

УДК 004.85:519.246

DOI [https://doi.org/10.24144/2616-7700.2026.49\(2\).155-160](https://doi.org/10.24144/2616-7700.2026.49(2).155-160)**О. С. Бурмей¹, П. П. Антосяк²**

¹ ДВНЗ «Ужгородський національний університет»,
аспірант кафедри системного аналізу і теорії оптимізації
oleksandr.burmei@uzhnu.edu.ua
ORCID: <https://orcid.org/0009-0004-6192-9742>

² ДВНЗ «Ужгородський національний університет»,
доцент кафедри системного аналізу і теорії оптимізації,
кандидат фізико-математичних наук
pavlo.antsiak@uzhnu.edu.ua
ORCID: <https://orcid.org/0000-0001-7131-1795>

РОЛЬ КОНСТРУЮВАННЯ ОЗНАК У ПРОГНОЗУВАННІ РОЗДРІБНОГО ПОПИТУ: ВІД СТАТИСТИЧНИХ БАЗОВИХ МОДЕЛЕЙ ДО НЕЙРОННИХ АНСАМБЛІВ

У статті представлено порівняльне дослідження методів прогнозування роздрібно-го попиту на наборі даних Kaggle Store Sales [13] (1 782 часових ряди, 54 магазини \times 33 товарні категорії). Зіставлено вісім методів — Seasonal Naive, Drift, Prophet, LightGBM, XGBoost, N-HiTS, PatchTST та ансамбль AutoGluon-TimeSeries — за точністю точкових прогнозів, імовірнісним калібруванням та виявленням стрибків попиту. За допомогою крос-валідації з ковзним вікном ($h = 15$ днів, 3 фолди) кожену модель оцінено у чотирьох конфігураціях виключення ознак. Градієнтний бустинг (LightGBM) досягає найкращої точності ($sMAPE_{nz} = 29,4\%$ проти $41,3\%$ для Seasonal Naive), зменшує похибку в промоційні періоди на 32% та забезпечує калібровані 80% -ві прогнозні інтервали (покриття = $79,8\%$). Моделі глибинного навчання досягають конкурентного $sMAPE_{nz}$ на високообігових категоріях ($13,6$ – $17,4\%$), але генерують менш калібровані інтервали. Абляційний аналіз показав, що ковзні ознаки забезпечують приріст $\approx 1,5$ п. п. $sMAPE$, а промоційні — $\approx 0,7$ п. п.

Ключові слова: прогнозування попиту, роздрібна торгівля, градієнтний бустинг, глибинне навчання, імовірнісне прогнозування, промоційні ефекти, часові ряди, аналіз внеску компонентів.

1. Вступ. Прогнозування роздрібно-го попиту є фундаментальною задачею управління ланцюгами постачання, що безпосередньо впливає на поповнення запасів, оптимізацію знижок та планування ресурсів [1, 2]. Поширення деталізованих даних з точок продажу стимулювало перехід від традиційних статистичних методів (ETS, ARIMA) до машинного навчання і, нещодавно, до глибинного навчання [3]. Конкурс M5 [4] на ієрархічних даних Walmart продемонстрував домінування ансамблів на основі градієнтного бустингу (GBDT), зокрема LightGBM [5], у рейтингу точності, а супутній конкурс M5 Uncertainty підкреслив важливість каліброваних прогнозних інтервалів [6]. Водночас архітектури PatchTST [7] та N-HiTS [8] продемонстрували високу якість на стандартних бенчмарках, хоча їхні переваги на реальних роздрібних даних із промоційними ефектами залишаються недостатньо дослідженими.

Водночас у літературі бракує контрольованих експериментів з виключенням ознак, що ізолюють внесок промоційних, ковзних та календарних ознак. Практичні застосування також потребують каліброваних прогнозних інтервалів та виявлення стрибків попиту — задач, яким приділяється недостатньо уваги. Це

дослідження заповнює цю прогалину: порівнює сім сімейств моделей за єдиним протоколом без витоків даних, кількісно оцінює маргінальний внесок груп ознак через чотири конфігурації виключення та проводить багатозадачне оцінювання (точкові прогнози, імовірнісна якість, виявлення стрибків).

2. Дані та методологія. Використано набір даних **Store Sales — Time Series Forecasting** [13] з платформи Kaggle, опублікований компанією Corporación Favorita у межах однойменного змагання та поширюваний відповідно до умов використання Kaggle (Kaggle Competition Rules), які дозволяють некомерційне використання у дослідницьких цілях: щоденні обсяги продажів для 54 магазинів за 33 товарними категоріями в Еквадорі ($N = 1782$ унікальних часових ряди, $\sim 4,5$ року, 2013–2017). Цільова змінна — щоденні продажі $y_t^{(s,f)}$ для кожної пари (магазин s , категорія f); екзогенні змінні — бінарний індикатор промоакції, прапорці свят, ціна на нафту та метадані магазинів (тип, кластер, місто). Під час навчання застосовується перетворення $\tilde{y} = \log(1 + y)$ з оберненим $\hat{y} = \exp(\tilde{y}) - 1$ перед обчисленням метрик.

Ознаки організовано в чотири групи: *лагові* (лаги за 1, 7, 14, 28 днів), *ковзні* (середнє, стд. відхилення, макс., мін. за вікна 7 і 28 днів із зсувом на 1 день), *календарні* (день тижня, місяць, прапорці свят) та *промоційні* (індикатор промоакції, промо-лаги, взаємодія промо \times день тижня). Додатково використано статичні ознаки магазинів (тип, кластер, місто).

Зіставлено вісім моделей: статистичні базові — Seasonal Naive, Drift, Prophet [9]; ML — LightGBM [5] (MAE + квантильна регресія, 50 спроб Optuna [12]), XGBoost [10]; DL — N-HiTS [8], PatchTST [7] (MQLoss, $q = 0,1; 0,5; 0,9$); AutoML — AutoGluon-TimeSeries [11]. Через обчислювальні обмеження DL та AutoGluon навчено на 10 найбільших категоріях (540 серій) і оцінено на останньому фолді.

Застосовано крос-валідацію з ковзним вікном (3 фолди, горизонт $h = 15$ днів). Метрики: точкові — MAE, RMSE, RMSLE, sMAPE_{nz} (sMAPE, обмежений спостереженнями з $y > 0$); імовірнісні — WIS для 80%-го інтервалу та емпіричне покриття; виявлення стрибків — PR-AUC з міткою $y_{\text{spike}} = 1$ при продажах $> P_{90}$ у межах кожної серії. Конфігурації виключення: **A** — без промоційних та святкових ознак; **B** — без ковзних ознак; **C** — лише статистичні моделі; **D** — повний набір ознак.

3. Результати. Таблиця 1 представляє результати на повній панелі з 1782 серій (конфігурація D, середнє за 3 фолди).

Таблиця 1.

Точкові прогнози на повній панелі (D, 1782 серії, середнє за 3 фолди)

Модель	sMAPE, %	sMAPE _{nz} , %	MAE	RMSE	RMSLE
LightGBM	54,2	29,4	57,2	212,6	0,379
XGBoost	53,0	29,5	58,6	217,1	0,380
Seasonal Naive	47,0	41,3	84,6	327,7	0,551
Drift	57,7	51,4	193,5	657,3	0,646
Prophet	65,0	47,3	389,0	2997,2	0,870

LightGBM досягає найнижчого sMAPE_{nz} (29,4%) — відносно покращення на 28,8% порівняно з Seasonal Naive (41,3%) та на 37,8% порівняно з Prophet

(47,3%). Розбіжність між необробленим sMAPE та sMAPE_{nz} підкреслює спотворення серіями з нульовими продажами, що обґрунтовує sMAPE_{nz} як основну метрику.

Таблиця 2 обмежує оцінювання підмножиною з 540 серій для безпосереднього порівняння з DL-моделями.

Таблиця 2.

Порівняння на підмножині 540 серій (останній фолд)

Модель	sMAPE _{nz} , %	MAE	RMSE	PR-AUC	Покр.80, %	WIS
LightGBM	13,6	177,1	404,0	0,486	79,8	1247,7
XGBoost	13,7	176,9	400,2	0,483	0,0	3537,4
AutoGluon	16,3	242,2	556,5	0,417	85,3	1546,2
N-HiTS	16,8	230,5	497,9	0,381	0,0	4610,3
PatchTST	17,4	239,5	510,1	0,461	0,0	4790,7
Seasonal Naive	18,2	248,5	592,3	0,411	0,0	4969,3
Prophet	32,7	1186,6	5434,3	0,445	93,1	15812,9
Drift	33,9	581,3	1188,5	0,690	0,0	11626,1

LightGBM домінує за sMAPE_{nz}, MAE та WIS. DL-моделі перевершують Seasonal Naive, але поступаються GBDT: AutoGluon 16,3%, N-HiTS 16,8%, PatchTST 17,4%. Квантильна регресія LightGBM генерує калібровані 80%-ві інтервали (покриття 79,8%), тоді як AutoGluon перепокриває (85,3%), Prophet — суттєво (93,1%), а DL-моделі без апостеріорного калібрування дають покриття $\approx 0\%$. LightGBM досягає найвищого PR-AUC (0,486) — на 18% більше за Seasonal Naive (рис. 1).

Figure 1: Model Comparison on Top-10 Families (540 series, h=15)

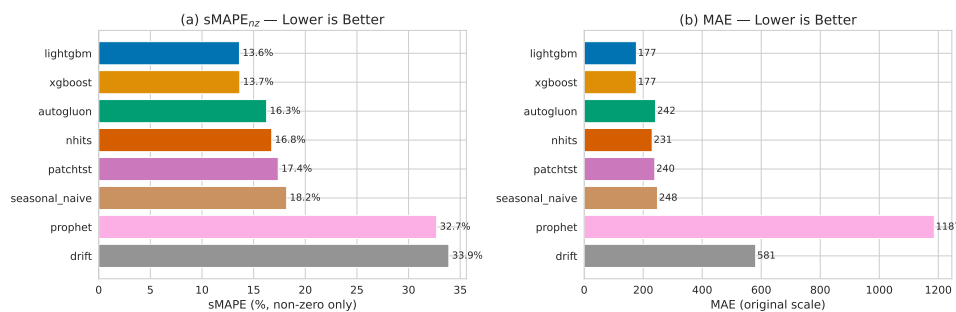


Рис. 1. Порівняння моделей за sMAPE_{nz}, MAE, PR-AUC та WIS на підмножині 540 серій (останній фолд)

У промоційні періоди GBDT зменшують sMAPE_{nz} на 32% відносно Seasonal Naive (19,2% проти 28,1%). DL-моделі мають нижчий sMAPE_{nz} у непромоційні періоди (14,5–16,7%), але суттєво вищий MAE під час промоакцій.

Таблиця 3 представляє результати аналізу виключення ознак для LightGBM на повній панелі.

Ковзні ознаки забезпечують приріст $\approx 1,5$ п. п. sMAPE_{nz} (30,9 \rightarrow 29,4), промоційні — $\approx 0,7$ п. п. (30,1 \rightarrow 29,4). Моделі лише зі статистичними методами (конфігурація C) суттєво слабші: Seasonal Naive 41,3%, Prophet 47,3%, Drift 51,4%.

Таблиця 3.

Аналіз виключення ознак LightGBM (середнє за 3 фолди, 1 782 серії)

Конфігурація	sMAPE _{nz} , %	MAE	RMSE	PR-AUC	WIS
D (Повний)	29,4	57,2	212,6	0,478	396,5
A (Без промо)	30,1	61,7	226,5	0,506	416,4
B (Без ковзних)	30,9	59,7	222,9	0,461	415,1

4. Обговорення та висновки. Результати підтверджують висновок конкурсу M5 [4]: добре налаштовані GBDT-моделі забезпечують найкращу точність на роздрібних даних. Перевага LightGBM пояснюється гнучкістю інтеграції гетерогенних ознак в єдиній моделі, ефективністю навчання на глобальній панелі при обмеженій кількості спостережень на серію та природною стійкістю до нульових продажів. DL-моделі досягають конкурентного sMAPE_{nz} на високообігових категоріях, але демонструють вищі абсолютні похибки та потребують апостеріорного калібрування інтервалів. Ансамбль AutoGluon (TFT: 53%, Chronos: 33%, ETS: 7%) є найбільш збалансованим серед DL, досягаючи покриття 85,3%.

Аналіз внеску компонентів виявляє ієрархію: лагові > ковзні > промоційні > календарні ознаки. Промоційні ознаки, попри скромний загальний внесок (0,7 п. п.), забезпечують 32%-ве зменшення похибки саме в промоційні періоди — найкритичніші для управління запасами.

Дослідження має обмеження: один набір даних, підвибірка для DL (540 серій, один фолд), відсутність явних коваріат у DL-моделях. Перспективні напрямки включають конформне прогнозування для калібрування DL-інтервалів та дослідження foundation-моделей часових рядів (Chronos, TimesFM).

Конфлікт інтересів

Автори заявляють, що не мають конфлікту інтересів щодо даного дослідження, включаючи фінансовий, особистий, авторський або будь-який інший, який міг би вплинути на дослідження, а також на результати, представлені в даній статті.

Фінансування

Дослідження було проведено без фінансової підтримки.

Доступність даних

Набір даних Store Sales — Time Series Forecasting [13] є відкритим і доступним на платформі Kaggle відповідно до умов використання Kaggle (Kaggle Competition Rules) для некомерційних дослідницьких цілей. Код експериментів, конфігурації та результати доступні в репозиторії проекту.

Використання штучного інтелекту

Автори підтверджують, що при створенні даної роботи вони не використовували технології штучного інтелекту.

Внесок авторів

Бурмей О. С.: концептуалізація, методологія, програмна реалізація, проведення експериментів, аналіз даних, написання — оригінальний проект. Антосяк П. П.: супервізія, методологічні консультації, написання — рецензування та редагування.

Авторські права ©



(2026). Бурмей О. С., Антосяк П. П. Ця робота ліцензується відповідно до Creative Commons Attribution 4.0 International License.

Список використаної літератури

1. Fildes, R., Ma, S., & Kolassa, S. (2022). Retail forecasting: Research and practice. *International Journal of Forecasting*, 38(4), 1283–1318. <https://doi.org/10.1016/j.ijforecast.2019.06.004>
2. Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting: Principles and Practice (3rd ed.)*. Melbourne: OTexts. Retrieved from <https://otexts.com/fpp3/>
3. Lim, B., & Zohren, S. (2021). Time-series forecasting with deep learning: A survey. *Philosophical Transactions of the Royal Society A*, 379(2194). <https://doi.org/10.1098/rsta.2020.0209>
4. Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2022). M5 accuracy competition: Results, findings, and conclusions. *International Journal of Forecasting*, 38(4), 1346–1364. <https://doi.org/10.1016/j.ijforecast.2021.11.013>
5. Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.-Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems*, 30, 3146–3154.
6. Gneiting, T., Balabdaoui, F., & Raftery, A. E. (2007). Probabilistic forecasts, calibration and sharpness. *Journal of the Royal Statistical Society: Series B*, 69(2), 243–268. <https://doi.org/10.1111/j.1467-9868.2007.00587.x>
7. Nie, Y., Nguyen, N. H., Sinthong, P., & Kalagnanam, J. (2023). A time series is worth 64 words: Long-term forecasting with transformers. In *Proceedings of ICLR*.
8. Challu, C., Olivares, K. G., Oreshkin, B. N., Garza, F., Mergenthaler-Canseco, M., & Dubrawski, A. (2023). N-HiTS: Neural hierarchical interpolation for time series forecasting. In *Proceedings of AAAI*, 37(6), 6989–6997. <https://doi.org/10.1609/aaai.v37i6.25854>
9. Taylor, S. J., & Letham, B. (2018). Forecasting at scale. *The American Statistician*, 72(1), 37–45. <https://doi.org/10.1080/00031305.2017.1380080>
10. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. (pp. 785–794). <https://doi.org/10.1145/2939672.2939785>
11. Shchur, O., Turkmen, A. C., Erickson, N., Shen, H., Shirkov, A., Hu, T., & Wang, B. (2023). AutoGluon-TimeSeries: AutoML for probabilistic time series forecasting. In *Proceedings of the 2nd International Conference on Automated Machine Learning (AutoML)*, PMLR, 224, 9/1–21.
12. Akiba, T., Sano, S., Yanase, T., Ohta, T., & Koyama, M. (2019). Optuna: A next-generation hyperparameter optimization framework. In *Proceedings of the 25th ACM SI-*

GKDD International Conference on Knowledge Discovery and Data Mining. (pp. 2623–2631).
<https://doi.org/10.1145/3292500.3330701>

13. Corporación Favorita, Cook, A., DanB, inversion, & Holbrook, R. (2021). *Store Sales – Time Series Forecasting* [Data set and competition]. Kaggle. License: Kaggle Competition Rules. Retrieved from <https://www.kaggle.com/competitions/store-sales-time-series-forecasting>

Burmei O. S., Antosiak P. P. The role of feature engineering in retail demand forecasting: from statistical baselines to neural ensembles.

This paper benchmarks retail demand forecasting methods on the Kaggle Store Sales dataset [13] (1,782 time series, 54 stores \times 33 product families). Eight methods — Seasonal Naive, Drift, Prophet, LightGBM, XGBoost, N-HiTS, PatchTST, and AutoGluon-TimeSeries — are compared on point accuracy, probabilistic calibration, and spike detection using rolling-origin cross-validation ($h = 15$ days, 3 folds) under four ablation configurations. LightGBM achieves the best accuracy ($\text{sMAPE}_{\text{IIZ}} = 29.4\%$ vs. 41.3% for Seasonal Naive), reduces promotional-period errors by 32%, and delivers well-calibrated 80% prediction intervals (coverage = 79.8%). Deep learning models achieve competitive $\text{sMAPE}_{\text{IIZ}}$ on top-10 families (13.6–17.4%) but produce less calibrated intervals. Ablation analysis shows rolling features contribute ≈ 1.5 sMAPE pp and promotional features ≈ 0.7 pp.

Keywords: demand forecasting, retail, gradient boosting, deep learning, probabilistic forecasting, promotional effects, time series, ablation study.

Отримано: 02.04.2026

Прийнято: 17.04.2026

Опубліковано: 30.04.2026