

УДК 004.85:[004.738.5:339]

**АНАЛІЗ МЕТОДІВ МАШИННОГО НАВЧАННЯ
ДЛЯ ПРОГНОЗУВАННЯ ПОВЕДІНКИ КОРИСТУВАЧІВ
В E-COMMERCE**

Писаренко К.К.

e-mail: katelyna.pysarenko1@nure.ua

Науковий керівник – к. пед. н., доц. Шеховцова В.І.

Харківський національний університет радіоелектроніки, каф. ІУС
м. Харків, Україна

In this study, machine learning methods for predicting user behavior in e-commerce systems are examined under conditions of rapidly increasing volumes of behavioral data. The main approaches to modeling customer behavior are reviewed, and a comparative evaluation of classification algorithms and ensemble methods is conducted. As an applied case, the prediction of a customer's probability of product return is considered. The results demonstrate the effectiveness of ensemble methods for improving predictive accuracy and supporting decision-making in e-commerce.

В умовах високої конкуренції e-commerce платформ особливої актуальності набуває застосування алгоритмів машинного навчання для вирішення конкретних прикладних задач, пов'язаних із підвищенням ефективності бізнес-процесів.

Метою дослідження є порівняльний аналіз методів машинного навчання для прогнозування поведінки користувачів у системах електронної комерції та обґрунтування використання ансамблевих алгоритмів для задачі прогнозування повернення товару.

Поведінка користувачів в електронній комерції характеризується динамічністю, швидкістю прийняття рішень, мобільністю та високим рівнем персоналізації. Користувачі мають доступ до широкого спектра альтернатив, порівнюють товари та взаємодіють із брендами через цифрові канали комунікації. У таких умовах здатність передбачати наступні дії клієнта на основі попередніх поведінкових сигналів стає стратегічно важливою для бізнесу. Прогнозування дозволяє не лише реагувати на попит, а й формувати його через персоналізовані рекомендації та адаптивні маркетингові механізми [1]. Для моделювання поведінки користувачів застосовуються алгоритми класифікації, регресії, кластеризації та глибокого навчання. Класифікаційні моделі використовуються для визначення ймовірності покупки або відтоку клієнта, регресійні – для оцінювання купівельної активності, а кластеризація забезпечує сегментацію аудиторії за поведінковими характеристиками [2]. Глибокі нейронні мережі, зокрема рекурентні архітектури, застосовують для аналізу послідовності дій користувача та прогнозування його наступного кроку в режимі реального часу. Особливе значення мають рекомендаційні системи, що поєднують алгоритми колаборативної та контентної

фільтрації з методами глибокого навчання, формуючи індивідуалізовані пропозиції відповідно до поведінкових патернів клієнта.

У сфері електронної комерції актуальною прикладною задачею є прогнозування ймовірності повернення товару покупцем. Високий рівень повернень збільшує логістичні витрати, ускладнює управління запасами та знижує рентабельність інтернет-магазинів. Оскільки це задача бінарної класифікації, основну увагу доцільно приділяти оцінюванню алгоритмів за їх класифікаційною ефективністю. Для її розв'язання застосовують різні алгоритми машинного навчання, зокрема логістичну регресію, дерева рішень, випадковий ліс (Random Forest), наївний байєсівський класифікатор, k-найближчих сусідів, AdaBoost та градієнтний бустинг.

Порівняльний аналіз методів виконано (див. табл. 1) на основі узагальнення результатів сучасних досліджень у сфері електронної комерції та машинного навчання [2–4].

Таблиця 1 – Порівняльна характеристика методів машинного навчання

Метод	Основні переваги	Основні обмеження	Критерії оцінювання					
			А	Б	В	Г	Д	Е
Логістична регресія	Простота реалізації; можливість інтерпретації коефіцієнтів; стабільність на малих вибірках	Обмежена здатність моделювати складні нелінійні залежності	4	3	2	5	3	5
Дерева рішень	Наочність структури рішень; відсутність вимог до масштабування ознак	Схильність до перенавчання; нестабільність при зміні даних	4	3	3	5	3	4
Random Forest	Висока точність; зменшення ризику перенавчання; стійкість до шуму	Зниження інтерпретованості; підвищені обчислювальні витрати	5	4	3	3	5	3
AdaBoost	Підвищення точності за рахунок послідовного комбінування моделей	Чутливість до викидів; складність налаштування параметрів	5	3	2	2	4	3
Gradient Boosting	Висока прогностична здатність; контроль складності моделі	Значні обчислювальні ресурси; ризик перенавчання без регуляризації	5	4	2	2	5	2
k-NN	Простота концепції; відсутність етапу явного навчання	Чутливість до розмірності простору ознак; зниження ефективності на великих вибірках	3	3	3	3	2	3
Наївний байєсівський класифікатор	Висока швидкість; ефективність при великій кількості ознак	Припущення незалежності ознак, що не завжди виконується	3	1	2	4	3	5

Оцінювання проведено за 5-бальною шкалою (1 – низький рівень відповідності критерію, 5 – високий) за такими критеріями:

А – придатність до задач класифікації; Б – придатність до регресійних задач; В – придатність до задач кластеризації; Г – інтерпретованість результатів; Д – стійкість до випадкових похибок у даних; Е – обчислювальна ефективність.

Як свідчать результати порівняльного аналізу, базові алгоритми класифікації, зокрема логістична регресія та дерева рішень, характеризуються високою інтерпретованістю та простотою реалізації, що робить їх придатними для побудови початкових моделей прогнозування [3]. Логістична регресія дозволяє отримувати ймовірнісну оцінку належності об'єкта до класу, тоді як дерева рішень забезпечують формування наочних правил прийняття рішень.

Водночас аналіз показує, що за ключовими критеріями класифікаційної ефективності та стійкості до випадкових похибок найвищі результати демонструють ансамблеві методи, зокрема Random Forest та алгоритми бустингу. Поєднання декількох моделей у межах ансамблю підвищує узагальнену здатність алгоритмів і зменшує вплив випадкових коливань вибірки, що є важливим у задачах прогнозування повернення товару [4]. Проведений аналіз підходів до розв'язання прикладної задачі прогнозування повернення товару свідчить про доцільність використання ансамблевих методів, зокрема випадкового лісу та алгоритмів бустингу, які забезпечують підвищення точності прогнозування, стійкість до випадкових похибок у даних та покращену здатність моделі до узагальнення. Застосування таких моделей створює можливості для зниження логістичних витрат, оптимізації управління запасами та підвищення ефективності бізнес-процесів e-commerce платформ.

Перспективними напрямками подальших досліджень є експериментальне оцінювання ефективності різних ансамблевих алгоритмів на реальних даних електронної комерції, аналіз впливу структури ознак на точність прогнозування, а також інтеграція прогнозних моделей у системи підтримки прийняття управлінських рішень.

Список використаних джерел:

1. Машинне навчання в електронній комерції – прогнозування поведінки клієнтів та персоналізація пропозиції // InPost. URL: <https://inpost.pl/ua/novyny-mashynne-navchannya-v-elektronniy-komertsiyi-prohnozuvannya-povedinky-kliyentiv-ta> (дата звернення: 20.02.2026).
2. Han J., Kamber M., Pei J. Data Mining: Concepts and Techniques. Waltham : Morgan Kaufmann, 2011. 740 p.
3. Rogers S., Girolami M. A First Course in Machine Learning. Boca Raton: Chapman & Hall/CRC, 2015. 306 p.
4. Zhang X., Guo F., Chen T., Pan L., Beliakov G., Wu J. A Brief Survey of Machine Learning and Deep Learning Techniques for E-Commerce Research // Journal of Theoretical and Applied Electronic Commerce Research. 2023. Vol. 18, No 4. P. 2188–2216.