

УДК 004.8:004.032.26:519.876

DOI [https://doi.org/10.24144/2616-7700.2026.49\(2\).205-211](https://doi.org/10.24144/2616-7700.2026.49(2).205-211)**В. І. Кудак¹, А. С. Вощепинець²**

¹ ДВНЗ «Ужгородський національний університет»,
аспірант кафедри системного аналізу та теорії оптимізації
volodymyr.kudak@uzhnu.edu.ua
ORCID: <https://orcid.org/0009-0008-1686-4686>

² ДВНЗ «Ужгородський національний університет»,
доцент кафедри системного аналізу та теорії оптимізації,
канд. фіз.-мат. наук
andrii.voshchepynets@uzhnu.edu.ua
ORCID: <https://orcid.org/0000-0001-8307-781X>

ПОРІВНЯЛЬНИЙ АНАЛІЗ MLP ТА DFSN ДЛЯ ПРОГНОЗУВАННЯ ЗАТРИМАНИХ КОНВЕРСІЙ

У роботі розглянуто проблему прогнозування в умовах затриманого зворотного зв'язку, що призводить до зміщення навчальних міток у часі та деградації якості прогнозу класичних моделей машинного навчання. Особливу увагу приділено проблемі label delay bias, яка виникає через те, що на момент формування навчальної вибірки частина позитивних подій ще не відбулася або не була зафіксована. У роботі було проведено порівняння якості прогнозу двох моделей: багатошарового перцептрона (Multilayer Perceptron, MLP) та Delayed Feedback Satellite Networks (DFSN), яка поєднує основну нейронну мережу з сателітними компонентами для врахування частково доступних даних.

Експерименти проведено на реальному промисловому наборі даних Criteo, що характеризується високою розмірністю ознак, розрідженістю та значним дисбалансом класів. Числові ознаки були стандартизовані, категоріальні — закодовані та представлені у вигляді векторних представлень.

Оцінювання моделей здійснювалося використовуючи метрику ROC-AUC. Результати показали, що DFSN суттєво перевершує класичний MLP у задачах із затриманими мітками, забезпечуючи меншу чутливість до неповноти міток на ранніх етапах навчання та вищу якість прогнозування.

Ключові слова: MLP, DFSN, delayed feedback, нейронні мережі.

1. Вступ. Сучасні цифрові платформи потребують ефективних підходів до прийняття рішень щодо оптимізації функціонування сервісів та покращення користувацького досвіду [1]. Традиційні статистичні методи аналізу даних мають низку обмежень, зокрема тривалий час накопичення достатнього обсягу даних та отримання надійних результатів через затримку появи цільових міток, високі обчислювальні витрати та знижену ефективність у випадках рідкісних подій, таких як низькі рівні конверсій [1, 2, 3]. Додаткову складність становить ефект затриманого зворотного зв'язку (delayed feedback) [3, 4, 5, 6], що може призводити до зміщення навчальних міток і зниження якості моделей.

У цьому контексті все більшої актуальності набувають методи машинного навчання, зокрема нейронні мережі [7], які дозволяють моделювати складні нелінійні залежності та враховувати динаміку даних у часі [4, 7]. Їх застосування відкриває можливості для прискорення процесу аналізу та підвищення точності прогнозування [8].

У статті досліджується застосування нейронних мереж для підвищення точності прогнозування рівня конверсій користувачів. У контексті розглянутої

задачі, під конверсією розуміємо цільову дію користувача, таку як покупка товару чи послуги. Особлива увага приділяється впливу затриманого зворотного зв'язку (delayed feedback).

2. Мета і завдання роботи. Метою даного дослідження є проведення порівняльного аналізу ефективності класичних багатопарових перцептронів (MLP) [7] та підходу з використанням Delayed Feedback Satellite Networks (DFSN) [8] у задачі прогнозування конверсій. Основна увага приділяється умовам, коли цільова подія є відкладеною у часі [6], що суттєво ускладнює процес навчання моделей [3, 6].

3. Опис даних. Для тренування моделей використано набір даних Criteo. Це широко використовуваний загальнодоступний набір даних для задачі із затримкою зворотного зв'язку, наприклад прогнозування конверсій у рекламних системах [9]. Датасет містить вибірку реального рекламного трафіку тривалістю 90 днів і налічує близько 16 млн записів.

Кожен запис відповідає одному натисканню (надалі — **клік**) по рекламному оголошенню після пошукового запиту. Після натискання користувач переходить на вебсторінку рекламодавця, де може переглянути характеристики продукту, його ціну та оформити замовлення. Далі можливі два сценарії:

- Цільова дія (надалі — конверсія) відбулася — після переходу на сайт користувач здійснив цільову дію, тобто здійснив покупку продукту. У такому випадку в даних фіксується:
 - мітка в записі `Sale = 1`;
 - час між кліком по рекламі та покупкою `time_delay_for_conversion`.
- Конверсія не відбулася — користувач перейшов за оголошенням, але не здійснив покупку. Наприклад, лише переглянув товар або залишив сайт. Тоді:
 - мітка в записі `Sale = 0` ;
 - поле `time_delay_for_conversion` має значення `-1`.

Таким чином, у датасеті відображено типовий ланцюжок взаємодії користувача з рекламою.

Для кожного кліку також доступна додаткова інформація така як, вартість товару, тип пристрою, а також числові й категоріальні характеристики користувача та продукту. До числових характеристик належать час кліку (`click_timestamp`), кількість кліків за попередній тиждень (`nb_clicks_1week`), ціна товару (`product_price`), сума покупки (`SalesAmountInEuro`) та затримка між кліком і конверсією (`time_delay_for_conversion`). До категоріальних ознак належать тип пристрою (`device_type`), вікова група (`product_age_group`), гендерна категорія (`product_gender`), бренд товару (`product_brand`), категорії товару (`product_category{1-7}`), країна продажу (`product_country`), а також анонімізовані ідентифікатори користувача, товару, продавця та аудиторії (`user_id`, `product_id`, `audience_id`, `product_title`).

3.1. Аналіз даних. Як показав аналіз даних (рис. 1), затримка конверсій може бути від декількох хвилин до декількох тижнів. Загалом лише близько 11% користувачів здійснили хоча б одну конверсію, причому, близько 65% конверсій відбуваються протягом першої доби після кліку на рекламне оголошення.

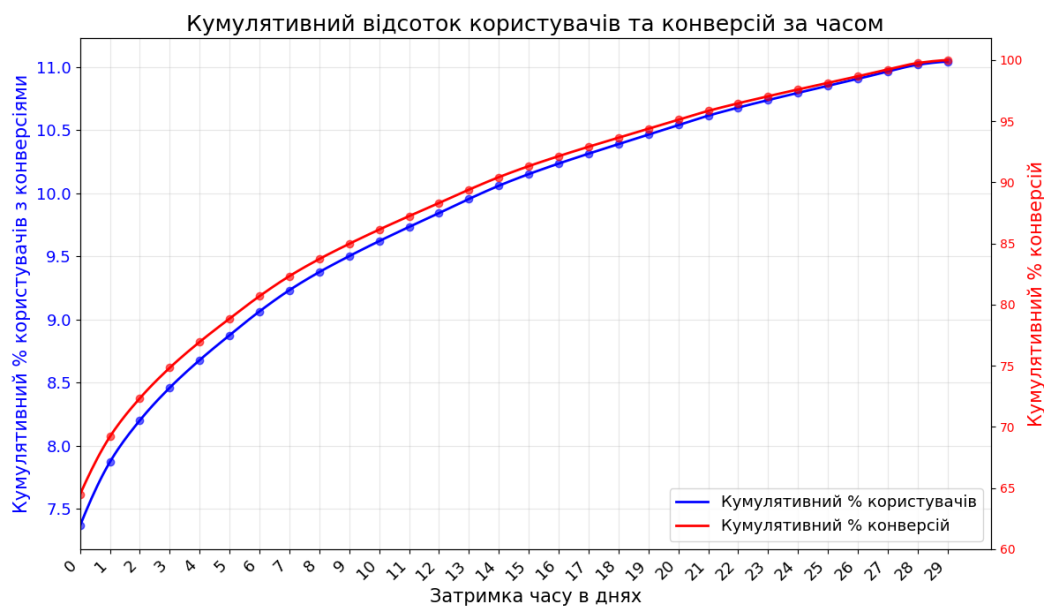


Рис. 1. Дисбаланс конверсій та затримка даних в датасеті Criteo.

Затриманий зворотний зв'язок призводить до того, що на момент формування навчального датасету частина конверсій ще не відбулась [3, 5], і відповідні записи помилково маркуються як негативні. Такий ефект формує зміщення навчальних міток (label delay bias) [4, 3, 5, 6], що негативно впливає на якість навчання традиційних моделей.

Дані в датасеті є високорозмірними та розрідженими [10], також спостерігається сильний дисбаланс класів [10], оскільки частка конверсій є значно меншою порівняно з тими де конверсія не відбулась.

4. Архітектура моделей та експериментальна частина. У дослідженні для побудови моделей прогнозування конверсій було використано засоби бібліотеки PyTorch. На етапі попередньої обробки даних числові ознаки були нормалізовані за допомогою стандартизації (zero mean, unit variance), що забезпечило стабільність градієнтного навчання [7]. Категоріальні ознаки були перетворені у числовий формат із використанням кодування міток (Label Encoding) [7]. Дані були розділені на навчальну та тестову вибірки у співвідношенні 80/20 зі збереженням балансу класів, розподілу значень категоріальних ознак та частки конверсій у кожній підвибірці.

Архітектура використаної у роботі моделі MLP [5, 7] представляє собою повнозв'язну нейронну мережу прямого поширення. Вхідний шар складається з 107 вузлів, з яких 3 для числових ознак та 104 (8×13) для закодованих категоріальних ознак представлених у вигляді 8-и розмірних векторів (Vector embedding) [7]. Після вхідного шару слідує два приховані шари розмірністю 128 та 64 нейронів відповідно. Як функцію активації було використано ReLU (Rectified Linear Unit) [7], що дозволяє ефективно моделювати нелінійні залежності у даних. Вихідний шар містить один нейрон що інтерпретується як ймовірність конверсії [7]. У процесі навчання застосовувався стандартний алгоритм зворотного поширення помилки (backpropagation) [7]. Алгоритм передбачає обчислення

градієнтів функції втрат та оновлення ваг мережі після обробки кожної підмножини навчальних прикладів (батч), що подаються моделі на одному кроці навчання. Розмір батчу складав 2048 записів.

Ключова особливість DFSN полягає у тому, що навчання відбувається асинхронно: базова модель відповідає за стабільне глобальне представлення залежностей, тоді як сателітна мережа динамічно адаптується до нових даних і компенсує затримку зворотного зв'язку [8]. Це дозволяє ефективно використовувати дані з неповністю сформованими мітками [3, 8] без очікування повного формування міток. Модель складається з трьох основних частин: базової мережі, сателітної мережі та модуля об'єднання.

Базова мережа виконує роль основного предиктора і навчалась на історичних даних, з остаточно сформованими мітками. Одна сателітна мережа була введена як допоміжний компонент і навчалась на проміжному сигналі, доступному раніше. Схему DFSN мережі представлено на Рис. 2.

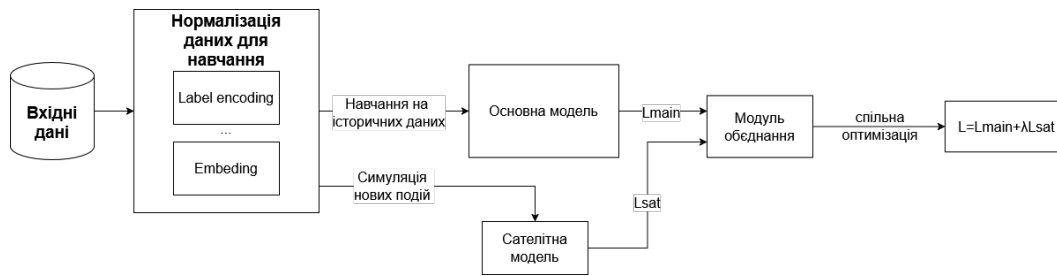


Рис. 2. Схема навчання DFSN моделі.

Сателітна мережа використовувала ті самі вхідні ознаки та векторні представлення, що і основна мережа, але для неї проводилась оптимізація власної функції втрат L_{sat} , що дозволило отримати додатковий навчальний сигнал у ситуації неповного або затриманого зворотного зв'язку. Вихід сателітної мережі використовувався як допоміжне представлення, що доповнює основну модель та її функцію втрат L_{main} . Спільне навчання, основної та сателітної мережі, здійснювалося шляхом мінімізації комбінованої функції втрат $L = L_{main} + \lambda L_{sat}$, де $\lambda \in [0, 1]$ — вага сателітного сигналу. Таким чином, навчання відбувалось в умовах частково спостережуваних міток, де сателітна нейронна мережа забезпечувала додатковий навчальний сигнал до моменту появи істинних міток конверсій для свіжих даних.

Як базова, так і сателітна мережа реалізовані у вигляді повнозв'язних нейронних мереж прямого поширення з однаковою структурою прихованих шарів. У якості функції втрат при навчання моделі була використана бінарна перехресна ентропія (binary cross-entropy), що вимірює різницю між бінарними мітками та прогнозованими моделлю ймовірностями:

$$L = -\frac{1}{N} \sum_{i=1}^N [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)],$$

де $y_i \in \{0, 1\}$ — бінарна мітка (конверсія відбулась чи ні), \hat{y}_i — ймовірність події прогнозована моделлю. N — кількість навчальних прикладів що подається

моделі з кожним батчем. Процес мінімізації функції втрат було виконано за допомогою оптимізатора Adam PyTorch, шляхом ітеративного оновлення ваг нейронної мережі методом градієнтного спуску.

5. Основний результат. Після завершення навчання, проведено оцінювання якості моделей на тестовій вибірці. Для оцінки якості прогнозу конверсій було використано метрику ROC-AUC (Receiver Operating Characteristic Area Under Curve) [2]. Для розрахунку використовувались ймовірності конверсій прогнозовані моделями. Розрахунок було проведено засобами sklearn.metrics.

Таблиця 1.

Результати оцінювання моделей на тестовій вибірці

Model	AUC-ROC
MLP	0.49
DFSN	0.72

ROC-AUC широко використовується для оцінки якості бінарних класифікаторів, та відображає здатність моделі розділяти класи (позитивний/негативний). ROC-AUC набуває значень від 0 до 1: відображає ймовірність того, що модель, отримавши випадково обрані позитивний і негативний приклади, надасть позитивному прикладу вищий рейтинг, ніж негативному, а значення близько 0.5 відповідає випадковому вгадуванню.

Значення ROC-AUC отримані для MLP та DFSN наведено у табл. 1. DFSN модель досягла значення 0.72, демонструючи задовільну здатність моделі прогнозувати позитивні приклади конверсій. В той же час, MLP модель досягла значення 0.49, що відповідає випадковому вгадуванню. Отримані результати узгоджуються з висновками наведеними у [8] де теж використовувався датасет Criteo у задачі прогнозування конверсій.

6. Висновки та перспективи подальших досліджень. DFSN архітектура продемонструвала вищу ефективність порівняно з базовою MLP моделлю в умовах затриманого зворотного зв'язку. Це пояснюється використанням сателітної мережі, яка формує додатковий навчальний сигнал на ранніх етапах після взаємодії користувача з системою та сприяє підвищенню точності прогнозування конверсій.

Подальші дослідження можуть бути спрямовані на застосування моделей які враховують механізмами затриманого зворотного зв'язку в задачах A/B тестування [1] та онлайн-експериментів, де важливо отримувати ранню оцінку ефективності варіантів в умовах частково доступних міток у момент навчання моделі. У цьому контексті перспективним є поєднання підходу DFSN із байєсівськими методами та Multi-Armed Bandit [11] алгоритмами для побудови адаптивних систем прийняття рішень у потокових даних.

Конфлікт інтересів

Автор заявляє, що не має 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

Оригінальні дані, представлені в дослідженні, відкрито доступні для академічних досліджень за адресою:

<https://ailab.criteo.com/criteo-sponsored-search-conversion-log-dataset>
під ліцензією Creative Commons CC-BY-NC-SA 4.0

Використання штучного інтелекту

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

Внесок авторів

В. І. Кудак: постановка проблеми, дослідження, пошук та аналіз даних, налаштування та тренування моделей нейронних мереж, інтерпретація результатів та підготовка публікації до друку, написання — рецензування та редагування.; А. С. Вощепинець: наукове керівництво, написання — рецензування та редагування.

Авторські права ©



(2026). Кудак В. І., Вощепинець А. С.
Ця робота ліцензується відповідно до Creative Commons Attribution 4.0 International License.

Список використаної літератури

1. Kohavi, R., Tang, D., & Xu, Y. (2020). *Trustworthy online controlled experiments: A practical guide to A/B testing*. Cambridge University Press. <https://doi.org/10.1017/9781108653985>
2. Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8), 861–874. <https://doi.org/10.1016/j.patrec.2005.10.010>
3. Li, H., Pan, F., Ao, X., Yang, Z., Lu, M., Pan, J., Liu, D., Xiao, L., & He, Q. (2021). Follow the prophet: Accurate online conversion rate prediction in the face of delayed feedback. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '21)*. (pp. 1902–1911). Association for Computing Machinery. <https://doi.org/10.1145/3404835.3463045>
4. Li, X., Chen, S., Xv, G., Zhang, L., Luo, M., Chan, Z., Sheng, X.-R., Zhu, H., Xu, J., & Lin, C. (2026). Delayed feedback modeling for post-click gross merchandise volume prediction: Benchmark, insights and approaches. *arXiv*. <https://doi.org/10.1145/3774904.379234>
5. Yang, J.-Q., & Zhan, D.-C. (2022). Generalized delayed feedback model with post-click information in recommender systems. In *Advances in Neural Information Processing Systems (NeurIPS 2022)*. <https://doi.org/10.48550/arXiv.2206.00407>
6. Chapelle, O. (2014). Modeling delayed feedback in display advertising. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery*

- and Data Mining (KDD '14)*. (pp. 1097–1105). Association for Computing Machinery. <https://doi.org/10.1145/2623330.2623634>
7. Bengio, Y., Courville, A., & Goodfellow, I. (2016). *Deep Learning*. MIT Press.
 8. Liu, Q., Li, H., Ao, X., Guo, Y., Dong, Z., Zhang, R., Chen, Q., Tong, J., & He, Q. (2023). Online conversion rate prediction via neural satellite networks in delayed feedback advertising. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '23)*. (pp. 1406–1415). Association for Computing Machinery. <https://doi.org/10.1145/3539618.3591747>
 9. McMahan, H. B., Holt, G., Sculley, D., Young, M., Ebner, D., Grady, J., Nie, L., Phillips, T., Davydov, E., Golovin, D., Chikkerur, S., Liu, D., Wattenberg, M., Hrafnkelsson, A. M., Boulos, T., & Kubica, J. (2013). Ad click prediction: A view from the trenches. In *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '13)*. (pp. 1222–1230). Association for Computing Machinery. <https://doi.org/10.1145/2487575.2488200>
 10. Cheng, H.-T., Koc, L., Harmsen, J., Shaked, T., Chandra, T., Aradhya, H., Anderson, G., Corrado, G., Chai, W., Ispir, M., Anil, R., Haque, Z., Hong, L., Jain, V., Liu, X., & Shah, H. (2016). Wide & Deep Learning for Recommender Systems. In *DLRS 2016: Workshop on Deep Learning for Recommender Systems*. ACM. <https://doi.org/10.1145/2988450.2988454>
 11. Li, L., Chu, W., Langford, J., & Schapire, R. E. (2010). A contextual-bandit approach to personalized news article recommendation. In *Proceedings of the 19th International Conference on World Wide Web (WWW '10)*. (pp. 661–670). Association for Computing Machinery. <https://doi.org/10.1145/1772690.1772758>

Kudak V. I., Voshchepynets A. S. A Comparative Analysis of MLP and DFSN for Predicting Delayed Conversions.

This paper addresses the problem of forecasting under delayed feedback conditions, which leads to a temporal shift in training labels and a degradation in the forecasting quality of classical machine learning models. Particular attention is paid to the problem of label delay bias, which arises because, at the time the training sample is formed, some positive events have not yet occurred or been recorded. The paper compares the prediction quality of two models: the Multilayer Perceptron (MLP) and Delayed Feedback Satellite Networks (DFSN), which combines a main neural network with satellite components to account for partially available data.

Experiments were conducted on the real-world industrial Criteo dataset, characterized by high feature dimensionality, sparsity, and significant class imbalance. Numerical features were standardized, while categorical features were encoded and represented as vector representations.

Model evaluation was performed using the ROC-AUC metric. The results showed that DFSN significantly outperforms the classical MLP in tasks with delayed labels, demonstrating lower sensitivity to label incompleteness during early training stages and higher prediction quality.

Keywords: MLP, DFSN, delayed feedback, neural networks.

Отримано: 03.04.2026

Прийнято: 19.04.2026

Опубліковано: 30.04.2026