

РОЗВИТОК КОМПЕТЕНТНОСТЕЙ СТУДЕНТІВ З ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ЧАСОВИХ РЯДІВ З ВИКОРИСТАННЯМ ПЛАТФОРМИ KAGGLE

Вінницький національний технічний університет

Анотація

Проаналізовано можливості використання платформи Kaggle для розвитку компетентностей студентів у галузі інтелектуального аналізу часових рядів. Розглянуто ключові функції та ресурси, які пропонує платформа. Охарактеризовано датасети та програми-ноутбуки. Наведено приклади задач, розв'язання яких дозволить здобувачам вищої освіти краще розвинути свої навички з інтелектуального аналізу часових рядів.

Ключові слова: компетентності студентів у галузі інформаційних технологій, інтелект. Аналіз часових рядів, Kaggle, датасет, ноутбук на Python.

Abstract

The possibilities of using the Kaggle platform for developing students' competencies in the field of time series analysis have been analyzed. Key features and resources offered by the platform are considered. Datasets and program notebooks are characterized. Examples of tasks are provided, the solutions of which will allow higher education students to better develop their skills in time series analysis.

Keywords: students' competencies in information technology, time series analysis, Kaggle, dataset, Python notebook.

Вступ

Однією з найбільш важливих компетентностей студентів ІТ-галузі (наприклад, спеціальності 124 Системний аналіз) є інтелектуальний аналіз даних. Водночас, багато процесів формалізуються як часові ряди. Одна з найбільших колекцій датасетів часових рядів у світі знаходиться на платформі Kaggle, яка станом на травень 2024 року містить біля 21 млн. акаунтів.

Метою даного дослідження є систематизація можливостей платформи Kaggle для розвитку компетентностей студентів з інтелектуального аналізу часових рядів.

Результати дослідження

Kaggle це одна з провідних онлайн-платформ для змагань з машинного навчання та аналізу даних, що надає широкі можливості для розвитку компетентностей студентів у галузі інтелектуального аналізу часових рядів. Вона забезпечує доступ до різноманітних наборів даних [1], інструментів для аналізу та співпраці з фахівцями з усього світу. Кожен набір даних — це спільнота, де можливо обговорювати дані, відкривати публічний код і методи, а також створювати власні проекти. Платформа також підтримує списки новин та підписки на користувачів для відстеження оновлень, а також інструменти пошуку наборів даних потрібної предметної галузі. Розглянемо основні можливості, які пропонує платформа Kaggle для студентів

Доступ до великих наборів даних: Kaggle надає доступ до численних публічних наборів даних, включаючи часові ряди різного характеру – від фінансових показників до екологічних даних [6]. Це дозволяє студентам працювати з реальними даними, що є важливим аспектом для розвитку практичних навичок. Найбільшою цінністю для навчального процесу є те, що більшість датасетів (особливо з Usability 10.0), мають багато публічних програм-ноутбуків на Python, які студенти можуть використати у навчальному процесі.

Приклад датасету наведено на рис.1.

Інтерактивні середовища для аналізу: Kaggle надає безкоштовні інтерактивні середовища для роботи з даними, такі як Kaggle Kernels (тепер Kaggle Notebooks) [2]. Ці середовища підтримують мови програмування Python і R, що дозволяє студентам писати та виконувати код безпосередньо на

платформі, використовуючи потужні бібліотеки для аналізу часових рядів, такі як Pandas, NumPy, Matplotlib, і TensorFlow. Крім того, користувачі мають доступ до безкоштовних обчислювальних ресурсів, включаючи CPU, GPU та TPU, що забезпечує можливість виконання складних обчислень та моделей машинного навчання без необхідності інвестувати у власне обладнання. Це робить платформу доступною та зручною для навчання і практики аналізу даних.

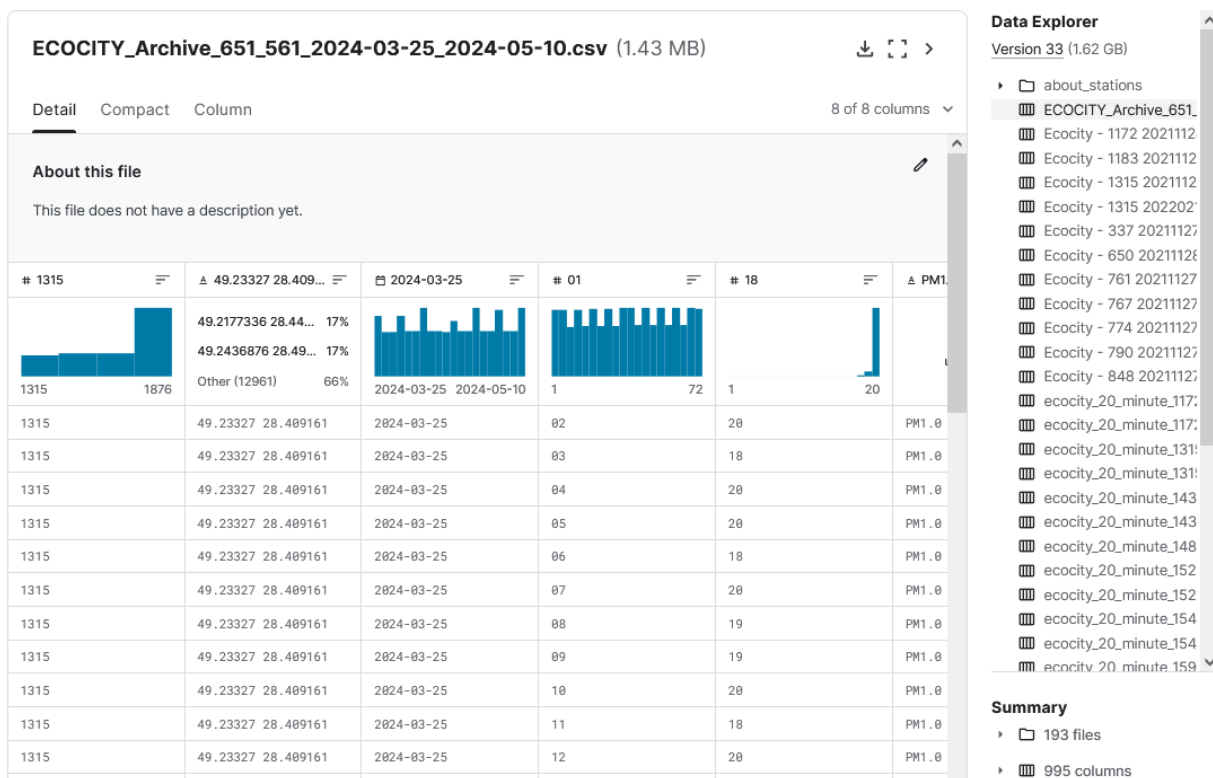


Рис.1. <https://www.kaggle.com/datasets/vbmokin/air-quality-monitoring-from-ecocity>

Зокрема, на рис. 2 наведений приклад ноутбука для регресійного аналізу прикладу даних з моніторингу якості атмосферного повітря за щодобовими багаторічними даними EcoCity на прикладі однієї станції та певного показника на ній (№ 650 «Турбів», показник «PM10»).

Навчальні ресурси та ком'юніті: Платформа має багату бібліотеку навчальних ресурсів, включаючи курси, tutorіали та документацію, що охоплюють всі аспекти аналізу часових рядів – від базових концепцій до передових методів. Крім того, активне ком'юніті Kaggle дозволяє студентам обмінюватися знаннями, задавати питання та отримувати поради від досвідчених аналітиків і дослідників.

Змагання та конкурси: Kaggle надає можливість участі у змаганнях з призовим фондом до 1,5 млн. дол. Серез таких конкурсів є й конкурси з тематики інтелектуального аналізу часових рядів: Рис.3.

Це сприяє розвитку навичок у конкурентному середовищі та стимулює до пошуку інноваційних рішень. В рамках таких змагань студенти можуть отримати цінні відгуки від експертів.

Портфоліо та працевлаштування: Участь у змаганнях та публікація своїх робіт на Kaggle допомагає студентам створити професійне портфоліо. Високі результати та активність на платформі можуть привернути увагу потенційних роботодавців і надати студентам конкурентну перевагу на ринку праці.

Висновки

Систематизовано можливості використання платформи Kaggle для розвитку компетентностей студентів у галузі інтелектуального аналізу часових рядів. Розглянуто ключові функції та ресурси, які пропонує платформа. Охарактеризовано датасети та програми-ноутбуки. Наведено приклади задач, розв'язання яких дозволить здобувачам вищої освіти краще розвинути свої навички з інтелектуального аналізу часових рядів.

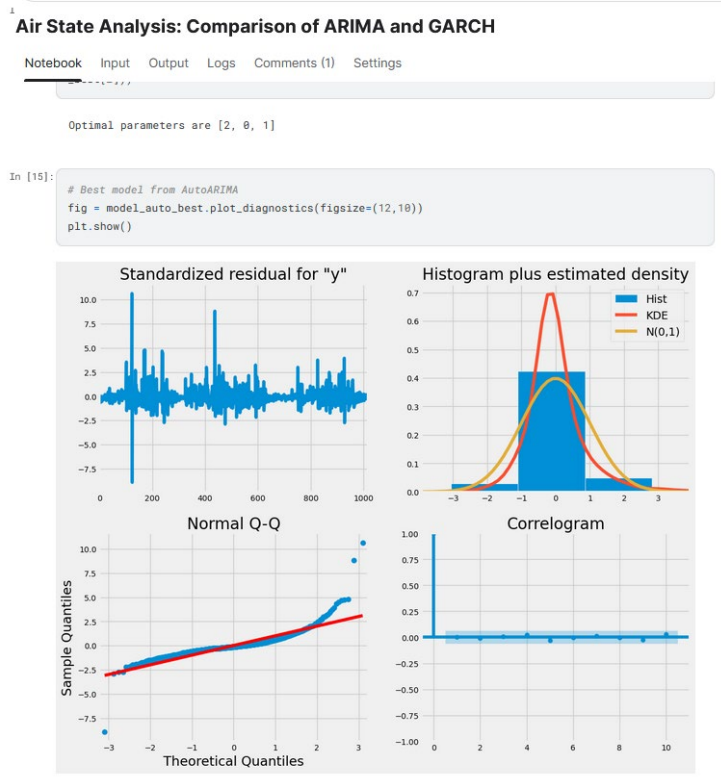


Рис.2. <https://www.kaggle.com/code/vbmokin/air-state-analysis-comparison-of-arma-and-garch>

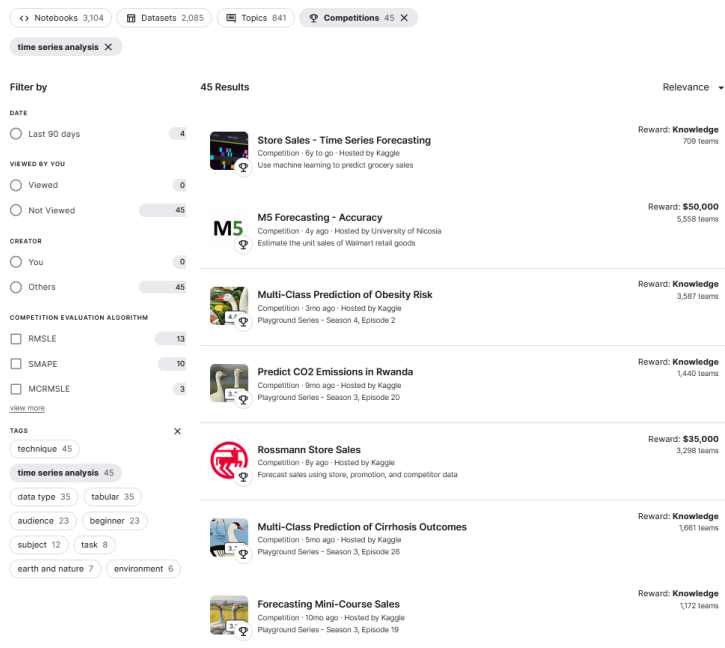


Рис.3. <https://www.kaggle.com/search?q=time+series+in%3Acompetitions>

Використання платформи Kaggle для інтелектуального аналізу часових рядів надає студентам унікальні можливості для розвитку їхніх компетентностей у цій галузі. Завдяки доступу до різноманітних наборів даних, участі у змаганнях та інтерактивним середовищам для аналізу, студенти можуть отримати цінний практичний досвід. Багатий набір навчальних ресурсів та активне ком'юніті сприяють глибокому розумінню теоретичних аспектів і застосуванню їх на практиці. Крім того, створення портфолію та можливість привернути увагу потенційних роботодавців додають важливу

мотивацію для активної участі у проектах на Kaggle.

Отже, впровадження платформи Kaggle у навчальний процес значно покращує підготовку студентів, забезпечуючи їм необхідні знання та навички для успішної кар'єри у сфері аналізу даних. Kaggle стає важливим інструментом для формування компетентних спеціалістів, готових до викликів сучасного ринку праці та здатних зробити вагомий внесок у розвиток технологій інтелектуального аналізу часових рядів.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Платформа Kaggle. Про датасети. URL: <https://www.kaggle.com/docs/datasets>.
2. Платформа Kaggle. Як використовувати Kaggle Notebooks. URL: <https://www.kaggle.com/docs/notebooks>.
3. Мокін В. Б., Дратований М. В. Наука про дані: машинне навчання та інтелектуальний аналіз даних : електронний навчальний посібник комбінованого (локального та мережевого) використання. Вінниця: ВНТУ, 2024. 258 с.
4. Копняк В.Є., Мокін В.Б. Дослідження проблем із гетероскедастичністю даних моніторингу якості атмосферного повітря. Матеріали ЛІІ науково-технічної конференції підрозділів Вінницького національного технічного університету (НТКПІ ВНТУ–2023), Вінниця, 21 – 23 червня 2023 р. Збірник наукових праць [Електронний ресурс], <https://press.vntu.edu.ua/index.php/vntu/catalog/view/788/1373/2632-1>, Вінниця: ВНТУ, 2023. (PDF, 3076 с.) ISBN 987-966-641-942-5, С. 320-324.
5. Шмундяк Д. О., Копняк В. Є. Метод ідентифікації локальних аномалій значень показників стану довкілля з використанням декомпозиції на півхвилі, Вісник Вінницького політехнічного інституту, вип. 1, с. 88–100, Лют. 2024, DOI <https://doi.org/10.31649/1997-9266-2024-172-1-88-100>.
6. Mokin Vitalii. Kopniak Volodymyr. Kaggle Notebook “Air State Analysis: Comparison of ARIMA and GARCH models”, 2023. URL: <https://www.kaggle.com/code/vbmokin/air-state-analysis-comparison-of-arima-and-garch>.

Копняк Володимир Євгенович – аспірант кафедри системного аналізу та інформаційних технологій, Вінницький національний технічний університет, Вінниця, vkopnyak@gmail.com

Кобилянська Ірина Миколаївна – к. пед. н, доцент, доцент кафедри безпеки життєдіяльності та педагогіки безпеки, Вінницький національний технічний університет, Вінниця, e-mail: irishakobilanska@gmail.com.

Volodymyr Ye. Kopniak – Postgraduate student of the Department of System Analysis and Information Technologies, Vinnytsia National Technical University, Vinnytsia, vkopnyak@gmail.com

Iryna M. Kobylyanska – Candidate of Sc. (Pedagogical), Associated Professor, Associated Professor of the Chair Security of Life and Safety Pedagogic, Vinnitsa National Technical University, Vinnytsia, e-mail: irishakobilanska@gmail.com