

ІДЕНТИФІКАЦІЯ ТА ВИБІР ОПТИМАЛЬНОЇ МОДЕЛІ ДЛЯ ПРОГНОЗУВАННЯ КОНЦЕНТРАЦІЇ НІТРАТІВ У РІЧКОВІЙ ВОДІ ПІВДЕННОГО БУГУ

Вінницький національний технічний університет

Анотація

Дана робота присвячена ідентифікації та вибору оптимальної моделі для прогнозування концентрації нітратів у річковій воді Південного Бугу, які було реалізовано на мові програмування Python. Було проведено попередній аналіз даних за допомогою більш складніших бібліотек, в порівнянні з аналогами. Для прогнозування концентрації нітратів було використано моделі для обробки і прогнозування часових рядів.

Ключові слова: інформаційна технологія, попередній аналіз даних, методи аналізу, прогнозування часових рядів.

Abstract

This paper is devoted to the identification and selection of the optimal model for forecasting the concentration of nitrates in the river water of the Southern Bug, which were implemented in the Python programming language. Preliminary data analysis was carried out with the help of more complex libraries, in comparison with analogues. Models for processing and forecasting time series were used to forecast the concentration of nitrates.

Keywords: information technology, preliminary data analysis, analysis methods, time series forecasting.

Вступ

На сьогоднішній день тема екології стоїть дуже гостро. Є досить актуальною, тому що з кожним роком рівень життя, та здоров'я людей погіршується. З'являються все нові і нові хвороби. До цього призводить забруднення планети, а саме повітря, ґрунту та води. Все це сильно впливає на стан здоров'я людей. На якість води виливають отруйні відходи; орошаючи поля добривами, людина сильно забруднює річкові та підземні води, які в подальшому потрапляють до споживання [1].

Дивлячись на дану проблему виникає необхідність у створенні інформаційної технології прогнозування забруднення нітратами річки Південний Буг. Вирішення даної проблеми є досить актуальним, оскільки є потреба в створенні точнішої, в порівнянні з аналогами, системи для прогнозування нітратів у річковій воді Південного Бугу.

Метою роботи є підвищення точності прогнозування концентрації нітратів у воді річки Південний Буг за допомогою інструментів інформаційних технологій.

Щоб досягнути поставлену мету, потрібно виконати такі задачі:

- проаналізувати предметну область,
- зробити розвідувальний аналіз якості річкових вод України;
- визначити набір найважливіших ознак для прогнозування;
- побудувати моделі для задачі прогнозування;
- провести прогнозування даних та визначити яка модель є найбільш точною.

Результати дослідження

В даному дослідженні розв'язувалась задача підвищення точності прогнозування концентрації нітратів у воді басейну річки Південний Буг на території Вінницької області, завдяки методам машинного навчання. Мета дослідження досягнута за рахунок використання складніших бібліотек для попереднього аналізу даних та прогнозування за рахунок складніших моделей з використанням часових рядів [2].

Оскільки ми знаходимося на території Вінницької області, тому доцільно буде робити прогноз концентрації нітраів у воді Південного бугу в місці питного водозабору міста Вінниці.

Під час дослідження та аналізу літератури за темою використання часових рядів для прогнозування концентрації нітратів вирішено об'єднати декілька моделей для більш ефективної роботи алгоритму [3].

Під час проведення прогнозу можна зрозуміти, що серед усіх моделей прогнозування найбільш ефективною є модель Facebook Prophet. Технологія прогнозування даних часових рядів, заснована на адитивній моделі, де нелінійні тенденції відповідають річній, тижневій та щоденній сезонності, плюс святкові ефекти, реалізована для мов програмування Python та R. При прогнозуванні на 4 доби модель показала за метрикою RMSE найкращу точність, яка склала 1,07.

Відображення рейтингу ефективності моделей прогнозування від найкращої до найгіршої зображено на рисунку 1.

	name_model	type_data	r2_score	rmse	mape
6	Prophet_12_days_4_order	valid	0.104793	1.107103	57.176322
7	Prophet_12_days_10_order	valid	0.082975	1.120513	60.580385
2	Prophet_4_days_4_order	valid	0.154377	1.076006	77.877611
5	Prophet_6_days_10_order	valid	0.003869	1.167843	80.763103
4	Prophet_6_days_4_order	valid	-0.003949	1.172417	84.931665
0	Prophet_3_days_4_order	valid	0.004941	1.167215	87.941357
3	Prophet_4_days_10_order	valid	0.126205	1.093783	88.03489
1	Prophet_3_days_10_order	valid	-0.018709	1.181004	93.723413
8	ARIMA_auto	valid	-1.988141	2.022679	187.254699
10	Support Vector Machines	valid	-8.648594	3.634615	333.814685
13	XGB Regressor	valid	-10.55456	3.977435	361.712793
11	Random Forest Regressor	valid	-14.494014	4.605835	410.13298
9	Linear Regression	valid	-19.572156	5.30721	481.138485
12	Bagging Regressor	valid	-23.563575	5.799255	532.046361

Number of models built - 14

Рис. 1. Рейтинг ефективності моделей прогнозування

Висновки

Під час виконання роботи було реалізовано кілька основних моделей для прогнозування концентрації нітратів у річковій воді Південного Бугу. Результати їхньої роботи були порівнянні між собою і було визначено найбільш ефективну модель.

В результаті дослідження було отримано найефективнішу модель Facebook Prophet, яка за метрикою RMSE показала найкращу точність.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Водні ресурси України [Електронний ресурс] – Режим доступу: https://uk.wikipedia.org/wiki/Водні_ресурси_України
2. Аналіз часових рядів [Електронний ресурс] – Режим доступу: https://uk.wikipedia.org/wiki/Аналіз_часових_рядів
3. Python [Електронний ресурс] – Режим доступу: <https://ru.wikipedia.org/wiki/Python>.

Лісовський Ростислав Русланович – студент групи 2ІСТ-21М, факультет інтелектуальних інформаційних технологій та автоматизації, Вінницький національний технічний університет, Вінниця, e-mail: rosteslavlisovski@gmail.com

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

Rostyslav Ruslanovych Lisovsky – student of the 2IST-21M group, faculty of intellectual information technologies and automation, Vinnytsia National Technical University, Vinnytsia, e-mail: dima.ishchuk00@gmail.com

Zhukov Serhii O. - Ph.D., Assistant Professor of the Department of Systems Analysis and Information Technology, Vinnytsia National Technical University, Vinnytsia, e-mail: sazhukov@gmail.com