

ПРОГНОЗУВАННЯ ПОВЕДІНКИ КОРИСТУВАЧІВ В ОНЛАЙН-МАГАЗИНАХ: ПОРІВНЯЛЬНИЙ АНАЛІЗ МОДЕЛЕЙ РЕКОМЕНДАЦІЙНИХ СИСТЕМ

Вінницький національний технічний університет

Анотація

У роботі розглядається проблема прогнозування поведінки користувачів в онлайн-магазинах на основі методів аналізу великих даних. Здійснено порівняльний аналіз поширених підходів до побудови рекомендаційних систем: контент-орієнтованих, колаборативної фільтрації, гібридних моделей та алгоритмів з використанням глибинного навчання. Оцінювання проводиться за показниками точності, стійкості до розрідженості даних, масштабованості та інтерпретованості результатів. Показано переваги та обмеження кожної групи моделей, а також окреслено напрями подальшого удосконалення рекомендаційних систем у сфері електронної комерції.

Ключові слова: рекомендаційні системи, Big Data, прогнозування поведінки користувачів, колаборативна фільтрація, контент-орієнтовані моделі, гібридні моделі, глибинне навчання, електронна комерція.

Abstract

The article addresses the problem of predicting user behavior in online stores using big data analytics. A comparative analysis of common approaches to building recommender systems is presented, including content-based models, collaborative filtering, hybrid approaches, and deep-learning-driven algorithms. The evaluation focuses on accuracy, robustness to data sparsity, scalability, and result interpretability. The study highlights the strengths and limitations of each model group and outlines future directions for improving recommender systems in e-commerce.

Keywords: recommender systems, big data, user behavior prediction, collaborative filtering, content-based models; hybrid models, deep learning, e-commerce.

Вступ

Стрімке зростання обсягів даних в електронній комерції зумовлює потребу у ефективних інструментах прогнозування поведінки користувачів. Онлайн-магазини накопичують інформацію про історію переглядів, покупки, рейтинги, кліки, час взаємодії з інтерфейсом та інші поведінкові характеристики. Однак без спеціалізованих методів аналізу ці дані залишаються малокорисними для підтримки управлінських рішень і персоналізації користувацького досвіду.

Рекомендаційні системи є ключовим механізмом підвищення лояльності клієнтів, збільшення конверсії та середнього чека, а також оптимізації маркетингових витрат. Попри значний прогрес у розвитку алгоритмів – від класичних методів колаборативної фільтрації до моделей глибинного навчання – актуальною залишається проблема вибору підходу, здатного забезпечити баланс між точністю прогнозування, масштабованістю, інтерпретованістю та стійкістю до розрідженості даних.

У зв'язку з цим метою роботи є виконання порівняльного аналізу різних типів моделей рекомендаційних систем для прогнозування поведінки користувачів в онлайн-магазинах, виявлення їх переваг і обмежень, а також окреслення напрямів подальшого удосконалення.

Результати дослідження

В онлайн-магазинах генеруються значні масиви цифрових даних, що виникають у процесі взаємодії користувачів із платформою – від переглядів товарів до здійснення транзакцій та публікації відгуків. Класифікація цих даних за типами та джерелами створює підґрунтя для подальшого аналізу та прогнозування поведінки користувачів. Онлайн-ритейл охоплює транзакційні записи, історію переглядів і кліків, дані про поведінку під час сесій, а також текстові відгуки і рейтинги товарів, кожен з яких має власний аналітичний потенціал для побудови моделей рекомендацій [1].

Поведінкові дані користувачів, зокрема інформація про переглянуті товари, відвідувані категорії, тривалість перебування на сторінці або додавання позицій до кошику, формують масиви так званого клікстріму (clickstream). Ці дані зазвичай зберігаються як хронологічні послідовності подій із часовими мітками і використовуються для моделювання інтересів та намірів користувачів у межах сесій. Аналіз

таких послідовних записів дозволяє виявляти поведінкові патерни та прогнозувати ймовірність майбутніх дій, наприклад додавання товару до кошика або завершення оформлення замовлення [2].

Окрім суто поведінкових ознак, платформи електронної комерції акумулюють демографічні дані користувачів, інформацію про попередні замовлення та історію взаємодій із платформою. Наприклад, набори даних (датасети), що містять атрибути товарів, цінові параметри, ідентифікатори користувачів і часові мітки подій, слугують основою для створення прогностичних моделей, таких як класифікація «купити/не купити» або формування переліку рекомендованих товарів на основі попередніх взаємодій [3].

Узагальнюючи, аналіз даних в e-commerce не обмежується лише окремими записами про покупців, а передбачає інтеграцію гетерогенних типів даних, що дає змогу будувати багатовимірні профілі користувачів і покращувати точність прогнозних моделей. Такий багатоканальний підхід є ключовим для персоналізації користувацького досвіду та реалізації ефективних рекомендаційних механізмів на великих платформах.

Класичні рекомендаційні системи розроблено з метою зменшення інформаційного перевантаження та допомоги користувачам у виборі товарів, що найбільш релевантні їхнім інтересам. Вони базуються на аналізі історичних взаємодій, оцінці подібності між об'єктами та суб'єктами, а також на використанні статистичних та евристичних методів. Зазначені системи стали фундаментом сучасних платформ електронної комерції, сприяючи підвищенню конверсії, зростанню середнього чека та покращенню користувацького досвіду (User Experience) [4].

Одним із поширеніших підходів є контент-орієнтовані рекомендації (content-based filtering). Система аналізує атрибути товарів та порівнює їх із тими, з якими користувач взаємодіяв раніше. Рекомендації формуються на основі міри подібності характеристик, розрахованою, наприклад, за допомогою TF-IDF або косинусної подібності. Перевагою методу є можливість персоналізації без використання даних інших користувачів, проте існує ризик «звуження» рекомендацій, коли система пропонує виключно однотипні товари [5].

Іншим базовим підходом є колаборативна фільтрація, що ґрунтується на гіпотезі, що користувачі з подібними вподобаннями в минулому здійснюватимуть аналогічний вибір у майбутньому. Рекомендації будуються не на властивостях товарів, а на основі матриці «користувач – товар», що фіксує рейтинги, факти купівель або кліки. Методи цієї групи включають підходи на основі сусідства (user-based / item-based) та матричну факторизацію. Колаборативна фільтрація ефективно виявляє приховані зв'язки між товарами, але вразлива до проблеми «холодного старту» (cold start) та розрідженості даних [5].

Матрична факторизація, як різновид колаборативної фільтрації, дозволяє здійснювати проєкцію великої розрідженої матриці взаємодій у простір прихованих факторів. Це дає змогу виявляти латентні інтереси користувачів і властивості товарів, які не задані явно. Метод набув популярності після конкурсу Netflix Prize і залишається еталоном (baseline) для порівняння нових алгоритмів. Проте його ефективність знижується за умов високого рівня шуму в даних або їхньої високої динамічності [6].

Для нівелювання недоліків окремих методів активно застосовуються гібридні рекомендаційні системи, що поєднують контент-орієнтований і колаборативний підходи. Це дозволяє зменшити вплив «холодного старту», покращити різноманітність рекомендацій та враховувати як атрибути товарів, так і колективні патерни поведінки. Гібридизація може відбуватися на рівні алгоритмів, ознак або фінального ранжування результатів [5].

Попри переваги, класичні методи стикаються зі спільними викликами: залежністю від якості даних, складністю оброблення неструктурованої інформації та обмеженням урахуванням контексту (час, типу пристрою, локації, поточного наміру). Ці обмеження згодом стимулювали перехід до більш складних моделей, зокрема методів глибинного навчання (Deep Learning) та контекстно-залежних рекомендацій [6].

З розвитком ресурсів обчислювальної техніки та алгоритмів машинного навчання класичні методи рекомендаційних систем доповнили й у багатьох випадках замінили сучасні моделі, здатні ефективно працювати з великими обсягами даних, враховувати складні поведінкові патерни та контекстні фактори. Ці моделі базуються як на традиційних алгоритмах машинного навчання, так і на архітектурах глибинного навчання.

Одним із провідних напрямів є використання нейронних мереж. На відміну від лінійної матричної факторизації, нейронні моделі навчаються виявляти латентні взаємозв'язки між користувачами й товарами через багатошарові нелінійні перетворення, що дозволяє моделі захоплювати складні патерни взаємодій. Зокрема, дослідження секвенційних моделей демонструють, що RNN- та Transformer-подібні архітектури ефективно моделюють послідовність поведінкових подій, що є критично важливим для прогнозування наступної дії користувача [7].

У порівняльному дослідженні сучасних підходів до рекомендаційних систем було показано, що нейронні рекомендаційні мережі часто перевищують за показниками точності традиційні методи,

особливо в умовах складних взаємодій та великих наборів даних. Проте їхня обчислювальна складність і потреба у великій кількості тренувальних прикладів робить такі моделі менш ефективними на «тонких» або розріджених вибірках без додаткових регуляризаційних технік [9].

Особливої актуальності набувають графові нейронні мережі (Graph Neural Networks, GNN), оскільки взаємодії в e-commerce (користувач – товар – атрибут) природним чином описуються графовими структурами. GNN дозволяють інтегрувати різномірні типи зв'язків та враховувати контекстні ознаки. Зокрема, моделі на основі multi-behavior GNN демонструють високу прогностичну здатність, оскільки одночасно аналізують кілька типів поведінкових сигналів (перегляд, клік, додавання в кошик, купівля) [8].

Перспективним напрямом є також інтеграція методів обробки природної мови (NLP) у рекомендаційні системи, зокрема для аналізу текстових відгуків і описів продуктів. Sentiment-aware recommender systems використовують оцінки, тон відгуків та семантичні фічі, що можуть істотно підвищувати точність передбачень щодо майбутньої поведінки користувача, оскільки ці дані несуть додаткову інформацію про вподобання, що не відображені у числових рейтингах [8].

Сучасні моделі також активно застосовують гібридні підходи, які поєднують глибинні мережі та класичну колаборативну фільтрацію, часто разом із контентними ознаками. Це дозволяє забезпечувати стабільність та інтерпретованість рекомендаційних рішень, уникаючи деяких обмежень окремих моделей. У таких гібридних системах результати від нейронних моделей інтегруються з традиційними метриками подібності або використовуються у якості додаткових ознак входу для класичних алгоритмів машинного навчання [7].

Варто також відзначити, що робота з великими й розрідженими даними залишається викликом для сучасних рекомендаційних моделей. Деякі дослідження пропонують комбінування технік попередньої обробки із навчанням моделей, які стійкі до «холодного старту» та розрідженості даних. Це має велике значення у практичних системах, де новий користувач або продукт можуть не мати достатньої кількості взаємодій для надійного прогнозування [9].

Таким чином, сучасні моделі рекомендаційних систем забезпечують потужні можливості для прогнозування поведінки користувачів у онлайн-магазинах, проте вибір конкретного підходу має враховувати специфіку даних, вимоги до точності, швидкості та здатність працювати з розрідженими вибірками. Більш складні архітектури, такі як нейронні мережі та графові моделі, мають високий потенціал, але потребують ретельної адаптації до предметної області та характеристик реальних e-commerce наборів.

Порівняльний аналіз моделей рекомендаційних систем вимагає формалізованого та відтворюваного підходу до оцінювання якості прогнозів поведінки користувачів. Така методика включає вибір відповідних наборів даних, формалізацію критеріїв оцінювання, техніки розбиття даних на тренувальні та тестові множини та обґрунтування вибору моделей для порівняння.

Для побудови експериментальної бази використовуються публічно доступні або синтетично згенеровані датасети електронної комерції, що містять історію взаємодій користувачів із товарами. Типовими компонентами таких наборів є ідентифікатори користувачів, товари, часові мітки подій (перегляди, покупки, рейтинги), а також потенційно демографічні або контекстні атрибути. Застосування стандартних датасетів, таких як Amazon Reviews або MovieLens, дозволяє забезпечити співставність результатів та відтворюваність експериментів. Важливо перевіряти структуру даних на предмет розрідженості (скільки пропущених значень у матриці користувач–товар), оскільки це безпосередньо впливає на поведінку моделей, особливо класичних методів колаборативної фільтрації [10].

Розбиття даних на тренувальну й тестову частини здійснюється за допомогою стратегій, що відповідають реаліям прогнозування поведінки. Найчастіше використовується time-based splitting (розділення за часовими мітками), коли події до певної дати формують навчальну множину, а наступні – тестову. Це імітує реальний сценарій прогнозування майбутньої поведінки на основі історичного контексту, знижуючи ризик «витоку даних». Альтернативно можуть застосовуватися K-fold cross-validation або leave-one-out стратегії для менших наборів даних, коли важлива стабільність метрик у різних підмножинах [10].

Ключовим елементом методики є вибір метрик оцінювання, що дозволяють кількісно порівнювати моделі. Для кількісного порівняння ефективності моделей використовуються метрики [11]:

- Precision@K і Recall@K, що вимірюють точність і повноту рекомендованих позицій у топ-K рекомендацій;
- NDCG (Normalized Discounted Cumulative Gain), що враховує як позицію у ранжованому списку, так і релевантність рекомендацій;
- MAP (Mean Average Precision) – усереднена точність для всіх релевантних позицій.

Ці метрики дозволяють оцінити не лише правильність передбачень, а й їхню корисність у контексті ранжування товарів, що є критичним у електронній комерції.

Оцінка моделей також повинна враховувати комп'ютерну складність та масштабованість, оскільки практичні застосування часто пов'язані з величезними обсягами даних. Для цього вимірюють час тренування, використання пам'яті та швидкість генерації рекомендацій у продуктивному середовищі. Такі показники особливо важливі при порівнянні класичних моделей з глибинними мережами або графовими методами, де витрати ресурсів можуть значно різнитися при порівнянних результатах [12].

Таким чином, порівняльний аналіз моделей рекомендаційних систем вимагає чіткої методологічної основи, що включає вибір наборів даних, розбиття на тренувальні й тестові множини, добір відповідних метрик та оцінювання не лише точності, а й їхньої продуктивності, масштабованості й стійкості до розрідженості. Тільки таке комплексне оцінювання дозволяє об'єктивно визначити сильні і слабкі сторони кожного підходу та коректно інтерпретувати результати в контексті прогнозування поведінки користувачів у онлайн-магазинах.

Порівняльні дослідження моделей рекомендаційних систем у сфері онлайн-торгівлі демонструють суттєві відмінності між алгоритмами за точністю прогнозування, стійкістю до розрідженості даних, здатністю працювати з новими користувачами та обчислювальною ефективністю. У більшості робіт як базові моделі використовуються алгоритми колаборативної фільтрації на основі сусідства (user-based та item-based), факторизаційні моделі (особливо matrix factorization) та сучасні підходи на основі глибинного навчання. Такий вибір є виправданим, оскільки ці групи методів представляють різні парадигми рекомендацій – від евристичних до латентно-факторних і нейронних [12].

У дослідженнях, де основним об'єктом порівняння є прогностична точність, моделі колаборативної фільтрації на основі сусідства зазвичай демонструють прийнятні результати на густих наборах даних, проте їх ефективність суттєво знижується за умов високої розрідженості. Це пов'язано з тим, що для пошуку «подібних» користувачів або товарів необхідна достатня кількість спільних взаємодій, чого часто не вистачає в реальних системах. У свою чергу, факторизаційні моделі, зокрема методи матричної факторизації, здатні виявляти латентні патерни взаємодій і тому зберігають стабільність навіть при значній розрідженості даних, що підтверджувалося, зокрема, у роботах, присвячених результатам конкурсу Netflix Prize [13].

Порівняння з нейронними рекомендаційними моделями показує, що останні дозволяють інтегрувати додаткові джерела інформації – часові характеристики, контекстні ознаки, текстові описи товарів, а іноді й поведінкові послідовності. У багатьох випадках це призводить до покращення метрик NDCG та MAP порівняно з класичними методами. Проте такі моделі потребують значних обчислювальних ресурсів, тривалого навчання й ретельного налаштування, що ставить під сумнів їх доцільність для невеликих або середніх онлайн-майданчиків [13].

Окрему увагу дослідники приділяють не лише точності, а й іншим властивостям рекомендацій – різноманітності, новизні, справедливості розподілу та уникненню «фільтраційних бульбашок». Дослідження засвідчують, що моделі, які оптимізовані виключно за критерієм точності, часто схильні рекомендувати популярні товари, ігноруючи нішеві позиції. Натомість гібридні або регуляризовані алгоритми уможливають досягнення балансу між точністю та різноманітністю, що підвищує загальну задоволеність користувачів і сприяє збільшенню довгострокових продажів [14].

Проведений аналіз демонструє, що онлайн-магазини часто зосереджуються лише на підвищенні точності рекомендацій, недостатньо враховуючи їхній вплив на поведінкові патерни користувачів. Варто наголосити, що важливим є не лише перелік запропонованих товарів, а й здатність системи формувати довіру, прозорість та відчуття контролю з боку покупця. Рекомендації, логіку яких користувач не може зрозуміти, поступово знижують інтерес до платформи, навіть за умови їхньої формальної точності.

Крім того, результати порівняння свідчать, що універсальної моделі не існує. Для невеликих магазинів складні глибинні моделі можуть бути економічно не вигідними та надлишковими, тоді як простіші підходи в поєднанні з якісною сегментацією клієнтів здатні забезпечити зіставну ефективність. Тому ключовим вбачається поетапне впровадження рекомендаційних систем із постійним моніторингом не лише точності прогнозування, а й бізнес-показників: конверсії, коефіцієнта утримання та рівня задоволеності клієнтів.

Висновки

У роботі проведено порівняльний аналіз основних підходів до побудови рекомендаційних систем для онлайн-магазинів. Встановлено, що кожен клас моделей орієнтований на специфічні типи даних та сценарії використання. Традиційні методи (контент-орієнтовані та колаборативна фільтрація) залишаються базисом більшості систем завдяки простоті, прозорості та низьким обчислювальним

витратам, проте їх використання обмежено проблемою «холодного старту» і недостатнім врахуванням контексту взаємодії.

Моделі матричної факторизації та гібридні підходи забезпечують оптимальний баланс між точністю та узагальненням, дозволяючи інтегрувати гетерогенні джерела даних (поведінкові журнали, атрибути товарів, історію покупок). Використання нейронних мереж, графових моделей та підходів, орієнтованих на аналіз відгуків, демонструє підвищення якості персоналізації, однак потребує значних обсягів даних, розгалуженої інфраструктури та ретельної валідації результатів.

Важливим результатом аналізу є висновок про те, що жодна модель не є універсальною. Оптимальне архітектурне рішення формується на перетині технічних можливостей платформи, обсягів доступних даних, бізнес-цілей та вимог до прозорості алгоритмів. Саме тому гібридні підходи, що поєднують різні алгоритми та дозволяють гнучко змінювати конфігурації, визначено як найбільш перспективні.

Окремої уваги потребує питання оцінювання ефективності. Традиційні метрики точності рекомендацій є необхідними, однак вони не завжди корелюють із реальним впливом на бізнес-результати. Пріоритетними стають показники конверсії, повторних покупок, тривалості взаємодії з платформою, а також рівень задоволеності користувачів. Таким чином, рекомендаційні системи слід розглядати не як суто технічний інструмент, а як стратегічний елемент клієнтського досвіду.

Актуальними залишаються етико-правові аспекти: захист персональних даних, уникнення надмірної персоналізації, забезпечення інтерпретованості алгоритмів. У перспективі ці чинники визначатимуть не лише технічні підходи, а й рівень довіри користувачів до цифрових платформ.

Підсумовуючи, результати дослідження підтверджують, що еволюція рекомендаційних систем у сфері e-commerce рухається в напрямі синергії аналітики великих даних, адаптивного машинного навчання та людиноцентрованого підходу. Подальші дослідження доцільно зосередити на емпіричному порівнянні гібридних систем, кореляції точності прогнозування з бізнес-ефектом та вдосконаленні механізмів прозорості й довіри до рекомендацій.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Ge Guo. A Study on User Behavior of E-commerce Platforms based on Data Analysis | Journal of Global Humanities and Social Sciences. BonFuturePress. URL: <https://ojs.bonfuturepress.com/index.php/GHSS/article/view/1418> (date of access: 24.12.2025).
2. Research Methodology for Analysis of E-Commerce User Activity Based on User Interest using Web Usage Mining | Journal of ICT Research and Applications. ITB Journal. URL: <https://journals.itb.ac.id/index.php/jictra/article/view/1921> (date of access: 24.12.2025).
3. GitHub - paulsamuel-w-e/E-commerce-Customer-Behaviour-Dataset. URL: <https://github.com/paulsamuel-w-e/E-commerce-Customer-Behaviour-Dataset> (date of access: 25.12.2025).
4. Croft W. B. Query Evolution. *SpringerLink*. URL: https://link.springer.com/chapter/10.1007/978-3-642-00958-7_1 (date of access: 25.12.2025).
5. Herlocker J. L. Evaluating collaborative filtering recommender systems. *ACM*. URL: <https://dl.acm.org/doi/10.1145/963770.963772> (date of access: 25.12.2025).
6. Yehuda Koren. MATRIX FACTORIZATION TECHNIQUES FOR RECOMMENDER SYSTEMS. *DataJobs.com: Analytics Jobs, Hadoop Jobs, DBA Jobs, Data Science Jobs*. URL: [https://datajobs.com/data-science-repo/Recommender-Systems-\[Netflix\].pdf](https://datajobs.com/data-science-repo/Recommender-Systems-[Netflix].pdf) (date of access: 24.12.2025).
7. He X. Neural Collaborative Filtering. *ACM*. URL: <https://dl.acm.org/doi/10.1145/3038912.3052569> (date of access: 26.12.2025).
8. Session-based Recommendations with Recurrent Neural Networks. *arXiv.org*. URL: <https://arxiv.org/abs/1511.06939> (date of access 26.12.2025).
9. Comprehensive Empirical Evaluation of Deep Learning Approaches for Session-based Recommendation in E-Commerce. *arXiv.org*. URL: <https://arxiv.org/abs/2010.12540> (date of access: 26.12.2025).
10. A Survey of Accuracy Evaluation Metrics of Recommendation Task. *ResearchGate*. URL: https://www.researchgate.net/publication/220320184_A_Survey_of_Accuracy_Evaluation_Metrics_of_Recommendation_Tasks (date of access: 26.12.2025).
11. Performance of recommender algorithms on top-n recommendation tasks. *ACM*. URL: <https://dl.acm.org/doi/10.1145/1864708.1864721> (date of access: 26.12.2025).
12. Recommender Systems Handbook. *SpringerLink*. URL: <https://link.springer.com/book/10.1007/978-1-4899-7637-6> (date of access: 26.12.2025).

13. MATRIX FACTORIZATION TECHNIQUES FOR RECOMMENDER SYSTEMS. *DataJobs.com: Analytics Jobs, Hadoop Jobs, DBA Jobs, Data Science Jobs*. URL: [https://datajobs.com/data-science-repo/Recommender-Systems-\[Netflix\].pdf](https://datajobs.com/data-science-repo/Recommender-Systems-[Netflix].pdf) (date of access: 26.12.2025).

14. Cayssol J. Ballistic transport through irradiated graphene. *arXiv.org*. URL: <https://arxiv.org/abs/1709.00090> (date of access: 24.12.2025).

Пінчук Дар'я Олександрівна – студентка групи 1KITS-22б, Факультет менеджменту та інформаційної безпеки, Вінницький національний технічний університет, м. Вінниця, e-mail: dashapinchukschool@gmail.com

Науковий керівник: **Грицак Анатолій Васильович** – кандидат технічних наук, доцент, доцент кафедри менеджменту та безпеки інформаційних систем, Вінницький національний технічний університет, м. Вінниця, e-mail: grytsak.a.v@gmail.com

Pinchuk Daria O. – student of group 1KITS-22b, Faculty of Management and Information Security, Vinnytsia National Technical University, Vinnytsia, e-mail: dashapinchukschool@gmail.com

Supervisor: **Hrytsak Anatoly V.** – Candidate of Technical Sciences, Associate Professor, Associate Professor of the Department of Management and Security of Information Systems Vinnytsia National Technical University, Vinnytsia, e-mail: grytsak.a.v@gmail.com