

АНАЛІЗ ТА ПРОГНОЗУВАННЯ ЯКОСТІ ВОДИ В РІЧЦІ ПІВДЕННИЙ БУГ НА ОСНОВІ ДАНИХ МОНІТОРИНГУ ЗА 2000 – 2023 РОКИ

Вінницький національний технічний університет

Анотація

В роботі розроблено технологію аналізу та прогнозування якості води в річці Південний Буг. Для розробки технології, було створено набір вхідних даних, що включає в себе дані моніторингу річки Південний Буг за 2000-2023 роки. Виконано прогнозування даних базуючись на моделі LinearRegression, проведено оцінку результатів.

Ключові слова: Південний Буг, датасет, аналіз даних, прогнозування, лінійна регресія, валідаційні дані.

Abstract

In the work, the technology of analysis and forecasting of water quality in the Southern Bug River was developed. To develop the technology, a set of input data was created, which includes monitoring data of the Southern Bug River for the years 2000-2023. Forecasting of data based on the LinearRegression model was performed, the results were evaluated.

.Keywords: Southern Bug, dataset, data analysis, forecasting, linear regression, validation data.

Актуальність дослідження

Проблема забруднення вод є однією з актуальних екологічних проблем сучасності, адже зростання кількості шкідливих речовин напряму впливає на екосистему всього басейну, де ці забруднення зростають. Збільшення вмісту шкідливих речовин в поверхневих водах негативно впливає на екосистему та спричиняє масове вимирання живих істот в забрудненому середовищі, а також негативно впливає на людську життєдіяльність та підвищує рівень захворюваності в світі [1-4]. Сучасними джерелами забруднення вод, є несанкціоновані скиди сміття в річки, скиди підприємств і т.п. [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.

Розробка модуля аналізу та прогнозування

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], було розроблено модуль аналізу та прогнозування даних, даний модуль виконується на платформі Kaggle, де здійснюються основні обрахунки а їх результати імпортуються в майбутню веб-систему. На рисунку 1 та 2 зображено дані, що буду застосовуватись під час розвідувального аналізу та прогнозування.

id	date	NH4	BSK5	Suspended	O2	N03	N02	S04	PO4
1	17.02.2000	0.33	2.77	12	12.3	9.5	0.057	154	0.454
1	11.05.2000	0.044	3	51.6	14.61	17.75	0.034	352	0.09
1	11.09.2000	0.032	2.1	24.5	9.87	13.8	0.173	416	0.2
1	13.12.2000	0.17	2.23	35.6	12.4	17.13	0.099	275.2	0.377
1	02.03.2001	0	3.03	48.8	14.69	10	0.065	281.6	0.134
1	07.06.2001	0.02	4.02	34	10.61	11.8	0.016	287	0.208
1	10.09.2001	0.863	3.91	147	10.96	20.5	0.284	595.2	0.674

Рисунок 1 – Огляд датасету «River Water Quality EDA and Forecasting»

38437 водосховище Саф'яни, 0 км, с. Саф'яни	UA_M5.3. Дунай	Лаборато 45.38672228.890194
38463 р. Південний Буг, 393 км, скид м. Ладизин ТОВ "Вінницька птахофабрика"	UA_M5.4. Південни	Лаборато 48.66183329.30500С
38465 р. Південний Буг, 0 км, с. Мигія, нижче м. Первомайськ, (Олександрівське водосховище), Націс	UA_M5.4. Південни	Лаборато 48.03219430.952167
38473 р. Прут, 864 км, с. Шепарівці, питний водозабір м. Коломия, КП "Коломияводоканал"	UA_M5.3. Дунай	Лаборато 48.55368324.95140Є
27099 канал Дніпро-Кривий Ріг, 43 км, Південне в-ще, питний в/з	UA_M5.1. Дніпро	Лаборато 47.78161133.55650Є
27118 р. Інгулець, 393 км, смт Петрово, Іскрівське вдсх., питний в/з селища	UA_M5.1. Дніпро	Лаборато 48.33482533.262072
27120 р. Інгулець, 335 км, м. Кривий Ріг, Карачунівське в-ще, питний в/з міста	UA_M5.1. Дніпро	Лаборато 47.90764233.28353Є
27122 р. Інгулець, 265 км, с. Андріївка	UA_M5.1. Дніпро	Лаборато 47.60314433.21230Є
27135 р. Південний Буг, 652 км, м. Хмільник, питний в/з, вище міста	UA_M5.2. Південни	Лаборато 49.55638527.91305Є
27136 р. Південний Буг, 607 км, с. Гушинці, нижче села, питний водозабір м.Калинівка	UA_M5.4. Південни	Лаборато 49.42111128.38250С
27137 р. Південний Буг, 582 км, м. Вінниця, Сабарівське вдсх., питний в/з міста, вище міста	UA_M5.4. Південни	Лаборато 49.27305Є28.47138Є
27138 р. Південний Буг, 570 км, смт Сабарів, (Сутиське водосховище), 500 м нижче скиду ВОКВП ВКП	UA_M5.4. Південни	Лаборато 49.19069428.443417
27140 р. Південний Буг, 413 км, с. Маньківка, вище села, питний в/з м.Ладизин	UA_M5.4. Південни	Лаборато 48.73888Є29.089167
27146 р. Південний Буг, 153 км, с. Олексіївка, питний в/з м. Південно-Українськ	UA_M5.4. Південни	Лаборато 47.86878Є31.119731
27147 р. Південний Буг, 136 км, у межах с. Олександрівка, Олександрівське вдсх.	UA_M5.4. Південни	Лаборато 47.70270Є31.25480Є

Рисунок 2 – Загальна структура даних Держводагентства

На рисунках 3 - 5 зображено елементи коду для об'єднання вхідних даних в один датафрейм для подальшої розробки модуля аналізу даних та прогнозування.

```
# 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('р. Південний Буг')]

    dfs.append(current_data)
```

Рисунок 3 – 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 5 – Імпорт даних датасету «River Water Quality EDA and Forecasting» та обробка даних Держводагентства

Після формування датасету, наступним кроком буде проведення розвідувального аналізу, побудова моделей, та їх застосування на даних, Приклад розбиття вхідних даних на тестові та тренувальні дані зображено на рисунку 6.

```
# Dividing data into training and test
train, test, target, target_test = train_test_split(df, target_data, test_size=0.4, random_state=0)
print(train.shape, test.shape)

(97, 8) (66, 8)
```

Рисунок 6 – Розбиття даних на тестові та тренувальні.

Для застосування тренувальних та тестових даних потрібно реалізувати та навчити моделі для виконання завдання прогнозування. Реалізацію та тренування однієї з моделей, а також її навчання та виведення результатів навчання зображено на рисунках 7 та 8.

```
[5]: # Linear Regression
lr = LinearRegression()
lr.fit(train, target_train)
```

Рисунок 7 – Побудова моделі LinearRegression, підстановка даних в модель

```
# Linear Regression
lr = LinearRegression()
lr.fit(train, target_train)

# Prediction for training data
y_train_lr = lr.predict(train)

# Accuracy of model
r2_score_acc = round(r2_score(target_train, y_train_lr), 2)
print(f'Accuracy of Linear Regression model training is {r2_score_acc}')

# Save to result dataframe
result.loc[result['model'] == 'Linear Regression', 'train_score'] = r2_score_acc

Accuracy of Linear Regression model training is 0.65

# Print rounded r2_lr = lr.predict(valid)
y_val_lr = lr.predict(valid)
r2_score_acc_valid = round(r2_score(target_valid, y_val_lr), 2)
result.loc[result['model'] == 'Linear Regression', 'valid_score'] = r2_score_acc_valid
print(f'Accuracy of Linear Regression model prediction for valid dataset is {r2_score_acc_valid}')

Accuracy of Linear Regression model prediction for valid dataset is 0.76
```

Рисунок 8 – Визначення точності для моделі LinearRegression

За результатами застосування моделі, точність LinearRegression для тренувальних даних складає 0.65 за оцінкою «r2_score» та 0.76 для валідаційних даних.

Порівняння відхилення значень спрогнозованих моделлю LinearRegression з тренувальними, валідаційними та тестовими даними зображено на рисунках 9 - 11.

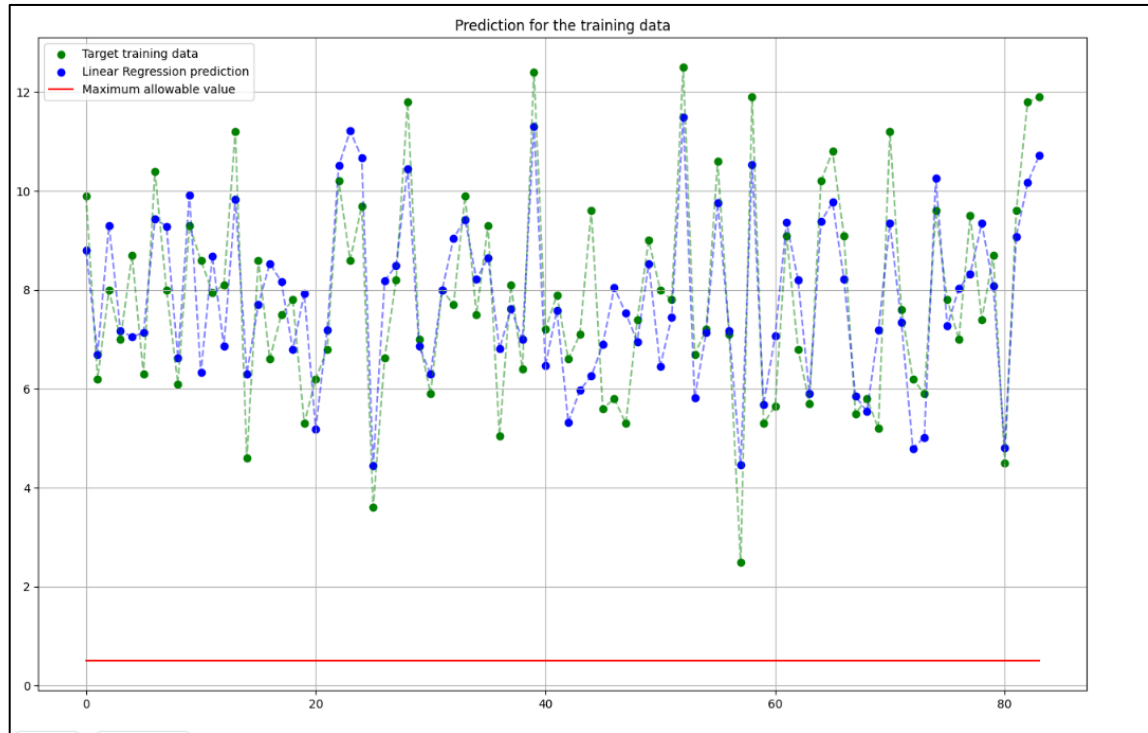


Рисунок 9 – Графік відхилення моделі LinearRegression від тренувальних даних

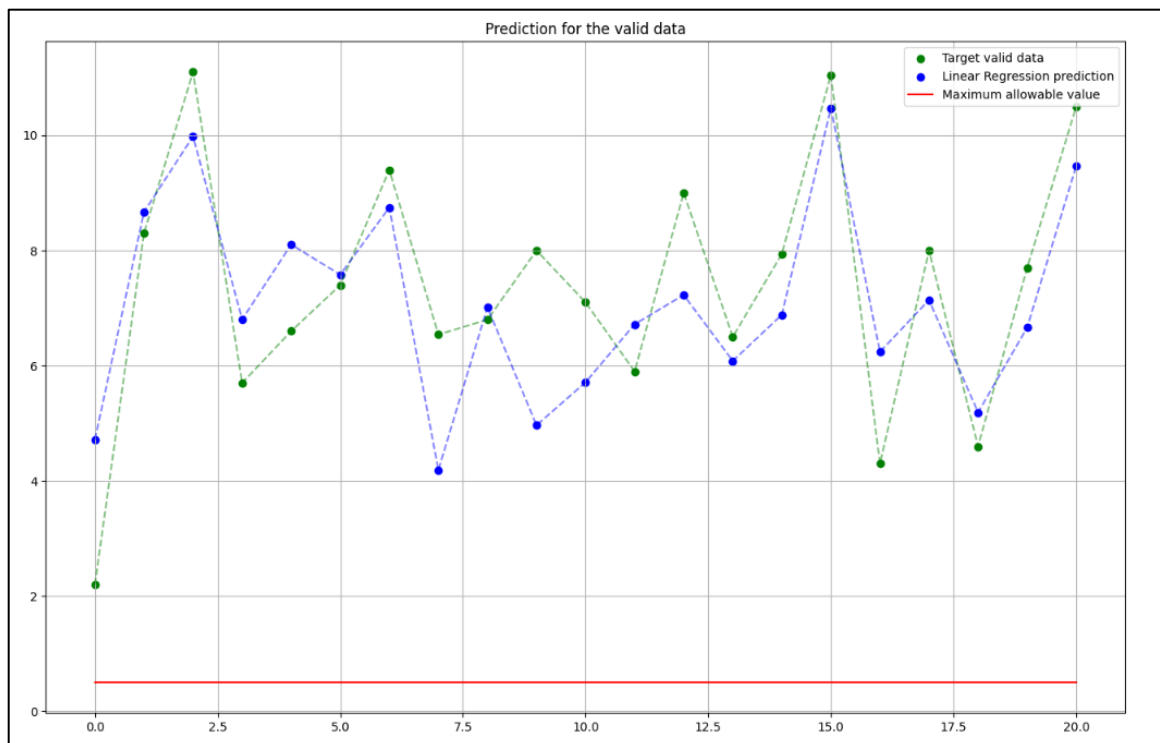


Рисунок 10 – Графік відхилення моделі LinearRegression від валідаційних даних

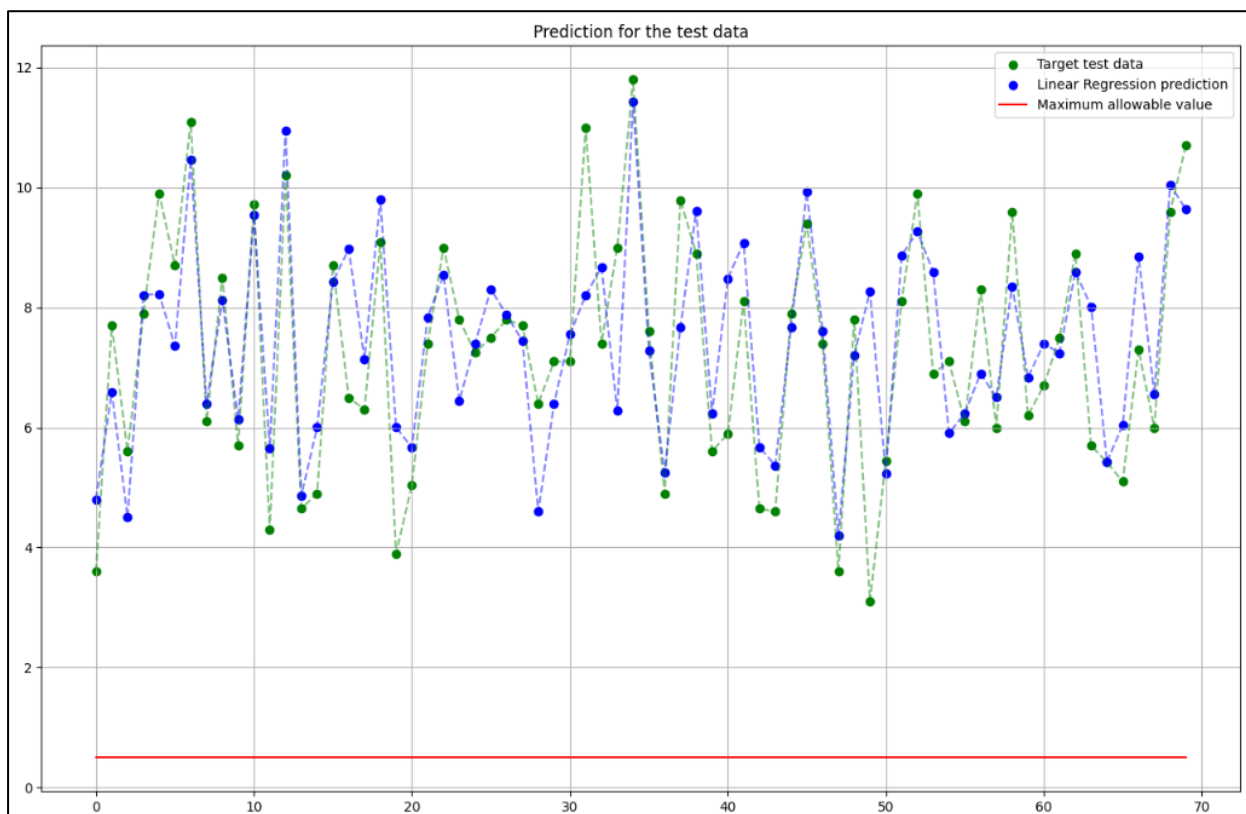


Рисунок 11 – Графік відхилення моделі LinearRegression від тестових даних

Висновки

Розроблено модуль аналізу та прогнозування даних для вхідних даних моніторингу Держводоагентства України, а також датасету «River Water Quality EDA and Forecasting». Результати аналізу та прогнозування планується імпортуватись в інформаційну систему аналізу та прогнозування якості води в річці Південний Буг.

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.

Крижановський Євгеній Миколайович – канд. техн. наук, доцент кафедри системного аналізу та інформаційних технологій, Вінницький національний технічний університет, Вінниця,

Марецький Денис Євгенійович – студент групи 2ІСТ-22м, Факультет інтелектуальних інформаційних технологій та автоматизації, Вінницький національний технічний університет, Вінниця, e-mail: 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;

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.