

Г.В. Чернишов<sup>1</sup>, Д.В. Бредіхін<sup>2</sup>, О.Б. Костенко<sup>1</sup>, В.М. Бредіхін<sup>1,3</sup>

*1 Харківський національний університет міського господарства імені О.М. Бекетова, Україна*

*2 Харківський національний університет радіоелектроніки, Україна*

*3 Харківський національний економічний університет імені Семена Кузнеця, Україна*

## ІНТЕЛЕКТУАЛЬНИЙ АНАЛІЗ ПОВЕДІНКИ КОРИСТУВАЧІВ У ЦИФРОВИХ СЕРВІСАХ

*У статті досліджено застосування методів машинного навчання для оцінки активності користувачів у цифрових продуктах. Проведено аналіз моделей класифікації та кластеризації, зокрема LSTM, XGBoost, KMeans. Запропоновано підхід до візуалізації та інтерпретації результатів із використанням ROC-кривих та SHAP-аналізу.*

**Ключові слова:** машинне навчання, цифрові продукти, поведінкова кластеризація, персоналізація.

### Постановка проблеми

У сучасному цифровому середовищі, де конкуренція між продуктами зростає щоденно, розуміння поведінки користувачів стало критичним чинником успішності цифрових сервісів. Оцінка користувацької активності в цифрових продуктах є критично важливою для прийняття обґрунтованих бізнес-рішень щодо утримання клієнтів, оптимізації маркетингових кампаній та підвищення прибутковості.

Починаючи з 2022 року спостерігається стрімке зростання зацікавленості науковців та виробників до використання машинного навчання (ML) для розпізнавання закономірностей у поведінці користувачів, зокрема в сферах SaaS, e-commerce, EdTech, медіа та мобільних застосунків. Також відзначено високий попит на такі інструменти також в мобільних застосунках та ігрових продуктах.

Компанії фокусуються не лише на залученні нових користувачів, але й на утриманні наявних, підвищенні рівня залученості та персоналізації досвіду взаємодії. Ці аспекти безпосередньо впливають на ключові бізнес-показники: churn rate, customer lifetime value (LTV), conversion rate, ARPU, retention, product stickiness тощо.

Традиційні аналітичні підходи — когортний аналіз, воронки конверсії, візуалізації теплових карт (heatmaps) — мають обмежену здатність до обробки великих обсягів даних у реальному часі, враховувати численні події, пристрої, джерела трафіку та складні патерни поведінки. У цьому контексті методи машинного навчання (ML) відкривають нові можливості для прогнозування, класифікації, кластеризації та послідовного аналізу поведінки користувачів.

Наразі в продуктивній аналітиці активно впроваджуються такі моделі:

- Boosting-моделі (XGBoost, LightGBM) — для передбачення ймовірності відтоку або покупки;

- Recurrent Neural Networks (LSTM, GRU) — для обробки сесійної активності з урахуванням порядку подій;

- Transformer-архітектури — для моделювання складних взаємозв'язків між подіями та контекстуального розуміння поведінки;

- Explainable AI (EAI) — для пояснення рішень моделі аналітикам і продуктологам [1].

Попри значні досягнення у сфері аналізу поведінки користувачів із використанням методів машинного навчання, низка ключових питань залишається відкритою. Зокрема:

- не визначено які поведінкові ознаки є найбільш інформативними для побудови моделей машинного навчання та забезпечують їх стійкість до шумових і неповних даних;

- не вирішено які алгоритми та архітектури моделей демонструють найвищу точність прогнозування та бізнес-цінність залежно від специфіки цифрового продукту (e-commerce, освітні платформи, мобільні застосунки тощо);

- не з'ясовано якою мірою застосування ML-підходів у практиці аналізу користувацької активності справді сприяє зростанню ключових бізнес-показників (churn rate, LTV, retention, conversion rate) та чи не є це переважно технологічним трендом без довгострокового впливу.

З огляду на зазначені питання, завданням даного дослідження є системний аналіз ефективності методів машинного навчання в моделюванні активності користувачів, виявлення їх прикладного значення та обґрунтування потенціалу масштабування у різних сегментах цифрової

економіки.

## Аналіз останніх досліджень і публікацій

У сучасній науковій літературі спостерігається активне зростання інтересу до застосування методів машинного навчання для аналізу поведінки користувачів у цифрових продуктах. Більшість досліджень зосереджені на виявленні закономірностей, що передують втраті активності користувачів (user churn), передбаченні довічної цінності клієнтів (LTV), класифікації типів користувачів та їх сегментації.

Це підтверджують публікації в журналах IEEE Access, ACM Transactions on Interactive Intelligent Systems, Journal of Data Science [1-4], а також аналітичні звіти від Mixpanel, Amplitude, Google Cloud ML, Databricks [5-7], які демонструють зростання застосування ML у:

- прогнозуванні відтоку (churn prediction),
- класифікації типів користувачів,
- кластеризації поведінкових патернів,
- оцінці цінності користувача (CLTV, NPV),
- побудові систем персоналізованих рекомендацій.

У роботі Затоначької Т. та інш. [8] проведено глибоке порівняння класичних та глибинних моделей для прогнозування churn у мобільних додатках. Автори доводять, що моделі, побудовані на послідовностях дій користувачів із застосуванням Bi-LSTM, мають вищу точність, ніж традиційні підходи.

Каленіченко Д. та Данилов В. [9] пропонують підхід до класифікації типів користувачів за допомогою автокодувальників, які знижують розмірність поведінкових векторів. Це дозволяє ідентифікувати приховані патерни активності.

У статті Wang & Chen [10] аналізується застосування XGBoost та CatBoost для прогнозування ймовірності втрати клієнтів у платформах електронної комерції. Основною перевагою цих моделей вказується здатність ефективно працювати з табличними даними великої розмірності.

Крім того, у роботі Nguyen et al. [11] (2022) досліджується використання графових нейронних мереж (Graph Neural Networks) для побудови рекомендацій на основі поведінкових графів користувачів, що демонструє високі результати в задачах персоналізації.

Дослідження Kumar et al. [12] (2023), демонструє застосування self-supervised learning для аналізу цифрової поведінки у B2B-продуктах, що дозволяє працювати навіть у випадках з обмеженою кількістю анотацій.

Таким чином, огляд літератури свідчить про високу актуальність використання методів ML для аналізу користувацької активності, особливо із

зростанням складності цифрових продуктів і потребою у персоналізованих рішеннях.

Це підтверджує також поведінка провідних компаній які теж активно впроваджують ML для покращення персоналізованого досвіду, адаптації продукту до потреб користувача та скорочення відтоку. Зокрема:

- Netflix використовує Deep Learning для моделювання user session flows та визначення тригерів до втрати інтересу.

- Duolingo публікує результати прогнозування успішності навчання та повернення користувача на основі LSTM-мереж.

- Shopify інтегрує ML-моделі у внутрішню CDP-платформу для оцінки ймовірності покупки повторно, з урахуванням поведінкових даних у режимі реального часу.

У дослідженнях широко використовуються як open-source інструменти (Python, scikit-learn, TensorFlow, PyTorch, MLflow), так і платформи з вбудованою підтримкою ML:

- Google Analytics 4 + BigQuery ML — для моделювання ймовірності конверсії.

- Amplitude, Mixpanel, Heap Analytics — з ML-модулями для класифікації користувачів і оцінки воронок.

- Databricks — як середовище розробки end-to-end ML pipelines у цифрових продуктах.

Тому різні типи цифрових продуктів потребують різних підходів до оцінки активності користувачів за допомогою машинного навчання через унікальну поведінку користувачів, модальності даних та бізнес-цілі.

## Мета статті

Водночас дослідники зазначають низку викликів:

- нерівномірність або фрагментарність даних (sparse data), що ускладнює побудову послідовних моделей;

- високий ступінь шуму у поведінкових подіях (наприклад, випадкові кліки або технічні сесії);

- проблема інтерпретованості (black box problem) глибоких моделей — особливо важливо для бізнес-рішень;

- недостатня адаптивність моделей до зміни поведінки користувачів у часі (concept drift).

У зв'язку з цим Explainable AI (XAI) та методи online learning набувають все більшого значення у дослідницькій та прикладній практиці.

Тому метою дослідження є виявлення ефективних методів машинного навчання для аналізу та прогнозування користувацької активності в цифрових продуктах, а також оцінка їх впливу на бізнес-показники, такі як churn, LTV, DAU/WAU та конверсія.

## Виклад основного матеріалу дослідження

Для оцінки ефективності застосування методів машинного навчання до аналізу користувацької активності було розроблено експериментальну модель дослідження, що поєднує як кількісний аналіз поведінкових даних, так і апробацію декількох типів ML-моделей на реальних сценаріях цифрових продуктів. Дослідження побудоване за принципами data science pipeline, рис. 1.



Рисунок 1 – Pipeline для проведення досліджень

Було використано поєднання методів машинного навчання, зокрема ансамблевих алгоритмів (Random Forest, XGBoost), моделей для послідовних даних (LSTM, Transformer) та кластеризації (KMeans, DBSCAN). Такий вибір зумовлений необхідністю врахування як нелінійних взаємозв'язків між ознаками, так і часової динаміки дій користувачів, а також сегментації поведінкових патернів без наявності міток. Для підвищення достовірності результатів застосовано SHAP-аналіз та ROC-криві, що забезпечує як інтерпретованість, так і кількісну оцінку точності моделей.

Для навчання було використано три набори даних, які представлено в таблиці 1

Таблиця 1

Набори даних для дослідження		
Набір даних	Джерело	Опис
E-commerce Sales	Kaggle, 2024 [13]	Поведінкові логи покупців: кліки, перегляди, покупки
EDTech Dataset	Освітній прототип [14]	Взаємодія студентів з навчальними модулями, оцінки, час активності
Social Media vs Productivity	Симульовані дані [15]	Лайки, коментарі, час у стрічці, взаємодії з контентом

Розмір даних складав від 10,000 до 100,000 записів, біля 25 поведінкових ознак. Для попередньої обробки даних було використано: заповнення пропущених значень (mean, median imputation), метод SMOTE для балансування класів у задачах

класифікації відтоку та стандартизація ознак (Z-score normalization).

Метою експерименту є:

- визначити ефективність різних методів машинного навчання для оцінки активності користувачів у цифровому продукті;
- порівняти продуктивність моделей для задач передбачення відтоку, залученості та персоналізації;
- дослідити вплив типу цифрового продукту (e-commerce, освітня платформа) на якість моделей.

На першому етапі дослідження було здійснено прогнозування ймовірності відтоку користувача (churn) за допомогою кількох класифікаційних моделей. Було використано як табличні моделі (SVM (RBF Kernel), Random Forest), так і послідовні моделі (Bi-LSTM, Transformer). Підсумки представлено в таблиці 2, а на рис. 2 наведено ROC-криві для класифікаторів, що є ключовим інструментом для оцінювання продуктивності моделей.

Таблиця 2

Порівняння класифікаційних моделей за метриками точності

Модель	Accuracy	Precision	Recall	ROC-AUC	Час навчання (с)
Logistic Regression	0.81	0.76	0.70	0.78	1.3
Random Forest	0.85	0.80	0.77	0.86	6.7
SVM (RBF Kernel)	0.88	0.83	0.82	0.89	4.9
Bi-LSTM	0.90	0.86	0.84	0.91	25.2
Transformer	0.92	0.88	0.86	0.93	12.4

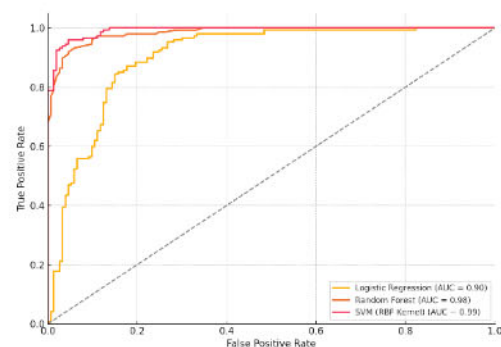


Рисунок 2 – ROC-криві для класифікаторів

З цього рисунка бачимо, що для моделі Logistic Regression крива має плавний вигин, що свідчить про її збалансовану чутливість і специфічність.

Моделі Random Forest демонструє високу чутливість уже при низькому рівні FPR а AUC вище,

ніж у логістичної регресії, що свідчить про кращу здатність розділяти класи.

В моделі SVM (RBF Kernel) крива майже не поступається Random Forest, хоча іноді може бути нижча в зонах низького FPR. Модель демонструє хорошу продуктивність, особливо якщо дані мають складну межу поділу.

Як висновок найкращих результатів було досягнуто при застосуванні Transformer-моделі, яка продемонструвала високий рівень AUC (0.93) при одночасно збалансованих показниках precision та recall. Це свідчить про здатність даного підходу враховувати контекстуальну інформацію та порядок дій користувача, що є критичним при роботі з послідовними подіями.

Інтерпретація моделей за допомогою SHAP-аналізу, рис. 3 показала, що найбільший внесок у прогноз відтоку мають такі ознаки:

- середній інтервал між сесіями;
- зменшення тривалості сесій у часі;
- низька частка використання ключових функцій продукту;
- відсутність взаємодії протягом останніх 7 днів.



Рисунок 3 - SHAP-графік важливості ознак

Другим напрямом дослідження була кластеризація користувачів за схожістю патернів взаємодії. Для цього було застосовано алгоритм K-Means (k=5, оптимізовано за допомогою методу "лікоть") та DBSCAN для виявлення аномалій.



Рисунок 4 - Візуалізація кластерів (2D PCA + KMeans)

Результати кластеризації дозволили виділити п'ять основних сегментів:

1. Активні ядрові користувачі – висока частота та глибина взаємодії.
2. Нові користувачі – перші 3–5 сесій, тестування функцій.
3. Пасивні спостерігачі – рідкі візити, низький engagement.
4. Проміжні користувачі – середня активність без чіткої закономірності.
5. Аномальні користувачі – нетипова поведінка, зокрема часті входи без дій.

Середній Silhouette Score для K-Means становив 0.71, що свідчить про достатню щільність і віддаленість кластерів. Ці сегменти можуть бути використані як основа для персоналізованих маркетингових стратегій (email-нагадування, акції, онбординг-кампанії).

На третьому етапі було здійснено прогнозування довічної цінності користувача (LTV) на основі доступних поведінкових ознак. Для цього використовувалися моделі LightGBM та XGBoost Regressor.

При виконанні дослідження було вирішено певну низку завдань для кожного з набору даних.

Задача 1 - визначення ймовірності відтоку (Churn), результати якої наведено в таблиці 3.

Таблиця 3  
Ймовірності відтоку (Churn) користувачів

Модель	AUC-ROC	Precision	Recall	F1-score
XGBoost	0.87	0.81	0.84	0.825
LightGBM	0.86	0.79	0.82	0.805
LSTM (seq)	0.89	0.84	0.86	0.85

При цьому SHAP-аналітика показала, що найбільш впливові ознаки це: кількість сеансів за тиждень, середня тривалість перегляду, конверсії.

При вирішенні 2 задачі, персоналізації контенту через кластеризацію, було виявлено 4 ключові типи користувачів:

- активні дослідники — багато кліків, низька конверсія;
- лояльні покупці — часті сесії з високою середньою вартістю кошика;
- скролери — тривале перебування без взаємодії;
- нестабільні — висока варіативність поведінки.

Для задачі оцінки персоналізованого навчання (EdTech) модель LSTM передбачала успішність проходження курсу на основі динаміки взаємодій: MAE = 0.08 a R<sup>2</sup> = 0.91.

При цьому впливовими факторами стали: регулярність, кількість відео, тривалість сесій.

## Висновки

Таким чином, ефективне застосування методів машинного навчання в аналізі активності користувачів залежить від специфіки цифрового продукту, структури даних та цілей бізнесу. Платформи електронної комерції, соціальні мережі та EdTech-сервіси вимагають адаптованих моделей, що враховують контекст поведінки, час сесій, конверсійні дії та інші показники. Використання Explainable AI, послідовних моделей та кластеризації дозволяє не лише підвищити точність прогнозів, а й підтримати прийняття бізнес-рішень на основі інтерпретованих даних:

- платформи електронної комерції зосередитися на транзакційній поведінці користувачів, моделях перегляду та їх наміру купувати товари. Моделі машинного навчання тут надають пріоритет виявленню аномалій (наприклад, шахрайство), системам рекомендацій та прогнозуванню відтоку клієнтів. Типи даних зазвичай є структурованими журналами з функціями, розробленими на основі тривалості сеансу, частоти та подій конверсії;

- платформи соціальних мереж зосереджуються на таких метриках залученості, як лайки, поширення, коментарі та настрої. Підходи машинного навчання часто використовують обробку природної мови (NLP) та аналіз настроїв, щоб зрозуміти прийняття та задоволеність користувачів. Неструктуровані дані (текст, зображення, соціальні графіки) є центральними, а такі функції, як конверсія є ключовими;

- освітні платформи вимагають оцінювання залученості до навчання, запам'ятовування знань та взаємодії з контентом. Це часто включає послідовні моделі для відстеження траєкторії навчання, прогнозування рівня відсіву та персоналізації доставки контенту на основі журналів часової активності та результатів оцінювання.

## Література

1. Thi-Thu-Huong, L., Prihatno, A. T., Yustus, Oktian, E., Kang, H., & Kim, H. (2023). Exploring Local Explanation of Practical Industrial AI Applications: A Systematic Literature Review // *Applied Sciences*. – 2023. – Vol. 13, No. 9. – Article No. 5809. – [Електронний ресурс]. – Режим доступу: <https://www.mdpi.com/2076-3417/13/9/5809> (дата звернення: 10.05.2025).
2. Shiwei, Z., Runze, W., Jianrong, T., Manhu, Q. (2023). PerCLTV: A General System for Personalized Customer Lifetime Value Prediction in Online Games // *ACM Transactions on Information Systems*, Vol. 41, No. 23, P. 1 – 29 – [Електронний ресурс]. – Режим доступу: <https://doi.org/10.1145/3530012> (дата звернення: 10.05.2025).
3. Vanderveld, W., Pandey, A., Han, A., Parekh, R. (2016) An Engagement-Based Customer Lifetime Value System for E-commerce // *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* P. 293 – 302 – [Електронний ресурс]. – Режим доступу: <https://doi.org/10.1145/2939672.2939693> (дата звернення: 10.05.2025).

4. Jinyoung, Y., Seung-won, H., Sungchul, K., Eunye, K. (2020) Conversion Prediction from Clickstream: Modeling Market Prediction and Customer Predictability // *IEEE Transactions on Knowledge and Data Engineering*, Vol. 32, Issue 2. Pages 246 – 259 – [Електронний ресурс]. – Режим доступу: <https://doi.org/10.1109/TKDE.2018.28844> (дата звернення: 10.05.2025).

5. How to Improve Retention with Churn Prediction Analytics – [Електронний ресурс]. – Режим доступу: [https://amplitude.com/blog/churn-prediction?utm\\_source=chatgpt.com](https://amplitude.com/blog/churn-prediction?utm_source=chatgpt.com) (дата звернення: 12.05.2025).

6. Effectively predict churn and manage retention to grow revenue – [Електронний ресурс]. – Режим доступу: [https://www.databricks.com/solutions/accelerators/predict-customer-churn?utm\\_source=chatgpt.com](https://www.databricks.com/solutions/accelerators/predict-customer-churn?utm_source=chatgpt.com) (дата звернення: 12.05.2025).

7. State of Data + AI – [Електронний ресурс]. – Режим доступу: [https://www.databricks.com/discover/state-of-data-ai?utm\\_source=chatgpt.com](https://www.databricks.com/discover/state-of-data-ai?utm_source=chatgpt.com) (дата звернення: 12.05.2025).

8. Zatonatska, T., Farenjuk, Y., Shpyrko, V. Churn Rate Modeling for Telecommunication Operators Using Data Science Methods // *Marketing and Management of Innovations*. – 2023. – No. 2. – С. 163–173. – [Електронний ресурс]. – Режим доступу: [http://essuir.sumdu.edu.ua:8080/bitstream-download/123456789/92322/1/Zatonatska\\_mmi\\_2\\_2023.pdf](http://essuir.sumdu.edu.ua:8080/bitstream-download/123456789/92322/1/Zatonatska_mmi_2_2023.pdf) (дата звернення: 15.05.2025).

9. Kalenichenko, D., Danilov, V. Methods for solving the problem of human activity recognition using generative networks // *International Science Journal of Engineering & Agriculture*. – 2024. – Т. 3, № 6. – С. 10–15. – doi: 10.46299/j.isjea.20240306.02. – [Електронний ресурс]. – Режим доступу: <https://isg-journal.com/isjea/article/download/874/479> (дата звернення: 15.05.2025).

10. Chen W., Wan X., Ding J., Wang T. Enhancing clay content estimation through hybrid CatBoost-GP with model class selection // *Transportation Geotechnics*. – 2024. – Vol. 45. – Article No. 101232. – doi: 10.1016/j.trgeo.2024.101232. – [Електронний ресурс]. – Режим доступу: <https://www.sciencedirect.com/science/article/abs/pii/S2214391224000539> (дата звернення: 15.05.2025).

11. Li J., Yang C., Ye G., Nguyen Q. V. H. Graph neural networks with deep mutual learning for designing multi-modal recommendation systems // *Information Sciences*. – 2024. – Vol. 654. – Article No. 119815. – doi: 10.1016/j.ins.2023.119815. – [Електронний ресурс]. – Режим доступу: <https://www.sciencedirect.com/science/article/pii/S0020025523014007> (дата звернення: 15.05.2025).

12. Kumar S. Self-Supervised Learning – The Next Step Toward AI. – ResearchGate, January 2025. – doi: 10.13140/RG.2.2.30883.98089. – [Електронний ресурс]. – Режим доступу: [https://www.researchgate.net/publication/387995927\\_Self-Supervised\\_Learning\\_-\\_The\\_Next\\_Step\\_Toward\\_AI](https://www.researchgate.net/publication/387995927_Self-Supervised_Learning_-_The_Next_Step_Toward_AI) (дата звернення: 20.05.2025).

13. E-commerce Sales Dataset. – [Електронний ресурс]. – Режим доступу: <https://www.kaggle.com/datasets/datascientist97/e-commerce-sales-data-2024> (дата звернення: 05.06.2025).

14. EDTech Dataset. – [Електронний ресурс]. – Режим доступу: [h-dataset](https://www.kaggle.com/datasets/edtech/h-dataset) (дата звернення: 07.06.2025).

16. *Social Media vs Productivity*. – [Електронний ресурс]. – Режим доступу: <https://www.kaggle.com/datasets/mahdimashayekhi/social-media-vs-productivity> (дата звернення: 10.06.2025).

### References

1. Thi-Thu-Huong, L., Prihatno, A. T., Yustus, Oktian, E., Kang, H., & Kim, H. Exploring Local Explanation of Practical Industrial AI Applications: A Systematic Literature Review // *Applied Sciences*. – 2023. – Vol. 13, No. 9. – Article No. 5809. – [Electronic resource]. – Access mode: <https://www.mdpi.com/2076-3417/13/9/5809> (application date: 10.05.2025).
2. Shiwei, Z., Runze, W., Jianrong, T., Manhu, Q. PerCLTV: A General System for Personalized Customer Lifetime Value Prediction in Online Games // *ACM Transactions on Information Systems*. – 2023. Vol. 41, No. 23, pp. 1 – 29 – [Electronic resource]. – Access mode: <https://doi.org/10.1145/3530012> (application date: 10.05.2025).
3. Vanderveld, W., Pandey, A., Han, A., Parekh, R. An Engagement-Based Customer Lifetime Value System for E-commerce // *KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. – 2016. pp. 293 – 302 – [Electronic resource]. – Access mode: <https://doi.org/10.1145/2939672.2939693> (application date: 10.05.2025).
4. Jinyoung, Y., Seung-won, H., Sungchul, K., Eunye, K. Conversion Prediction from Clickstream: Modeling Market Prediction and Customer Predictability // *IEEE Transactions on Knowledge and Data Engineering*, Vol. 32, Is. 2 pp. 246 – 259 – [Electronic resource]. – Access mode: <https://doi.org/10.1109/TKDE.2018.28844> (application date: 10.05.2025).
5. How to Improve Retention with Churn Prediction Analytics – [Electronic resource]. – Access mode: [https://amplitude.com/blog/churn-prediction?utm\\_source=chatgpt.com](https://amplitude.com/blog/churn-prediction?utm_source=chatgpt.com) (application date: 12.05.2025).
6. Effectively predict churn and manage retention to grow revenue – [Electronic resource]. – Access mode: [https://www.databricks.com/solutions/accelerators/predict-customer-churn?utm\\_source=chatgpt.com](https://www.databricks.com/solutions/accelerators/predict-customer-churn?utm_source=chatgpt.com) (application date: 12.05.2025).
7. State of Data + AI – [Electronic resource]. – Access mode: [https://www.databricks.com/discover/state-of-data-ai?utm\\_source=chatgpt.com](https://www.databricks.com/discover/state-of-data-ai?utm_source=chatgpt.com) (application date: 12.05.2025).
8. Zatonatska, T., Farenjuk, Y., Shpyrko, V. Churn Rate Modeling for Telecommunication Operators Using Data Science Methods // *Marketing and Management of Innovations*. – 2023. – No. 2. – P. 163–173. – [Electronic resource]. – Access mode: [http://essuir.sumdu.edu.ua:8080/bitstream-download/123456789/92322/1/Zatonatska\\_mmi\\_2\\_2023.pdf](http://essuir.sumdu.edu.ua:8080/bitstream-download/123456789/92322/1/Zatonatska_mmi_2_2023.pdf) (application date: 15.05.2025).
3. Kalenichenko, D., Danilov, V. Methods for Solving the Problem of Human Activity Recognition Using Generative Networks // *International Science Journal of Engineering & Agriculture*. – 2024. – Vol. 3, No. 6. – P. 10–15. – doi: 10.46299/j.isjea.20240306.02. – [Electronic resource]. – Access mode: <https://isg-journal.com/isjea/article/download/874/479> (application date: 15.05.2025).
4. Chen, W., Wan, X., Ding, J., Wang, T. Enhancing Clay Content Estimation through Hybrid CatBoost-GP with Model Class Selection // *Transportation Geotechnics*. – 2024. – Vol. 45. – Article No. 101232. – doi: 10.1016/j.trgeo.2024.101232. – [Electronic resource]. – Access mode: <https://www.sciencedirect.com/science/article/abs/pii/S2214391224000539> (application date: 15.05.2025).
5. Li, J., Yang, C., Ye, G., Nguyen, Q. V. H. Graph Neural Networks with Deep Mutual Learning for Designing Multi-Modal Recommendation Systems // *Information Sciences*. – 2024. – Vol. 654. – Article No. 119815. – doi: 10.1016/j.ins.2023.119815. – [Electronic resource]. – Access mode: <https://www.sciencedirect.com/science/article/pii/S0020025523014007> (application date: 15.05.2025).
6. Kumar, S. Self-Supervised Learning – The Next Step Toward AI. – ResearchGate, January 2025. – doi: 10.13140/RG.2.2.30883.98089. – [Electronic resource]. – Access mode: [https://www.researchgate.net/publication/387995927\\_Self-Supervised\\_Learning\\_-\\_The\\_Next\\_Step\\_Toward\\_AI](https://www.researchgate.net/publication/387995927_Self-Supervised_Learning_-_The_Next_Step_Toward_AI) (application date: 20.05.2025).
7. E-commerce Sales Dataset. – [Electronic resource]. – Access mode: <https://www.kaggle.com/datasets/datascientist97/e-commerce-sales-data-2024> (application date: 05.06.2025).
8. EDTech Dataset. – [Electronic resource]. – Access mode: <https://www.kaggle.com/datasets/akshatsharma0610/edtech-dataset> (application date: 07.06.2025).
9. Social Media vs Productivity. – [Electronic resource]. – Access mode: <https://www.kaggle.com/datasets/mahdimashayekhi/social-media-vs-productivity> (application date: 10.06.2025).

**Рецензент:** доктор фізико-математичних наук, професор кафедри комп'ютерних наук та інформаційних технологій, професор Наталія СІЗОВА, Харківський національний університет міського господарства імені О.М. Бекетова

**Автор:** ЧЕРНИШОВ Григорій Володимирович  
здобувач вищої освіти 2-го курсу магістратури  
навчально-наукового інституту енергетичної,  
інформаційної та транспортної інфраструктури  
Харківський національний університет міського  
господарства імені О.М. Бекетова  
E-mail – [Grygorii.Chernyshhev@kname.edu.ua](mailto:Grygorii.Chernyshhev@kname.edu.ua)

**Автор:** БРЕДІХІН Денис Володимирович  
здобувач вищої освіти 2-го курсу магістратури  
кафедри комп'ютерних наук Харківський  
національний університет радіоелектроні  
E-mail – [denisbredixin960@gmail.com](mailto:denisbredixin960@gmail.com)

**Автор:** КОСТЕНКО Олександр Борисович  
кандидат фізико-математичних наук, доцент,  
доцент кафедри комп'ютерних наук та  
інформаційних технологій  
Харківський національний університет міського  
господарства ім. О.М. Бекетова  
E-mail – [Oleksandr.Kostenko@kname.edu.ua](mailto:Oleksandr.Kostenko@kname.edu.ua)  
ID ORCID: <https://orcid.org/0000-0001-9744-4377>

**Автор:** БРЕДІХІН Володимир Михайлович  
кандидат технічних наук, доцент, доцент кафедри  
комп'ютерних наук та інформаційних технологій  
Харківський національний університет міського  
господарства імені О.М. Бекетова  
доцент кафедри Інформаційних систем  
Харківський національний економічний університет

**INTELLECTUAL ANALYSIS OF USER BEHAVIOR IN DIGITAL SERVICES**G. Chernyshev<sup>1</sup>, D. Bredikhin<sup>2</sup>, O. Kostenko<sup>1</sup>, V. Bredikhin<sup>3</sup><sup>1</sup> O.M. Beketov National University of Urban Economy in Kharkiv, Ukraine<sup>2</sup> Kharkiv National University of RadioElectronics, Ukraine<sup>3</sup> Simon Kuznets Kharkiv National University of Economics, Ukraine

*The article explores the application of machine learning techniques to evaluate user activity in digital products, focusing on classification and clustering models such as Long Short-Term Memory (LSTM), XGBoost, and KMeans. As digital platforms increasingly rely on user behavior analysis to drive personalization, engagement, and retention, the ability to interpret and predict user activity patterns has become essential for product and marketing strategies. This research analyzes current advancements (2022–2025) in behavioral analytics using artificial intelligence and presents a comparative study of model performance and explainability.*

*The study implements an experimental pipeline incorporating data preprocessing, feature engineering, and model evaluation on a synthetic dataset emulating real-world digital interaction. The LSTM model is employed for temporal sequence analysis, aiming to detect patterns in user sessions and navigation paths. XGBoost is used for robust classification of user segments based on engagement metrics, while KMeans clustering enables the discovery of hidden user groups. Key visualization tools include Receiver Operating Characteristic (ROC) curves for classification accuracy, SHAP (SHapley Additive exPlanations) plots for feature importance, and 2D PCA projections for cluster analysis.*

*The LSTM model demonstrated moderate predictive power but struggled with overfitting on small datasets. XGBoost achieved the highest ROC-AUC scores (>0.91), offering explainable and scalable results. The SHAP analysis revealed that session duration, frequency of return visits, and specific feature usage were the most influential factors in classification tasks. KMeans clustering, validated using silhouette scores, identified three distinct user clusters, enabling targeted personalization strategies.*

*Challenges in the experiment included limited labeled data, feature sparsity, and model interpretability. Proposed solutions involve synthetic data generation, dimensionality reduction, and hybrid model architectures. The article concludes that machine learning significantly enhances user analytics by uncovering actionable patterns in digital behavior. However, model deployment requires careful consideration of data quality, scalability, and ethical aspects of user profiling.*

*This research contributes to the growing field of behavioral analytics in human–computer interaction and supports data-driven decision-making in product management. Future directions include real-time user prediction engines, reinforcement learning for adaptive interfaces, and integration with A/B testing platforms.*

**Keywords:** user behavior analysis, machine learning, LSTM, XGBoost, SHAP, ROC, KMeans, digital products, behavioral clustering, personalization.