

ПОРІВНЯННЯ МОДЕЛЕЙ LSTM ТА GRU ДЛЯ ПРОГНОЗУВАННЯ ЗАЛИШКІВ У БУСТИНГОВІЙ ГЕТЕРОСКЕДАСТИЧНІЙ EGARCH-МОДЕЛІ

¹Вінницький національний технічний університет

Анотація

Здійснено порівняльний аналіз моделей LSTM та GRU для прогнозування залишків у бустинговій гетероскедастичній моделі EGARCH. Результати показали, що LSTM забезпечує кращу точність, ніж GRU на прикладі прогнозування концентрації пилу Сахари в атмосферному повітрі м. Вінниці за даними мережі громадського моніторингу EcoCity.

Ключові слова: прогнозування часових рядів, бустинг, гетероскедастична модель, модель LSTM, модель GRU, прогнозування якості атмосферного повітря, пил Сахари, EcoCity.

Abstract

A comparative analysis of LSTM and GRU models for forecasting residuals in the boosting heteroskedastic EGARCH model was carried out. The results showed that LSTM provides better accuracy than GRU on the example of forecasting the concentration of Sahara dust in the atmospheric air of Vinnytsia, according to the data of the EcoCity public monitoring network.

Keywords: time series forecasting, boosting, heteroskedastic model, LSTM model, GRU model, air quality forecasting, Sahara dust, EcoCity.

Вступ

У статті [1] було здійснено побудову бустингової гетероскедастичної EGARCH-моделі (бустингова кооперація гетероскедастичної моделі і дерев рішень) для оцінювання дисперсії ряду даних щодо якості атмосферного повітря, а також застосовано LSTM-модель для прогнозування залишків з метою підвищення точності прогнозу. Однак, відкритим залишилось питання чи є модель LSTM (з англ. "Long Short-Term Memory" – довга короткочасна пам'ять) оптимальним вибором, адже для прогнозування залишків, ще популярною є модель GRU (з англ. "Gated Recurrent Unit" - керований рекурентний блок) [2].

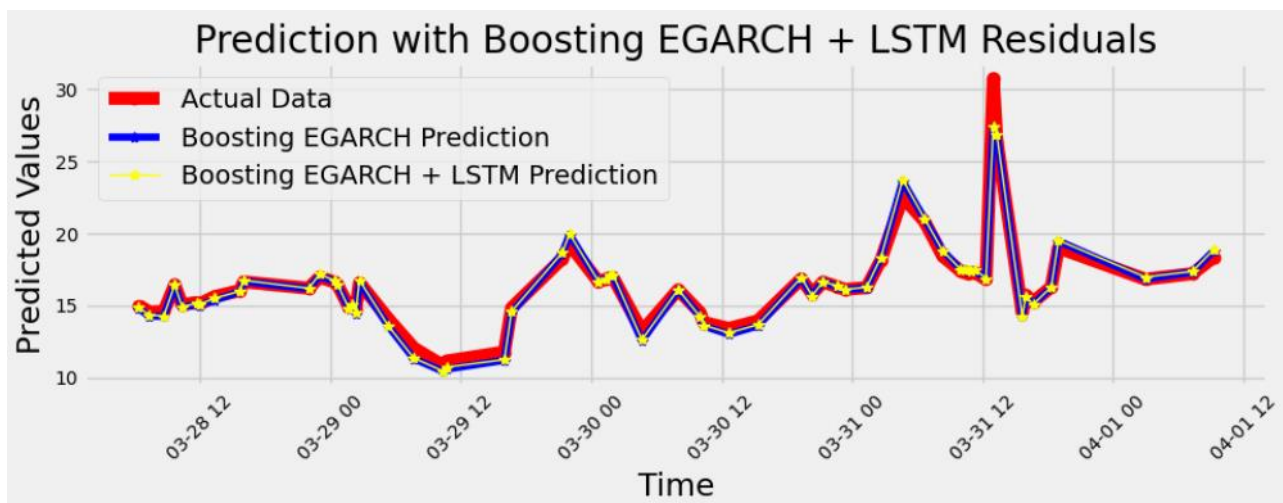
Мета дослідження – порівняння точності моделей LSTM та GRU для прогнозування залишків бустингової гетероскедастичної EGARCH-моделі на прикладі моделювання поширення пилу Сахари в атмосферному повітрі м. Вінниці весною 2024 р., наведеного у статті [1] та у ноутбучі [3]. Дані для цього прикладу отримано з датасету [4], який наповнений даними, отриманими із системи «Кабінет дослідника» з використанням авторизованого доступу від ВНТУ до даних моніторингу якості атмосферного повітря від громадського проєкту EcoCity та україно-чеського проєкту «Чисте повітря для України» [5].

Результати дослідження

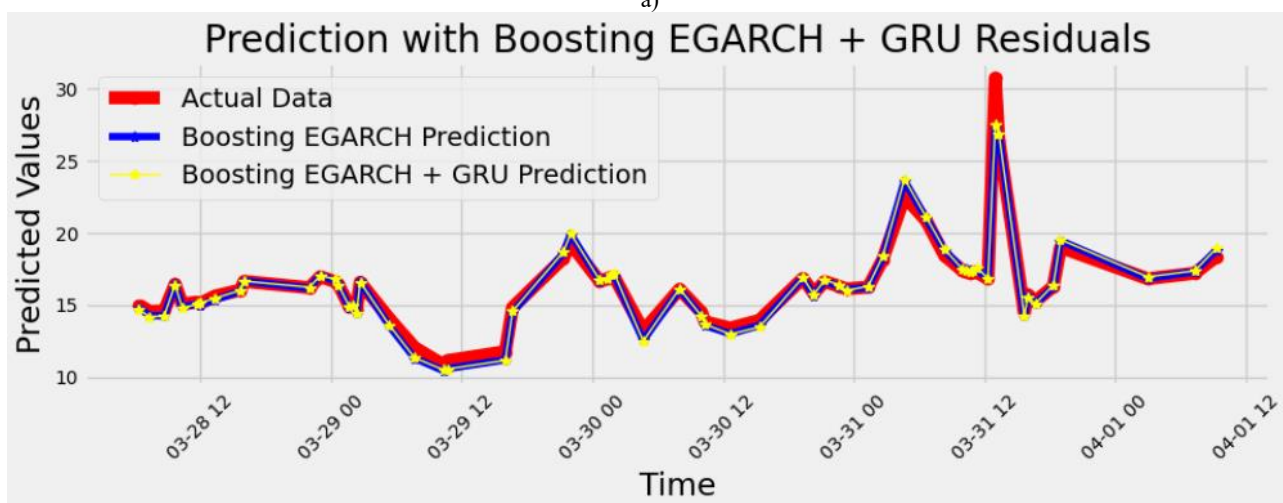
У Kaggle побудовано 2 ноутбуки [3, 6] на Python для моделювання концентрації пилу Сахари за показником «PM1» в атмосферному повітрі м. Вінниці весною 2024 р.:

- «[Sahara's Dust in UA: Boosting EGARCH + LSTM](#)»
- «[Sahara's Dust in UA: Boosting GARCH + GRU](#)»

На рис. 1 наведено результати прогнозування за моделями LSTM, GRU, а у таблиці 1 – точність прогнозування за метриками RMSE (з англ. "Root Mean Squared Error" - корінь середньоквадратичної помилки), R² score (коефіцієнт детермінації), MAPE ("Mean Absolute Percentage Error" - середня абсолютна відносна похибка).



а)



б)

Рис. 1. Результат прогнозування концентрації пилу «PM1» в атмосферному повітрі м. Вінниця весною 2024 р. за бустинговою гетероскедастичною моделлю EGARCH та моделями LSTM (а) та GRU (б) для прогнозування залишків

Таблиця 1 – Точність прогнозування залишків бустингової гетероскедастичної EGARCH-моделі за моделями LSTM та GRU

Назва моделі	RMSE	R ² score	MAPE
Boosting EGARCH + LSTM	0,520504	0,970454	0,015607
Boosting EGARCH + GRU	0,531728	0,969166	0,017449
Boosting EGARCH (базова)	0,525977	0,969829	0,015924

Як показує аналіз, модель GRU забезпечує нижчу точність та адекватність, ніж модель LSTM, отже вибір моделі у статті [1] був оптимальним.

Висновки

Проведено порівняння моделей LSTM та GRU для прогнозування залишків бустингової гетероскедастичної EGARCH-моделі зі статті [1], що довело оптимальність моделі LSTM за всіма метриками.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Копняк В. Є., Мокін В. Б., Жуков С. О., Варчук І. В., Скринник Т. В., Метод бустингу гетероскедастичних моделей для прогнозування концентрації пилу Сахари в атмосферному повітрі України, Наукові праці ВНТУ [Електронний ресурс]. Вип. 2, Лип 2024. Режим доступу: doi <https://doi.org/10.31649/2307-5376-2024-2-28-38>
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. Дратований – Вінниця : ВНТУ, 2024. – 258 с. – Режим доступу: <https://docs.vntu.edu.ua/card.php?id=8163>

3. Mokin Vitalii, Kopniak Volodymyr. Sahara's Dust in UA: Boosting EGARCH + LSTM, June 2024, Kaggle Notebook, URL: <https://www.kaggle.com/code/vbmokin/sahara-s-dust-in-ua-boosting-egarch-lstm>
4. Mokin Vitalii, Shmundiak Dmytro, Kopniak Volodymyr. Air Quality Monitoring from EcoCity, May 2024, Kaggle Dataset. URL: <https://www.kaggle.com/datasets/vbmokin/air-quality-monitoring-from-ecocity>
5. Eco-City. Громадський моніторинг стану якості повітря [Електронний ресурс] – Режим доступу до ресурсу: <https://reborn.eco-city.org.ua/>.
6. Mokin Vitalii, Kopniak Volodymyr. Sahara's Dust in UA: Boosting GARCH + GRU, June 2025, Kaggle Notebook, URL: <https://www.kaggle.com/code/vbmokin/sahara-s-dust-in-ua-boosting-garch-gru>

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

Мокін Віталій Борисович – д-р. техн. наук, проф., завідувач кафедри системного аналізу та інформаційних технологій, Вінницький національний технічний університет, Вінниця, e-mail: vbmokin@vntu.edu.ua

Kopniak Volodymyr Y. – Postgraduate student of the Department of System Analysis and Information Technologies, Vinnytsia National Technical University, Vinnytsia, vkopnyak@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: vbmokin@vntu.edu.ua