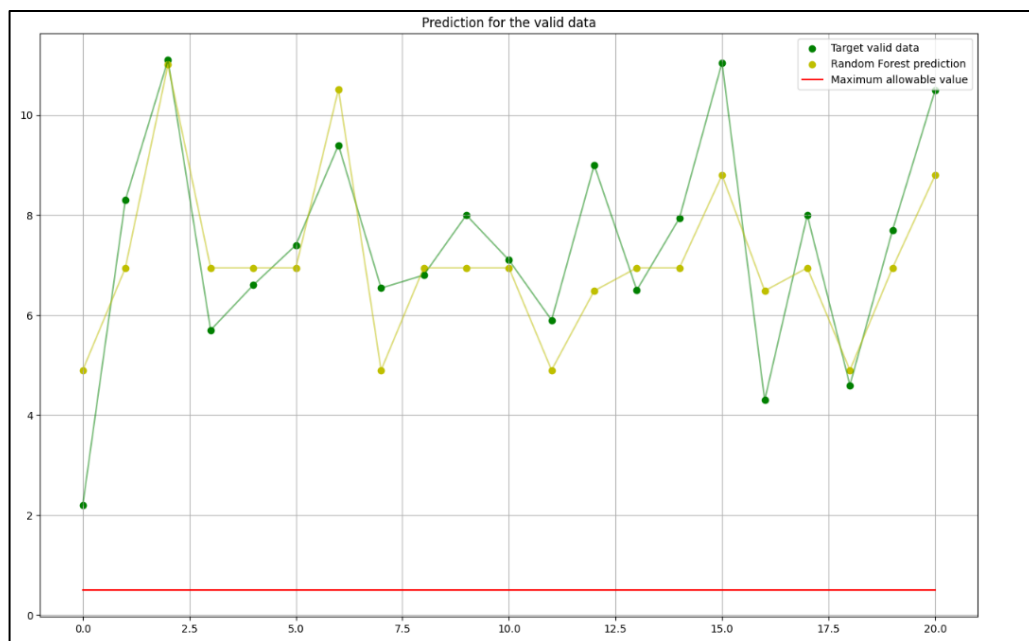


Рисунок 12 – Графік прогнозування моделі RandomForestRegressor для валідних даних

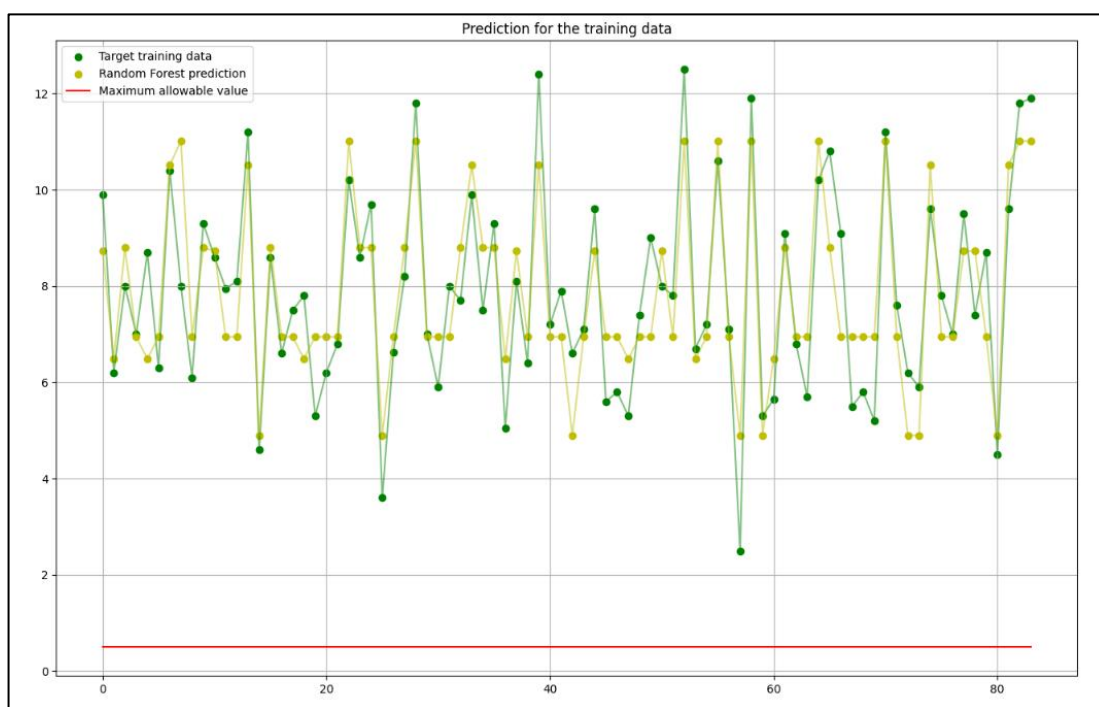


Рисунок 13 – Графік прогнозування моделі RandomForestRegressor для тренувальних даних

## Висновки

Створено інформаційну системи аналізу та прогнозування якості води в річці Південний Буг. Створено сайт, на якому відображаються результати аналізу, а також вихідні дані, на основі яких було проведено дослідження.

## СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. В. Б. Мокін, Б. І. Мокін, О. В. Дезірон, М. Я. Бабич, В. К. Гамлявий, Ю. С. Гавриков, Н. В. Тананчук, М. П. Боцула, Є. М. Крижановський, Ю. М. Коновалюк, А. Р. Ящолт / Система прийняття управлінських рішень керівниками водогосподарських організацій для басейну річки Південний Буг з використанням геоінформаційних технологій. - Методичний посібник. / Під ред. В.Б. Мокіна. - Вінниця: УНІВЕРСУМ-Вінниця, 2009. - 244 с.
2. Mokin V. B. Kaggle Dataset «River Water Quality EDA and Forecasting» – версія датасету – 2021 р.: [Електронний ресурс]. URL: <https://www.kaggle.com/datasets/vbmokin/wq-southern-bug-river-01052021>.
3. Дані державного моніторингу поверхневих вод: [Електронний ресурс]. – URL: <https://data.gov.ua/dataset/surface-water-monitoring>
4. Геоінформаційні системи в екології. – Електронний навчальний посібник / В. Б. Мокін, Є. М. Крижановський / Під ред. Крижановського Є. М. – Вінниця : ВНТУ, 2014. – 192 с.

**Євгеній Миколайович Крижановський** – канд. техн. наук, доцент кафедри системного аналізу та інформаційних технологій, Вінницький національний технічний університет, Вінниця, e-mail: [kruzhan@gmail.com](mailto:kruzhan@gmail.com);

**Марецький Денис Євгенійович** – студент групи 2ІСТ-22м, Факультет інтелектуальних інформаційних технологій та автоматизації, Вінницький національний технічний університет, Вінниця, e-mail: [deniszhvan@gmail.com](mailto:deniszhvan@gmail.com);

**Evgeniy Kryzhanovsky M.** – Cand. Sc. (Eng), Department of Systems Analysis and Information Technology, Vinnytsia National Technical University, Vinnytsia, e-mail: [kruzhan@gmail.com](mailto:kruzhan@gmail.com);

**Maretskyi Denys E.** - student of 2IST-22m group, Faculty of Intelligent Information Technologies and Automation, Vinnytsia National Technical University, Vinnytsia, e-mail: [deniszhvan@gmail.com](mailto:deniszhvan@gmail.com);