

НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ «ОДЕСЬКА ПОЛІТЕХНІКА»
МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ «ОДЕСЬКА ПОЛІТЕХНІКА»
МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

*Кваліфікаційна наукова
праця на правах рукопису*

ІВАНОВ ДМИТРО ВЯЧЕСЛАВОВИЧ

УДК 004.891:004.65:339.138(045)

ДИСЕРТАЦІЯ

**ІНТЕЛЕКТУАЛЬНИЙ АНАЛІЗ ДАНИХ В2В-КЛІЄНТІВ ДЛЯ ПЕРСОНАЛІЗАЦІЇ
КОНТЕНТУ СИСТЕМ ЕЛЕКТРОННОЇ КОМЕРЦІЇ**

Спеціальність: 122 – «Комп'ютерні науки»

Галузь знань: 12 – «Інформаційні технології»

Подається на здобуття наукового ступеня доктора філософії

Дисертація містить результати власних досліджень. Використання ідей, результатів і текстів інших авторів мають посилання на відповідне джерело.

_____ Д. В. Іванов

Науковий керівник:

Арсирій Олена Олександрівна, доктор технічних наук, професор

Одеса – 2026

АНОТАЦІЯ

Іванов Д.В. Інтелектуальний аналіз даних B2B-клієнтів для персоналізації контенту систем електронної комерції. – Кваліфікаційна наукова праця на правах рукопису.

Дисертація на здобуття наукового ступеня доктора філософії за спеціальністю 122 – Комп’ютерні науки. – Національний університет «Одеська політехніка», МОН України, Одеса, 2026.

У **вступі** обґрунтовано актуальність створення персоналізованого контенту систем електронної комерції для підвищення своєчасності формування предикативних пропозицій у відповідності до класів B2B-клієнтів за рахунок інтелектуального аналізу масивів транзакційно-повелінкових даних. Визначено **об’єкт дослідження** як процес персоналізації контенту в системах B2B електронної комерції з урахуванням транзакційно-поведінкових даних клієнтів. **Предметом дослідження** є моделі та методи інтелектуального аналізу транзакційно-поведінкових даних B2B-клієнтів систем електронної комерції. Сформульовано **мету дослідження**, яка полягає в вирішенні науково-прикладної задачі персоналізації контенту систем електронної комерції за рахунок підвищення ефективності формування предикативних пропозицій у відповідності до класів B2B-клієнтів шляхом розробки інтелектуального ціннісно-зваженого темпорального підходу до аналізу масивів транзакційно-поведінкових даних. Досягнення зазначеної мети дослідження пов’язано з вирішенням наступних задач: проаналізувати існуючі технології персоналізації контенту в системах електронної комерції та визначити шляхи підвищення їх ефективності для B2B-клієнтів в сучасних умовах; розробити ціннісно-зважений темпоральний підхід до інтелектуального аналізу масивів транзакційно-повелінкових даних B2B клієнтів; розробити метод класифікації B2B клієнтів для предикативної персоналізації контенту систем електронної комерції; провести моделювання стратегій персоналізації контенту за рахунок підвищення ефективності формування предикативних пропозицій порівняно з класичними підходами та

інтегрувати розроблений аналітичний модуль в систему рекомендації для реальної B2B-платформи.

У **першому розділі** проаналізовано стан існуючих технологій персоналізації контенту в системах електронної комерції та визначено шляхи підвищення їх ефективності для B2B-клієнтів в сучасних умовах. Обґрунтовано специфіку B2B-сегмента в умовах динамічного ринку. Показано, що понад 60 % українських B2B-компаній зазначають, що основною перешкодою для переходу в онлайн є неможливість забезпечити клієнта релевантним контентом, а саме індивідуальними цінами, технічною документацією в реальному часі та предикативними автоматизованими рекомендаціями щодо здійснення нових транзакцій для поповнення складів. Дослідження показують, що лише 12 % вітчизняних B2B-платформ використовують елементи інтелектуального аналізу даних для персоналізації пропозицій, тоді як у світі цей показник перевищує 50 %.

Результати проведеного аналізу B2B-сегмента електронної комерції за схемою «*Wholesale, Marketing, Customer, Collaborative*» свідчать, що на відміну від B2C, де домінують емоційні та разові покупки, B2B-транзакції характеризуються циклічністю, раціональністю та тривалим життєвим циклом клієнта. Це доводить, що персоналізація в B2B-системах має базуватися на глибокому інтелектуальному аналізі ціннісних та темпоральних параметрів та докорінно відрізнятися від B2C-систем, де контент генерується переважно шляхом статистичного аналізу частотних покупок та загальних поведінкових шаблонів масового споживача.

Проаналізовано традиційні методи взаємодії з клієнтами в системах електронній комерції на основі створення рекомендацій та зауважено, що функціонал модулів колаборативної та контентної фільтрації при врахуванні операційної складності B2B-процесів має певні недоліки. Так виявлено, що класичні рекомендаційні системи часто ігнорують часову динаміку та специфічну цінність товарів, що призводить до низької релевантності пропозицій у сферах із рідкісними, але високовартісними закупівлями. Крім того показано, що незважаючи на спроби створення складних гібридних рекомендаційних систем, як надбудови над B2B-системою електронної комерції шляхом комбінування колаборативного та

контентного підходів залишаються недоліки ігнорування економічної цінності та стабільності циклу транзакцій та не гарантується своєчасність рекомендацій для бізнес-клієнта. Таким чином, аналіз технологій персоналізації контенту за допомогою базових технологій рекомендаційних систем виявив системний дефіцит інструментів для врахування ціннісно-темпоральної специфіки B2B-транзакцій. Це обґрунтовує необхідність використання методів інтелектуального аналізу транзакційно-поведінкових даних B2B-клієнтів для побудови асоціацій та послідовностей.

Однак, проведений компаративний аналіз методів ІАД, спрямованих на побудову асоціацій та послідовностей у транзакційних масивах дозволив встановити наступне:

- класичні методи аналізу асоціативних правил (*Association Rule Mining ARM*), зокрема Apriori, FP-Growth та Eclat, орієнтовані на знаходження статичних зв'язків товарів в транзакції. При цьому враховується лише частота появи предметів у транзакціях, що повністю нівелює інформацію про його вартість чи обсяг закупівлі. Як наслідок, стратегічно важливі, але рідкісні B2B-товари відсікаються, тоді як дешеві супутні товари з високою частотою стають основою моделі. Це все демонструють повну «фінансову сліпоту», класичних методів *ARM*;

- класичні методи аналізу послідовних шаблонів (*Sequential Pattern Mining, SPM*), такі як GSP, PrefixSpan та SPADE, спрямовані на виявлення кореляцій між транзакціями на тлі хронології подій. При цьому вони частково вирішують проблему врахування фактору часу, проте характеризуються високим рівнем «ідентифікаційної дискретності», розглядаючи клієнта як абстрактну послідовність різних ідентифікаторів, що не дає можливість виявити індивідуальний ритм транзакцій та призводить до «темпоральної сліпоти» *SPM*-методів.

Таким чином, виявлено, що існуючий інструментарій ІАД, а саме *ARM* та *SPM*, потребує адаптації шляхом інтеграції фінансових показників корисності та часових інтервалів між транзакціями.

Наприкінці розділу на основі аналізу можливостей розповсюдженого галузевого стандарту, а саме є моделі RFM (*Recency, Frequency, Monetary*) щодо сегментації клієнтів в B2B-середовищі також доведено її суттєві обмеження. Так

показник Recency (давність) фіксує лише статичну точку на осі часу, не враховуючи волатильність індивідуального закупівельного циклу, а показник Monetary (вартість) оцінює валовий дохід, ігноруючи специфічну корисність (Utility) та стабільність патернів споживання. Таким чином класична модель RFM-аналізу орієнтована на ретроспективний опис стану («що відбулося»), що не дозволяє будувати предикативні моделі CLV (Customer Lifetime Value) з високою точністю для B2B-клієнтів із нерівномірним ритмом активності.

Узагальнюючи результати аналізу стану та тенденцій галузі, встановлено критичну необхідність розробки нових підходів до персоналізації B2B-контенту. Обґрунтовано доцільність переходу від статичної сегментації до побудови ціннісно-зваженого темпорального профілю клієнта. Це дозволить подолати фінансову та темпоральну сліпоту існуючих моделей та забезпечити перехід від констатації минулих транзакцій до предикативного управління комунікаційними тригерами на основі прогнозованого закупівельного циклу, фінансової ємності та стабільності поведінки кожного конкретного контрагента.

У **другому розділі** розроблено ціннісно-зважений темпоральний підхід до інтелектуального аналізу масивів транзакційно-повелінкових даних B2B клієнтів. Результатом проведення системного дослідження існуючих методів інтелектуального аналізу асоціативних правил та послідовних шаблонів в накопичених масивах B2B транзакцій є встановлення *фундаментальних методичних обмежень*, які унеможливають їх безпосереднє застосування для створення персоналізованого контенту у системах електронної комерції. Обґрунтовано існування п'яти пов'язаних між собою критичних розривів (обмежень) між математичними припущеннями методів пошуку асоціацій та послідовностей та природою B2B-предметної галузі, а саме це «*фінансова сліпота*» бінарних моделей, наявність «*комбінаторного вибуху*» при зваженому аналізі, «*темпоральна сліпота*» статичних моделей, наявність *ідентифікаційної дискретності* та *відсутність диференційованої стратегії взаємодії з клієнтами*. Встановлено, що зв'язок цих обмежень має ієрархічну структуру. Тобто спроба подолати фінансову сліпоту класичних моделей ІАД шляхом додавання вартісних показників активує проблему комбінаторного вибуху.

Своєю чергою, перехід до аналізу темпоральних інтервалів транзакційних даних вимагає не лише нової архітектури обробки, а й усунення ідентифікаційної дискретності для B2B клієнтів, оскільки розрахунок циклічності можливий лише на цілісних, а не фрагментованих часових рядах. Таким чином, фінальне обмеження, відсутність диференційованої стратегії взаємодії з клієнтами, тобто стратегічна монотонність є результатом дефіциту даних на всіх попередніх етапах аналізу. Тобто в B2B-системі електронної комерції не можливо диференціювати (визначати класи) клієнтів без ідентифікації їхніх індивідуальних фінансово-темпоральних патернів.

Таким чином, подолання зазначених обмежень є науковою основою для розробки ціннісно-зваженого темпорального аналізу, математичний апарат якого умовно поділений на дві складові частини: модель ціннісно-зваженої транзакції та метод розростання ціннісних патернів (*Utility Pattern Growth, UP-Growth*), а також модель темпоральної послідовності та метод темпорального аналізу ціннісних патернів (*Utility-Temporal Pattern Growth, UT- Growth*).

Розроблено модель ціннісно-зваженої B2B-транзакції шляхом додавання до класичного визначення транзакції формалізацію індикатора сумарної корисності (*Transaction Utility*) – $U(T)$ визначеної транзакції, як суми добутків кількості придбаних одиниць товарних складових на їх індивідуальну ціну. Для подолання ефекту комбінаторного вибуху, який виникає за використання індикатора $U(T)$ формалізовано визначення транзакційно-зваженої корисності (*Transaction-Weighted Utilization*) – $TWU(X)$, як сумарну вартість усіх транзакцій бази, які містять визначену транзакцію. З використанням моделі ціннісно-зваженої B2B-транзакції розроблено метод розростання ціннісних патернів *UP-Growth*. Показано, що його використання, дозволяє усунути «фінансову сліпоту» класичного підходу до пошуку асоціативних патернів в транзакційних даних. Так на *першому проході* при побудові *UP*-дерева в вузлах накопичується не кількість входжень товарної позиції, а їх грошова вартість. Використання механізму *TWU* під час *другого проходу* відновлює властивість антимонотонності і забезпечує обчислювальну стабільність навіть при аналізі довгих B2B-накладних. Результатом роботи методу є впорядкований перелік ціннісних патернів – комбінацій товарів, які формують основну частку виторгу підприємства.

Розроблено модель темпоральної B2B-послідовності шляхом додавання до класичного визначення послідовності часу виникнення транзакцій, що дозволило отримати вектор міжзакупівельних інтервалів (*Inter-Purchase Time, IPT*) для довільного B2B-клієнта. Це спільно з розрахованим безрозмірним показником – коефіцієнтом варіації (*Coefficient of Variation, CV*) дає можливість визначити стабільність циклу транзакцій, долаючи таким чином «темпоральну сліпоту» та ідентифікаційну дискретність класичного методу пошуку шаблонів в послідовностях клієнтів. З використанням моделі темпоральної B2B-послідовності розроблено метод темпорального аналізу ціннісних патернів *UT- Growth*, які попередньо отримано за методом *UP-Growth*. Показано, що використання методу *UT- Growth* дозволяє усунути «темпоральну сліпоту» класичного підходу до пошуку послідовних патернів в транзакційних даних через побудову *IPT*-векторів та їх подальшого аналізу за критерієм *CV* для довільного B2B-клієнта. Наведено інтерпретацію критерію *CV* в залежності від встановленого порогу стабільності. Значення порогу не є фіксованим і може налаштовуватися залежно від специфіки галузі та прийнятності підприємства до похибки прогнозу.

Як загальний підсумок розділу показано що використання запропанованого двоетапного конвеєру методів *UP-Growth* та *UT- Growth* дозволяє отримати перелік персоналізованих тригерів, завдяки якому для кожного контрагента та кожної фінансово значущої та циклічно стабільної транзакції визначається момент отримання предикативної пропозиції. Як підтвердження наведено детальний розгляд двох типових сценаріїв для B2B систем електронної комерції – «регулярна сезонна закупівля» та «нерегулярна закупівля». Саме можливість їх реалізації перетворює B2B-платформу з реактивного каталогу на інтелектуальну систему рекомендацій, здатну передбачати потреби клієнтів, щодо поповнення складських запасів і взаємодіяти з ними у найбільш сприятливий момент.

У **третьому розділі** розроблено метод класифікації B2B клієнтів для предикативної персоналізації контенту систем електронної комерції. Доведено існування трьох критичних обмежень використання класичного *RFM*-аналізу для класифікації B2B-клієнтів. Показано, використання параметру давності *R* дозволяє

фіксувати лише час, що минув з останньої транзакції. При цьому не будується вектор межзакупівельних патернів IPT , що робить показник R «темпорально сліпим» до індивідуального ритму B2B-бізнесу. Використання параметру частоти F дозволяє оцінювати загальну кількість транзакцій за період. При цьому ігнорується стабільність транзакцій, індикатором якої є коефіцієнт варіації CV , розрахований на основі IPT . Параметр грошового обсягу транзакції M розраховується як загальна сума витрат клієнта. Але в такому вигляді він не дає інформацію про структуру транзакцій. При цьому не враховується транзакційно-зважена корисність (TWU). Це не дозволяє виділити патерни товарів-драйверів, які забезпечують стабільність обороту.

Для подолання зазначених недоліків стандартного RFM -аналізу формалізовано ціннісно-зважений темпоральний профіль клієнта (*Utility-Weighted Temporal Customer Profile, UW-TCP*) у вигляді кортежу значень наступних ознак: середній цикл закупівлі (значення часу між повторними транзакціями), середня стабільність циклу транзакцій (коефіцієнт варіації) та середня грошова вартість транзакції

З метою формування предикативних пропозицій для B2B клієнтів розроблено метод класифікації, який є третьою складовою ціннісно-зваженого темпорального аналізу транзакційно-поведінкових даних B2B клієнтів системи електронної комерції. Він функціонує виключно як надбудова над запропонованими методами *UP-Growth* та *UT-Growth*. Використання методу класифікації B2B клієнтів базується на ознаках ціннісно-зваженого темпорального профілю B2B клієнта, які є агрегатами вже наявних проміжних результатів, що робить метод класифікації B2B-клієнтів *третім етапом* конвеєра формування предикативних пропозицій для B2B-клієнтів. Визначення номерів класів B2B-клієнтів інтегрується у систему формування комунікаційних тригерів. Це забезпечує змістовний зв'язок між класом клієнта та стратегією взаємодії з ним через математично обґрунтоване призначення ширини вікна тригеру відповідно до цінності клієнта а не через суб'єктивне визначення порогів сегментації як в стандартному RFM -аналізі.

У **четвертому розділі** виконано експериментальну перевірку ефективності запропонованого ціннісно-зваженого темпорального підходу для формування предикативних пропозицій у B2B-платформі електронної комерції. Розроблені моделі

та методи розглянуто як складові рекомендаційного модуля, інтегрованого в контур роботи реальної B2B-платформи «База взуття». Формалізовано взаємодію платформи, аналітичного конвеєра та інтерфейсу менеджера, у межах якої історія замовлень, дані про клієнтів, товарні позиції, артикули, ціни та суми транзакцій використовуються для формування активних персоналізованих рекомендацій.

Експериментальне дослідження проведено на масиві транзакційно-поведінкових даних за 2022–2025 роки, який охоплював 29,35 тис. замовлень, 2,53 тис. клієнтів, 56,20 тис. унікальних артикулів і загальний обсяг замовлень 416,28 млн грн. Після попередньої обробки, формування транзакційного подання та фільтрації надто довгих замовлень для подальшого аналізу використано 28,53 тис. замовлень, що дозволило зменшити комбінаторну складність без втрати основної частини експериментального масиву.

Для оцінювання ефективності проведено порівняльне експериментальне дослідження, у межах якого послідовно зіставлено чотири сценарії: базовий частотний S0, ціннісно-зважений S1, ціннісно-темпоральний S2 та інтегрований S3, що додатково враховує клас B2B-клієнта. Результати показали послідовне зростання якості рекомендацій: конверсія збільшилася з 3,91 % у S0 до 5,39 % у S1, 12,69 % у S2 та 15,91 % у S3. Одночасно кількість сформованих рекомендацій скоротилася з 920 до 220, тобто на 76,09 %, що свідчить про зменшення надмірного комунікаційного навантаження на менеджерів і клієнтів.

Найкращий результат отримано для інтегрованого сценарію S3. Він сформував 220 рекомендацій, з яких 35 виявилися успішними, забезпечив середню виручку на одну рекомендацію 6,91 тис. грн та охоплену тестову виручку 1,52 млн грн. Порівняно з базовим сценарієм S0, кількість рекомендацій зменшилася на 700, однак кількість успішних звернень майже не змінилася. При цьому значення метрики *ARPR* (ефективність однієї рекомендації) збільшилось із 1,21 тис. грн до 6,91 тис. грн, тобто приблизно у 5,7 раза. Це підтвердило, що повна модель відсікає значну частину нерелевантних рекомендацій та підвищує грошову цінність кожної.

Економічна оцінка також засвідчила перевагу інтегрованого підходу. За умови вартості одного комунікаційного звернення 150 грн операційні витрати зменшилися

зі 138,00 тис. грн у базовому сценарії до 33,00 тис. грн у сценарії S3. Вартість однієї успішної рекомендації знизилася з 3,83 тис. грн до 0,94 тис. грн, а чистий фінансовий результат інтегрованої моделі становив 1,49 млн грн. Отримані результати доводять, що запропонований підхід забезпечує перехід від масового формування рекомендацій до точнішої, більш своєчасної та економічно обґрунтованої персоналізації контенту B2B-системи електронної комерції.

Архітектурно розроблений прикладний рекомендаційний модуль розширює функціональність B2B-платформи без зміни її основних транзакційних процесів. Програмну реалізацію виконано у вигляді аналітичного модуля, який отримує дані про клієнтів, замовлення та товарні позиції, застосовує розроблений триетапний конвеєр і формує активні предикативні рекомендації для менеджера. Для демонстрації практичного використання результатів реалізовано інтерфейсні екрани головної панелі менеджера, активних рекомендацій та персональної пропозиції B2B-клієнту.

Наукова новизна одержаних результатів полягає в наступному:

- удосконалено модель ціннісно-зваженої транзакції, яка на відміну представлення у вигляді переліка товарних пропозицій (SKU), додатково включає обсяг SKU та індивідуальну ціну, що дозволяє в подальшому використовувати їх добуток для розрахунку транзакціно-зваженої корисності (TWU) та ідентифікувати транзакції не лише за критерієм повторюваності, а і за економічної значущості для подолання обмеження «фінансової сліпоты»;

- отримав подальший розвиток метод розростання цінністних патернів *UP-Growth* за рахунок використання показника транзакціно-зваженої корисності (TWU) для розрахунку значень вузлів дерева *UP-Growth* замість частотних лічильників класичного *FP-Growth*, що дозволило відновити властивість антимонотоності долаючи обмеження «комбінаторного вибуху», а також змінити пріоритетність генерації асоціативних правил, фокусуючи систему персоналізації на економічно значущих транзакціях;

- удосконалено модель темпоральної послідовності, яка на відміну представлення у вигляді кортежів ідентифікаційно дискретних транзакцій враховує

наявність детермінованого ідентифікатора B2B-клієнта та додатково включає час виникнення транзакції, що дозволяє створити вектор міжтранзакційних інтервалів (*IPT*) та використовувати його статистичні характеристики для розрахунку ймовірної дати активації предикативної пропозиції для подолання обмежень «темпоральної сліпоти» та ідентифікаційної дискретності;

– отримав подальший розвиток метод темпорального аналізу ціннісних патернів *UT- Growth*, який спільно з моделлю темпоральної послідовності дозволяє розрахувати коефіцієнт варіації *CV* вектору міжтранзакційних інтервалів, що дає можливість визначити дату предикативної пропозиції B2B-клієнту формуючі предикативні тригери вигляду «клієнт – товарний набір – дата звернення» для підвищення рівня персоналізації контенту B2B-систем електронної комерції.

– вперше розроблено метод класифікації B2B клієнтів систем електронної комерції, який на відміну від існуючого RFM-аналізу базується на ціннісно-зваженому темпоральному профілі клієнта сформованому у результаті реалізації двоетапного конвеєру аналізу транзакційно-поведінкових даних B2B клієнта. Це дозволило реалізувати трьох етапний конвеєр формування індивідуальних предикативних пропозицій в залежності від визначеного класу клієнта та підвищити фінальну ефективність предикативної персоналізації контенту

Практичне значення результатів дослідження полягає у розробленні програмного аналітичного модуля для формування персоналізованих предикативних пропозицій у системах B2B електронної комерції. Основою модуля є триетапний конвеєр, який поєднує ціннісно-зважений аналіз товарних патернів, темпоральну перевірку стабільності закупівельних циклів та класифікацію B2B-клієнтів за ціннісно-зваженим темпоральним профілем. Це дозволяє формувати пропозиції не лише на основі товарної релевантності, а й з урахуванням фінансової значущості патерну, прогнозованого моменту звернення та класу клієнта.

Практичну апробацію результатів виконано шляхом інтеграції розробленого конвеєра у вигляді Python-модуля в контур роботи B2B-платформи «База взуття». Реалізований модуль використовує дані про клієнтів, замовлення, товарні позиції, артикули, ціни та суми транзакцій, після чого формує активні рекомендації для

менеджера. Для демонстрації прикладного використання результатів розроблено інтерфейсні екрани головної панелі менеджера, активних рекомендацій та персональної пропозиції B2B-клієнту.

Ефективність запропонованого підходу підтверджено експериментальним порівнянням базового частотного сценарію з повним інтегрованим триетапним конвеєром з використанням адаптованих бізнес-орієнтованих метрик HR, VC та ARPR. Отримані результати свідчать про підвищення точності та своєчасності надання індивідуальних предикативних пропозицій B2B-клієнтам, зменшення кількості нерелевантних рекомендацій і підвищення практичної цінності кожного комунікаційного звернення.

Науково-практичну цінність дисертаційного дослідження становлять триетапний конвеєр формування індивідуальних предикативних пропозицій, методика його інтеграції в B2B-платформу та система адаптованих метрик оцінювання ефективності рекомендацій. Запропонований ціннісно-зважений темпоральний підхід до інтелектуального аналізу транзакційно-поведінкових даних B2B-клієнтів впроваджено в діяльність ФОП (NEWOPT) та ТОВ «СТІЛЛІ», а також використано в навчальному процесі та при виконанні НДР Національного університету «Одеська політехніка» на кафедрі інформаційних систем.

Ключові слова: інтелектуальний аналіз даних; B2B-системи електронної комерції, методи класифікації, персоналізація контенту, рекомендаційні системи, транзакційно-поведінкові данні.

Список публікацій здобувача за темою дисертації

1. Арсірій О., Іванов Д. Моделі афінитивного аналізу транзакційних і поведінкових даних клієнтів систем B2B електронної комерції для створення персоналізованого контенту. Праці Одеського політехнічного університету. 2025. 2(72) С 90-101(<https://doi.org/10.15276/opu.2.72.2025>)
<https://pratsi.op.edu.ua/index.php/pratsi/article/view/549/511> ISSN 2076-2429 (print)
2. Arsirii, O., & Ivanov, D. (2026). A Model and Method of Transactional-behavioral Data Mining for B2B Content Personalization. Electronics and Control Systems, 1(87), 14–23. <https://doi.org/10.18372/1990-5548.87.20881> ISSN 1990-5548
3. Arsirii, O., & Ivanov, D. (2026). Data Mining of B2b Transactions Using a High-utility Temporal Approach. Electronics and Control Systems, 2(88), 16–31. <https://doi.org/10.18372/1990-5548.88.20958> ISSN 1990-5548
4. Olena O. Arsirii, Dmitriy V. Ivanov, Oksana Yu. Babilunha Analysis of models, methods and technologies for creating personalized content in B2B e-commerce systems. (2025). Інформатика. <https://doi.org/10.15276/ict.02.2025.41> ISSN 2522-1523 (Online)
5. Арсірій О. О., Любомська О. М., Руденко О. В., Іванов Д. В. Гібридна рекомендаційна система для підтримки UI/UX дизайнерів. Матеріали міжнародної конференції Інформатика. Культура. Техніка. 2024; Том 1 № 1: 29–35 <https://ict.od.ua/index.php/journal/en/article/view/93> <https://doi.org/10.15276/.01.2024.03> ISSN 2522-1523 (Online)
6. Сіроштан М.В., Іванов Д.В., Арсірій О.О. Дослідження ефективності методів машинного навчання на прикладі задачі класифікації предметів одягу. Project, Program, Portfolio Management. РЗМ-2024: Тези доповідей IX Міжнародної науково-практичної конференції : [у 2т.]. Відповідальний за випуск П.О. Тесленко Том 1. Одеса: ІШІР, 2024 С. 201-204 <https://rep.nuos.edu.ua/server/api/core/bitstreams/f2c763dc-2728-47ca-af64-31618bd2a659/content> ISSN 2522-9435
7. Іванов Д.В. Арсірій О.О. Класифікація B2B-клієнтів системи електронної комерції за ціннісно-зваженим темпоральним профілем. Матеріали XVI міжнародної конференції "Modern Information Technology 2026", ІКС, Національний університет "Одеська політехніка", Одеса, Україна, 14-15 травня, 2026. – С. 118-121. https://drive.google.com/drive/folders/1uXN7b84231YhSfT_tY6l9eMhPjDrIBKJ ISSN 2708-311X

ABSTRACT

Ivanov D.V. B2B Customer Data Mining for Content Personalization in E-commerce Systems. – Qualification scientific work in the form of manuscript.

Thesis for the PhD degree in specialty 122 Computer science. – Odessa Polytechnic National University, Ministry of Education and Science of Ukraine, Odesa, 2026.

The Introduction substantiates the relevance of creating personalized content in e-commerce systems to improve the timeliness of generating predictive offers in accordance with B2B customer classes through the data mining of transactional and behavioral datasets. **The object of the study** is defined as the process of content personalization in B2B e-commerce systems based on the transactional and behavioral data of customers. **The subject of the study** consists of models and methods for the data mining of transactional and behavioral data of B2B customers in e-commerce systems. **The purpose of the study** is formulated as solving the scientific and applied problem of content personalization in e-commerce systems by increasing the efficiency of generating predictive offers in accordance with B2B customer classes through the development of an intelligent utility-weighted temporal approach to the analysis of transactional and behavioral datasets. Achieving this research purpose involves **solving the following tasks**: to analyze existing content personalization technologies in e-commerce systems and determine ways to increase their efficiency for B2B customers under modern conditions; to develop a utility-weighted temporal approach to the data mining of transactional and behavioral datasets of B2B customers; to develop a B2B customer classification method for predictive content personalization in e-commerce systems; to model content personalization strategies by increasing the efficiency of predictive offer generation compared to classical approaches and to integrate the developed analytical module into the recommendation system for a real-world B2B platform

The first chapter analyzes the state of existing content personalization technologies in e-commerce systems and identifies ways to increase their efficiency for B2B customers under modern conditions. The specifics of the B2B segment in a dynamic market environment are substantiated. It is shown that more than 60 % of Ukrainian B2B companies

state that the main obstacle to moving online is the inability to provide customers with relevant content, namely personalized prices, real-time technical documentation, and predictive automated recommendations for executing new transactions to replenish warehouse inventories. Studies indicate that only 12 % of domestic B2B platforms utilize data mining elements to personalize offers, whereas globally this figure exceeds 50 %.

The results of the analysis of the B2B e-commerce segment according to the «Wholesale, Marketing, Customer, Collaborative» framework indicate that, unlike B2C, where emotional and one-time purchases dominate, B2B transactions are characterized by cyclicity, rationality, and a long customer lifetime value. This proves that personalization in B2B systems must be based on deep data mining of utility and temporal parameters, fundamentally differing from B2C systems where content is generated primarily through the statistical analysis of purchase frequencies and general behavioral patterns of the mass consumer.

Traditional methods of customer interaction in e-commerce systems based on recommendation generation are analyzed, noting that the functionality of collaborative and content-based filtering modules has certain disadvantages when considering the operational complexity of B2B processes. Thus, it is revealed that classical recommendation systems often ignore temporal dynamics and the specific value of goods, leading to low relevance of offers in sectors with rare but high-value purchases. Furthermore, it is shown that despite attempts to create complex hybrid recommendation systems as add-ons to B2B e-commerce systems by combining collaborative and content-based approaches, this approach also suffers from ignoring economic utility and transaction cycle stability, failing to guarantee the rationality of recommendations for a business customer.

Consequently, the analysis of content personalization technologies using basic recommendation system frameworks revealed a systemic deficit of tools for incorporating the utility-temporal specifics of B2B transactions. This justifies the necessity of employing data mining methods on B2B customer transactional and behavioral data to discover associations and sequences. However, a comparative analysis of data mining methods aimed at identifying associations and sequences in transactional datasets allowed for the establishment of the following:

– it was found that classical association rule mining (ARM) methods, particularly the Apriori, FP-Growth, and Eclat algorithms, focus on finding static relationships between items in a transaction. In doing so, only the frequency of items appearing in transactions is considered, completely neutralizing information regarding their cost or purchase volume. As a result, strategically important but rare B2B products are discarded, while cheap high-frequency corporate accessories become the core of the model. All of this demonstrates the complete «financial blindness» of classical data mining methods;

– it was found that classical sequential pattern mining (SPM) methods, such as GSP, PrefixSpan, and SPADE, aim to discover correlations between transactions against a chronological backdrop of events. While they partially address the time factor, they are characterized by a high level of "identification discreteness," treating the customer as an abstract sequence of distinct identifiers. This makes it impossible to detect individual transaction rhythms, leading to the «temporal blindness» of SPM methods.

Thus, it is revealed that the existing data mining toolkit requires adaptation through the integration of financial utility metrics and time intervals between transactions. At the end of the chapter, based on an analysis of the capabilities of the widespread industry standard, namely the RFM (Recency, Frequency, Monetary) model for customer segmentation in the B2B environment, its significant limitations are proven. Specifically, the Recency metric captures only a static point on the timeline without considering the volatility of the individual purchasing cycle, while the Monetary metric evaluates gross revenue, ignoring specific utility and the stability of consumption patterns.

Thus, the classical RFM analysis model is oriented toward a retrospective description of state («what happened»), which prevents building predictive CLV (Customer Lifetime Value) models with high accuracy for B2B customers with irregular activity rhythms. Summarizing the results of the state and trend analysis of the industry, a critical need to develop new approaches to B2B content personalization is established. The expediency of shifting from static segmentation to constructing a utility-weighted temporal customer profile is substantiated. This will allow overcoming the financial and temporal blindness of existing models, ensuring a transition from merely stating past transactions to the predictive

management of communication triggers based on the forecasted purchasing cycle, financial capacity, and behavioral stability of each specific counterparty.

The second chapter develops a utility-weighted temporal approach to the data mining of transactional and behavioral datasets of B2B customers. The systemic investigation of existing association rule mining and sequential pattern mining methods in accumulated B2B transaction arrays resulted in establishing fundamental methodical limitations that prevent their direct application for creating personalized content in e-commerce systems.

The existence of five critical gaps (limitations) between the mathematical assumptions of association and sequence mining methods and the nature of the B2B domain is substantiated, namely: the «financial blindness» of binary models, the occurrence of a «combinatorial explosion» during weighted analysis, the «temporal blindness» of static models, the presence of identification discreteness, and the lack of a differentiated customer interaction strategy. It is established that these limitations are conceptually interconnected and their relationship possesses a hierarchical structure.

That is, an attempt to overcome the financial blindness of classical data mining models by adding cost metrics activates the combinatorial explosion problem. In turn, moving to the analysis of temporal intervals in transactional data requires not only a new processing architecture but also the elimination of identification discreteness for B2B customers, since calculating cyclicity is possible only on integral rather than fragmented time series. Thus, the final limitation – the lack of a differentiated customer interaction strategy, i.e., strategic monotony – is a direct result of data deficits at all previous stages of analysis. In other words, in a B2B e-commerce system, it is impossible to differentiate (determine classes of) customers without identifying their individual financial and temporal patterns.

Therefore, overcoming these limitations serves as the scientific foundation for developing a utility-weighted temporal analysis, the mathematical apparatus of which is conventionally divided into two components: a utility-weighted transaction model and a utility pattern growth method (UP-Growth), alongside a temporal sequence model and a method for the temporal analysis of utility patterns. A utility-weighted B2B transaction model has been developed by adding the formalization of a total transaction utility indicator

$U(T)$ to the classical transaction definition, calculated as the sum of the products of the quantities of purchased item components and their individual prices. To overcome the combinatorial explosion effect that arises when using this utility indicator, the definition of Transaction-Weighted Utilization $U(T)$ is formalized as the total value of all transactions in the database that contain the specified transaction. Using the utility-weighted B2B transaction model, the UP-Growth utility pattern mining method was developed. It is shown that its application eliminates the «financial blindness» of the classical approach to discovering associative patterns in transactional data.

Thus, during the first pass of UP-tree construction, nodes accumulate the monetary value of items rather than their support count. Utilizing the $U(T)$ mechanism during the second pass restores the downward closure (anti-monotonicity) property, ensuring computational stability even when analyzing long B2B invoices. The output of the method is an ordered list of utility patterns – combinations of goods that generate the main share of the enterprise's revenue. A temporal B2B sequence model has been developed by adding the transaction occurrence time to the classical sequence definition, which allowed obtaining an *Inter-Purchase Time (IPT)* vector for an arbitrary B2B customer. This, in conjunction with a calculated dimensionless indicator – the Coefficient of Variation CV – provides the capability to determine transaction cycle stability, thereby overcoming the «temporal blindness» and identification discreteness of classical sequential pattern mining.

Using the temporal B2B sequence model, a method for the temporal analysis of utility patterns obtained via the UP-Growth method was developed. It is shown that the implementation of this method eliminates the «temporal blindness» of the classical approach to sequential pattern mining by constructing *IPT* vectors and their subsequent evaluation against the CV criterion for any B2B customer. An interpretation of the CV criterion depending on the established stability threshold is provided. The value of the threshold is not fixed and can be adjusted depending on industry specifics and the enterprise's tolerance for prediction errors. In most practical applications, the recommended value is in a range from 0,2 to 0,5. For utility-weighted sequential patterns that pass the verification check under the CV criterion, the date of the predictive offer to the B2B customer is calculated as the moment of communication trigger activation, factoring in the preventiveness coefficient,

which determines the warning depth. As a general summary of the chapter, it is demonstrated that using the proposed two-stage pipeline consisting of the UP-Growth utility pattern mining method and its temporal analysis yields a list of personalized triggers.

Thanks to this, the moment for receiving a predictive offer is determined for each counterparty and each financially significant and cyclically stable transaction. As confirmation, a detailed examination of two typical scenarios for B2B e-commerce systems – «regular seasonal purchasing» and «irregular purchasing» – is provided. It is precisely the capability to implement these scenarios that transforms a B2B platform from a reactive catalog into an intelligent recommendation system capable of anticipating customer needs regarding warehouse inventory replenishment and interacting with them at the most favorable moment.

The third chapter develops a B2B customer classification method for predictive content personalization in e-commerce systems. The existence of three critical limitations of using classical RFM analysis for B2B customer classification is proven. It is shown that using the Recency parameter R only captures the time elapsed since the last transaction. In doing so, an IPT vector of inter-purchase patterns is not constructed, making the R metric «temporally blind» to the individual rhythm of B2B operations. Using the Frequency parameter F evaluates the total number of transactions over a period, yet ignores transaction stability, the indicator of which is the *Coefficient of Variation* CV calculated based on IPT . The Monetary volume parameter M is calculated as the total expenditure of the customer, but in this form, it provides no information on transaction structure. Furthermore, it does not account for Transaction-Weighted Utilization TWU , which prevents isolating the driver-product patterns that ensure turnover stability.

To overcome these drawbacks of standard RFM analysis, a utility-weighted temporal customer profile (UW-TCP) is formalized as a tuple of features.

To generate predictive offers for B2B customers, a classification method was developed, which constitutes the third component of the utility-weighted temporal analysis of B2B customer transactional and behavioral data in an e-commerce system. It functions exclusively as an add-on over the proposed UP-Growth utility pattern mining method and the method for their temporal analysis. The application of the B2B customer classification

method is based on the features of the B2B customer's utility-weighted temporal profile, which are aggregates of already available intermediate results, making the classification method the third stage of the pipeline for generating predictive offers for B2B customers. The assignment of B2B customer class IDs is integrated into the communication trigger generation system. This provides a meaningful link between the customer class and the interaction strategy through a mathematically substantiated assignment of the trigger window width according to customer value, rather than through a subjective definition of segmentation thresholds as seen in standard RFM analysis.

The fourth chapter presents an experimental validation of the effectiveness of the proposed utility-weighted temporal approach for generating predictive offers in a B2B e-commerce platform. The developed models and methods are considered as components of a recommendation module integrated into the operational workflow of the real B2B platform “Baza Vzuttya”. The interaction between the platform, the analytical pipeline, and the manager interface is formalized, where order history, customer data, product items, SKUs, prices, and transaction amounts are used to generate active personalized recommendations.

The experimental study was conducted on a transactional-behavioral dataset for 2022–2025, which included 29.35 thousand orders, 2.53 thousand customers, 56.20 thousand unique SKUs, and a total order volume of UAH 416.28 million. After preprocessing, forming the transactional representation, and filtering excessively long orders, 28.53 thousand orders were used for further analysis, which made it possible to reduce combinatorial complexity without losing the main part of the experimental dataset.

To evaluate effectiveness, the reverse ablation study methodology was tested. Four scenarios were sequentially compared: the baseline frequency-based scenario S0, the utility-weighted scenario S1, the utility-temporal scenario S2, and the integrated scenario S3, which additionally takes into account the B2B customer class. The results showed a consistent improvement in recommendation quality: conversion increased from 3.91 % in S0 to 5.39 % in S1, 12.69 % in S2, and 15.91 % in S3. At the same time, the number of generated recommendations decreased from 920 to 220, that is, by 76.09 %, indicating a reduction in excessive communication load on managers and customers.

The best result was obtained for the integrated scenario S3. It generated 220 recommendations, 35 of which were successful, ensured an average revenue per recommendation of UAH 6.91 thousand, and covered test revenue of UAH 1.52 million. Compared with the baseline scenario S0, the number of recommendations decreased by 700, while the number of successful interactions remained almost unchanged. At the same time, ARPR increased from UAH 1.21 thousand to UAH 6.91 thousand, that is, approximately 5.7 times. This confirmed that the full model filters out a significant share of irrelevant recommendations and increases the practical value of each interaction.

The economic evaluation also demonstrated the advantage of the integrated approach. Assuming the cost of one communication interaction is UAH 150, operational costs decreased from UAH 138.00 thousand in the baseline scenario to UAH 33.00 thousand in scenario S3. The cost of one successful recommendation decreased from UAH 3.83 thousand to UAH 0.94 thousand, while the net financial result of the integrated model amounted to UAH 1.49 million. The obtained results prove that the proposed approach enables a transition from mass recommendation generation to more accurate, timely, and economically justified personalization of B2B offers.

The place of the developed system in the architecture of the B2B platform was defined as an applied recommendation module that extends the platform's functionality without changing its core transactional processes. The software implementation was carried out as an analytical module that receives data on customers, orders, and product items, applies the developed three-stage pipeline, and generates active predictive recommendations for the manager. To demonstrate the practical use of the results, interface screens were implemented for the manager dashboard, active recommendations, and a personalized offer for a B2B customer.

The scientific novelty of the obtained results is as follows:

- Improved: The utility-weighted transaction model. Unlike traditional representations in the form of a flat stock-keeping unit (SKU) list, it additionally incorporates SKU volume and individual price. This enables the subsequent use of their product to calculate Transaction-Weighted Utilization (\$TWU\$) and identifies transactions

not only by the repetition criterion but also by their economic significance, thereby overcoming the "financial blindness" limitation;

- Further developed: The utility pattern growth (UP-Growth) method. By utilizing the Transaction-Weighted Utilization (\$TWU\$) metric to compute the values of UP-tree nodes instead of the frequency counters used in classical FP-Growth, the downward closure (anti-monotonicity) property has been restored, overcoming the "combinatorial explosion" limitation. Furthermore, this shifts the priority of association rule generation, focusing the personalization system on economically significant transactions;

- Improved: The temporal sequence model. Unlike traditional representations in the form of tuples of identificationally discrete transactions, it accounts for a deterministic B2B customer identifier and additionally includes the transaction timestamp. This allows for the creation of an Inter-Purchase Time *IPT* vector and the utilization of its statistical characteristics to calculate the probable activation date of a predictive offer, overcoming the limitations of "temporal blindness" and identification discreteness;

- Further developed: The method for the temporal analysis of UP-Growth utility patterns. Working in conjunction with the temporal sequence model, it enables the calculation of the Coefficient of Variation *CV* of the Inter-Purchase Time vector. This provides the capability to determine the exact date of a predictive offer to a B2B customer, generating predictive triggers in the form of «customer – product set – interaction date» to increase the level of content personalization in B2B e-commerce systems;

- Developed for the first time: A B2B customer classification method for e-commerce systems. Unlike the existing RFM analysis, it is based on a utility-weighted temporal customer profile formed through a two-stage analysis pipeline of B2B customer transactional and behavioral data. This enabled the implementation of a three-stage pipeline for generating individual predictive offers depending on the determined customer class, ultimately increasing the final accuracy of predictive content personalization.

The practical significance of the research results lies in the development of a software analytical module for generating personalized predictive offers in B2B e-commerce systems. The module is based on a three-stage pipeline that combines utility-weighted analysis of product patterns, temporal verification of purchasing cycle stability,

and classification of B2B customers based on a utility-weighted temporal profile. This makes it possible to generate offers not only on the basis of product relevance, but also taking into account the financial significance of the pattern, the predicted moment of interaction, and the customer class.

The practical approbation of the results was carried out by integrating the developed pipeline as a Python module into the operational workflow of the B2B platform «Baza Vzuttya». The implemented module uses data on customers, orders, product items, SKUs, prices, and transaction amounts, after which it generates active recommendations for the manager. To demonstrate the applied use of the results, interface screens were developed for the manager dashboard, active recommendations, and a personalized offer for a B2B customer.

The effectiveness of the proposed approach was confirmed by an experimental comparison of the baseline frequency-based scenario and the full integrated model using adapted business-oriented metrics. The obtained results indicate an increase in the accuracy and timeliness of providing individual predictive offers to B2B customers, a reduction in the number of irrelevant recommendations, and an increase in the practical value of each communication interaction.

The scientific and practical value of the dissertation research consists of the three-stage pipeline for generating individual predictive offers, the methodology for its integration into a B2B platform, and the system of adapted metrics for evaluating recommendation effectiveness. The proposed utility-weighted temporal approach to the data mining of B2B customers' transactional and behavioral data has been implemented into the operations of Sole Proprietor (NEWOPT) and LLC "STILLI", and has also been utilized in the educational process and during the execution of R&D (Research and Development) projects at the Department of Information Systems of Odesa National Polytechnic University.

Keywords: data mining; B2B e-commerce systems; B2B customer classification; content personalization; recommendation systems; transactional and behavioral data.

List of publications of the applicant on the topic of the dissertation

1. Arsirii, O., & Ivanov, D. (2025). Models of affinity analysis of B2B e-commerce customers' transactional and behavioral data for personalized content creation]. *Pratsi Odeskoho Politekhnicnoho Universitetu*, 2(72), 90–101.
<https://doi.org/10.15276/opu.2.72.2025>
2. Arsirii, O., & Ivanov, D. (2026). A Model and Method of Transactional-behavioral Data Mining for B2B Content Personalization. *Electronics and Control Systems*, 1(87), 14–23. <https://doi.org/10.18372/1990-5548.87.20881>
3. Arsirii, O. O., & Ivanov, D. V. (2026) Data mining of B2B customer transactional data based on a utility-weighted temporal approach]. *Elektronika ta Systemy Upravlinnia*, 2(88), 16–31. <https://doi.org/10.18372/1990-5548.88.20958>
4. Arsirii, O. O., Ivanov, D. V., & Babilunha, O. Yu. (2025). Analysis of models, methods and technologies for creating personalized content in B2B e-commerce systems. *Informatyka. Kultura. Tekhnika*, 2, 273–278. <https://doi.org/10.15276/ict.02.2025.41>
5. Arsirii, O. O., Liubomska, O. M., Rudenko, O. V., & Ivanov, D. V. (2024). Hibrydna rekomendatsiina systema dlya pidtrymky UI/UX dyzaineriv [A hybrid recommendation system for supporting UI/UX designers]. *Materialy mizhnarodnoi konferentsii "Informatyka. Kultura. Tekhnika"*, 1(1), 29–35.
<http://dspace.opu.ua/jspui/handle/123456789/14629>
6. Siroshtan, M. V., Ivanov, D. V., & Arsirii, O. O. (2024). Investigation of machine learning methods efficiency on the example of clothing items classification problem. *Project, Program, Portfolio Management (P3M-2024): Tezy dopovidei IX Mizhnarodnoi naukovo-praktichnoi konferentsii*, 1, 201–204.
<https://rep.nuos.edu.ua/server/api/core/bitstreams/f2c763dc-2728-47ca-af64-31618bd2a659/content>
7. Ivanov, D. V., & Arsirii, O. O. (2026). B2B customer classification in an e-commerce system based on a utility-weighted temporal profile. *Materialy XVI mizhnarodnoi konferentsii "Suchasni Informatsiini Tekhnologii 2026" (Modern Information Technology 2026)*, Odesa, Ukraine, 118–121.
https://drive.google.com/drive/folders/1uXN7b84231YhSfT_tY6I9eMhPjDrIBKJ

ЗМІСТ

Вступ.....	27
1 Аналіз стану та тенденцій персоналізації B2B-контенту в системах електронної комерції.....	34
1.1 Аналіз тенденцій розвитку B2B систем електронної комерції	34
1.2 Особливості B2B та B2C транзакцій.....	38
1.3 Огляд існуючих технологій персоналізації та рекомендаційних систем	41
1.4 Аналіз існуючих методів пошуку асоціацій та послідовностей.....	49
1.5 Аналіз сегментації клієнтів за RFM-аналізом	66
1.6 Висновки до першого розділу	69
2 Ціннісно-зважений темпоральний підхід до інтелектуального аналізу транзакційних даних B2B-клієнтів.....	72
2.1 Методичні обмеження класичних методів створення асоціативних правил та послідовних шаблонів у контексті транзакційних масивів B2B	72
2.2 Ціннісно-зважений аналіз транзакційних даних B2B клієнтів.....	81
2.3 Темпоральний аналіз транзакційних даних B2B клієнтів.....	89
2.4 Зведена характеристика ціннісно-зваженого темпорального аналізу	95
2.5 Висновки до другого розділу	99
3 Класифікація B2B-клієнтів за ціннісно-зваженим темпоральним профілем.....	103
3.1 Методичні обмеження стандартного RFM-аналізу до кластеризації B2B-клієнтів.....	104
3.2 Висновки до третього розділу	114
4 Дослідження ефективності рекомендацій для персоналізації контенту на B2B-платформі електронної комерції	117
4.1 Характеристика експериментальних даних та їх попередня обробка	117
4.2 Метрики оцінювання ефективності рекомендацій	126
4.3 Сценарії проведення комплексного експерименту.....	128
4.4 Порівняння конфігурацій рекомендаційного підходу.....	133
4.5 Аналіз чутливості темпоральних параметрів рекомендаційної моделі.....	139

4.6 Аналіз приросту ефективності запропонованих моделей і методів	143
4.7 Економічна оцінка використання рекомендаційного модуля в B2B-платформі	148
4.8 Інтерпретація та обговорення результатів експериментального дослідження...	151
4.9 Характеристика взаємодії B2B-платформи електронної комерції та рекомендаційної системи	154
4.10 Програмний модуль підтримки персоналізованих B2B-пропозицій.....	157
4.11 Висновки до четвертого розділу	161
Висновки	164
Список використаних джерел	168
Додаток А Акт впровадження результатів на B2B-платформі «База взуття»	180
Додаток Б Акт впровадження результатів у НВП «Каре»	181
Додаток В Довідка про використання результатів дисертаційної роботи у навчальному процесі Національного університету «Одеська політехніка»	182
Додаток Г Довідка про використання результатів дисертаційної роботи у навчальному процесі Національного університету «Одеська політехніка»	1823
Додаток Д Фрагменти вихідного коду	184
Додаток Е Список публікацій здобувача	205
Додаток Ж Відомості про апробацію результатів дисертаційної роботи	206

ВСТУП

Аналіз специфіки глобального ринку електронної комерції показує, що середньорічний темп приросту B2B-сегмента за прогнозами провідних аналітичних агентств (Gartner, Statista) становитиме близько 18,2% у найближчі п'ять років. В Україні, попри виклики воєнного часу, також спостерігається інтенсивна трансформація закупівельних процесів, де понад 60% суб'єктів господарювання переходять на цифрові платформи для оптимізації ланцюгів постачання. З іншого боку більш детальні дослідження показують, що такий приріст ринку електронної комерції є значною мірою результатом так званого «цифрового примусу», що ринок зростає екстенсивно за рахунок збільшення кількості клієнтів, а не інтенсивно за рахунок збільшення якості обслуговування. Тобто релокація підприємств, розрив старих логістичних ланцюгів та перехід на віддалену роботу змусили компанії переносити свою діяльність в онлайн. При цьому більшість платформ надають своїм клієнтам в якості персоналізованого контенту послуги «вітрини» або «калькулятора» майже ігноруючи можливості побудови так званих інтелектуальних патернів поведінки. Якщо така персоналізація для сегмента B2C електронної комерції, де на поведінку клієнта впливають емоційні чинники, є більш прийнятною, то у B2B персоналізація повинна трансформуватися в систему рекомендацій, яка базується на показниках повернення інвестицій (ROI) та підтримці операційної стабільності партнерських відносин шляхом формування предикативних пропозицій. Це зумовило актуальність критичного перегляду існуючих моделей, методів та технологій рекомендації для персоналізації контенту систем електронної комерції на основі інтелектуального аналізу транзакційно-поведінкових даних клієнтів. Аналіз показав, що класичні рекомендаційні системи, побудовані на принципах колаборативної фільтрації та асоціативного аналізу (Apriori, FP-Growth, Eclat), а також методи сегментації (RFM-аналіз), демонструють високу ефективність у B2C-системах з однорідними масивами даних. Натомість у B2B-середовищі ці методи стикаються з проблемою методичної обмеженості, оскільки вони орієнтовані на виявлення частотних залежностей, ігноруючи при цьому ціннісну вагу окремих позицій та

індивідуальну циклічність клієнтських закупівель, крім того вони не адаптовані до роботи з розрідженими даними B2B-транзакцій та тому мають високу обчислювальну складність. Як наслідок, це унеможлиблює предикативну персоналізацію контенту B2B-систем електронної комерції, яка є основою раціонального обґрунтування для рекомендаційних рішень.

Таким чином незважаючи на високі темпи кількісного зростання сегмента, якість функціонування B2B-систем електронної комерції залишається на рівні механічного обслуговування транзакцій. Тобто спостерігається технічне протиріччя між технологічними можливостями накопичення великих даних та реальним рівнем їх використання для створення персоналізованого контенту. Крім того існує наукова суперечність між необхідністю переходу до інтенсивних методів утримання клієнтів та відсутністю в існуючих системах електронної комерції інструментів для інтелектуального прогнозування закупівельного циклу. Це призводить до зниження лояльності та неефективного управління запасами у B2B-клієнтів.

Таким чином, розробка моделей та методів інтелектуального аналізу даних B2B-клієнтів для персоналізації контенту систем електронної комерції є актуальною як з технічної так із наукової точки зору.

Зв'язок роботи з науковими програмами, темами, планами. Дисертаційне дослідження виконано у відповідності до пріоритетних напрямків науково-дослідних робіт Національного університету «Одеська політехніка», згідно координаційних планів Міністерства освіти і науки України, зокрема, в рамках наукових досліджень за держбюджетними науково-дослідними роботами (НДР): «Методи та засоби інформаційного забезпечення діагностування, вдосконалення та експлуатації енергетичного обладнання» № 164-145, 2021-2026 р.р. (№ держреєстрації 0119U002004) та «Розробка нейромережових технологій для практичних застосунків» № 254-145, 2025-2030 р.р. (№ держ. реєстрації 01240005141).

Метою дослідження полягає в вирішенні науково-прикладної задачі персоналізації контенту систем електронної комерції за рахунок підвищення ефективності формування предикативних пропозицій у відповідності до класів B2B-

клієнтів шляхом розробки інтелектуального ціннісно-зваженого темпорального підходу до аналізу масивів транзакційно-повелінкових даних.

Для досягнення поставленої мети визначено комплекс взаємопов'язаних задач:

- проаналізувати існуючі технології персоналізації контенту в системах електронної комерції та визначити шляхи підвищення їх ефективності для B2B-клієнтів в сучасних умовах;

- розробити ціннісно-зважений темпоральний підхід до інтелектуального аналізу масивів транзакційно-поведінкових даних B2B клієнтів;

- розробити метод класифікації B2B клієнтів для предиктивної персоналізації контенту систем електронної комерції;

- провести симуляційне моделювання стратегій персоналізації контенту для оцінки їхньої ефективності порівняно з класичними підходами та інтегрувати розроблений програмний аналітичний модуль в систему рекомендації для реальної B2B-платформи.

Об'єктом дослідження є процес персоналізації контенту в системах B2B електронної комерції з урахуванням транзакційно-поведінкових даних клієнтів.

Предметом дослідження є моделі та методи інтелектуального аналізу транзакційно-поведінкових даних B2B-клієнтів систем електронної комерції.

Методи дослідження. Для розв'язання поставлених у роботі задач було застосовано комплекс загально наукових методів, що ґрунтуються на фундаментальних положеннях комп'ютерних наук та теорії інтелектуального аналізу даних. Зокрема: методи системного та компаративного аналізу використані при дослідженні існуючих технологій персоналізації контенту, аналізі особливостей функціонування B2B-платформ та визначенні критичних обмежень сучасних рекомендаційних систем; методи теорії інтелектуального аналізу даних, а саме пошуку асоціативних правил та послідовних шаблонів застосовано при розробці ціннісно-зваженого темпорального підходу для виявлення прихованих закономірностей у масивах транзакційно-поведінкових даних; методи математичної статистики використані для розрахунку характеристик закупівельного циклу (середнього інтервалу, відхилення, коефіцієнта варіації) та оцінки стабільності

поведінки B2B-клієнтів; методи математичного моделювання та теорії класифікації покладені в основу розробки методу побудови ціннісно-зваженого темпорального профілю клієнта та подальшого розподілу контрагентів за класами для предиктивної персоналізації; методи симуляційного моделювання та об'єктно-орієнтованого програмування застосовані для перевірки ефективності розроблених стратегій персоналізації, проведення порівняльних експериментів із класичними підходами та програмної реалізації аналітичного модуля для реальної B2B-платформи.

Наукова новизна отриманих результатів полягає у вирішенні науково-прикладної задачі підвищення ефективності взаємодії B2B-клієнтів шляхом предиктивної персоналізації контенту систем електронної комерції за рахунок розробки моделей та методів інтелектуального ціннісно-зваженого темпорального аналізу масивів транзакційно-повелінкових даних:

1) *удосконалено* модель ціннісно-зваженої транзакції, яка на відміну представлення у вигляді переліка товарних пропозицій (SKU), додатково включає обсяг SKU та індивідуальну ціну, що дозволяє в подальшому використовувати їх добуток для розрахунку транзакційно-зваженої корисності (TWU) та ідентифікувати транзакції за не лише за критерієм повторюваності, а і за економічної значущості для подолання обмеження «фінансової сліпоты»;

2) *отримав подальший розвиток* метод розростання цінністних патернів (Utility Pattern Growth) за рахунок використання транзакційно-зваженої корисності (TWU) для розрахунку значень вузлів дерева UP-Growth замість частотних лічильників класичного FP-Growth, що дозволило відновити властивість антимонотоності долаючи обмеження «комбінаторного вибуху», а також змінити пріоритетність генерації асоціативних правил, фокусуючи систему персоналізації на економічно значущих транзакціях;

3) *удосконалено* модель темпоральної послідовності, яка на відміну представлення у вигляді кортежів ідентифікаційно дискретних транзакцій враховує наявність детермінованого ідентифікатора B2B-та додатково включає час виникнення транзакції, що дозволяє створити вектор міжтранзакційних інтервалів (IPT) та використовувати його статистичні характеристики для розрахунку ймовірної дати

активації предикативної пропозиції для подолання обмежень «темпоральної сліпоты» та ідентифікаційної дискретності;

4) *отримав подальший розвиток* метод темпорального аналізу ціннісно-патернів UT-Growth, який спільно з моделлю темпоральної послідовності дозволяє розрахувати коефіцієнт варіації (CV) вектору міжтранзакційних інтервалів, що дає можливість визначити дату предикативної пропозиції B2B-клієнту формуючі предикативні тригери вигляду «клієнт – товарний набір – дата звернення» для підвищення рівня персоналізації контенту B2B-систем електронної комерції;

5) *вперше розроблено* метод класифікації B2B клієнтів систем електронної комерції, який на відміну від існуючого RFM-аналізу базується на ціннісно-зваженому темпоральному профілі клієнта сформованому у результаті реалізації двоетапного конвеєру аналізу транзакційно-поведінкових даних B2B клієнта. Це дозволило реалізувати трьох етапний конвеєр формування індивідуальних предикативних пропозицій в залежності від визначеного класу клієнта та підвищити фінальну точність предикативної персоналізації контенту.

Практичне значення одержаних результатів. Практичне значення результатів дослідження полягає в розробленні програмного аналітичного модуля для формування персоналізованих предикативних пропозицій у системах B2B електронної комерції. Основою модуля є триетапний конвеєр, який поєднує ціннісно-зважений аналіз товарних патернів, темпоральну перевірку закупівельних циклів та класифікацію B2B-клієнтів. Це дозволяє формувати індивідуальні пропозиції з урахуванням класу клієнта, фінансової значущості товарного патерну та прогнозованого моменту звернення.

Практичну апробацію результатів виконано шляхом інтеграції розробленого конвеєра у вигляді Python-модуля в контур роботи B2B-платформи «База взуття». Ефективність запропонованого підходу підтверджено порівняльним експериментальним дослідженням базового частотного сценарію та повної інтегрованої моделі за системою адаптованих бізнес-орієнтованих метрик: частка влучань рекомендацій, конверсія рекомендацій, цінність успішних рекомендацій, охоплення виручки рекомендаціями та середня виручка на одну рекомендацію.

Отримані результати свідчать про підвищення точності та своєчасності надання індивідуальних предикативних пропозицій B2B-клієнтам, зменшення кількості нерелевантних рекомендацій і підвищення практичної цінності комунікаційного звернення.

Науково-практичну цінність дисертаційного дослідження становлять триетапний конвеєр формування індивідуальних предикативних пропозицій, методика його інтеграції в B2B-платформу та система адаптованих метрик оцінювання ефективності рекомендацій. Запропонований ціннісно-зважений темпоральний підхід до інтелектуального аналізу транзакційно-поведінкових даних B2B-клієнтів впроваджено в діяльність ФОП (NEWOPT) та ТОВ «СТІЛЛІ», а також використано в при виконанні НДР та навчальному процесі Національного університету «Одеська політехніка» на кафедрі інформаційних систем.

Особистий внесок здобувача. Основні наукові положення, висновки та практичні рекомендації, що містяться в дисертації та виносяться на захист, отримано здобувачем особисто у період навчання в аспірантурі з 2023 по 2025 рр. У роботі [1] та роботах [4-7], які написано у співавторстві внесок здобувача полягає у проведенні системного аналізу існуючих технології персоналізації контенту в системах електронної комерції та визначенні шляхів підвищення їх ефективності для B2B-клієнтів; у роботі [2] автором запропоновано модель ціннісно-зваженої транзакції метод розростання цінністних патернів при проведенні інтелектуального аналізу масивів транзакційно-поведінкових B2B-даних; у роботі [3] автором розроблено модель темпоральної послідовності та метод темпорального аналізу цінністних патернів UT-Growth для підвищення рівня персоналізації контенту B2B-систем електронної комерції; у роботах [4, 7], написаних у співавторстві, автором запропоновано метод класифікації B2B клієнтів систем електронної комерції, який ціннісно-зваженому темпоральному профілі клієнта.

Апробація результатів дисертації. Основні положення та практичні результати дисертаційної роботи докладалися та одержали схвальні відгуки на міжнародних наукових конференціях, які проводились на базі Національного університету «Одеська політехніка» навчально-науковими інститутами

комп'ютерних систем та штучного інтелекту та робототехніки. А саме: XVI Modern Information Technology 2026 / Сучасні Інформаційні Технології 2026 (м. Одеса, 14-15 травня 2026 р.); I міжнародна конференція Інформатика. Культура. Техніка 2024/ Informatics. Culture. Technology 2024 (м. Одеса, 26-27 вересня 2024 р.) та II міжнародна конференція Інформатика. Культура. Техніка 2025/ Informatics. Culture. Technology 2025 (м. Одеса, 16-17 жовтня 2025 р.), IX Міжнародна науково-практичної конференції Project, Program, Portfolio Management. P3M-2024

Публікації. Основні результати дисертаційної роботи викладено в 7 публікаціях, з них: 3 статті у наукових фахових виданнях України з технічних наук та 4 публікацій у працях і матеріалах наукових конференцій.

Структура і обсяг роботи. Дисертаційна робота складається зі вступу, чотирьох розділів, висновків, списку використаних джерел та додатків. Загальний обсяг роботи містить 206 сторінок, в тому числі 25 рисунків, 39 таблиць та 7 додатків. Список використаних джерел налічує 112 найменувань. Додатки включають акт та довідки про впровадження результатів дисертаційної роботи, а також фрагменти вихідного коду.

1 АНАЛІЗ СТАНУ ТА ТЕНДЕНЦІЙ ПЕРСОНАЛІЗАЦІЇ B2B-КОНТЕНТУ В СИСТЕМАХ ЕЛЕКТРОННОЇ КОМЕРЦІЇ

В розділі показано, що зростання фінансової значущості та структурної складності сегмента ринку систем B2B електронної комерції, викликало необхідність підвищення його ефективності. Визначено концептуальну схему транзакцій B2B та порівняно системи B2B та B2C за основними показниками. Визначено специфіку персоналізації в B2B системах електронної комерції. Встановлено, що на відміну від B2C, де домінують емоційні чинники, у B2B персоналізація повинна базуватися на раціональних розрахунках, показниках повернення інвестицій (ROI) та підтримці операційної стабільності партнерських відносин. Це зумовило актуальність критичного перегляду існуючих моделей, методів та технологій рекомендації для персоналізації контенту систем електронної комерції на основі інтелектуального аналізу транзакційно-поведінкових даних клієнтів. Аналіз показав, що класичні рекомендаційні системи, побудовані на принципах колаборативної фільтрації та асоціативного аналізу (Apriori, FP-Growth, Eclat), а також методи сегментації (RFM-аналіз), демонструють високу ефективність у B2C-системах з однорідними масивами даних. Натомість у B2B-середовищі ці методи стикаються з проблемою методичної обмеженості, оскільки вони орієнтовані на виявлення частотних залежностей, ігноруючи при цьому ціннісну вагу окремих позицій та індивідуальну циклічність клієнтських закупівель, крім того вони не адаптовані до роботи з розрідженими даними B2B-транзакцій та тому мають високу обчислювальну складність. Як наслідок, це унеможливлює предикативну персоналізацію контенту B2B-систем електронної комерції, яка є основою раціонального обґрунтування для рекомендаційних рішень.

1.1 Аналіз тенденцій розвитку B2B систем електронної комерції

Сучасні B2B системи електронної комерції функціонують у висококонкурентному інтернет середовищі, де від якості контенту залежить

швидкість і точність задоволення потреб оптових клієнтів. Сучасні стандарти якості контенту базуються на персоналізованому підході до його створення. А необхідність ефективної персоналізації контенту B2B-систем підтверджується величезними фінансовими обсягами цього сегмента. Критична важливість ефективної B2B-персоналізації підтверджується масштабними фінансовими показниками цього сегмента.

Світовий ринок B2B електронної комерції, оцінений у 11,54 трлн дол. США у 2024 році, значно перевершує ринок B2C (\$6,55 трлн дол. США) [1]. Прогноз його зростання до \$60,62 трлн дол. США до 2034 року вказує на те, що навіть незначне підвищення ефективності взаємодії з клієнтами за допомогою персоналізації матиме мультитрильйонний економічний ефект [2]. Вагомий внесок у таку динаміку вносять саме цифрові платформи (маркетплейси), які на сьогодні агрегують понад 70 % усіх B2B-запитів. Згідно з аналітичними даними, обсяг продажів через B2B-маркетплейси зростає в середньому на 20-25 % щорічно [3], що вдвічі перевищує темпи зростання загальних прямих цифрових продажів. Очікується, що до 2027 року понад 50 % усіх корпоративних закупівель здійснюватимуться через консолідовані торгові майданчики [4].

На сьогодні до Топ-10 світових B2B-маркетплейсів, що формують ландшафт сучасної цифрової торгівлі, належать: Alibaba.com, яка є лідером за обсягом транзакцій (рис. 1.1), Amazon Business, з найшвидшими темпами впровадження персоналізації, eWorldTrade, Global Sources, ThomasNet, Made-in-China, Mercado Libre, DHGate, BizVibe та Kompass [5].

Роль цих платформ полягає у переході від моделі «простого каталогу» до інтелектуальних рекомендаційних систем, де персоналізація контенту стає ключовим інструментом утримання клієнта (Retention) та збільшення середнього чека (AOV) у раціональному сегменті закупівлі [6]. Тобто на світовому рівні фіксується тенденція, коли B2B сектор переживає «другу хвилю» цифровізації, де фокус зміщується з простої наявності вебресурсу на III-персоналізацію та автоматизацію закупівель [7]. Для підтвердження наведемо деякі статистичні данні, щодо збільшення обсягів ринку, цифрової домінації, зміна очікувань B2B-клієнтів та підвищення ролі III:

– очікується, що до кінця 2026 року глобальний ринок B2B e-commerce сягне \$36 трлн, а у 2027 році подолає відмітку в \$40 трлн (CAGR ~14.5 %). Для порівняння, це в кілька разів перевищує обсяги B2C ринку;

– за прогнозами *Gartner*, до кінця 2026 року 85 % усіх B2B транзакцій здійснюватимуться в цифровому форматі [8];

– близько 75 % B2B-клієнтів готові змінити постачальника, якщо інша платформа запропонує зручніший та більш персоналізований інтерфейс (Amazon-like experience) [9];

– у 2025 році понад 60 % B2B платформ вже інтегрували інструменти штучного інтелекту, причому 38 % з них використовують ШІ для товарних рекомендацій [10].

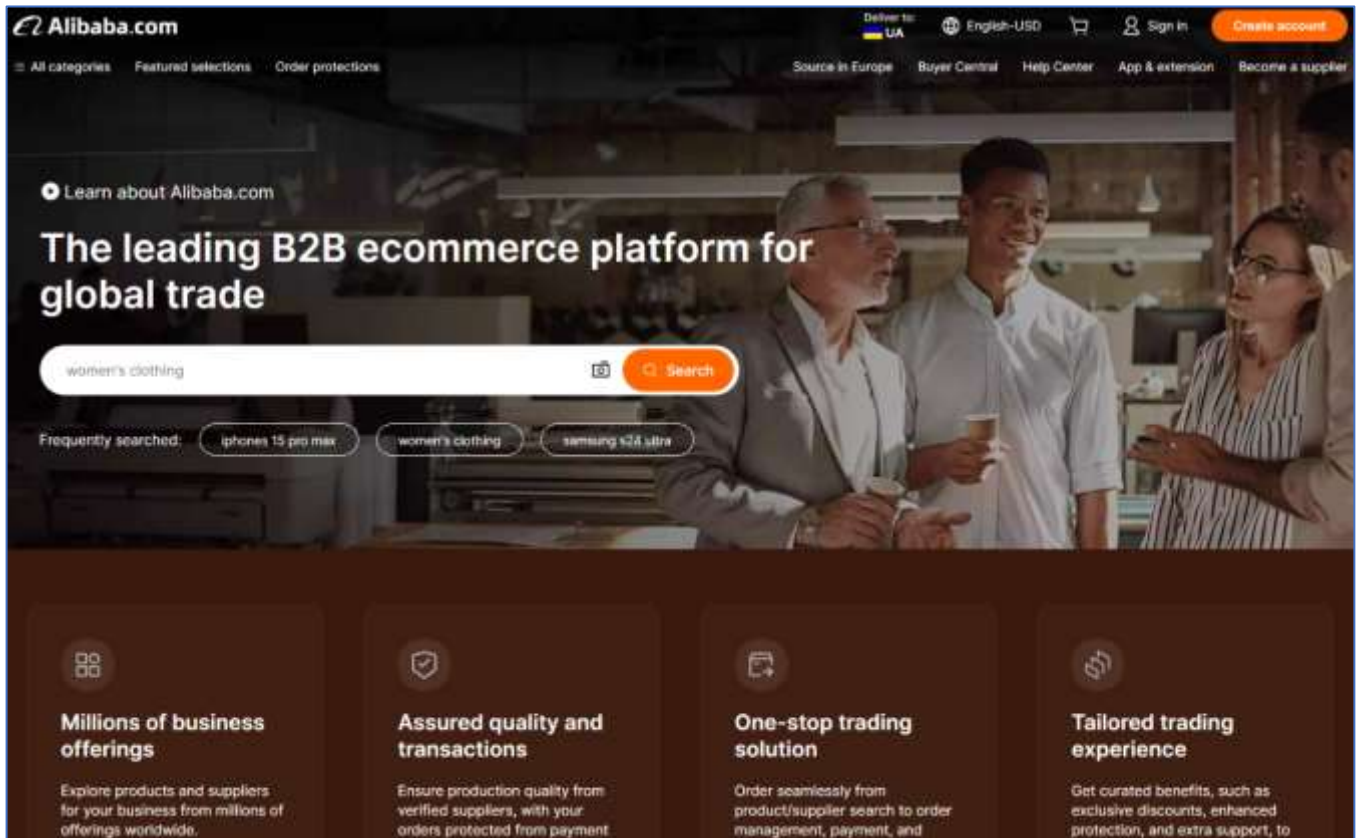


Рисунок 1.1 – Головна сторінка платформи «Alibaba.com»

На фоні стрімкого зростання B2B сегменту, вітчизняний ландшафт електронної комерції демонструє специфічну асиметрію розвитку [11]. Статистика підтверджує той факт, що B2B-сегмент електронної комерції в Україні зростає значно повільніше,

ніж світовий, а ринок традиційно надає перевагу розвиненим B2C-платформам, таким як Rozetka, Prom.ua, OLX та Allo.ua. Якщо у світі частка e-commerce у загальному B2B-обороті наближається до 70 %, то в Україні цей показник, за різними оцінками, ледь сягає 10 %. Більшість оптових операцій досі проводяться через традиційні канали (телефон, месенджери, особисті зустрічі). Типовим представником компаній, що працюють у цьому складному середовищі, є B2B SERVICE™, офіційний імпортер та ексклюзивний дистриб'ютор господарських товарів повсякденного попиту (рис. 1.2).

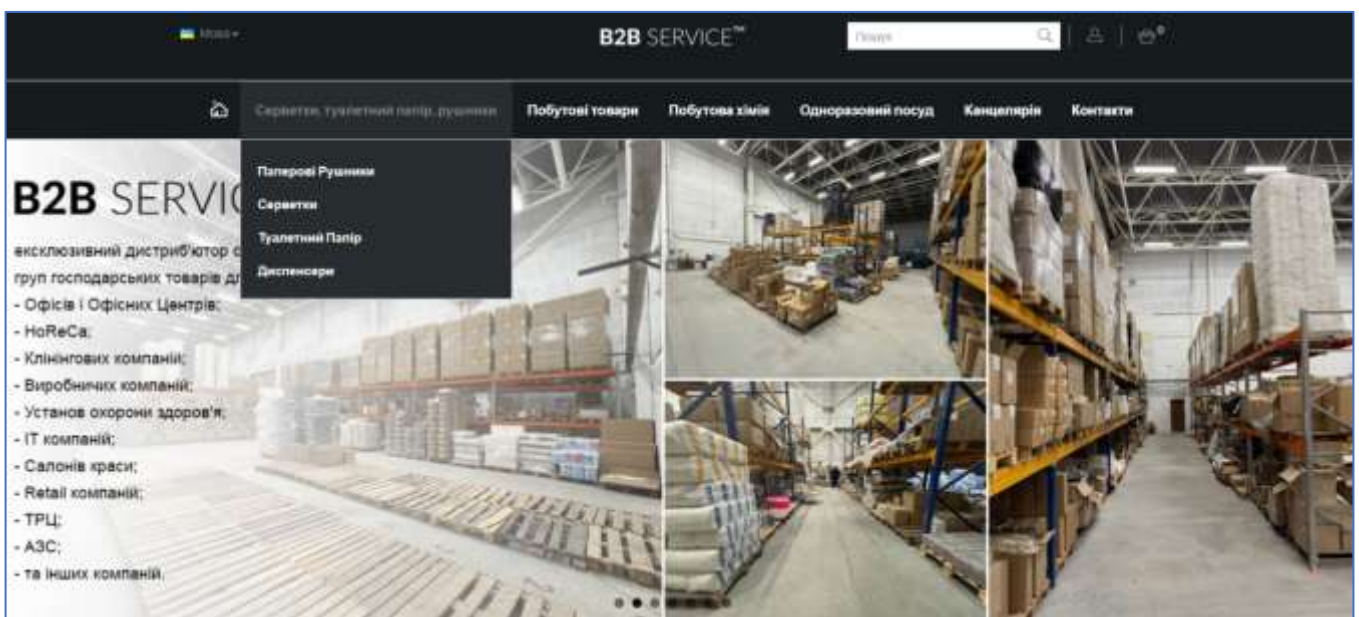


Рисунок 1.2 – Головна сторінка платформи «B2B SERVICE™»

Понад 60% українських B2B-компаній зазначають, що основною перешкодою для переходу в онлайн є неможливість забезпечити клієнта релевантним контентом: індивідуальними цінами, технічною документацією в реальному часі та предиктивними автоматизованими рекомендаціями щодо поповнення складів [12]. Дослідження показують, що лише 12 % вітчизняних B2B-платформ використовують елементи інтелектуального аналізу даних для персоналізації пропозицій, тоді як у світі цей показник перевищує 50 %.

Попри складні умови, український e-commerce демонструє високу динаміку, особливо в сегменті малого та середнього бізнесу (ФОП), які є основними клієнтами

таких B2B-платформ, як «База взуття». Так в цьому сегменті можна констатувати сплеск реєстрацій, орієнтацію на внутрішній ринок та перехід від оптових ринків у формат у формат маркетплейсів. Наведено приклади за 2025 рік, тенденції якого аналізуються:

- протягом 2025 року в Україні зафіксовано справжній бум нових онлайн-бізнесів. Кількість нових реєстрацій ФОП у сфері онлайн-торгівлі за 9 місяців 2025 року зросла на 62% порівняно з аналогічним періодом 2024 року [13];

- статистика показує, що 78% українських інтернет-магазинів працюють виключно на внутрішній ринок, що створює високу конкуренцію за лояльність локального клієнта [14];

- ключовим драйвером є перехід оптових ринків (Одеса «7-й кілометр», Хмельницький) у формат маркетплейсів. Це робить B2B-платформи електронної комерції малого та середнього бізнесу (ФОП) типовим та репрезентативними представниками сучасної української B2B екосистеми.

1.2 Особливості B2B та B2C транзакцій

Розглянемо особливості B2B та B2C транзакцій більш детально. Система B2B електронної комерції (B2B E-commerce System) це комплексна онлайн-платформа або набір взаємопов'язаних технологічних рішень [15], призначених для здійснення комерційних операцій, обміну інформацією та автоматизації бізнес-процесів між двома або більше юридичними особами.

На відміну від роздрібної торгівлі, транзакції B2B передбачають складний обмін товарами, послугами чи інформацією між детермінованими бізнес-клієнтами – юридичними особами (підприємствами) [16].

Структуру такої взаємодії показано на схемі (рис. 1.3), яка базується на чотирьох доменах Wholesale (Оптова торгівля), Marketing (Маркетинг), Customer (Клієнт) та Collaborative (Співпраця). Детальні відмінності між цими двома моделями функціонування систем електронної комерції за ключовими показниками наведено в таблиці 1.1.

Детальний розгляд представленої схеми дозволяє виділити специфічні аспекти B2B-транзакцій. Згідно концепції транзакцій першим аспектом є «*Оптова торгівля (Wholesale)*». Ця бізнес-модель дозволяє ритейлерам придбати продукцію за нижчою ціною і перепродати її споживачам із націнкою. Тому в блоці увага акцентується на масштаб та ціннісну неоднорідність транзакцій. Оптова торгівля оперує великими партіями SKU та складними ціновими матрицями, що вимагає від аналітичних модулів врахування «економічної ваги» кожної позиції [17].

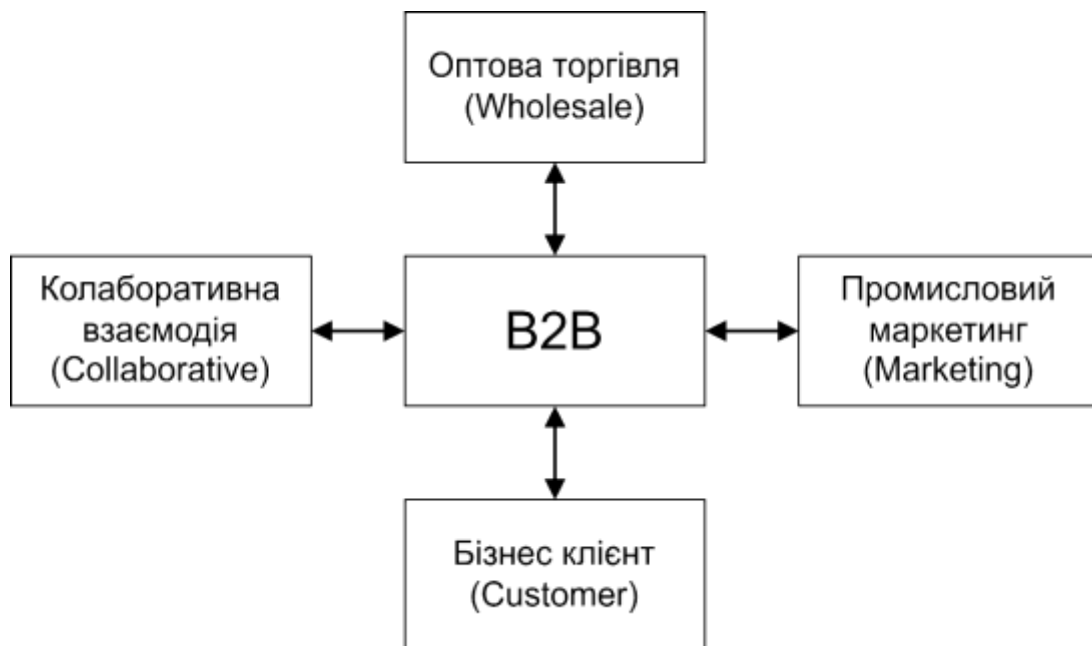


Рисунок 1.3 – Схема взаємодії суб'єктів у системах B2B електронної комерції

Наступним значущим аспектом є «*Промисловий маркетинг (Marketing)*». B2B-маркетингові стратегії зосереджені на побудові відносин з іншими підприємствами, а не з індивідуальними споживачами [18]. Тому блок трансформується у стратегію побудови довгострокових відносин. Це передбачає довший цикл продажу, де ключовим завданням персоналізації стає супровід складних процесів прийняття рішень. Ще одним суб'єкт схеми представлено в блоці «*Бізнес-клієнт (Customer)*». У B2B-контексті клієнтами є інші компанії (підприємства), а не фізичні особи. Такі клієнти мають відмінні потреби та купівельні поведінки порівняно з індивідуальними споживачами. Бізнес-клієнти характеризуються раціональною купівельною поведінкою є детермінованими тому система отримує можливість формувати

індивідуальні темпоральні профілі (ритми закупівель) на основі їхньої попередньої активності [19].

Таблиця – 1.1 Порівняння систем B2B та B2C систем електронної комерції

Назва показника	B2B (Business-to-Business)	B2C (Business-to-Consumer)
Цільова аудиторія	Детерміновані бізнес-клієнти (юридичні особи –, підприємства, великі корпорації, державні установи, неприбуткові організації). Кількість покупців менша	Масовий (індивідуальний) споживач. Кількість покупців значно більша та вони не детерміновані
Цінність транзакції	Висока вартість, більші обсяги, менша частота	Низька вартість, менші обсяги, вища частота
Мотивація покупки	Раціональна: орієнтована на ефективність бізнесу, зменшення витрат, збільшення прибутку, довгострокові цілі, розрахунок ROI	Емоційна/Імпульсивна: задоволення особистих потреб, бажань, на основі ціни, бренду, трендів
Тривалість процесу прийняття рішень	Довший і складніший: залучає кількох зацікавлених сторін, переговори, можливе тестування, контракти	Коротший і простіший: часто одноосібне рішення, швидка покупка
Цикл продажу	Довший (часто від 1 до 3 місяців, або більше)	Коротший
Особливості відносин з клієнтами	Довгострокові партнерства: акцент на побудові довіри, підтримці, індивідуальних рішеннях	Короткострокові, транзакційні
Особливості маркетингу	Фокус на логіці, цінності, ROI, персоналізовані пропозиції, прямі продажі, контент-маркетинг	Фокус на емоціях, іміджі, рекламі, соціальних мережах, акціях, зручності
Методи оплати	Складніші методи (кредитні лінії, рахунки-фактури, банківські перекази, відстрочки платежу). Вищий пріоритет безпеки	Швидкі та зручні методи (кредитні/дебетові картки, мобільні гаманці, електронні платежі)
Налаштування товару/послуги	Високий ступінь кастомізації, спеціалізовані рішення	Стандартизовані «готові» продукти

Блок «*Колаборативна взаємодія (Collaborative)*» є ключовим елементом B2B-відносин. Підприємства часто тісно працюють разом для досягнення спільних цілей. Такий підхід до співпраці сприяє довгостроковому партнерству та інноваціям. Вона передбачає тісну інтеграцію бізнес-процесів постачальника та замовника, де персоналізація контенту перетворюється на інструмент проактивної підтримки партнерства.

Таким чином, аналіз наведеної схеми «*Wholesale, Marketing, Customer, Collaborative*» доводить, що персоналізація в B2B має базуватися на глибокому інтелектуальному аналізі ціннісних та темпоральних параметрів [20]. Це докорінно відрізняє її від B2C-систем, де контент генерується переважно шляхом статистичного аналізу частотних покупок та загальних поведінкових шаблонів масового споживача.

1.3 Огляд існуючих технологій персоналізації та рекомендаційних систем

Визначені у попередньому підрозділі особливості B2B-транзакцій, зокрема їхня висока цінність та темпоральна детермінованість, висувають специфічні вимоги до технологій персоналізації контенту в системах електронної комерції. Оскільки традиційні методи взаємодії з клієнтами в електронній комерції часто не враховують операційну складність B2B-процесів, виникає необхідність аналізу існуючих технологій персоналізації контенту та рекомендаційних систем (РС) на предмет їхньої адаптивності до умов оптової торгівлі. Тобто метою даного підрозділу є аналіз є виявлення технологічних прогалин у класичних підходах, що перешкоджають реалізації персоналізації контенту для забезпечення ефективної B2B-взаємодії типу «вчасно та релевантно».

1.3.1 Загальна характеристика рекомендаційних систем

Рекомендаційна система (РС) – це інтелектуальна система (підклас системи фільтрації інформації), яка сприяє формуванню пропозицій, на основі аналізу

релевантної інформації, отриманої з великих даних про предметну область, наприклад ретейл, колекціонування, оренда житла, пошук роботи, тощо [21].

РС працює на основі аналізу поведінки користувача, яка певною мірою характеризує його уподобання. РС – це програмні засоби, що аналізують інформацію про користувачів та об'єктів, зокрема товари або контент, з метою пропозиції споживачам персоналізованих рекомендацій [22]. Це важливий компонент багатьох *онлайн-сервісів*, таких як електронна комерція, потокові медіа та соціальні мережі

У загальному вигляді, РС можна описати наступним чином: припустимо, що U – множина користувачів, I – множина об'єктів, R – множина відомих оцінок об'єктів, та $u_{i,j} \in R$, $u_i \in U$, $j \in I$ – оцінка користувача u_i об'єкту j . Необхідно розрахувати значення оцінки для пари (u', j') , для якої $u' \in U$, $j' \in I$, $u_{i',j'} \notin R$.

Процесну логіку та схему взаємодії системи електронної комерції з рекомендаційним модулем (рис. 1.4) опишемо наступним чином. Згідно з наведеною схемою, життєвий цикл формування рекомендацій складається з наступних етапів:

1) отримання вхідних даних транзакції. Система електронної комерції акумулює сирі дані про взаємодію клієнтів з платформою (історія замовлень, перегляди, транзакції) та передає їх до аналітичного контуру РС;

2) інтелектуальний аналіз даних (ІАД). На основі отриманих масивів даних транзакцій виконується попередня обробка, фільтрація та виявлення прихованих закономірностей. Саме на цьому етапі реалізуються алгоритми Data Mining, що дозволяють виокремити значущі патерни з розріджених даних;

3) створення моделі рекомендацій. Сформовані закономірності формалізуються у вигляді математичної моделі, яка здатна генерувати прогнози для нових пар «користувач-об'єкт»;

4) оцінювання якості моделі. Виконується перевірка адекватності та точності отриманої моделі. У разі незадовільних показників (шлях «-») система повертається до етапів отримання даних, переналаштування параметрів ІАД або створення моделі. При позитивному результаті (шлях «+») модель вважається готовою до експлуатації;

5) використання моделі рекомендацій. Фінальний етап, на якому персоналізований контент інтегрується безпосередньо в інтерфейс системи електронної комерції для представлення кінцевому клієнту.

Важливою особливістю схеми є наявність контуру зворотного зв'язку. Реакція клієнтів на надані рекомендації стає новими вхідними даними для етапу 1, забезпечуючи безперервне самонавчання системи [23].

Зважаючи на подальший матеріал дослідження блок *ІАД* (етап 2) є центральним компонентом системи, де реалізуються авторські методи ціннісно-зваженого аналізу, що дозволяють нівелювати недоліки класичних підходів фільтрації. Крім того слід наголосити, що науковий пошук у даній роботі зосереджений на вдосконаленні не лише блоку *ІАД* (етап 2), де формуються ціннісно-зважені бази знань, а й безпосередньо процесу створення моделі рекомендацій (етап 3).

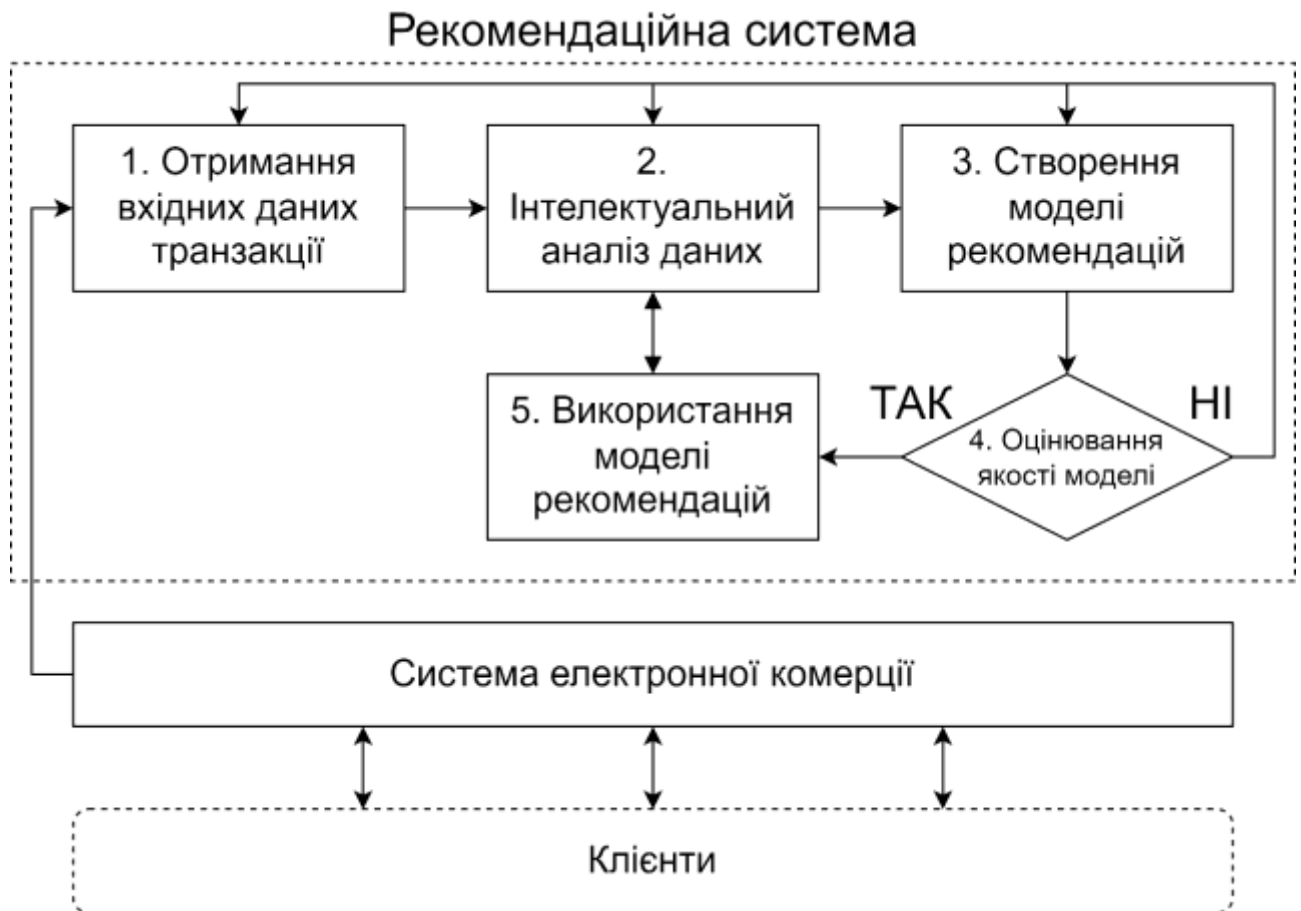


Рисунок 1.4 – Схема взаємодії системи електронної комерції та рекомендаційної системи

1.3.2 Аналіз базових моделей фільтрації контенту та обмеження їх застосування в B2B-середовищі

Для реалізації етапу створення моделі рекомендацій (етап 3 на рис. 1.4) у сучасній практиці рекомендаційних систем, які використовуються для персоналізації контенту систем електронної комерції виділяють три основні методичні підходи: колаборативну фільтрацію, контентну фільтрацію та гібридні моделі. Розглянемо більш детально кожний підхід [24].

Рекомендаційні системи засновані на колаборативній фільтрації (Collaborative Filtering) базується на ідентифікації подібності між профілями користувачів або об'єктів на основі їхньої попередньої активності. Колаборативна фільтрація – це метод рекомендаційних систем, який базується на аналізі взаємодій між користувачами та предметами [25]. Він шукає схожості між користувачами та предметами, щоб зробити рекомендації. Основна ідея полягає в тому, що якщо два користувачі мають схожі смаки або інтереси, то їм, ймовірно, сподобається той же предмет. Рекомендації будуються виходячи з попередньої реакції користувача на предмети у системі. Це можуть бути як оцінки за якоюсь шкалою, так і просто інформація про користування певними предметами. При цьому, рекомендація надається з урахуванням поведінки інших користувачів системи.

Основа колаборативної фільтрації – матриця, де по одній осі розташовані користувачі, а по іншій – предмети [26]. У клітинках матриці міститься інформація про те, як той чи інший користувач оцінив певний предмет.

При цьому, відсутність значення означає, що користувач до цього моменту ніяк не взаємодіяв з предметом. Ціллю роботи системи є заповнення цих пустих клітинок матриці. Існує два підходи до заповнення пустих клітинок: на основі користувачів (User-Based Collaborative Filtering) та на основі предметів (Item-Based Collaborative Filtering).

Підхід колаборативної фільтрації на основі користувачів реалізовано наступним чином: серед усіх користувачів необхідно виділити n найбільш схожих на ключового та на основі їх наявних оцінок зробити прогноз, щодо відсутніх оцінок.

(рис. 1.5, а). Підхід колаборативної фільтрації на *основі предметів* реалізовано наступним чином: серед усіх предметів необхідно виділити *n* найбільш схожих на ключовий та на основі їх наявних оцінок зробити прогноз щодо відсутніх оцінок (рис. 1.5, б)

Тобто маємо матрицю оцінок між 4 користувачами та 4 предметами. За підходом *за користувачами* прогнозом оцінки Користувача 4 Предмету 2 є значення «3», а за підходом *за предметами* Предмет 2 найбільше схожий Предмет 4, тому прогнозоване значення оцінки дорівнює «2».

	Предмет 1	Предмет 2	Предмет 3	Предмет 4
Користувач 1	3	3	5	3
Користувач 2	1	5	2	5
Користувач 3			2	
Користувач 4	3	?	5	2

а

	Предмет 1	Предмет 2	Предмет 3	Предмет 4
Користувач 1	3	3	5	3
Користувач 2	1	5	2	5
Користувач 3			2	
Користувач 4	3	?	5	2

б

Рисунок 1.5 – Візуалізація матриці колаборативної фільтрації (а – за користувачами; б – за предметами)

Для розрахунків оцінок *схожості між користувачами або предметами* для подальшого передбачення застосовують міри подібності [27]. Серед найбільш популярних мір подібності виділяють:

– косинусну міру, що являє собою кут між векторами оцінок:

$$\cosSim(u_i, u_j) = \cos(u_i, u_j) = \frac{\sum_{k=1}^n u_{i,k} u_{j,k}}{\sqrt{\sum_{k=1}^n u_{i,k}^2} \sqrt{\sum_{k=1}^n u_{j,k}^2}}; \quad (1.1)$$

– скоригований косинус подібності:

$$adjustedCosSim(u_i, u_j) = \frac{\sum_{k=1}^n (u_{i,k} - \bar{u}_k)(u_{j,k} - \bar{u}_k)}{\sqrt{\sum_{k=1}^n (u_{i,k} - \bar{u}_k)^2} \sqrt{\sum_{k=1}^n (u_{j,k} - \bar{u}_k)^2}}; \quad (1.2)$$

– критерій кореляції Пірсона, який показує наскільки добре два набори даних лягають на пряму:

$$PearsonSimilarity(u_i, u_j) = \frac{\sum_{k=1}^n (u_{i,k} - \bar{u}_i)(u_{j,k} - \bar{u}_j)}{\sqrt{\sum_{k=1}^n (u_{i,k} - \bar{u}_i)^2} \sqrt{\sum_{k=1}^n (u_{j,k} - \bar{u}_j)^2}}; \quad (1.3)$$

– коефіцієнт Жаккара – міра спільної частини двох множин, поділена на міру об'єднання множин:

$$JaccardSimilarity(u_i, u_j) = \frac{|u_i \cap u_j|}{|u_i \cup u_j|}; \quad (1.4)$$

де u_i, u_j – вектори користувачів i та j ;

n – кількість предметів;

$\bar{u}_i$ – середні оцінки користувачів i та j ;

$\bar{u}_k$ – середня оцінка предмета k .

Колаборативна фільтрація може бути дуже ефективним методом рекомендацій, особливо коли у вас багато даних про взаємодію користувачів з предметами. Однак воно має обмеження «холодного старту» (*Cold start*) для нових користувачів або предметів і проблеми зі розрахунками до великих обсягів даних [28].

Необхідно зазначити, що у B2B-середовищі підхід колаборативної фільтрації стикається з проблемою розрідженості даних (data sparsity) [29]. Вузька спеціалізація та унікальність закупівельних стратегій B2B-клієнтів (підприємств) призводять до

того, що спільні патерни між різними клієнтами є мінімальними. Тобто пошук «схожого сусіда» (user-based) часто виявляється нерелевантним, оскільки потреби бізнес-клієнта детерміновані не ринковою модою, а індивідуальними технологічними циклами, що робить статистичне усереднення поведінки груп користувачів малоефективним.

Рекомендаційні системи засновані на контентній фільтрації (Content-based Filtering) спрямовані на аналіз характеристик самих товарів (SKU), які клієнт купував раніше, для формування пропозицій зі схожими атрибутами. Розглянемо більш загальну характеристику таких рекомендаційних систем.

Контентна фільтрація (*Content-based filtering*) – це стратегія персоналізації, яка базується на аналізі внутрішніх характеристик об'єктів та побудові профілю вподобань користувача на базі атрибутів товарів, з якими він взаємодіяв раніше [30].

Для глибшого розуміння цього підходу, розглянемо його ключові механізми та специфічні обмеження для B2B-сегмента. В основі контентної фільтрації лежить порівняння двох векторів (профілів). Перший вектор – *вектор характеристик об'єкта (Item Profile)* містить набір ознак $j \in I$, таких як категорія, бренд, технічні параметри, ціновий діапазон тощо. Другий вектор *профіль користувача (User Profile)* – це агрегований вектор ознак усіх об'єктів, які користувач u_i оцінив позитивно або придбав. Рекомендація формується шляхом розрахунку міри схожості наприклад, косинусної відстані за формулою (1.1) між профілем користувача профілями нових об'єктів. Якщо вектори близькі, система вважає, що товар буде релевантним виробу користувача.

Для реалізації фільтрування на основі контенту рекомендаційні системи використовують спеціалізовані методи [31]:

- методи обробки природної мови (NLP-аналіз) використовуються для семантичного аналізу текстового контенту, такого як технічні характеристики продуктів, відгуки користувачів або анотації до об'єктів [32]. Це дозволяє системі розуміти контекстну подібність товарів поза межами простого збігу назв;

- методи комп'ютерного зору (Computer Vision) застосовуються для аналізу візуального вмісту зображень і відео [33]. Вони дозволяють ідентифікувати об'єкти,

сцени або специфічні атрибути (наприклад, дизайн чи колірну гаму), що критично для галузей, де візуальна складова є визначальною (мода, архітектура);

- цифрова обробка та аналіз аудіосигналів (Audio Content Analysis) передбачає вилучення низькорівневих ознак аудіофайлів (темپ, спектральні характеристики) та високорівневих дескрипторів (жанр, настрій) для формування рекомендацій у медіа-сервісах;

- атрибутивний аналіз та інтелектуальна обробка метаданих (Metadata Analysis) базується на використанні структурованих характеристик предметів (категорії, теги, ключові слова, артикули). Це найбільш розповсюджений метод у B2B-системах, де товари класифікуються за чіткими ієрархічними довідниками та технічними специфікаціями»;

До переваг контентної фільтрації відносимо: відсутність проблеми «холодного старту» для товарів; прозорість (Explainability), тобто системі легко «пояснити» створену рекомендацію (наприклад, «ви отримали цю пропозицію, бо раніше купували шкіряне взуття цього бренду»); незалежність від інших користувачів, тобто на відміну від колаборативної фільтрації, рекомендації для конкретного клієнта не залежать від дій інших суб'єктів платформи.

Незважаючи на зазначені переваги контентної фільтрації та логічність використання, для систем B2B електронної комерції цей підхід до побудови рекомендацій має критичні недоліки.

Наприклад, контентна фільтрація аналізує *що* купує клієнт, але ігнорує динаміку *коли* він це робить [34]. У B2B, де закупівлі циклічні (наприклад, сезонне оновлення колекції взуття), система може пропонувати релевантний товар у момент, коли склад клієнта вже заповнений або коли бюджет ще не сформований. Такій недолік має назву «темпоральна сліпота».

Ще одним недоліком є обмеженість спеціалізації (*Overspecialization*). Тобто система схильна рекомендувати лише ті товари, які дуже схожі на вже придбані. Це створює «інформаційну бульбашку», заважаючи клієнту побачити нові супутні категорії, що важливо для розширення асортименту B2B-менеджерами.

Критичним для B2B є також ігнорування економічної ваги (*Utility*). За цим підходом однаково оцінюється схожість для дешевих витратних матеріалів та дорогого обладнання, якщо їхні описові атрибути збігаються. Він не враховує прибутковість або фінансову значущість транзакції для бізнесу.

Гібридні рекомендаційні системи (Hybrid Systems) – системи, які комбінують у собі різні підходи до рекомендацій з метою забезпечення більш точних і релевантних рекомендацій для користувачів [35]. Вони поєднують в собі переваги різних підходів, маючи на меті уникнути їхніх недоліків.

Наприклад гібридні системи можуть комбінувати колаборативну фільтрацію та фільтрацію на основі контенту, використовуючи як інформацію про взаємодії користувачів з предметами, так і характеристики самих предметів для забезпечення рекомендацій. Тобто система може враховувати як схожість між користувачами, так і схожість між предметами для здійснення рекомендацій. Але, незважаючи на комбінування підходів використання гібридних рекомендаційних систем як надбудови над B2B-системою електронної комерції також має недоліки ігнорування економічної цінності та стабільності циклу транзакцій. А висока складність моделі не гарантує раціональності рекомендації для бізнес-клієнта. Таким чином, аналіз базових технологій персоналізації контенту за допомогою базових технологій рекомендаційних систем виявив системний дефіцит інструментів для врахування ціннісно-темпоральної специфіки B2B-транзакцій [36]. Це обґрунтовує необхідність використання методів Data Mining (виявлення асоціацій та послідовностей), що буде розглянуто далі.

1.4 Аналіз існуючих методів пошуку асоціацій та послідовностей

1.4.1 Базові визначення аналізу транзакційно-поведінкових даних

Для формалізації задачі виявлення прихованих закономірностей у B2B-системах необхідно визначити базовий математичний апарат аналізу транзакційних масивів [37].

В теорії транзакція T – це набір об’єктів (товарів, послуг, дій), які відбулися одночасно або в межах одного комерційного акту. Прикладом транзакції у B2C є один чек у супермаркеті чи одна корзина на сайті. У B2B системі транзакція T це:

- одне оптове замовлення від компанії-клієнта (рахунок-фактура);
- набір товарів, замовлених у межах одного сезону закупівель;
- комплекс дій користувача-закупника (наприклад, перегляд каталогу, запит на індивідуальну ціну, оформлення пробної партії) в межах одного переговорного циклу.

Визначення 1.1. *Транзакція (Transaction)*. Формально множина товарних позицій I визначається як :

$$I = \{i_1, i_2 \dots i_n\}, \quad (1.5)$$

де i_n – ідентифікатор товарної позиції (SKU), яка доступна для замовлення в системі електронної комерції.

Тоді *транзакція* T – це *логічна одиниця роботи*, що відбувається в один момент часу t , яка має унікальний ідентифікатор TID , є підмножиною I , тобто $T \subseteq I$ та відображає факт одноразової закупівлі m товарів клієнтом (рис. 1.6, а) [38]:

$$T = \{i_1, i_2 \dots i_m\}. \quad (1.6)$$

База транзакцій D складається з набору таких записів (рис. 1.6, б).

Визначення 1.2. *Послідовність (Sequence)*. Формально *послідовність* s_j в системі електронної комерції – це впорядкований за часом журнал активності одного клієнта (або Buying Center), що складається з набору транзакцій T , які відбулися в різні моменти часу [39]:

$$s_j = \langle T_{t_1}, T_{t_2}, \dots, T_{t_k} \rangle \quad (1.7)$$

де T_{t_i} – це B2B-транзакція (1.6), яка відбулася в момент часу t_i , при цьому маємо впорядкування за часом $t_1 < \dots < t_k$.

Аналіз послідовностей дозволяє досліджувати динаміку зміни споживчих інтересів у часі (рис. 1.6, в). На рисунку 1.6 показні рівні представлення транзакційних даних:

а) База транзакцій D із зазначенням складу замовлень;

б) Матричне представлення транзакцій та розрахунок базової підтримки $\text{Supp}(i_k)$;

в) Формування бази послідовностей S за ідентифікаторами клієнтів (CID).

TID (ID чека)	CID (ID клієнта)	Дата	Склад транзакції (T)
1	C-101	2026-05-01	{A, B}
2	C-102	2026-05-02	{B, D}
3	C-103	2026-05-15	{A, C}
4	C-104	2026-05-16	{A, B, E}
5	C-105	2026-05-20	{D}

а

Елементи I_k	i_1	i_2	i_3	i_4	i_5
TID/Назва	A	B	C	D	E
1	1	1	0	0	0
2	0	1	0	1	0
3	1	0	1	0	0
4	1	1	0	0	1
5	0	0	0	1	0
$\text{Supp}(i_k)$	60 %	60 %	20 %	40 %	20 %

б

CID (ID клієнта)	Послідовність (S)	Склад (з прив'язкою до TID)
C-101	<{A, B}, {A, C}, {D}>	Транзакції 1, 3, 5
C-102	<{B, D}>	Транзакція 2
C-103	<{A, B, E}>	Транзакція 4

в

Рисунок 1.6 – Рівні представлення транзакційних даних

Асоціативне правило (АП) (Association Rule AR) в B2B відображає одночасну присутність товарів, послуг або дій у межах однієї транзакції $T(1)$. Нехай I – множина всіх елементів (SKU/послуг/дій), а A та B – два набори елементів, такі що $A \subset I$, $B \subset I$, $A \cap B = \emptyset$. Асоціативне правило формулюється наступним чином :

$$AR = A \rightarrow B. \quad (1.8)$$

Тобто, *якщо* в оптовому замовленні (транзакції) присутній набір елементів A , *тоді* існує висока ймовірність того, що в цьому ж замовленні є набір елементів B .

Послідовний шаблон (ПШ) (Sequential Pattern SP) у B2B відображає впорядковану за часом появу транзакцій або поведінкових подій (3), що формують ланцюжок активності клієнта. Нехай S – множина всіх B2B-послідовностей клієнтів s_j (3), α та β – дві послідовності транзакцій або подій $\alpha = \langle T_{t_a} \rangle$, $\beta = \langle T_{t_b} \rangle$. Послідовний шаблон формулюється у вигляді :

$$SP = \alpha \Rightarrow \beta. \quad (1.9)$$

Тобто, *якщо* в момент часу t_a клієнт здійснив подію або набір транзакцій α , *тоді* існує висока ймовірність того, що в наступний момент часу t_b , $t_a < t_b$ він здійснить подію або набір транзакцій β . Позначка « $\Rightarrow$ » використовується, що підкреслити саме часову послідовність. Визначимо метрики якості, які характеризують зв'язок між наборами товарів, встановлений за допомогою побудованих АП під час проведення афінитивного аналізу.

Підтримка (*support*) показує, як часто набір $\{A \cup B\}$ з'являється у загальній базі B2B-транзакцій D . Тоді для АП $A \rightarrow B$ підтримка *support* $S_{AR}(A \rightarrow B)$ виглядає :

$$S_{AR}(A \rightarrow B) = S_{AR}(A \cup B) = \frac{|\{T \in D | A \cup B \subseteq T\}|}{|D|}, \quad (1.10)$$

де $|\{T \in D | A \cup B \subseteq T\}|$ – загальна кількість транзакцій які містять A та B ;
 $|D|$ – загальна кількість транзакцій в базі B2B.

Достовірність (confidence) АП $A \rightarrow B$ показує ймовірність того, що якщо клієнт купив A , він також купить B у цій же транзакції. Це пряма міра прогностичної цінності набору A для продажу B в межах одного оптового замовлення D . Тоді для АП $A \rightarrow B$ *confidence* $C_{AR}(A \rightarrow B)$ обчислюється наступним чином :

$$C_{AR}(A \rightarrow B) = \frac{S_{AR}(A \cup B)}{S_{AR}(A)}. \quad (1.11)$$

Для В2В-систем оцінка якості (*Support, Confidence*) АП має враховувати не лише кількість транзакцій $|D|$, а й сумарну фінансову вартість або обсяг елементів A та B у цих транзакціях, тобто буди зваженою (*Weighted*) згідно визначення (3).

Ліфт (Lift) показує, наскільки часто набір $\{A \cup B\}$ з'являється частіше, ніж очікувалося, якщо A та B були незалежними [40]. Тоді для $A \rightarrow B$, метрика *Lift* $L_{AR}(A \rightarrow B)$ визначається наступним чином :

$$L_{AR}(A \rightarrow B) = \frac{C_{AR}(A \rightarrow B)}{S_{AR}(A)} = \frac{S_{AR}(A \cup B)}{S_{AR}(A) \cdot S_{AR}(B)} \quad (1.12)$$

Якщо $L_{AR} > 1$, тоді існує позитивна кореляція (раціональний сенс) у спільній купівлі; якщо $L_{AR} < 1$, тоді існує негативна кореляція (товари заміни). Це ключова метрика для В2В, оскільки вона виключає тривіальні правила, які є частими просто через високу загальну популярність одного із товарів.

Лeverедж (Leverage) вимірює різницю між фактичною частотою спільної появи наборів A та B і тією частотою, яка очікувалася б, якби A та B були повністю незалежними один від одного. Тоді для $A \rightarrow B$, метрика *Leverage* $Le_{AR}(A \rightarrow B)$ визначається наступним чином [41]:

$$Le_{AR}(A \rightarrow B) = S_{AR}(A \cup B) - S_{AR}(A) \cdot S_{AR}(B). \quad (1.13)$$

Якщо $Le_{AR} = 0$, тоді елементи A та B є незалежними. Їхня спільна поява в замовленнях є суто випадковою. Якщо $Le_{AR} > 0$, між елементами існує логічний

сильний зв'язок, тобто вони з'являються разом частіше, ніж очікувалося. Це вказує на сильний логічний зв'язок. У В2В це сигнал для створення бандлів або спільної логістики. Якщо $Le_{AR} < 0$, тоді елементи з'являються разом рідше, ніж очікувалося. Це може свідчити про те, що товари є конкурентами або взаємозамінними.

Конвікція (Conviction) показує, як часто АП (5) дає неправильні прогнози. Вимірює залежність набору A від набору B . Тоді для $A \rightarrow B$, метрика *Conviction* $V_{AR}(A \rightarrow B)$ обчислюється наступним чином :

$$V_{AR}(A \rightarrow B) = \frac{1 - S_{AR}(B)}{1 - C_{AR}(A \rightarrow B)} \quad (1.14)$$

Високе значення V_{AR} вказує на те, що правило дуже сильне, і набір A справді «примушує» з'являтися B . Це корисно для фільтрації правил, коли A є дуже популярним.

Визначимо метрики якості, які характеризують зв'язок між послідовними шаблонами SP (6). Вони застосовуються до бази послідовностей S , коли подія β відбувається після події α .

Метрика *Підтримка (support)* для SP обчислюється наступним чином:

$$S_{SP}(\alpha \Rightarrow \beta) = \frac{|\{s \in S | \alpha \text{ передуює } \beta \text{ у послідовності } s\}|}{|S|}. \quad (1.15)$$

Метрика *Достовірність (confidence)* SP обчислюється наступним чином:

$$C_{SP}(\alpha \Rightarrow \beta) = \frac{S_{SP}(\alpha \Rightarrow \beta)}{S_{SP}(\alpha)}. \quad (1.16)$$

Метрика *Ліфт (Lift)* для SP обчислюється наступним чином:

$$L_{SP}(\alpha \Rightarrow \beta) = \frac{S_{SP}(\alpha \Rightarrow \beta)}{S_{SP}(\alpha) \cdot S_{SP}(\beta)}. \quad (1.17)$$

Метрика *Лeverедж* (*Leverage*) для SP обчислюється наступним чином:

$$Le_{SP}(\alpha \Rightarrow \beta) = S_{SP}(\alpha \Rightarrow \beta) - S_{SP}(\alpha) \cdot S_{SP}(\beta). \quad (1.18)$$

Метрика *Конвікція* (*Conviction*) для SP обчислюється наступним чином:

$$V_{SP}(\alpha \Rightarrow \beta) = \frac{1 - S_{SP}(\beta)}{1 - C_{SP}(\alpha \Rightarrow \beta)}. \quad (1.19)$$

1.4.2 Компаративний аналіз класичних методів пошуку асоціативних правил

Пошук асоціативних правил (*Association Rule Mining, ARM*) є фундаментальною задачею інтелектуального аналізу даних, що спрямована на виявлення сталих зв'язків між об'єктами у великих базах транзакцій D [42].

Основним завданням методів цього класу є знаходження всіх наборів предметів (itemsets), чия підтримка $\text{Support } S_{AR}(A \rightarrow B)$ (1.10) перевищує заданий поріг min_supp .

Фундаментальний інструментарій для пошуку таких частотних наборів (*Itemset Mining*) базується на трьох ключових підходах, що базуються на різних стратегіях обходу простору станів. Зокрема це відомий алгоритм Apriori, який здійснює перебір кандидатів з подальшим відкиданням непопулярних наборів [43], алгоритм FP-Growth, який будує FP-Tree для подальшого пошуку без генерації кандидатів [44] та алгоритм Eclat, який шукає популярні набори через перетин попередньо побудованих TID-списків [45].

1. *Алгоритм Apriori (ітеративний підхід)*. Apriori (апріорная приналежність, apriori property) застосовується до баз даних транзакцій. Транзакційна або операційна БД (Transaction database) (ТБД) являє собою двовимірну таблицю, яка складається з номерів транзакції (TID) та переліку товарів, які входять до складу цієї транзакції (див. рис. 1.6, а, б). Алгоритм працює в два етапи:

- пошук всіх наборів елементів, підтримка яких більше або дорівнює $\min_supp$, так званих *frequent itemset*;
- розробка *ARM* на основі популярних наборів з достовірністю більшою або рівною $\min_conf$.

Пошук *frequent itemset* виконується на основі перебору всіх можливих наборів елементів із ТБД потребує $O(2^{|I|})$ операцій, де $|I|$ – кількість елементів, при цьому із зростанням числа елементів в I , число потенційних наборів елементів зростає експоненційно. Для зниження розмірності простору пошуку використовують властивість антимонотонності підтримки наборів елементів, що полягає в тому, що підтримка будь-якого набору елементів не може перевищувати мінімальної підтримки будь-якого з його піднаборів. Якщо всі можливі набори елементів з I можна представити у вигляді решітки зв'язків, яка починається з порожнього набору, потім на 1 рівні розташовуються 1-елементні набори, на 2-му – 2-елементні і т.д. На k -му рівні представлені k -елементні набори, пов'язані з усіма своїми $(k - 1)$ -елементними піднаборами (рис. 1.7).

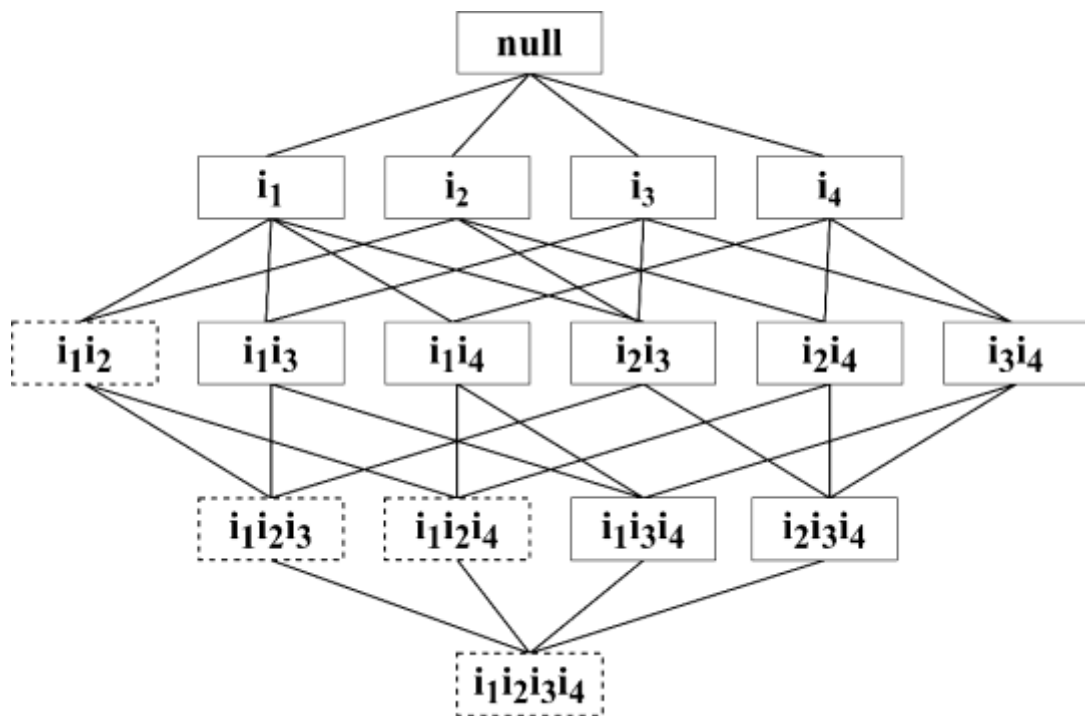


Рисунок 1.7 – Решітка зв'язків можливих наборів елементів з ТБД для $I = \{i_1, i_2, i_3, i_4\}$

Припустимо, що набір з елементів $\{i_1 i_2\}$ згідно (1.10) має підтримку нижче заданого порогу та, відповідно, не є популярним. Тоді, відповідно до властивості антимонотонності, всі його піднабори також не є популярними і відкидаються. На рисунку 1 – це частина решітки, починаючи з $\{i_1 i_2\}$. Таким чином, будь k -елементний набір буде популярним тоді і тільки тоді, коли всі його $(k - 1)$ -елементні піднабори будуть популярними

Розробка *ARM* на основі знайдених популярних наборів *frequent itemset* виконується після розрахунку підтримки та достовірності, використовуючи (1.10) і (1.11) для всіх імплікацій типу $A \rightarrow B$. При цьому в якості A використовуються всі можливі популярні та непорожні $(k - 1)$ -елементні піднабори $itemset_{k-1}$ популярного k -елементного набору $itemset_k$. В якості B використовується різниця R між $itemset_k$ і всіма $itemset_{k-1}$. Наприклад (див. рис. 1.7), для набору $itemset_3 = \{i_1 i_3 i_4\}$ піднабором будуть $itemset_2 = \{\{i_1 i_3\}, \{i_1 i_4\}, \{i_3 i_4\}\}$, а різницями $R = \{\{i_4\}, \{i_3\}, \{i_1\}\}$ відповідно, тоді всі імплікації $A \rightarrow B$ будуть мати вигляд, як $A \rightarrow B = \{(\{i_1 i_3\} \rightarrow \{i_4\}), (\{i_1 i_4\} \rightarrow \{i_3\}), (\{i_3 i_4\} \rightarrow \{i_1\})\}$. При цьому імплікація $A \rightarrow B$ буде відноситися до АП тоді і тільки тоді, коли $S_{AR}(A \rightarrow B) > \min_supp$ и $C_{AR}(A \rightarrow B) > \min_conf$.

Таким чином на основі властивості антимонотонності алгоритм використовує стратегію пошуку в ширину (BFS) та ітеративно генерує кандидатів довжиною $k + 1$ на основі частих наборів довжиною k . Із цього висновку випливає обмеження використання *Apriori* в B2B-системі електронної комерції, а саме :

- проблема генерації кандидатів, тобто на розріджених даних (*sparse data*), характерних для B2B (див. рис. 1.7), кількість потенційних кандидатів зростає експоненціально, що призводить до надмірного споживання пам'яті;

- необхідність багаторазове сканування бази, тобто для перевірки кожного рівня кандидатів алгоритм потребує повного сканування бази транзакцій D , що є недопустимим для систем з великими даними.

2. Алгоритм *FP-Growth* (підхід на основі деревоподібних структур). Оригінальна назва *FP-Growth* це алгоритм зростання частих шаблонів (*Frequent*

Pattern Growth). Поняття Frequent Pattern, частий шаблон або частий набір у контексті аналізу даних – це часто повторювані комбінації елементів у транзакціях. Growth – зростання, розширення. Тобто алгоритм FP-Growth поступово будує (вирощує) часті шаблони на основі FP-дерева. Основна ідея алгоритму полягає в тому, що замість повного перебору всіх можливих комбінацій товарів (Apriori), FP-Growth будує компактну структуру даних FP-дерево (Frequent Pattern Tree) та на його основі рекурсивно шукає часті шаблони [46]. Основні етапи роботи FP-Growth:

1) обчислення підтримки кожного товару. Проходимо всі транзакції (див. рис. 1.6). Визначаємо підтримку для кожного товару $\text{Support } S_{AR}(A \rightarrow B)$ (1.10). Відкидаємо товари, які не проходять min_supp ;

2) побудова FP-дерева. Упорядковуємо товари в кожній транзакції за спаданням підтримки. Проходимо транзакції одну за одною. Додаємо товари як вузли в дерево. Однакові префікси транзакцій об'єднуємо в спільні гілки. До вузлів додаємо лічильники підтримки. В результаті маємо компакте FP-дерево, що зберігає всі транзакції у вигляді префіксів;

3) побудова умовних баз та умовних FP-дерев. Для кожного товару (знизу вгору) формуємо умовну базу частих наборів (*frequent itemset*) – список усіх шляхів у дереві, що містять цей товар. Для кожної умовної бази будуємо умовне FP-дерево. Рекурсивно повторюємо процес;

4) пошук частих наборів. Комбінуємо результати умовних дерев. Збираємо всі часті набори, підтримка яких $\geq \text{min_supp}$.

Розглянемо приклад бази, яка складається із п'яти умовних транзакцій. Результати пошуку асоціативних правил (частих наборів) за алгоритмом *FP-Growth* наведені в таблиці 1.2

На кроці 4 на основі умовних FP-дерев, побудованих для кожного товару на кроці 3 знаходимо всі часті набори, які містять цей товар. Знайдені часті набори виглядають наступним чином: $\{E, B\}$ 2; $\{D, B\}$ 3; $\{D, B, A\}$ 2; $\{C, B\}$ 3; $\{C, B, A\}$ 2; $\{A, B\}$ 3. Таким чином, за *FP-Growth* відбувається відмова від явної генерації кандидатів на користь стисненого представлення даних. Тобто дані з бази транзакцій (див. рис 1.6) імпортуються у компактну структуру *FP-дерево* (Frequent Pattern Tree).

Пошук частих наборів здійснюється шляхом рекурсивного розбиття бази на умовні бази (conditional databases). Проте даний підхід має суттєві обмеження для використання в B2B-системах електронної комерції, а саме:

- нерелевантність частотних лічильників (Frequency-Utility Trade-off) є основним обмеженням FP-Growth тому, що дерево будується виключно на основі лічильників частоти появи предметів у транзакціях. У вузлах дерева зберігається лише кількість входжень товару, що повністю нівелює інформацію про його вартість чи обсяг закупівлі. Як наслідок, стратегічно важливі, але рідкісні B2B-товари відсікаються на етапі побудови дерева, тоді як дешеві супутні товари з високою частотою стають основою моделі;

- складність оновлення дерева, тобто при появі нових транзакцій FP-дерево часто потребує повної перебудови, що ускладнює роботу в режимі реального часу;

- високі вимоги до оперативної пам'яті, тобто хоча алгоритм працює швидше за Apriori, але при великій кількості унікальних SKU (як у випадку широкого асортименту B2B-платформи) дерево може не поміститися в оперативну пам'ять.

3. *Алгоритм Eclat (вертикальний підхід)*. Eclat (*Equivalence Class Transformation*, перетворення класів еквівалентності) це алгоритм для пошуку частих наборів товарів у транзакційних даних, який працює через перетини списків ідентифікаторів транзакцій, на відміну від перебору кандидатів у Apriori чи побудови дерев у FP-Growth. Назва алгоритму *Eclat* відображає його здатність трансформувати базу даних у вертикальний формат та декомпонувати простір пошуку на незалежні класи еквівалентності на основі спільних префіксів товарних наборів. Основна ідея алгоритму полягає в тому щоб зберігати для кожного товару *список ідентифікаторів транзакцій* (TID-листи), де цей товар зустрічається, і знаходити часті набори шляхом перетину цих списків. Основні етапи роботи Eclat:

- побудова TID-списків. Для товару визначаємо всі транзакції, де він присутній;
- рекурсивне злиття TID-списків. Перетинаємо списки для комбінацій товарів, тобто перетин $TID(A) \cap TID(B) = TID(A, B)$. Підраховуємо підтримку для кожного перетину. Якщо підтримка $\geq \min_supp$ то перетин зберігаємо;

– формування частих наборів. Рекурсивно об'єднуємо товари та їх TID-списки, поки є часті комбінації.

Таблиця 1.2 – Результати пошуку асоціативних правил алгоритму *FP-Growth*

Результати кроків 1 та 2 (розрахунок підтримки та побудова <i>FP-дерева</i>)			
Транзакція Товари	Розрахунок S_{AR} $\min_supp=2$	Впорядкування за S_{AR}	FP-дерево
T1A, B, D, E T2B, C, D T3A, B, C, E T4B, E T5A, B, C, D	$A \rightarrow 3$ $B \rightarrow 5$ $C \rightarrow 3$ $D \rightarrow 3$ $E \rightarrow 3$	T1B, A, D, E T2B, C, D T3B, A, C, E T4B, E T5B, A, C, D	<pre> (корінь) ├── B (5) │ ├── A (3) │ │ ├── D (1) │ │ └── C (2) │ │ ├── E (1) │ │ └── D (1) │ ├── C (1) │ │ └── D (1) │ └── E (1) </pre>
Результати кроку 3 (<i>Побудова умовних баз та умовних FP-дерев</i>)			
Порядок $B (5) > A (3) > C (3) > D (3) > E (3)$			
Шляхи до вузла <i>E</i>	Шляхи до вузла <i>D</i>	Шляхи до вузла <i>C</i>	Шляхи до вузла <i>A</i>
<pre> (корінь) ├── B (5) │ ├── A (3) │ │ ├── D (1) │ │ └── C (2) │ │ ├── E (1) │ │ └── D (1) │ ├── C (1) │ │ └── D (1) │ └── E (1) </pre>	<pre> (корінь) ├── B (5) │ ├── A (3) │ │ ├── D (1) │ │ └── C (2) │ │ ├── E (1) │ │ └── D (1) │ ├── C (1) │ │ └── D (1) │ └── E (1) </pre>	<pre> (корінь) ├── B (5) │ ├── A (3) │ │ ├── D (1) │ │ └── C (2) │ │ ├── E (1) │ │ └── D (1) │ ├── C (1) │ │ └── D (1) │ └── E (1) </pre>	<pre> (корінь) ├── B (5) │ ├── A (3) │ │ ├── D (1) │ │ └── C (2) │ │ ├── E (1) │ │ └── D (1) │ ├── C (1) │ │ └── D (1) │ └── E (1) </pre>
Умовна БД (<i>E</i>)	Умовна БД (<i>D</i>)	Умовна БД (<i>C</i>)	Умовна БД (<i>A</i>)
$\{B, A, C\} : 1$ $\{B\} : 1$	$\{B, A\} : 1$ $\{B, C\} : 1$ $\{B, A, C\} : 1$	$\{B, A\} : 2$ $\{B\} : 1$	$\{B\} : 3$
<i>FP-дерево</i> (<i>E</i>)	<i>FP-дерево</i> (<i>D</i>)	<i>FP-дерево</i> (<i>C</i>)	<i>FP-дерево</i> (<i>A</i>)
<pre> (корінь) ├── B (2) </pre>	<pre> корінь ├── B (3) │ ├── A (2) │ │ └── C (1) │ └── C (1) </pre>	<pre> (корінь) ├── B (3) │ └── A (2) </pre>	<pre> (корінь) ├── B (3) </pre>

Розглянемо попередній приклад бази, яка складається із п'яти умовних транзакцій. Результати пошуку асоціативних правил (частих наборів) за алгоритмом *Eclat* наведені в таблиці 1.3.

Таблиця 1.3 – Результати пошуку асоціативних правил за кроками алгоритму *Eclat*

Результати кроку 1. Побудова <i>TID</i> -списків			
Транзакція Товари	TID-списки для кожного товару		
T1A, B, D, E	Товар	TID-список	Підтримка
T2B, C, D	A	{T1, T3, T5}	3
T3A, B, C, E	B	{T1, T2, T3, T4, T5}	5
T4B, E	C	{T2, T3, T5}	3
T5A, B, C, D	D	{T1, T2, T5}	3
	E	{T1, T3, T4}	3
Результати кроку 2. Рекурсивне злиття TID-списків (min_support = 2)			
Перетин TID-списків для пар товарів	Перетин TID-списків для трійок товарів		
$A \& B \rightarrow \{T1, T3, T5\} \cap \{T1, T2, T3, T4, T5\} = \{T1, T3, T5\} \rightarrow 3$ $A \& C \rightarrow \{T1, T3, T5\} \cap \{T2, T3, T5\} = \{T3, T5\} \rightarrow 2$ $A \& D \rightarrow \{T1, T3, T5\} \cap \{T1, T2, T5\} = \{T1, T5\} \rightarrow 2$ $A \& E \rightarrow \{T1, T3, T5\} \cap \{T1, T3, T4\} = \{T1, T3\} \rightarrow 2$ $B \& C \rightarrow \{T1, T2, T3, T4, T5\} \cap \{T2, T3, T5\} = \{T2, T3, T5\} \rightarrow 3$ $B \& D \rightarrow \{T1, T2, T3, T4, T5\} \cap \{T1, T2, T5\} = \{T1, T2, T5\} \rightarrow 3$ $B \& E \rightarrow \{T1, T2, T3, T4, T5\} \cap \{T1, T3, T4\} = \{T1, T3, T4\} \rightarrow 3$ $C \& D \rightarrow \{T2, T3, T5\} \cap \{T1, T2, T5\} = \{T2, T5\} \rightarrow 2$ $C \& E \rightarrow \{T2, T3, T5\} \cap \{T1, T3, T4\} = \{T3\} \rightarrow 1$ $D \& E \rightarrow \{T1, T2, T5\} \cap \{T1, T3, T4\} = \{T1\} \rightarrow 1$	$A, B, C \rightarrow \{T1, T3, T5\} \cap \{T2, T3, T5\} \cap \{T1, T2, T3, T4, T5\} = \{T3, T5\} \rightarrow 2$ $A, B, D \rightarrow \{T1, T3, T5\} \cap \{T1, T2, T5\} \cap \{T1, T2, T3, T4, T5\} = \{T1, T5\} \rightarrow 2$ $B, C, D \rightarrow \{T2, T3, T5\} \cap \{T1, T2, T5\} \cap \{T1, T2, T3, T4, T5\} = \{T2, T5\} \rightarrow 2$		

Далі наведено детальний приклад операції злиття. Знаходимо всі транзакції, де одночасно присутні всі товари з певної комбінації. Наприклад. Якщо $TID(A) = \{T_1, T_3, T_5\}$, а $TID(B) = \{T_1, T_2, T_3, T_4, T_5\}$, тоді $TID(A) \cap TID(B) = \{T_1, T_3, T_5\}$. Це

означає, що товари A і B одночасно купували в транзакціях T_1 , T_3 і T_5 . Якщо додаємо третій товар $TID(C) = \{T_2, T_3, T_5\}$ то для трьох множин маємо $TID(A) \cap TID(B) \cap TID(C) = \{T_3, T_5\}$. Тобто маємо транзакції, де одночасно присутні товари A , B , C .

Враховуючи результати рекурсивного злиття TID-списків та $\min_support = 2$, сформовано часті набори, які виглядають: $\{A, B\}$ 3; $\{A, C\}$ 2; $\{A, D\}$ 2; $\{A, E\}$ 2; $\{B, C\}$ 3; $\{B, D\}$ 3; $\{B, E\}$ 3; $\{C, D\}$ 2; $\{A, B, C\}$ 2; $\{A, B, D\}$ 2; $\{B, C, D\}$ 2.

Таким чином, на відміну від попередніх методів, Eclat використовує вертикальне представлення бази транзакцій (tid-lists). Тобто для кожного товару $i \in I$ зберігається список ідентифікаторів транзакцій TID , у яких він присутній.

Пошук частих наборів здійснюється через операцію перетину множин (set intersection). Наприклад, щоб знайти підтримку набору $\{A, B\}$, алгоритму достатньо знайти перетин списків $TID(A)$ та $TID(B)$. Використання алгоритму має переваги при використанні в системах електронної комерції, такі як швидкість розрахунку підтримки та ефективність на вузьких вибірках коли потрібно знайти асоціації для конкретного переліку стратегічних клієнтів.

Разом з тим обмеження його використання для аналізу транзакцій B2B клієнтів полягають в надмірному використанні пам'яті при побудові довгих списки TID та ігноруванні кількісних та вартісних показників транзакції.

Результати порівнянь розглянутих алгоритмів пошуку асоціативних правил наведені в таблиці 1.4.

1.4.3 Компаративний аналіз класичних методів пошуку послідовних шаблонів

На відміну від аналізу асоціативних правил, що розглядає транзакції як незалежні події, пошук послідовних патернів (*Sequential Pattern Mining, SPM*) спрямований на виявлення кореляцій між закупівлями з урахуванням фактору часу [47]. Основним завданням методів цього класу є знаходження впорядкованих наборів подій (sequences) що визначено за формулою (1.7), чия підтримка $S_{SP}(\alpha \Rightarrow \beta)$ визначається за формулою (1.15) та у базі послідовностей перевищує заданий поріг

min_supp. Як було зазначено у підрозділі 1.4.1, одиницею аналізу тут виступає історія активності конкретного клієнта (див. рис. 1.5, б).

Таблиця 1.4 – Результати аналізу алгоритмів пошуку асоціацій у контексті B2B

Критерій порівняння	Apriori	FP-Growth	Eclat
Формат даних	Горизонтальний	Деревоподібний	Вертикальний (TID-lists)
Стратегія пошуку	В ширину (BFS)	В глибину (DFS)	В глибину (DFS)
Основна операція	Сканування бази	Рекурсивний обхід	Перетин множин
Генерація кандидатів	Так (надлишкова)	Ні	Ні
Ефективність на розріджених даних	Низька (багато нулів)	Середня	Висока
Критичне обмеження для B2B	Експоненціальний ріст: велика кількість SKU призводить до обчислювального колапсу.	Частотна пастка: ігнорує ціну товару; вузли дерева зберігають лише лічильники входжень.	Пам'ять: при великій кількості транзакцій вертикальні списки переповнюють RAM.
Врахування Utility (прибутку)	Відсутнє	Відсутнє	Відсутнє

Фундаментальний інструментарій для пошуку таких патернів еволюціонував паралельно з методами ARM, адаптуючи три ключові підходи до обробки часових ланцюжків. Так алгоритм GSP (Generalized Sequential Pattern) є розвитком ідеології Apriori [48]. Він базується на властивості антимонотонності та використовує ітеративну генерацію кандидатів-послідовностей із подальшим скануванням бази. Алгоритм PrefixSpan представляє групу методів зростання патернів (як і FP-Growth) [49]. Він відмовляється від генерації кандидатів на користь рекурсивного побудови проєкцій бази даних для кожного частого префікса. Алгоритм SPADE [50] за аналогією до Eclat, використовує вертикальний формат представлення даних, де замість TID-списків застосовуються ID-lists із часовими мітками, а пошук здійснюється через операції перетину множин.

Тому на відміну від підрозділу 1.4.2, де було детально розглянуто механізми генерації асоціативних правил, матеріал даного підрозділу наведено у стислому варіанті, із фокусом на принципових відмінностях алгоритмів та їхній здатності обробляти динамічні дані.

Об'єктом аналізу тут виступає не окрема транзакція, а послідовність S (1.7), тобто впорядкований за часом журнал активності клієнта, що об'єднує набір транзакцій (див. рис. 1.6). Для B2B-платформ такий підхід дозволяє перейти від аналізу статичного «кошика» до прогнозування циклічності закупівель. Стисло розглянемо основні класичні методи у цій сфері:

1. *Алгоритм GSP (Generalized Sequential Pattern)*. Є прямим розвитком ідеології Apriori для часових послідовностей. Також ітеративно генерує кандидатів довжиною $k + 1$, з'єднуючи часті послідовності довжиною k . При аналізі довгих клієнтських історій (див. рис. 1.6, а, в) кількість кандидатів стає надмірною, що робить алгоритм занадто повільним для систем з великою кількістю CID.

2. *Алгоритм PrefixSpan (Prefix-projected Sequential Pattern mining)*. Вважається найбільш ефективним серед класичних методів, оскільки відмовляється від генерації кандидатів на користь методу проектування баз даних. PrefixSpan є логічним продовженням ідеї FP-Growth для ситуацій, де порядок подій має значення. FP-Growth стискає базу транзакцій в дерево, щоб не генерувати кандидатів. PrefixSpan використовує схожу ідею «стиснення», але замість побудови одного величезного дерева він створює «проекції» бази даних для кожного префікса. Тобто згідно з алгоритмом будуються локальні проекції бази для кожного частого товару або набору («префікса»), що дозволяє швидко виявляти наступні елементи в ланцюжку. Основою побудови обох алгоритмів FP-Growth та PrefixSpan є принцип «розділяй і володарюй» (Pattern-Growth). PrefixSpan є логічним продовженням ідеї FP-Growth для ситуацій, де порядок подій має значення. Для B2B алгоритм виявляє лише *порядок* подій, але повністю ігнорує реальний часовий інтервал (IPT) між ними.

3. *Алгоритм SPADE (Sequential Pattern Discovery using Equivalence classes)*. Використовує вертикальне представлення даних (аналогічно алгоритму Eclat). Базується на операціях перетину списків ідентифікаторів транзакцій та часових міток.

Але вимагає значних обсягів RAM для зберігання вертикальних структур при великій кількості клієнтів та SKU. Наведемо результати порівнянь розглянутих алгоритмів пошуку послідовних шаблонів в таблиці 1.5.

Таблиця 1.5 – Результати аналізу алгоритмів пошуку послідовних шаблонів у контексті B2B

Критерій порівняння	GSP	PrefixSpan	SPADE
Принцип дії	Генерація кандидатів	Проектування баз даних	Вертикальний аналіз
Ефективність на розріджених даних	Низькая	Високая	Середня
Врахування прибутку (Utility)	Відсутнє	Відсутнє	Відсутнє
Врахування часового інтервалу (IPT)	Відсутнє	Відсутнє	Відсутнє
Темпоральна сліпота (часова розірваність)	Наявна (ігнорує інтервали)	Наявна (фіксує лише порядок)	Наявна (обробляє лише мітки)
Ідентифікаційна дискретність	Висока	Середня (через префікси)	Середня

Таким чином існуючі методи SPM (Sequential Pattern Mining) є архітектурними спадкоємцями методів ARM (Association Rule Mining). Вони успадкували сильні сторони (швидкість, роботу з пам'яттю), але, на жаль, успадкували і головні недоліки ігнорування фінансової цінності (Utility) та точних часових інтервалів (IPT) [51].

Незважаючи на здатність алгоритмів SPM (PrefixSpan, SPADE) виявляти порядок подій, їм властива «темпоральна сліпота» статичних моделей. Вона проявляється у тому, що методи фіксують лише факт слідування однієї події за іншою, ігноруючи реальну тривалість інтервалів між ними. Для B2B-систем така «сліпота» є критичною, оскільки вона не дозволяє відрізнити регулярні закупівлі від епізодичних, що робить неможливим точне прогнозування дати наступного замовлення.

Наявність ідентифікаційної дискретності означає, що алгоритми розглядають кожну транзакцію клієнта як ізольовану подію, та не здатні враховувати накопичувальний досвід взаємодії з конкретним суб'єктом господарювання. В умовах B2B, де ідентифікація клієнта за CID є ключем до розуміння його потреб, такий дискретний підхід призводить до втрати контексту довгострокових відносин. Середній рівень ідентифікаційної дискретності алгоритмів PrefixSpan та SPADE зумовлений тим, що попри успішне групування транзакцій за ідентифікатором замовника, ці методи продовжують розглядати клієнта як абстрактну послідовність символів. Вони ігнорують внутрішню логіку суб'єкта наприклад, категорію, обсяг капіталу а також зовнішню логіку (часові ритми), що робить отримані патерни статичними та відірваними від реальних бізнес-процесів

1.5 Аналіз сегментації клієнтів за RFM-аналізом

Для класифікації споживачів за їхньою транзакційною активністю використовується модель RFM-аналізу (Recency – давність, Frequency – частота, Monetary – грошова цінність), яка базується на трьох ретроспективних показниках: часі з моменту останньої покупки, загальній кількості замовлень та їхній сумарній вартості [52].

Сегментація клієнтської бази в системах електронної комерції – це процес диференціації суб'єктів (клієнтів) на групи або сегменти на основі визначених характеристик, таких як транзакційна активність, поведінкові патерни, демографічні дані та специфічні потреби. Основним завданням цього процесу є глибинне розуміння структури споживання для розробки точних і ефективних маркетингових стратегій.

Найбільш розповсюдженим інструментом для вирішення задачі сегментації у системах електронної комерції є RFM-аналіз (Recency, Frequency, Monetary) [53]. Він є галузевим стандартом та базується на аналізі трьох ключових факторів, що характеризують транзакційно-поведені дані клієнта:

- *Recency* (Давність) – часовий інтервал з моменту останньої транзакції клієнта;

– *Frequency* (Частота) – інтенсивність здійснення покупок за аналізований період (аналог показника підтримки *Supp*, розглянутого в підрозділі 1.4.2);

– *Monetary* (Грошова цінність) – сукупний обсяг фінансових витрат клієнта на товари або послуги.

Теоретичне обґрунтування моделі RFM базується на принципі Парето, тобто 80% прибутку компанії зазвичай приносять 20 % клієнтів [54]. Сегментація за цією моделлю дозволяє ідентифікувати ці 20 % та диференціювати маркетингові зусилля для різних груп споживачів.

Для ілюстрації принципу дії моделі в контексті аналізу транзакцій розглянемо задачу формування індивідуальних рекомендацій (SMS, e-mail) для B2C-сегмента. Технічне завдання полягає у класифікації користувачів для визначення оптимальних пропозицій для кожної групи. Згідно з методикою RFM, за кожним із факторів проводиться ранжування клієнтів залежно від рівня їхньої лояльності. Кількість рангів залежить від розміру бази даних, проте класичний підхід передбачає виділення трьох рівнів (табл. 1.6).

Таблиця 1.6 – Матриця клієнтів за результатами RFM

Ранг	Давність (Recency)	Частота (Frequency)	Сума (Monetary)
1	Недавній клієнт	Висока частота	Велика сума
2	«Сплячий» клієнт	Середня частота	Середня сума
3	Давній клієнт	Одинична покупка	Мала сума

Після присвоєння відповідних балів формується матриця з 27 сегментів, які згодом можуть бути об'єднані в укрупнені кластери (наприклад, «VIP-клієнти», «Лояльні», «Група відтоку») [55]. Такий вид аналізу дозволяє виявити клієнтів, які демонструють стабільні послідовні патерни (висока частота та давність), а також ідентифікувати суб'єктів, що потребують стимулювання спеціальними пропозиціями для відновлення купівельної активності.

Згідно з фундаментальними роботами А. Хьюза та Р. Блаттберга [56], модель RFM виступає надійним «проксі-показником» для розрахунку CLV (Customer

Lifetime Value), тобто сукупного чистого прибутку, який компанія очікує отримати від клієнта за весь період взаємодії з ним. Якщо RFM оцінює поточну лояльність на основі минулих дій, то CLV дозволяє перевести ці дані у прогностичну площину, визначаючи майбутню цінність клієнта для бізнесу. CLV розраховується як [57]:

$$CLV = (AOV \times F) \times L, \quad (1.20)$$

де AOV – середній чек (Average Order Value), фінансовий еквівалент параметра Monetary (M), тобто сума, що відповідає рангам 1 (велика), 2 (середня) або 3 (маленька);

F – прямий аналог параметра Frequency (F) є показником інтенсивності взаємодії клієнта з платформою, задається як кількість транзакцій за базовий період (наприклад, рік), що відповідає рангам від «часто» до «давно не купував»;

L – період, протягом якого клієнт залишатиметься активним (Customer Lifespan), є аналогом параметра Recency (R). Показник, який перетворює поточну прибутковість у стратегічну цінність клієнта, для рангу 1 (недавній) цей показник є максимальним, для рангу 3 (давній) – мінімальним або нульовим через високу ймовірність відтоку.

Ілюстрації зв'язків матриці клієнтів за RFM-аналізом та значеннями проксі-показника CLV та наведено в таблиці 1.7.

Таблиця 1.7 – Класифікація рангів у моделі RFM та зв'язок із CLV

Ранг (R-F-M)	Характеристика сегмента	Середній чек (M), грн	Частота за рік (F)	Прогноз «життя» (L)	Прогнозний CLV, грн
1-1-1	VIP (Champions)	50 000	12 разів	3 роки	1 800 000
2-2-2	Лояльні (Loyal)	20 000	4 рази	1 рік	80 000
3-3-3	Група відтоку (At Risk)	5 000	1 раз	0.5 року	2 500

Таким чином, після присвоєння рангів формується матриця сегментів, що дозволяє виявити клієнтів, які демонструють стабільні послідовні патерни, та ідентифікувати суб'єктів, що потребують стимулювання. Проте, у контексті B2B-

систем, класичний RFM та базові моделі CLV демонструють «темпоральну сліпоту», оскільки фіксують лише статичні зрізи даних [58]. Вони ігнорують динамічні інтервали між транзакціями (IPT) та специфічну цінність (Utility) кожного окремого елемента в послідовності. Це обґрунтовує необхідність розробки ціннісно-зваженого темпорального профілю клієнта, який буде детально представлений у Розділі 3, як інструменту для подолання обмежень стандартної сегментації.

1.6 Висновки до першого розділу

У першому розділі проведено всебічний аналіз стану та тенденцій персоналізації контенту в B2B-системах електронної комерції. Отримані результати дозволяють зробити наступні висновки.

Обґрунтовано специфіку B2B-сегмента в умовах динамічного ринку. Показано, що понад 60 % українських B2B-компаній зазначають, що основною перешкодою для переходу в онлайн є неможливість забезпечити клієнта релевантним контентом, а саме індивідуальними цінами, технічною документацією в реальному часі та предикативними автоматизованими рекомендаціями щодо здійснення нових транзакцій для поповнення складів. Дослідження показують, що лише 12 % вітчизняних B2B-платформ використовують елементи інтелектуального аналізу даних для персоналізації пропозицій, тоді як у світі цей показник перевищує 50 %.

Результати проведеного аналізу B2B-сегмента електронної комерції за схемою «*Wholesale, Marketing, Customer, Collaborative*» свідчать, що на відміну від B2C, де домінують емоційні та разові покупки, B2B-транзакції характеризуються циклічністю, раціональністю та тривалим життєвим циклом клієнта. Це доводить, що персоналізація в B2B-системах має базуватися на глибокому інтелектуальному аналізі ціннісних та темпоральних параметрів та докорінно відрізнятися від B2C-систем, де контент генерується переважно шляхом статистичного аналізу частотних покупок та загальних поведінкових шаблонів масового споживача.

Проаналізовано традиційні методи взаємодії з клієнтами в системах електронній комерції на основі створення рекомендацій та зауважено, що функціонал

модулів колаборативної та контентної фільтрації при врахуванні операційної складності B2B-процесів має певні недоліки. Так виявлено що класичні рекомендаційні системи часто ігнорують часову динаміку та специфічну цінність товарів, що призводить до низької релевантності пропозицій у сферах із рідкісними, але високовартісними закупівлями. Крім того показано, що незважаючи на спроби комбінування колаборативного та контентного підходів при створенні гібридних рекомендаційних систем, як надбудови над B2B-системою електронної комерції, висока складність моделі такої моделі також має недоліки ігнорування економічної цінності та стабільності циклу транзакцій та не гарантує раціональності рекомендації для бізнес-клієнта. Таким чином, аналіз базових технологій персоналізації контенту за допомогою базових технологій рекомендаційних систем виявив системний дефіцит інструментів для врахування ціннісно-темпоральної специфіки B2B-транзакцій. Це обґрунтовує необхідність використання методів інтелектуального аналізу транзакційно-поведінкових даних B2B-клієнтів для виявлення асоціацій та послідовностей.

Проведений компаративний аналіз методів ІАД, спрямованих на виявлення асоціацій та послідовностей у транзакційних масивах дозволив встановити наступне:

- виявлено, що класичні методи аналізу асоціативних правил (*Association Rule Mining ARM*), зокрема алгоритми Apriori, FP-Growth та Eclat, орієнтовані на знаходження статичних зв'язків товарів в транзакції. При цьому враховується лише частота появи предметів у транзакціях, що повністю нівелює інформацію про його вартість чи обсяг закупівлі. Як наслідок, стратегічно важливі, але рідкісні B2B-товари відсікаються, тоді як дешеві супутні товари з високою частотою стають основою моделі. Це все демонструють повну «фінансову сліпоту», класичних методів ІАД;

- виявлено, що класичні методи аналізу послідовних шаблонів (*Sequential Pattern Mining, SPM*), такі як GSP, PrefixSpan та SPADE, спрямовані на виявлення кореляцій між транзакціями на тлі хронології подій. При цьому вони частково вирішують проблему врахування фактору часу, проте вони характеризуються високим рівнем «ідентифікаційної дискретності», розглядаючи клієнта як абстрактну

послідовність різних ідентифікаторів, що не дає можливість виявити індивідуальний ритм транзакцій на призводить до «темпоральної сліпоти» *SPM*-методів.

Таким чином, виявлено, що існуючий інструментарій ІАД потребує адаптації шляхом інтеграції фінансових показників корисності та часових інтервалів між транзакціями.

Дослідження методів сегментації клієнтів на основі ретроспективних даних показало, що найбільш розповсюдженим галузевим стандартом є модель RFM (Recency, Frequency, Monetary). Проте доведено, що класичне застосування RFM-аналізу в B2B-середовищі має суттєві обмеження. Так показник Recency (давність) фіксує лише статичну точку на осі часу, не враховуючи волатильність індивідуального закупівельного циклу, а показник Monetary (вартість) оцінює валовий дохід, ігноруючи специфічну корисність (Utility) та стабільність патернів споживання. Таким чином класична модель RFM-аналізу орієнтована на ретроспективний опис стану («що відбулося»), що не дозволяє будувати предикативні моделі CLV (Customer Lifetime Value) з високою точністю для B2B-клієнтів із нерівномірним ритмом активності.

Узагальнюючи результати аналізу стану та тенденцій галузі, встановлено критичну необхідність розробки нових підходів до персоналізації B2B-контенту. Обґрунтовано доцільність переходу від статичної сегментації до побудови ціннісно-зваженого темпорального профілю клієнта. Це дозволить подолати фінансову та темпоральну сліпоту існуючих моделей та забезпечити перехід від констатації минулих транзакцій до предикативного управління комунікаційними тригерами на основі прогнозованого закупівельного циклу, фінансової ємності та стабільності поведінки кожного конкретного контрагента.

2 ЦІННІСНО-ЗВАЖЕНИЙ ТЕПОРАЛЬНИЙ ПІДХІД ДО ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ТРАНЗАКЦІЙНИХ ДАНИХ B2B- КЛІЄНТІВ

У розділі обґрунтовано наявність п'яти методичних обмежень класичних методів пошуку асоціативних правил та послідовних шаблонів у контексті інтелектуального аналізу масивів транзакційних B2B-даних. Для подолання перших чотирьох обмежень, а саме фінансової сліпоти, наявності комбінаторного вибуху темпоральної сліпоти та ідентифікаційної дискретності запропоновано ціннісно-зважений темпоральний підхід до інтелектуального аналізу транзакційних даних B2B-клієнтів. Формально запропонований підхід складається із двох складових: моделі та методу ціннісно-зваженого аналізу та моделі та методу темпорального аналізу. Практичну апробацію розроблених моделей та методів запропонованого ціннісно-зваженого темпорального підходу поведено з використанням транзакціо-поведінкових даних клієнтів B2B-системи електронної комерції «База взуття». Отримані результати свідчать про підвищення релевантності персоналізованого контенту, трансформуючи його з суто інформаційного у дієві рекомендації щодо прийняття раціональних рішень B2B-клієнтом.

2.1 Методичні обмеження класичних методів створення асоціативних правил та послідовних шаблонів у контексті транзакційних масивів B2B

Ефективність предиктивної персоналізації контенту систем електронної комерції у сегменті B2B безпосередньо визначається здатністю аналітичних методів трансформувати накопичені транзакційні дані на прийнятні для практики знання. Метою проведення системного аналізу існуючих аналітичних підходів є встановлення їх *фундаментальних методичних обмежень* до інтелектуального аналізу асоціативних правил та послідовних шаблонів в накопичених масивах B2B

транзакцій, які унеможлиблюють їх безпосереднє застосування для створення персоналізованого контенту у системах електронної комерції.

Методи пошуку асоціативних правил та послідовних шаблонів формувались в контексті В2С-задач роздрібної торгівлі [59]. Їх аналітична база будується на понятті транзакції як бінарного кошика покупок, тобто факту наявності або відсутності товару в кошику. Задача формально описується таким чином: дано множину товарних позицій $I = \{i_1, i_2, \dots, i_n\}$ та базу транзакцій $D = \{T_1, T_2, \dots, T_m\}$, де кожна транзакція $T_j \subseteq I$. Потрібно знайти асоціативні правила вигляду $A \rightarrow B$, де $A \subset I$, $B \subset I$, $A \cap B = \emptyset$. Ключовою метрикою, навколо якої вибудовується весь апарат класичних методів є частотна підтримка (*Frequency Support*). Для набору товарів $X \subseteq I$ вона визначається як частка транзакцій, які містять цей набір, у загальній кількості транзакцій бази D :

$$Supp(X) = \frac{|\{T \in D | X \subseteq T\}|}{|D|}, \quad (2.1)$$

де X – досліджуваний асоціативний набір товарів (itemset);

D – повна база транзакцій підприємства за аналізований період;

$|D|$ – загальна кількість транзакцій у базі;

T – окремий елемент бази даних (транзакція, рахунок-фактура або накладна);

$\{T \in D | X \subseteq T\}$ – підмножина транзакцій, у яких набір X присутній як ціле.

Послідовність $C_j = \langle T_1, T_2, \dots, T_k \rangle$ – це впорядкований за часом журнал активності одного клієнта C_j (або Buying Center), що складається з набору транзакцій T , які відбулися в різні моменти часу, при цьому $1 < 2 < \dots < k$. Тоді база даних послідовностей $C = \{C_1, C_2, \dots, C_n\}$ – множина всіх таких клієнтських історій. Кажуть, що шаблон (послідовність) S міститься у послідовності клієнта C_j ($S \subseteq C_j$), якщо всі елементи шаблону зустрічаються в історії клієнта з дотриманням того самого відносного порядку. Відповідно, підтримка шаблону S розраховується як:

$$Supp(S) = \frac{|\{C_j \in C | S \subseteq C_j\}|}{|C|}, \quad (2.2)$$

де S – досліджувана послідовність набору товарів (itemset);

C – множина клієнтських історій підприємства за аналізований період;

$|C|$ – загальна кількість клієнтів у базі;

C_j – повна історія подій (транзакцій, рахунків-фактур або накладних) ідентифікованого клієнта;

$C_j \in C | S \subseteq C_j$ – підмножина клієнтів, у яких послідовність S присутня як ціле.

З формули (2.1) випливає той факт, що кожна транзакція T має однакову алгоритмічну вагу незалежно від її фінансової ємності [60]. Такий підхід є прийнятним у сегменті B2C, коли середній чек клієнта відрізняється несуттєво, а закупівлі здійснюються хаотично і часто. З формули (2.2) випливає, що класичний підхід до аналізу послідовностей базується виключно на факті та порядку здійснення транзакцій, ігноруючи при цьому часові інтервали між ними.

Для методів пошуку шаблонів не має значення, який час пройшов між транзакціями день або рік [61]. В будь-якому випадку шаблон вважається ідентичним. Отже, формули (2.1) та (2.2) демонструють, що традиційний аналітичний апарат розглядає кожну транзакцію T_i та кожного клієнта C_j як рівнозначні одиниці, незалежно від фінансової ємності та ритмічності їх транзакцій. Такий підхід є прийнятним у масовому сегменті B2C, де закупівлі часто мають хаотичний або емоційний характер. Однак у специфічних умовах B2B-взаємодії це призводить до суттєвих викривлень результатів аналізу. Системний розгляд існуючих підходів дозволяє виділити п'ять фундаментальних методичних обмежень, що перешкоджають ефективній персоналізації контенту в системах B2B електронної комерції. Це п'ять критичних розривів (обмежень) між математичними припущеннями методів пошуку асоціацій та послідовностей та природою предметної галузі, а саме по-перше «фінансова сліпота» бінарних моделей; по-друге наявність «комбінаторного вибуху» при зваженому аналізі; по-третє – «темпоральна сліпота» статичних моделей; по-четверте наявність ідентифікаційної

дискретності; по-п'яте відсутність диференційованої стратегії взаємодії з клієнтами. Розглянемо кожне із зазначених обмежень більш детально.

Перше обмеження – «фінансова сліпота» бінарних моделей. Транзакція у базі даних сегмента B2B на відміну від B2C не є простим переліком товарних пропозицій, а формалізується як багатопараметрична функція [62]:

$$T = f(i_k, q_k, p_k) \quad (2.3)$$

де T – транзакція, яка відображає зв'язок між:

i_k – ідентифікаторо товарної позиції (SKU);

q_k – кількістю одиниць;

p_k – індивідуальною ціною за одиницю, яку узгоджено з конкретним контрагентом.

При перетворенні транзакції у бінарний вектор для подальшої обробки за допомогою класичних методів, наприклад Apriori, FP-Growth та Eclat Apriori, із кортежу зникають параметри q_k та p_k , тобто саме ті, що визначають реальну фінансову цінність операції для підприємства.

Дисперсія сум транзакцій у B2B є надзвичайно високою, тобто закупівля дрібних витратних матеріалів та придбання великої партії стратегічного товару алгоритмічно сприймаються класичним методом як рівнозначні події [64]. У результаті система ідентифікує патерни за лише критерієм повторюваності, а не економічної значущості. Це призводить до явища «фінансової сліпоти», яке систематично знецінює рідкісні, проте стратегічно важливі оптові закупівлі. Для ілюстрації наслідків такого підходу порівняємо два типи закупівельних патернів, які показано в таблиці 2.1. З даних таблиці видно, що класичний метод FP-Growth розпізнає Патерн А як значущий через його масовість, тоді як Патерн Б, який генерує у 14 разів більше виручки відхиляється як статистично незначущий шум.

Друге обмеження – наявність «комбінаторного вибуху» при зваженому аналізі. Спроба безпосередньо замінити частотну підтримку на фінансово зважену стикається з обчислювальною перешкодою. У B2B-сегменті поширені так звані

«чеки-кити» – масштабні замовлення, що містять десятки різних артикулів. Якщо один такий чек самостійно перевищує поріг мінімальної фінансової підтримки, то метод розглядає всі можливі підмножини товарів із цієї транзакції [63].

Таблиця 2.1 – Порівняльна оцінка патернів при частотному та ціннісному підходах

Параметр порівняння	Патерн А (Витратні матеріали)	Патерн Б (Технологічне обладнання)
Склад асоціативного патерну	{Рукавиці захисні, Очищувач}	{Дизель-генератор, Масильна речовина}
Кількість виявлених випадків	148	6
Середня вартість набору, грн	180	62 000
Частотна підтримка Supp, %	29,6 % (визнано значущим)	1,2 % (відхилено як шум)
Сумарний фінансовий результат, грн	1 118 880	15 624 000

Кількість таких підмножин C визначається як:

$$C(k) = 2^k - 1, \quad (2.4)$$

де k – кількість унікальних артикулів у межах однієї транзакції;

C – загальна кількість підмножин, що підлягають генерації та верифікації.

Порівняння значень (2.4) для типових довжин транзакцій B2C і B2B наведено в таблиці 2.2.

Третє обмеження – «темпоральна сліпота» статичних моделей. Класичні методи розглядають базу транзакцій як статичний зріз, де хронологія подій не має значення для алгоритму [65]. Результатом є правило вигляду $A \rightarrow B$, яке констатує лише факт сумісності товарів, але не дає відповіді на питання, коли саме виникає повторна потреба клієнта.

У сегменті B2B закупівлі є раціональними, плановими і підпорядкованими внутрішнім бізнес-циклам підприємства-покупця: виробничому процесу,

регламентам обслуговування обладнання, бюджетним циклам, сезонності [66]. Якщо клієнт заковує набір товарів X кожні 45 днів, пропозиція, зроблена на 10-й або на 80-й день після попередньої закупівлі, буде однаково нерелевантною. Відсутність обліку часового інтервалу між транзакціями Δt позбавляє систему можливості розрізнити ці ситуації, що призводить до генерації рекомендаційних тригерів у нерелевантні моменти часу. Отримуючи безсистемний потік пропозицій, контрагенти швидко набувають байдужості до комунікацій платформи – ефект, відомий у літературі як Notification Fatigue [70].

Таблиця 2.2 – Залежність обчислювальної складності від довжини транзакції

Кількість позицій у чеку k	Кількість комбінацій $2^k - 1$	Типовий сегмент	Обчислювальні наслідки
5	31	B2C (звичайний кошик)	Миттєва обробка
10	1 023	B2C (велика закупівля)	Незначне навантаження
20	1 048 575	B2B (мале замовлення)	Суттєве навантаження на RAM
30	1 073 741 823	B2B (типова накладна)	Критичний збій (Out of Memory)
50	$> 10^{15}$	B2B (крупна поставка)	Неможливість обробки

Четверте обмеження – наявність ідентифікаційної дискретності. В основі аналізу послідовностей лежить побудова цілісної траєкторії поведінки клієнта [67]. В сегменті B2C ця цілісність часто порушується через ідентифікаційну дискретність, яка визначається розрив логічного зв'язку між транзакційними сесіями одного й того самого клієнта. Оскільки ідентифікація в B2C сегменті часто є стохастичною (базується на тимчасових файлах cookies або цифрових відбитках пристрою Fingerprinting), історія транзакцій «розсипається» на фрагменти, які метод обробки послідовностей сприймає як дії різних анонімних користувачів. Феномен фрагментації транзакційних послідовностей унеможливорює розрахунок ключових темпоральних метрик, таких як типовий (середній) середній інтервал між подіями μ та коефіцієнт варіації *Coefficient of Variation, CV*. Розглянемо приклад. Припустимо, що клієнт здійснює циклічну закупівлю товарів з інтервалом у 14 днів (табл. 2.3).

Таблиця 2.3 – Порівняння стохастичної та детермінованої ідентифікації при обробці послідовностей

Сценарій В2С		Сценарій В2В	
ID	Подія	ID	Подія
C_1	$\langle\{A\}\rangle$	C_1	$\langle\{A\}, \{B, C\}, \{A, C\}\rangle$
C_2	$\langle\{B, C\}\rangle$		
C_3	$\langle\{A, C\}\rangle$		

У сценарії В2С якщо клієнт здійснив першу покупку з ноутбука, другу після очищення кешу браузера, а третю зі смартфона без авторизації, база послідовностей S зафіксує три незалежні одноразові події для різних C (наприклад, C_1, C_2, C_3). Для класичного методу такий патерн виглядатиме як хаотичний попит трьох різних осіб, що робить предиктивну персоналізацію неможливою.

У сценарії В2В завдяки детермінованій ідентифікації (обов'язкова авторизація контрагента для отримання доступу до гуртових цін, складських залишків та фінансових взаєморозрахунків), усі три транзакції будуть жорстко прив'язані до єдиного профілю клієнта. Це дозволяє системі сформувати цілісну послідовність $C_j = \langle T_1, T_2, T_3 \rangle$, де $1 < 2 < 3$ та точно розрахувати ритмічність бізнес-процесів В2В-клієнта. Тобто детермінована ідентифікація (через особистий кабінет), невикористаний класичними методами потенціал для більш точного темпорального аналізу.

П'яте обмеження – відсутність диференційованої стратегії взаємодії з клієнтами. Навіть якщо припустити усунення перших чотирьох обмежень, класичні методи залишають поза увагою принципово важливий аспект В2В-комерції – різні контрагенти мають кардинально різну цінність для підприємства і потребують різних стратегій взаємодії [68].

Класичний підхід до сегментації клієнтів – RFM-аналіз (Recency–Frequency–Monetary) оперує ретроспективними показниками: давністю останньої покупки, кількістю транзакцій та сумарним обсягом [69]. Ці показники описують минуле, але не прогнозують майбутню закупівельну поведінку.

Унаслідок цього рекомендаційна система застосовує однакову глибину попередження та ширину комунікаційного вікна до всіх клієнтів незалежно від їхнього профілю.

Преміум-клієнт з великим середнім чеком і жорстко регулярним циклом закупівель та дрібний контрагент із хаотичними замовленнями отримують ідентичне за форматом звернення – що є неоптимальним з точки зору рентабельності витраченого ресурсу. Для підприємства важливо не лише знайти стабільний патерн і визначити момент звернення, але й адаптувати параметри цього звернення до фінансового профілю конкретного клієнта: цінний контрагент потребує точного попадання у момент потреби (вузьке вікно, висока точність), тоді як менш цінний допускає ширше охоплення з більшим допуском. Відсутність такої диференціації призводить до розпорошення комунікаційного ресурсу і зниження загальної рентабельності відділу продажів.

Концептуально описані обмеження є взаємопов'язаними і мають ієрархічну структуру [71]. Спроба подолати фінансову сліпоту класичних моделей шляхом додавання вартісних показників неминуче активує проблему комбінаторного вибуху. Своєю чергою, перехід до аналізу темпоральних інтервалів вимагає не лише нової архітектури обробки, а й усунення ідентифікаційної дискретності, оскільки розрахунок циклічності можливий лише на цілісних, а не фрагментованих часових рядах.

Таким чином, фінальне обмеження відсутність диференційованої стратегії взаємодії з клієнтами, тобто стратегічна монотонність є результатом дефіциту даних на всіх попередніх етапах: система не може диференціювати клієнтів, поки не навчиться ідентифікувати їхні індивідуальні фінансово-темпоральні патерни.

Схематично характер інформаційних втрат при обробці B2B-транзакцій класичними методами наведено на рисунку 2.1.

Зіставлення виявлених обмежень дозволяє сформулювати п'ять вимоги до математичного апарату інтелектуального аналізу транзакційних даних клієнтів B2B, що розробляється далі.

По-перше, пропонується замінити частотний критерій значущості на фінансово зважений, де вага кожної транзакції пропорційна добутку кількості та вартості товарів.

По-друге, забезпечити механізми, що запобігають комбінаторному вибуху при аналізі довгих B2B-транзакцій без втрати економічно значущих патернів.

По-третьє, інтегрувати темпоральну компоненту, тобто запропонувати оцінку стабільності часових інтервалів між закупівлями для генерації проактивних рекомендацій у вікні максимальної ймовірності повторної потреби клієнта.

По-четверте, врахувати що перехід від ймовірнісної до детермінованої ідентифікації B2B клієнтів дозволяє консолідувати фрагментовані транзакційні сесії у цілісні багаторічні послідовності, що забезпечує високу достовірність розрахунку індивідуальних закупівельних траєкторій.

По-п'яте ввести класифікацію клієнтів за циклічними та фінансовими характеристиками закупівельної поведінки для диференційованого визначення параметрів комунікаційного звернення.



Рисунок 2.1 – Структура втрат інформаційних параметрів B2B-транзакції при застосуванні класичних методів

Зазначені вимоги є науковою основою для розробки ціннісно-зваженого темпорального аналізу, математичний апарат якого описано далі у цьому розділі, а

також моделі та методу класифікації B2B-клієнтів для предиктивної персоналізації контенту систем електронної комерції, які складають основу четвертого розділу даного дисертаційного дослідження.

2.2 Ціннісно-зважений аналіз транзакційних даних B2B клієнтів

В попередньому підрозділі було встановлено, що фундаментальним недоліком класичних методів пошуку асоціативних правил є бінарна модель транзакції, згідно якої фіксується лише факт присутності товару у чеку та ігнорується кількість і вартість придбаних позицій. Для B2B-сегменту це призводить до «фінансової сліпоти», тобто система виявляє часті, але дешеві транзакції, ігноруючи рідкісні, проте стратегічно значущі оптові замовлення. Розроблений математичний апарат ціннісно-зваженого аналізу, який складається із ціннісно-зваженої моделі B2B транзакції та методу розростання ціннісних патернів UP-Growth (Utility Pattern Growth) для пошуку асоціативних правил дозволяють враховувати фінансову цінність транзакцій.

2.2.1 Розробка ціннісно-зваженої моделі B2B транзакції

Розглянемо розширене формальне визначення B2B-транзакції. На відміну від класичної моделі, де $T \subseteq I$ є підмножиною артикулів, у розробленій математичній моделі ціннісно-зважена транзакція представлена кортежем, що включає не лише ідентифікатор товарної позиції, але й кількісно-вартісні параметри. Елемент транзакції визначається наступним виразом:

$$t_j = \langle i_k, q_k, p_k \rangle, \quad (2.5)$$

де i_k – ідентифікатор товарної позиції (SKU);

q_k – кількість одиниць товару (ящиків, штук, одиниць обліку), придбаних у межах цієї транзакції;

p_k – індивідуальна ціна за одиницю товару для даного контрагента, що враховує узгоджені знижки, умови постачання та обсягові коефіцієнти.

На основі параметрів q_k і p_k вводиться функція корисності (utility function), що характеризує фінансовий внесок конкретної товарної позиції у межах транзакції [72]. Корисність елемента i_k у транзакції T визначається наступним виразом:

$$u(i_k, T) = q_k \cdot p_k, \quad (2.6)$$

де $u(i_k, T)$ – корисність товарної позиції i_k у транзакції T .

Сумарна корисність транзакції (Transaction Utility) – це повна фінансова вартість усього замовлення, що обчислюється як сума внесків усіх товарних позицій за наступним виразом [73]:

$$U(T) = \sum_{i_k \in T} u(i_k, T) = \sum_{i_k \in T} q_k \cdot p_k, \quad (2.7)$$

де $U(T)$ – сумарна корисність (грошова вартість) транзакції T ;

$i_k \in T$ – товарна позиція, що належить до множини елементів транзакції T .

Саме значення $U(T)$ із формули (2.7) стає вагою транзакції у процесі пошуку асоціативних правил на основі побудови ціннісно-зваженого дерева. У класичному методі пошуку асоціативних правил (АП) за допомогою частотного дерева FP-Growth, кожна транзакція вносить однаковий інкремент «+1» до лічильника вузла незалежно від її вартості. У запропонованому далі методі пошуку АП – UP-Growth замість «+1» вузол накопичує значення сумарної корисності транзакції $U(T)$. Таким чином транзакція стає ціннісно-зваженою.

Але інтуїтивно зрозуміла ідея заміни частотної підтримки на фінансову стикається з серйозною теоретичною перешкодою [74]. У класичному методі FP-Growth ефективне відсікання нерелевантних гілок пошуку (прунінг) ґрунтується на властивості *антимонотонності* (*Downward Closure Property*), яка полягає в тому, що підтримка (2.1) будь-якої підмножини $A \subset X$ не може бути меншою за

підтримку самого набору X , тобто $Supp(A) \geq Supp(X)$. Це дозволяє одразу відкидати цілі гілки пошукового простору, не перебираючи кожну комбінацію.

У ціннісно-зваженому аналізі ця властивість не виконується. Невеликий набір товарів A може зустрічатися переважно у дешевих транзакціях, тоді як ширший набір $X \supset A$ – у дорогих.

У такому разі реальна фінансова корисність $U(X)$ може перевищувати $U(A)$, що руйнує антимонотонність і позбавляє можливості безпечно відсікати гілки пошуку.

Наслідком є неконтрольоване зростання числа кандидатів, показане у (2.4), що призводить до виникнення комбінаторного вибуху, який є другим обмеженням вже не частотного, а зваженого аналізу.

Для вирішення цієї проблеми до моделі ціннісно-зваженої транзакції (2.5)-(2.7) вводиться поняття транзакційно-зваженої корисності (Transaction-Weighted Utilization, TWU). TWU набору X розраховується як сумарна вартість усіх транзакцій бази D , що містять цей набір, і слугує консервативною верхньою оцінкою реальної корисності набору, що відображено у наступному виразі [75]:

$$TWU(X) = \sum_{\substack{T \in D \\ X \subseteq T}} U(T), \quad (2.8)$$

де $TWU(X)$ – транзакційно-зважена корисність набору X ;

$\{T \in D | X \subseteq T\}$ – підмножина транзакцій бази D , що містять усі елементи набору X одночасно;

$U(T)$ – повна вартість транзакції за виразом (2.7).

Ключова теоретична властивість TWU полягає у відновленні *антимонотонності* у зваженому просторі: для будь-якої підмножини $A \subset X$ виконується $TWU(A) \geq TWU(X)$ [76].

Це дозволяє безпечно відсікати нерелевантні гілки: якщо $TWU(X) < \text{min_utility}$, жодна надмножина набору X не може бути високоціннісним патерном

і подальший її аналіз не проводиться. Завдяки цьому метод зберігає обчислювальну ефективність FP-Growth, але у ціннісно значущому просторі пошуку.

Таким чином, отримала подальшого розвитку ціннісно-зважена модель B2B транзакції (2.5)-(2.8) за рахунок додавання сумарної корисності визначеної транзакції, як суми добутків кількості придбаних одиниць товарних складових визначеної транзакції на індивідуальну ціну та транзакційно-зваженої корисності, як сумарна вартість усіх транзакцій бази, які містять визначену транзакцію, що дозволило удосконалити метод пошуку ціннісно-зважених АП.

2.2.2 Розробка метода розростання цінністних патернів UP-Growth

Розглянемо детально процес пошуку АП за допомогою *методу розростання цінністних патернів UP-Growth* [77]. Цей процес складається із двох етапів: побудови ціннісного дерева патернів та рекурсивного виявлення всіх гілок дерева з сумарною цінністю, що перевищує експериментально заданий поріг підтримки (2.1) – $\min_utility$

Етап 1. Побудова ціннісного дерева патернів (Utility Pattern Tree, UP-Tree) згідно розробленого *методу розростання цінністних патернів UP-Growth (Utility Pattern Growth)*. відбувається у два проходи по базі транзакцій.

Під час *першого проходу* для кожного артикулу i_k обчислюється $TWU(i_k)$ за виразом (2. 8). Позиції, у яких $TWU(i_k) < \min_utility$, виключаються з подальшого аналізу. Знайдені таким чином артикули транзакції впорядковуються за спаданням TWU . Таким порядком пріоритетів позначається $<_U$ та визначає *структуру гілок дерева*.

Під час *другого проходу* кожна транзакція T вставляється у дерево. артикули впорядковуються відповідно до $<_U$, після чого здійснюється обхід відповідної гілки.

Принципова відмінність метода *UP-Growth* від *FP-Growth* полягає в тому, що значення вузол i_k не збільшується на «1», а накопичує $U(T)$ за виразом (2.7).

Формально значення вузла після вставлення n транзакцій, що проходять через нього, визначається наступним виразом:

$$NodeValue(i_k) = \sum_{j=1}^n U(T_j) \cdot 1[i_k \in path(T_j)], \quad (2.9)$$

де $NodeValue(i_k)$ – накопичене значення вузла для артикулу i_k ;

$U(T_j)$ – повна вартість j -ої транзакції за виразом (2.6);

$1[\cdot]$ – індикаторна функція, що дорівнює 1, якщо i_k присутній у шляху транзакції T_j , та 0 – в іншому випадку.

Етап 2. Після побудови *UP-дерева* метод рекурсивно виявляє всі набори товарів, сумарна фінансова цінність яких перевищує поріг *min_utility*. Для кожного артикулу (від найменш до найбільш цінного за $<_U$) формується умовна база потенційної цінності, з якої будується *умовне UP-дерево*. Набір X визнається ціннісним патерном (High-Utility Itemset) і включається до результату, якщо виконується наступна умова:

$$U(X) = \sum_{\substack{T \in D \\ X \subseteq T}} u(X, T) \geq min_utility, \quad (2.10)$$

де $U(X)$ – фактична сумарна корисність набору X по всій базі транзакцій;

$u(X, T) = \sum_{i_k \in X} u(i_k, T)$ – корисність набору X у конкретній транзакції T ;

min_utility – встановлений поріг мінімальної фінансової значущості, грн.

Для ілюстрації роботи методу розглянемо тестовий приклад, де база D містить п'ять транзакцій $\{T_1, \dots, T_5\}$ з артикулами $\{A, B, C, D, E\}$ та встановленим порогом $min_utility = 130$ грн. Вхідні дані та розраховані значення TWU для кожного артикулу наведені в таблицях 2.4 та 2.5.

З даних таблиць видно, що артикул E , присутній у трьох транзакціях, виключається, оскільки сукупна вартість транзакцій, де він зустрічається (125 грн),

не досягає порогу. Решта артикулів впорядковуються за спаданням TWU: $B(240) > C(165) > D(155) > A(150)$.

Таблиця 2.4 – Вхідні дані тестового прикладу ($min_utility = 130$ грн)

TID	Склад транзакції (SKU, кількість, ціна, грн)	$U(T)$, грн
T_1	$\{A(1,10), B(1,20), D(1,5), E(1,5)\}$	40
T_2	$\{B(2,20), C(1,10), D(1,5)\}$	55
T_3	$\{A(1,10), B(1,20), C(1,10), E(2,5)\}$	50
T_4	$\{B(1,20), E(3,5)\}$	35
T_5	$\{A(2,10), B(1,20), C(1,10), D(2,5)\}$	60

Таблиця 2.5 – Розрахунок TWU та відбір артикулів

Артикул	Транзакції, що містять	TWU, грн	Статус
B	$T_1 + T_2 + T_3 + T_4 + T_5$	240	✓ Включено
C	$T_2 + T_3 + T_5$	165	✓ Включено
D	$T_1 + T_2 + T_5$	155	✓ Включено
A	$T_1 + T_3 + T_5$	150	✓ Включено
E	$T_1 + T_3 + T_4$	125	✗ Виключено ($125 < 130$)

У цьому порядку товари вставляються у UP-дерево, що забезпечує максимальне перекриття гілок та мінімізує розмір структури. Покроковий процес побудови UP-дерева для роботи метода розростання ціннісних патернів UP-Growth наведено в таблиці 2.6.

На кроці генерації умовних баз аналіз здійснюється знизу вгору за $<_U$ (від A до B). Для набору $\{B, C\}$ фактична корисність дорівнює $U(T_2) + U(T_3) + U(T_5) = 55 + 50 + 60 = 165$ грн, що перевищує поріг 130 грн. Набір $\{B, C\}$ визнається ціннісним патерном і включається до результату.

Порівняно з класичним FP-Growth, де цей самий набір міг би поступитися місцем набору, що містить дешевший, але частіший артикул E , метод UP-Growth гарантує пріоритет фінансово значущих асоціацій.

Таблиця 2.6 – Покроковий процес побудови UP-дерева

TID	Впорядкована послідовність (за $<_U$)	Дія з деревом
T_1	$B \rightarrow D \rightarrow A$	Нова гілка: $B(40) \rightarrow D(40) \rightarrow A(40)$
T_2	$B \rightarrow C \rightarrow D$	$B: 40 + 55 = 95$; нова гілка $C(55) \rightarrow D(55)$
T_3	$B \rightarrow C \rightarrow A$	$B: 95 + 50 = 145$; $C: 55 + 50 = 105$; нова гілка $C(105) \rightarrow A(50)$
T_4	B	$B: 145 + 35 = 180$
T_5	$B \rightarrow C \rightarrow D \rightarrow A$	$B: 180 + 60 = 240$; $C: 105 + 60 = 165$; $D: 55 + 60 = 115$; нова гілка $D(115) \rightarrow A(60)$

Для чіткого розуміння внесеного удосконалення систематизуємо принципи відмінності між класичним методом FP-Growth та запропонованим UP-Growth у таблиці 2.7.

Як видно з таблиці, зміна є точковою і методологічно обґрунтованою. Єдина трансформація це заміна одиничного інкременту лічильника на реальну вартість транзакції зі збереженням усієї архітектури дерева та механізму умовних баз [78]. Це означає, що метод UP-Growth успадковує обчислювальну ефективність FP-Growth, водночас перебудовуючи критерій значущості патернів відповідно до фінансової логіки B2B-комерції.

Для ілюстрації практичного ефекту цієї зміни розглянемо узагальнений сценарій на основі тестового прикладу. Якби застосовувався класичний FP-Growth із порогом підтримки $min_sup = 3$ транзакції (тобто набір має з'являтися щонайменше у 3 з 5 транзакцій), найзначущішим результатом стала б асоціація за

участі артикулу B , який входить у всі п'ять транзакцій. Проте цей результат не враховує, наскільки вагомою є кожна з транзакцій.

Таблиця 2.7 – Порівняльна характеристика методів FP-Growth та UP-Growth

Компонент методу	FP-Growth (класичний)	UP-Growth (запропонований)
Одиниця виміру у вузлі дерева	Лічильник входжень (+1)	Вартість транзакції $U(T)$, грн
Критерій значущості набору	Частотна підтримка $Supp(X)$	Сумарна корисність $U(X) \geq min_utility$
Антимонотонна верхня оцінка	$Supp$ (антимонотонна за означенням)	TWU (антимонотонна у зваженому просторі)
Впорядкування артикулів у дереві	За спаданням частоти	За спаданням TWU
Врахування кількості товарів q_k	Ні	Так
Врахування ціни товарів p_k	Ні	Так
Цільовий результат	Часті набори (Frequent Itemsets)	Фінансово значущі набори (High-Utility Itemsets)
Пріоритет у B2B	Масові дешеві товари	Оптові замовлення з максимальним виторгом

Метод UP-Growth із порогом $min_utility = 130$ грн натомість виявляє набір $\{B, C\}$ з фактичною корисністю 165 грн як пріоритетний, навіть якщо частота входження C менша за B , оскільки транзакції, де присутній C , є фінансово найвагомішими ($T_2 = 55$ грн, $T_3 = 50$ грн, $T_5 = 60$ грн). Відмінність результатів двох підходів наведена на рисунку 2.2.



Рисунок 2.2 – Порівняльна схема пріоритетів у класичному FP-Growth та запропонованому UP-Growth

Таким чином, для реалізації ціннісно-зваженого підходу до інтелектуального аналізу транзакційних B2B-даних розроблено модель ціннісно-зваженої транзакції та метод розростання цінністних патернів UP-Growth. Ці розробки складають перший та другий пункти наукової новизни.

Перший пункт наукової новизни. Удосконалено модель ціннісно-зваженої транзакції, яка на відміну представлення у вигляді переліка товарних пропозицій (SKU), додатково включає обсяг SKU та індивідуальну ціну, що дозволяє в подальшому використовувати їх добуток для розрахунку транзакційно-зваженої корисності (TWU) та ідентифікувати транзакції за не лише за критерієм повторюваності, а і за економічної значущості для подолання обмеження «фінансової сліпоти»

Другий пункт наукової новизни. Отримав подальший розвиток метод розростання цінністних патернів (Utility Pattern Growth) за рахунок використання транзакційно-зваженої корисності (TWU) для розрахунку значень вузлів дерева UP-Growth замість частотних лічильників класичного FP-Growth, що дозволило відновити властивість антимонотоності долаючи обмеження «комбінаторного вибуху», а також змінити пріоритетність генерації асоціативних правил, фокусуючи систему персоналізації на економічно значущих транзакціях.

2.3 Темпоральний аналіз транзакційних даних B2B клієнтів

В цьому підрозділі описано створення другої компоненти запропонованого ціннісно-зваженого темпорального підходу до інтелектуального аналізу B2B транзакцій, а саме темпоральній верифікації. Її завданням є встановлення регулярності або випадковості виявлення методом UP-Growth ціннісно-зваженого патерну. Якщо його поява в UP-дереві для конкретного контрагента є регулярною, тоді необхідно розрахувати оптимальний момент наступної пропозиції, як складовою до рекомендаційного контенту. Розглянемо наступні пояснення.

Ціннісно-зважені патерни (АП), які знайдено в транзакційних даних B2B клієнтів за допомогою методу UP-Growth надають відповіді на питання «що

купують разом» та виявляють фінансово значущі товарні комбінації. Однак для побудови предикативної системи рекомендацій у B2B-сегменті цього недостатньо. Не менш важливо відповісти на питання «коли це запропонувати». Рекомендація, зроблена B2B платформою поза циклом реальної потреби клієнта, не лише марна, але й руйнує довіру контрагента. Надлишок нерелевантних пропозицій (Notification Fatigue) призводить до ігнорування будь-яких повідомлень від постачальника, включаючи своєчасні та корисні.

На користь створення предикативної системи рекомендації свідчить ще факт принципово іншої природи B2B транзакцій у порівнянні з B2C. B2B транзакції є раціональними, плановими та підпорядкованими внутрішнім бізнес-процесам підприємства-покупця [79]. Виробничий регламент, графік технічного обслуговування обладнання, сезонність асортименту, бюджетні цикли – все це формує *стійкі часові патерни повторних B2B-транзакцій*. Ці патерни є прихованою структурою транзакційного масиву, яку класичні методи, що розглядають базу даних як статичний зріз, виявити не здатні.

2.3.1 Темпоральна структура B2B-послідовностей

Формальною основою темпорального аналізу є поняття послідовності B2B транзакцій (2.2). Для подолання обмежень темпоральної сліпоты та ідентифікаційної дискретності введемо поняття темпоральної послідовності транзакцій для клієнта C_j – це впорядкований за часом ланцюжок транзакцій, що відображає хронологію його закупівельної діяльності за наступним виразом:

$$C_j = \langle T_{\tau_1}, T_{\tau_2}, \dots, T_{\tau_m} \rangle, \quad \tau_1 < \tau_2 < \tau_m, \quad (2.11)$$

де T_{τ_i} – транзакція, здійснена у момент часу τ_i ;

m – загальна кількість транзакцій клієнта за аналізований період.

Послідовний шаблон (Sequential Pattern, SP) у B2B описує впорядковану за часом появу подій або транзакцій, що з достатньою регулярністю повторюється в

множині клієнтських послідовностей [80]. Якщо $\alpha = \langle T_{\tau_a} \rangle$ та $\beta = \langle T_{\tau_b} \rangle$ – дві підпослідовності, де $\tau_a < \tau_b$, то послідовний шаблон формується як:

$$SP = \alpha \Rightarrow \beta. \quad (2.12)$$

Тобто, якщо у момент часу τ_a клієнт здійснив транзакцію або набір транзакцій α , існує висока ймовірність того, що у наступний момент $\tau_b > \tau_a$ він здійснить транзакцію або набір транзакцій β . Знак $\Rightarrow$ підкреслює саме часову, а не логічну послідовність, на відміну від оператора $\rightarrow$ в асоціативних правилах.

Виявлення послідовного шаблону $\alpha \Rightarrow \beta$ лише підтверджує факт часового зв'язку між подіями – транзакціями. Для створення предикативних рекомендацій необхідна більш конкретна інформація, а саме через скільки днів після α зазвичай настає β . Для формалізації цієї характеристики вводиться поняття *вектору міжзакупівельних інтервалів (Inter-Purchase Time, IPT)*. Нехай для ціннісно-зваженого патерну X , виявленого методом UP-Growth, та клієнта C_j зафіксована наступна послідовність дат τ транзакцій цього набору:

$$T(X, C_j) = \{\tau_1, \tau_2, \dots, \tau_n\}, \tau_1 < \tau_2 < \dots < \tau_n, \quad (2.13)$$

де n – кількість закупівель набору X клієнтом C_j за аналізований період; мінімальне значення $n \geq 3$ є умовою застосування подальшого аналізу, оскільки для отримання хоча б двох інтервалів необхідно щонайменше три закупівлі. Тоді вектор міжзакупівельних інтервалів $IPT = (\delta_1, \delta_2, \dots, \delta_{n-1})$ обчислюється так:

$$\delta_i = \tau_{i+1} - \tau_i, \quad i = 1, 2, \dots, n-1, \quad (2.14)$$

де δ_i – тривалість i -го інтервалу між суміжними закупівлями набору X клієнтом C_j , у днях; значення $\delta_i > 0$ за означенням, оскільки транзакції впорядковані хронологічно.

Для отриманого таким чином вектору міжзакупівельних інтервалів клієнта C_j розрахуємо статистичні характеристики: середнє значення тривалості інтервалу, середньоквадратичне відхилення та коефіцієнт варіації [81].

Середнє значення тривалості інтервалів μ характеризує типовий (середній) цикл транзакційного набору для даного клієнта і розраховується за наступним виразом [82]:

$$\mu = \frac{1}{n-1} \sum_{i=1}^{n-1} \delta_i, \quad (2.15)$$

де μ – середній цикл закупівлі, у днях;

$n - 1$ – кількість спостережуваних інтервалів.

Середньоквадратичне відхилення σ характеризує мінливість циклу – наскільки окремі закупівлі відхиляються від середнього значення [83]:

$$\sigma = \sqrt{\frac{1}{n-2} \sum_{i=1}^{n-1} (\delta_i - \mu)^2}, \quad (2.16)$$

де σ – середньоквадратичне відхилення інтервалів, у днях; вираз (2.15) використовує незміщену оцінку дисперсії з дільником $(n - 2)$, оскільки кількість спостережень дорівнює $n - 1$ інтервалам, а оцінка μ є вільним параметром.

Середнє μ (2.15) та відхилення σ (2.16) самі по собі не дають відповіді на питання, чи є цикл достатньо стабільним для прогнозування. Значення $\sigma = 10$ днів при середньому циклі $\mu = 30$ днів характеризує значну нестабільність, тоді як те саме відхилення $\sigma = 10$ днів при $\mu = 120$ днів – незначне коливання.

Для порівнянної оцінки стабільності циклів різної тривалості застосовується безрозмірний показник стабільності циклу – *коефіцієнт варіації* (*Coefficient of Variation, CV*), який визначається наступним виразом [84]:

$$CV(X, C_j) = \frac{\sigma}{\mu}, \quad (2.17)$$

де $CV(X, C_j) \in [0, +\infty)$ – відносна мінливість інтервалів між закупівлями набору X для клієнта C_j ;

σ та μ – визначені виразами (2.15) та (2.16) відповідно.

Коефіцієнт варіації є центральним критерієм темпорального аналізу, тобто якщо $CV(X, C_j) \leq \gamma$, де γ – встановлений поріг стабільності, тоді транзакційний цикл вважається достатньо регулярним для включення до системи так званих *предикативних тригерів*. Якщо $CV(X, C_j) > \gamma$, тоді транзакції здійснюються нерегулярно, то надійне визначення послідового шаблону для прогнозування моменту наступної транзакції є неможливим. Інтерпретація значень CV стосовно управлінських рішень у B2B наведена в таблиці 2.8.

Таблиця 2.8 – Інтерпретація значень коефіцієнта варіації для B2B-закупівель

Діапазон CV	Характеристика циклу	Типовий B2B-сценарій	Рішення методу
$0,00 \leq CV \leq 0,15$	Жорстко регулярний	Щомісячна закупівля витратних матеріалів за довгостроковим контрактом	Генерувати тригер
$0,15 \leq CV \leq 0,30$	Стабільний з незначними відхиленнями	Квартальна закупівля основного асортименту з похибкою ± 2 тижні	Генерувати тригер
$0,30 \leq CV \leq 0,50$	Помірно нестабільний	Сезонні закупівлі з бюджетною залежністю або зміщенням строків	Генерувати тригер з розширеним вікном $k < 0,5$
$CV > 0,50$	Хаотичний	Разові або ситуативні замовлення без вираженої регулярності	Виключити з проактивних тригерів

Значення порогу γ налаштовується залежно від специфіки галузі та прийнятності підприємства до похибки прогнозу (як правило, $\gamma \in [0,3; 0,5]$) [85].

2.3.2 Розрахунок дати комунікаційного тригера та вікна предикативної пропозиції

Для цінсно-зважених послідовних патернів, що пройшли перевірку за виразом (2.17), за методом темпорального аналізу розраховується оптимальний момент активації *комунікаційного тригера*, а саме дату, у яку система електронної комерції має ініціювати предикативну пропозицію B2B клієнту. При цьому враховується, що пропозиція має надійти не у день очікуваної транзакції, а з певним упередженням. Тобто B2B-клієнту потрібен час на розгляд, погодження та оформлення замовлення. Дата комунікаційного тригера $T_{trigger}$ визначається наступним виразом [86]:

$$T_{trigger} = T_{last} + \mu - k \cdot \sigma, \quad (2.18)$$

де $T_{trigger}$ – розрахункова дата активації предикативної пропозиції;

T_{last} – дата останньої фактичної закупівлі набору X клієнтом C_j ;

μ – середній цикл за виразом (2.15), у днях;

σ – середньоквадратичне відхилення за виразом (2.16), у днях;

$k \in [0; 2]$ – коефіцієнт предикативності, значення якого визначає глибину вікна попередження.

Економічний зміст коефіцієнта k визначається наступним чином. При $k = 0$ тригер активується рівно у момент середньої очікуваної дати наступної транзакції, тобто коли потреба вже виникла [87]. При $k = 1$ тригер спрацьовує за σ днів до цього моменту – це стандартне вікно завчасного сповіщення, розраховане на клієнтів з типовим циклом узгодження замовлення. При $k = 2$ пропозиція надходить за 2σ днів, що доцільно для дорогих і тривалих транзакцій, де процес прийняття рішення займає кілька тижнів (наприклад, закупівля промислового обладнання). Вибір k здійснюється підприємством на основі галузевої специфіки та емпіричних даних про середній час від отримання пропозиції до підписання

замовлення. Вираз (2.17) визначає також «вікно активної пропозиції», а саме інтервал часу $[T_{trigger}; T_{last} + \mu + \sigma]$, протягом якого ймовірність закупівлі набору X є максимальною [88]. Пропозиції, надіслані до $T_{trigger}$, потрапляють у «передчасну зону», де потреба ще не актуальна. Пропозиції, надіслані після верхньої межі вікна – у «запізнілу зону», де клієнт, найімовірніше, вже звернувся до конкурента або відклав закупівлю. Концепцію вікна предикативної пропозиції ілюструє рисунок 2.3 та таблиця 2.9.



Рисунок 2.3 – Концепція вікна предикативної пропозиції на основі розподілу ймовірності повторної закупівлі

Таблиця 2.9 – Опис зон вікна предикативної пропозиції

Зона	Інтервал	Характеристика
Передчасна	до $T_{trigger}$	Потреба ще не актуальна; пропозиція ігнорується
Активна (оптимум)	$[T_{trigger}; T_{last} + \mu + \sigma]$	Максимальна ймовірність прийняття
Запізніла	після $T_{last} + \mu + \sigma$	Клієнт або вже закупив в іншого, або відклав

2.4 Зведена характеристика ціннісно-зваженого темпорального аналізу

Поєднання методу UP-Growth (підрозділ 2.2) з методом темпорального аналізу утворює цілісний аналітичний конвеєр, кожен з етапів якого виконує чітко визначену функцію. Перший етап відповідає на питання «що запропонувати» – виявляє фінансово значущі товарні комбінації. Другий відповідає на питання «кому і коли» – ідентифікує тих клієнтів, для яких даний патерн є регулярним, і визначає оптимальний момент пропозиції. Послідовна взаємодія обох етапів наведена на рисунку 2.4.

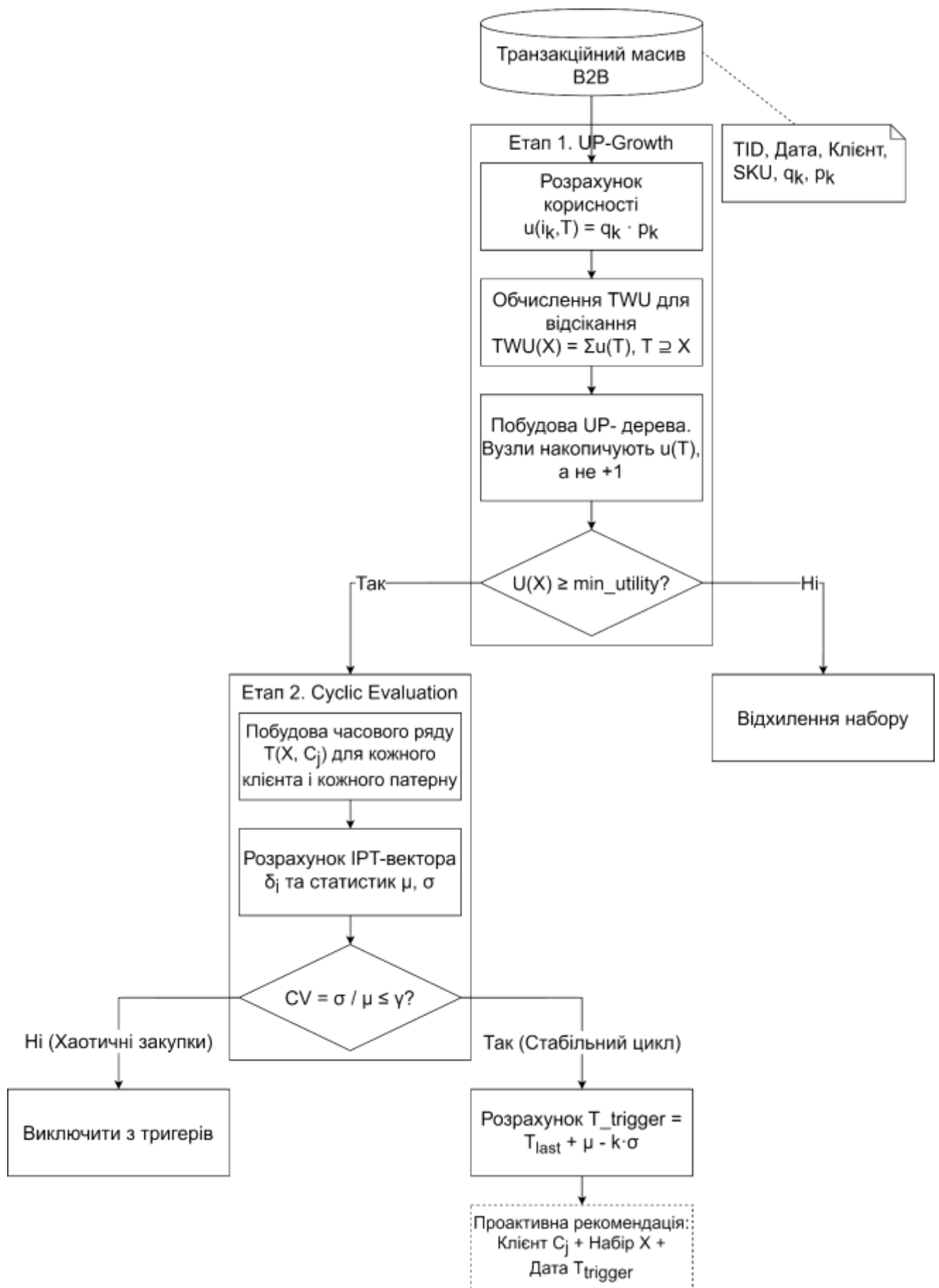


Рисунок 2.4 – Повна архітектура ціннісно-зваженого темпорального аналізу

Двоетапна архітектура вирішує проблему обчислювальної складності, описану у підрозділі 2.1: темпоральний аналіз (Етап 2) застосовується не до мільйонів можливих комбінацій, а лише до відносно невеликого переліку ціннісно-зважених патернів, що пройшли фінансовий відбір на Етапі 1 [89]. Це робить метод практично прийнятним для реальних транзакційних масивів. Для демонстрації змістовного значення виразів (2.11)–(2.18) розглянемо два типові сценарії, характерних для B2B-систем електронної комерції у сфері оптової торгівлі.

Сценарій 1. Регулярна сезонна закупівля. Клієнт C_1 закупив набір $X = \{\text{зимові чоботи Art. 501, засіб захисту від солі}\}$ у такі дати: $t_1 = 15.10.2022$, $t_2 = 12.10.2023$, $t_3 = 18.10.2024$. Вектор інтервалів: $\delta_1 = 362$ дні, $\delta_2 = 371$ день. Розрахунок: $\mu = 366,5$ дні, $\sigma \approx 6,4$ дні, $CV \approx 0,017$. Значення $CV \ll \gamma$ свідчить про жорстко регулярний річний цикл. При $k = 1$ дата тригера для наступного сезону: $T_{trigger} = 18.10.2024 + 367 - 6 \approx 14.10.2025$. Система завчасно сформує і запропонує клієнту оптовий бандл із зимового асортименту [90].

Сценарій 2. Нерегулярна закупівля. Клієнт C_2 закупив набір $Y = \{\text{чоловічі кросівки Art. 403}\}$ у такі дати: $t_1 = 05.02.2023$, $t_2 = 18.05.2023$, $t_3 = 11.01.2024$. Вектор інтервалів: $\delta_1 = 102$ дні, $\delta_2 = 238$ днів. Розрахунок: $\mu = 170$ днів, $\sigma \approx 96$ днів, $CV \approx 0,56$. Значення $CV > \gamma = 0,5$ вказує на хаотичний характер закупівель – набір Y для клієнта C_2 виключається з проактивних тригерів. Можливе пояснення: закупівля кросівок залежить від кон'юнктурних факторів (наявність акцій, зміна попиту) і не підпорядкована стійкому циклу.

Ці два сценарії наочно демонструють ключову перевагу темпорального аналізу [91]. Використання того підходу дозволяє відокремити системні бізнес-цикли від ситуативних замовлень, дозволяючи зосередити ресурс відділу продажів виключно на тих взаємодіях, де предикативна пропозиція має реальний шанс на успіх.

Таким чином, запропонований ціннісно-зважений темпоральний підхід, математичний апарат якого формалізований виразами (2.3)–(2.18), усуває чотири перших із п'яти наявних обмежень класичних підходів, виявлених у підрозділі 2.1:

«фінансову сліпоту» через заміну частотного критерію на транзакційно-зважену корисність (TWU); проблему комбінаторного вибуху через механізм використання TWU в методі UP-Growth для відновлення властивості антимонотоності; «темпоральну сліпоту» через аналіз IPT -векторів і критерій CV та ідентифікаційну дискретність через наявність детермінованого ID B2B-клієнта.

Зведена характеристика ціннісно-зваженого темпорального аналізу, відносно класичного FP-Growth наведена в таблиці 2.10.

Таблиця 2.10 – Порівняльна характеристика класичного FP-Growth та ціннісно-зваженого темпорального аналізу

Характеристика	FP-Growth	UT-Growth
Критерій значущості патерну	Частота Supp	Фінансова утиліті $U(X)$, грн
Врахування q_k та p_k	Ні	Так
Захист від комбінаторного вибуху	Тільки за рахунок $antimonot. Supp$	TWU + ціннісний фільтр Етапу 1
Аналіз часових циклів	Ні	Так (IPT , μ , σ , CV)
Персоналізація до клієнта	Ні	Так (індивідуальний цикл C_j)
Момент пропозиції	Не визначається	$T_{trigger} = T_{last} + \mu + k\sigma$
Тип результату	Статичне правило $A \rightarrow B$	Динамічний тригер (клієнт · набір · дата)
Застосовність до B2B	Обмежена	Цільова

Результатом роботи конвеєра який складається із двох запропонованих методів, а розростання ціннісних патернів (Utility Pattern Growth) та темпорального аналізу – Utility Pattern Growth + Temporal – (Utility-Temporal Growth UT-Growth) є перелік персоналізованих тригерів вигляду $\langle C_j, X, T_{trigger} \rangle$ для кожного контрагента, кожного фінансово значущого та циклічно стабільного

товарного набору визначено оптимальний момент отримання предикативної пропозиції. Це перетворює B2B-платформу з реактивного каталогу на інтелектуальну систему підтримки продажів, здатну передбачати потреби клієнтів і взаємодіяти з ними у найбільш сприятливий момент.

Таким чином, для додавання темпоральної складової до ціннісно-зваженого підходу розроблено модель темпоральної послідовності та метод темпорального аналізу ціннісно-зважених патернів. Ці розробки складають третій та четвертий пункти наукової новизни.

Третій пункт наукової новизни. Удосконалено модель темпоральної послідовності, яка на відміну представлення у вигляді кортежів ідентифікаційно дискретних транзакцій враховує наявність детермінованого ідентифікатора B2B-та додатково включає час виникнення транзакції, що дозволяє створити вектор міжтранзакційних інтервалів (*IPT*) та використовувати його статистичні характеристики для розрахунку ймовірної дати активації предикативної пропозиції для подолання обмежень «темпоральної сліпоты» та ідентифікаційної дискретності

Четвертий пункт наукової новизни. Отримав подальший розвиток метод темпорального аналізу цінністних патернів UP-Growth, який спільно з моделлю темпоральної послідовності дозволяє розрахувати коефіцієнт варіації (*CV*) вектору міжтранзакційних інтервалів, що дає можливість визначити дату предикативної пропозиції B2B-клієнту формуючі предикативні тригери вигляду «клієнт – товарний набір – дата звернення» для підвищення рівня персоналізації контенту B2B-систем електронної комерції.

2.5 Висновки до другого розділу

Результатом проведення системного дослідження існуючих методів інтелектуального аналізу асоціативних правил та послідовних шаблонів в накопичених масивах B2B транзакцій є встановлення їх *фундаментальних методичних обмежень*, які унеможливають їх безпосереднє застосування для створення персоналізованого контенту у системах електронної комерції.

Обґрунтовано існування п'яти критичних розривів (обмежень) між математичними припущеннями методів пошуку асоціацій та послідовностей та природою B2B-предметної галузі, а саме це «фінансова сліпота» бінарних моделей, наявність «комбінаторного вибуху» при зваженому аналізі, «темпоральна сліпота» статичних моделей, наявність ідентифікаційної дискретності та відсутність диференційованої стратегії взаємодії з клієнтами.

Встановлено, виявлені обмеження існуючих методів інтелектуального аналізу транзакційних даних концептуально взаємопов'язані та їх зв'язок має ієрархічну структуру. Тобто спроба подолати фінансову сліпоту класичних моделей шляхом додавання вартісних показників активує проблему комбінаторного вибуху. Своєю чергою, перехід до аналізу темпоральних інтервалів транзакційних даних вимагає не лише нової архітектури обробки, а й усунення ідентифікаційної дискретності для B2B клієнтів, оскільки розрахунок циклічності можливий лише на цілісних, а не фрагментованих часових рядах. Таким чином, фінальне обмеження, відсутність диференційованої стратегії взаємодії з клієнтами, тобто стратегічна монотонність є результатом дефіциту даних на всіх попередніх етапах аналізу. Тоді в системі електронної комерції B2B не можливо диференціювати (визначати класи) клієнтів без ідентифікації їхніх індивідуальних фінансово-темпоральних патернів.

Прагнення до подолання зазначених обмежень є науковою основою для розробки ціннісно-зваженого темпорального аналізу, математичний апарат якого умовно поділений на дві складові частини: модель ціннісно-зваженої транзакції та метод розростання ціннісних патернів (Utility Pattern Growth), а також модель темпоральної послідовності та метод темпорального аналізу ціннісних патернів UP-Growth.

Модель ціннісно-зваженої B2B-транзакції формалізовано виразами (2.5)-(2.8). До класичного визначення транзакції додано формалізацію сумарної корисності (*Transaction Utility*) – $U(T)$ визначеної транзакції, як суми добутків кількості придбаних одиниць товарних складових p_k на їх індивідуальну ціну q_k . Для подолання ефекту комбінаторного вибуху, який виникає за використання

індикатора сумарної корисності формалізовано визначення транзакційно-зваженої корисності (*Transaction-Weighted Utilization*) – $TWU(X)$, як сумарна вартість усіх транзакцій бази, які містять визначену транзакцію

Показано, що використання методу розростання ціннісних патернів UP-Growth, дозволяє усунути «фінансову сліпоту» класичного підходу до пошуку асоціативних патернів в танзакціних даних через зміну одного ключового параметру. На першому проході при побудові UP-дерева в вузлах накопичується не кількість входжень товарної позиції, а їх грошова вартість. Використання механізм TWU під час другого проходу відновлює властивість антимонотонності і забезпечує обчислювальну стабільність навіть при аналізі довгих B2B-накладних. Результатом роботи методу є впорядкований перелік ціннісних патернів – комбінацій товарів, що формують основну частку виторгу підприємства.

Модель темпоральної B2B-послідовності формалізовано виразами (2.11)-(2.18) До класичного визначення послідовності додано час виникнення транзакцій τ_i , що дозволило отримати вектор міжзакупівельних інтервалів (*Inter-Purchase Time, IPT*) для довільного B2B-клієнта, що спільно з розрахованим безрозмірний показником – коефіцієнтом варіації (*Coefficient of Variation, CV*) дає можливість визначити стабільність циклу транзакцій, долаючи таким чином темпоральну сліпоту та ідентифікаційну дискретність класичного методу пошуку шаблонів в послідовностях клієнтів.

Показано, що використання методу темпорального аналізу ціннісних патернів отриманих за UP-Growth дозволяє усунути «темпоральну сліпоту» класичного підходу до пошуку послідовних патернів в танзакціних даних через побудову IPT -векторів та їх подальший аналіз за критерієм CV для довільного B2B-клієнта. Наведено інтерпретацію критерію CV в залежності від встановленого порогу стабільності γ . Значення порогу γ не є фіксованим і може налаштовуватися залежно від специфіки галузі та прийнятності підприємства до похибки прогнозу. У більшості практичних застосувань рекомендоване значення $\gamma \in [0,3; 0,5]$. Для ціннсно-зважених послідовних патернів, що пройшли перевірку за критерієм CV , розраховується дата предикативної пропозиції B2B клієнту як оптимальний

момент активації комунікаційного тригера $T_{trigger}$ з врахуванням коефіцієнт превентивності $k \in [0; 2]$, що визначає глибину попередження.

Як загальний підсумок розділу показано що використання запропанованого двоетапного конвеєру методів розростання цінністних патернів UP-Growth та їх темпорального аналізу дозволяє отримати перелік персоналізованих тригерів вигляду $\langle C_j, X, T_{trigger} \rangle$ завдяки якому для кожного контрагента та кожної фінансово значущої та циклічно стабільної транзакції визначено оптимальний момент отримання предикативної пропозиції. Як підтвердження наведено детальний розгляд двох типових сценаріїв для B2B систем електронної комерції – «регулярна сезонна закупівля» та «нерегулярна закупівля» Саме можливість їх реалізації перетворює B2B-платформу з реактивного каталогу на інтелектуальну систему підтримки продажів, здатну передбачати потреби клієнтів і взаємодіяти з ними у найбільш сприятливий момент.

3 КЛАСИФІКАЦІЯ В2В-КЛІЄНТІВ ЗА ЦІННІСНО-ЗВАЖЕНИМ ТЕМПОРАЛЬНИМ ПРОФІЛЕМ

Ціннісно-зважений темпоральний підхід до інтелектуального аналізу транзакційно-поведінкових даних В2В клієнтів, який розроблено у другому розділі, формує для кожного клієнта набір стабільних товарних патернів з індивідуальними характеристиками закупівельного циклу: середнім інтервалом μ , відхиленням σ та коефіцієнтом варіації CV .

Ці характеристики є побічним результатом отриманими за методом темпорального аналізу поведінкових послідовностей В2В- клієнтів. Проте вони несуть у собі значно більше аналітичної цінності, ніж просте відсіювання нестабільних патернів.

За використанням цих статистичних характеристик створюється по суті унікальний «поведінковий портрет» кожного В2В-клієнта платформи електронної комерції для відповіді на важливе бізнесове запитання наскільки регулярно та на яку суму здійснює покупки такий клієнт. Мова іде про формування так званого *ціннісно-зваженого темпорального профілю клієнта (UW-TCP Utility-Weighted Temporal Customer Profile)*.

Крім того зауважимо, що у підрозділі 2.3.3 значення коефіцієнта превентивності k у формулі визначення комунікаційного тригера (2.18) є однаковим для всіх клієнтів.

Це означає, що система електронної комерції звертається до преміум-клієнта з великим середнім чеком і до дрібного нерегулярного замовника з однаковою глибиною попередження, що є методично необґрунтованим. Тобто різні типи контрагентів потребують встановлення різних стратегій взаємодії.

Для цінних клієнтів з великим чеком критична точність моменту предикативної пропозиції, тоді як для менш цінних допустиме ширше вікно з більшим охопленням. Формалізація цієї відмінності через класифікацію клієнтів є предметом цього підрозділу.

3.1 Методичні обмеження стандартного *RFM*-аналізу до кластеризації B2B-клієнтів

Класичним методом класифікації клієнтів у комерційних системах є *RFM*-аналіз (*Recency–Frequency–Monetary*), що характеризує клієнта за трьома показниками: давністю (новизною) останньої покупки R , частотою транзакцій F та сумарним грошовим обсягом M .

Більш детально переваги та недоліки *RFM*-аналізу розглянуто в розділі 1 даної роботи. Незважаючи на широке застосування у B2C-сегменті, при спробі прямого використання R - F - M -показників для класифікації клієнтів в B2B-системах електронної комерції знову виникає три критичних обмеження, що впливають із розробленої в Розділі 2 структури інформаційних втрат, яку ми формували при аналізі обмежень класичних методів пошуку асоціативних та поведінкових патернів (див. рис. 2.1). Розглянемо виникаючі обмеження більш детально:

1) при проведенні *RFM*-аналізу використання параметру давності – R дозволяє фіксувати лише час, що минув з останньої транзакції [92]. Але для B2B-клієнта з циклом транзакції в 90 днів (наприклад, сезонне взуття) значення $R = 60$ днів є ознакою «нормального очікування», а для клієнта з тижневим циклом така перерва в послідовності транзакції є ознакою його «втрати». Таким чином, при проведенні *RFM*-аналізу не будується вектор межакупівельних патернів IPT (2.14), що робить його «темпорально сліпим» до індивідуального ритму бізнесу;

2) при проведенні *RFM*-аналізу використання параметру частоти – F дозволяє оцінювати загальну кількість транзакцій за період [93]. Але два B2B-клієнта з однаковим показником $F = 10$ можуть мати принципово різну поведінку, наприклад один замовляє товар щопонеділка з високою стабільністю, а інший зробив 10 замовлень за один тиждень і зник. Таким чином при проведенні *RFM*-аналізу ігнорується стабільність транзакцій, індикатором якої є коефіцієнт варіації CV (2.17), розрахований на основі IPT ;

3) при проведенні *RFM*-аналізу параметр грошового обсягу транзакції – M розраховується як загальна сума витрат клієнта [94]. Але в такому вигляді він не

дає інформацію про структуру транзакцій, тобто склад транзакції містить один дорогий товар, чи багато дрібних SKU. Таким чином при проведенні *RFM*-аналізу не враховується транзакційно-зважена корисність *TWU*, що не дозволяє виділити патерни товарів-драйверів, які забезпечують стабільність обороту. Для подолання зазначених недоліків стандартного *RFM*-аналізу запропоновано розраховувати ціннісно-зважений темпоральний профіль клієнта (*UW-TCP* Utility-Weighted Temporal Customer Profile), який в подальшому використовуємо для кластеризації/класифікації B2B-клієнтів системи електронної комерції. Порівняння *RFM*-аналізу *UW-TCP* наведено в таблиці 3.1.

Таблиця 3.1 – Порівняльна характеристика *RFM*-аналізу профілю *UW-TCP* B2B-клієнта

Критерій	<i>RFM</i> -підхід (B2C)	<i>UW-TCP</i> (B2B)
Ознака «давність»	Дні від останньої покупки R	Відхилення від очікуваної дати циклу
Ознака «частота»	Кількість транзакцій F	Стабільність циклу $CV = \sigma/\mu$
Ознака «цінність»	Сумарна виручка за період M	Середня вартість транзакції $\overline{U}(T)$
Часовий компонент	Ретроспективний (минуле)	Предикативна пропозиція
Прогностичний потенціал	Низький (описує минуле)	Високий (описує майбутню поведінку)
Адаптація до B2B	Часткова	Цільова

Таким чином, заміна *RFM*-ознак на характеристики *UW-TCP*, які вже обчислені у рамках методу темпорального аналізу (підрозділ 2.3), дозволяє побудувати класифікатор, що є природним продовженням методів ціннісно-зваженого темпорального підходу.

3.1.1 Формування ціннісно-зваженого темпорального профілю B2B-клієнтів

Для кожного ідентифікованого клієнта C_j , який має послідовність із щонайменше min_dates стабільних транзакцій хоча б одного ціннісного патерну,

формується індивідуальний вектор ознак $f_j = (\bar{\mu}_j, \overline{CV}_j, \overline{U(T)}_j)$ [95]. До такого аналізу включаються лише ті патерни, які за результатами темпоральної перевірки були визнані достатньо регулярними, тобто мають значення коефіцієнта варіації $CV \leq \gamma$ і можуть бути використані для формування предикативних тригерів. Це означає, що профіль клієнта будується не на всій історії його замовлень, а лише на тих товарних наборах, для яких виявлено стійку повторюваність у часі та економічну значущість для B2B-платформи.

Кожна компонента вектора f_j агрегується по всіх стабільних патернах клієнта. Показник $\bar{\mu}_j$ характеризує узагальнений середній цикл повторних закупівель клієнта, тобто відображає типовий часовий інтервал, через який він повертається до закупівлі значущих товарних наборів. Показник $\overline{CV}_j$ описує середній рівень стабільності таких закупівельних циклів і дозволяє відрізнити клієнтів із регулярною поведінкою від клієнтів із більш хаотичними або ситуативними замовленнями.

Показник $\overline{U(T)}_j$ відображає середню фінансову ємність транзакцій клієнта, пов'язаних зі стабільними ціннісними патернами, і тому характеризує його економічну значущість для платформи. Таким чином, сформований вектор f_j є ціннісно-зваженим темпоральним профілем B2B-клієнта. На відміну від класичних ретроспективних показників сегментації, такий профіль поєднує три важливі характеристики закупівельної поведінки: періодичність, стабільність і фінансову цінність.

Ознака 1. Середній цикл закупівлі $\bar{\mu}_j$ характеризує типову тривалість між повторними закупівлями стабільних патернів клієнта і визначається так [96]:

$$\bar{\mu}_j = \frac{1}{|\mathcal{P}_j^*|} \sum_{X \in \mathcal{P}_j^*} \mu(X, C_j), \quad (3.1)$$

де $\mathcal{P}_j^*$ – множина стабільних ціннісних патернів клієнта C_j , тобто патернів, для яких $CV(X, C_j) \leq \gamma$;

$\mu(X, C_j)$ – середній цикл закупівлі набору X клієнтом C_j , розрахований за виразом (2.15).

Ознака $\bar{\mu}_j$ відображає горизонт планування клієнта: невелике значення $\bar{\mu}_j < 30$ (днів) характерне для клієнтів із частими поповненнями складу, велике $\bar{\mu}_j < 90$ (днів) – для сезонних або великих оптових замовників.

Ознака 2. Середня стабільність циклу $\overline{CV}_j$ характеризує регулярність закупівельної поведінки клієнта і визначається наступним виразом [97]:

$$\overline{CV}_j = \frac{1}{|\mathcal{P}_j^*|} \sum_{X \in \mathcal{P}_j^*} CV(X, C_j), \quad (3.2)$$

де $CV(X, C_j)$ – коефіцієнт варіації інтервалів для патерну X та клієнта C_j , розрахований за виразом (2.17).

Оскільки до множини $\mathcal{P}_j^*$ включаються лише патерни з $CV < \gamma$, значення $\overline{CV}_j$ знаходиться у діапазоні $[0; \gamma]$. Клієнт з $\overline{CV}_j \approx 0$ демонструє жорстко регулярні закупівлі (наприклад, виробничий регламент), клієнт із $\overline{CV}_j \approx \gamma$ – помірно нестабільний цикл.

Ознака 3. Середня вартість транзакції $\overline{U(T)}_j$ характеризує фінансову ємність клієнта і визначається наступним виразом [98]:

$$\overline{U(T)}_j = \frac{1}{|\mathcal{P}_j^*|} \sum_{X \in \mathcal{P}_j^*} \overline{U(T)}_j(X, C_j), \quad (3.3)$$

де $\overline{U(T)}_j(X, C_j)$ – середня вартість транзакції з патерном X , для клієнта C_j ;

$U(T)$ – повна вартість транзакції за виразом (2.5).

Ознака $\overline{U(T)}_j$ відображає типовий обсяг замовлення клієнта у грошовому вимірі та є фінансово зваженим аналогом показника M у RFM. Процес отримання трьох ознак UW-TCP профілю клієнта за запропонованим ціннісно-зваженим

темпоральним конвеєром (див. рис. 2.4) та їх використання для класифікації (кластеризації) ілюструє рисунок 3.1.

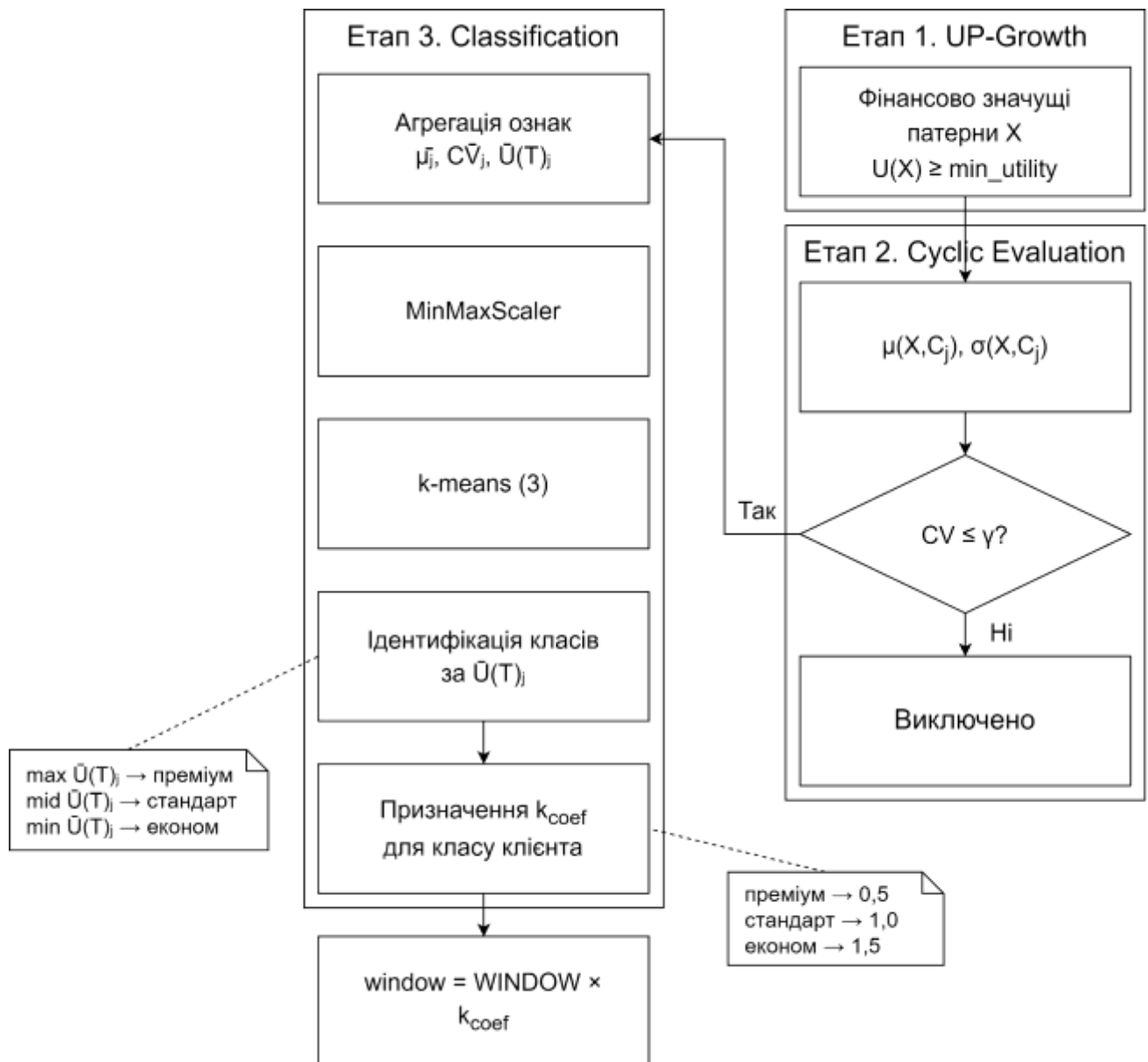


Рисунок 3.1 – Архітектура трьох етапного конвеєра класифікації B2B-клієнтів для формування предикативних пропозицій

3.1.2 Етапи нормалізації ознак профілю та кластеризації B2B-клієнтів

Як бачимо на рисунку 3.1 значення отриманих трьох середніх ознак, а саме циклу закупівлі $\bar{\mu}_j$, коефіцієнта варіації $\overline{CV}_j$ та грошової вартості тарнзакції $\overline{U}(T)_j$,

мають принципово різні масштаби. Значення $\bar{\mu}_j$ вимірюється у днях (порядок десятків), значення $\overline{CV}_j$ є безрозмірним дробом від 0 до 0,5, а значення $\overline{U(T)}_j$ може сягати десятків тисяч гривень. Безпосереднє застосування алгоритму кластеризації в нашому випадку це відомий *k-means*, до масштабно різнорідних ознак призводить до домінування ознаки з найбільшим абсолютним розкидом $\overline{U(T)}_j$, що спотворює результати. Для усунення цього ефекту застосовується мінімаксна нормалізація (*Min-Max Scaling*), коли кожна ознака приводиться до діапазону за наступним виразом, де в якості значень функції f_{ij} використовуються значення $\bar{\mu}_j$, $\overline{CV}_j$, $\overline{U(T)}_j$

$$\tilde{f}_{ij} = \frac{f_{ij} - \min_j(f_i)}{\max_j(f_i) - \min_j(f_i)}, \quad (3.4)$$

де $\tilde{f}_{ij}$ – нормалізоване значення i -ї ознаки для клієнта C_j ;

$\tilde{f}_{ij}$ – вихідне значення;

$\max_j(f_i)$ та $\min_j(f_i)$ – мінімальне та максимальне значення ознаки f_i по всій множині класифікованих клієнтів.

Після нормалізації сформована матриця ознак $\tilde{F} \in \mathbb{R}^{n \times 3}$, де n – кількість клієнтів, яких відібрано ціннісно-зваженим темпоральним конвеєром, передається на вхід алгоритму кластеризації *k-means*. Згідно алгоритму мінімізується сумарна внутрішньокластерна дисперсія за наступним виразом:

$$J = \sum_{l=1}^K \sum_{C_j \in \mathcal{C}_l} \|\tilde{\mathbf{f}}_j - \mathbf{v}_l\|^2, \quad (3.5)$$

де J – цільова функція (сумарна внутрішньокластерна інерція);

$K = 3$ – кількість кластерів;

$\mathcal{C}_l$ – множина клієнтів, що належать до l -го кластера;

$\mathbf{v}_l$ – центроїд l -го кластера;

$\|\cdot\|$ – евклідова відстань у нормалізованому просторі ознак.

Із теорії відомо, що використання алгоритму *k-means* потребує апіорного знання кількості кластерів [99]. В нашому випадку вибір $K = 3$ кластерів обґрунтовується наступним чином. З управлінської точки зору три кластери є природним розподілом для множини B2B клієнтів системи електронної комерції – це *Преміум*, *Стандарт* та *Економ* кластери. Проведені експерименти показали, збільшення кількості кластерів понад 3 при відносно невеликій кількості класифікованих клієнтів (визначається обсягом стабільних патернів) призводить до нестійкої кластеризації. Зменшення до 2 кластерів не дозволяє відокремити середній сегмент зі стандартною стратегією взаємодії.

3.1.3 Визначення індивідуального вікна комунікаційного тригера за класами

За результатами кластеризації кожному B2B-клієнту в системі електронної комерції призначається значення коефіцієнту k [100]. Із виразу (2.28) попереднього розділу який визначає дату активації комунікаційного тригера $T_{trigger}$ впливає, значення суттєво впливає на призначення дати предикативної пропозиції. В підрозділі 2.3.3 дата предикативної пропозиції визначалась на основі розподілу ймовірності повторної закупівлі (див. рис. 2.3). Опис зон вікна предикативної пропозиції наведений в таблиці 2.9. Як бачимо, значення коефіцієнту k призначалося експериментальним чином та не залежало від кластеру B2B-клієнта.

Після проведення процедури кластеризації отримано можливість більш обґрунтованого визначення ширини вікна комунікаційного тригера $window_j$ у формулі перевірки активності тригера за наступним виразом:

$$window_j = WINDOW \times k_j, \quad (3.6)$$

де $WINDOW$ – ширина базового вікна тригера в днях;

k_j – коефіцієнт, призначений клієнту C_j на основі визначеного кластеру.

Ідентифікація кластерів після проведення етапу кластеризації здійснюється автоматично за значенням $\overline{U(T)}$ центроїду, тобто кластер з найвищим значенням $\overline{U(T)}$ визнається як клас преміум, з найнижчим вважається належним до економічного класу [101]. Залежність ширини вікна комунікаційного триггеру від класів B2B-клієнтів та її інтерпретація наведена в таблиці 3.2.

Таблиця 3.2 – Залежність ширини вікна комунікаційного триггеру від класів B2B-клієнтів та її інтерпретація

Клас	Ідентифікація	k	Ширина вікна	Інтерпретація ширини вікна
Преміум	Найвищий $\overline{U(T)}$	0,5	$0,5 \times WINDOW$	Висока цінність угоди вимагає точного потрапляння у момент потреби; зайве звернення поза вікном дорожче хибного мовчання
Стандарт	Середній $\overline{U(T)}$	1,0	$1,0 \times WINDOW$	Базова стратегія взаємодії без коригувань
Економ	Найнижчий $\overline{U(T)}$	1,5	$1,5 \times WINDOW$	Нижча цінність угоди допускає ширше охоплення з дещо меншою точністю

На рисунку 3.2 показаний зв'язок двох стратегій побудови ширини вікна комунікаційного триггера для надання предикативної пропозиції у системі електронної комерції B2B клієнту. Згідно Стратегії 3 використовується вікно фіксованого розміру, за Стратегією 4 ширина вікна формується індивідуально та залежить від класу до якого належить B2B-клієнт за процедурами кластеризації та ідентифікації. Як бачимо, відмінність розрахунку дати надання предикативної пропозиції є суттєвою.

Згідно Стратегії 3 використовується вікно фіксованої ширини, наприклад 14 днів $WINDOW = 14$ днів. За Стратегією 4 використовується індивідуальна ширина вікна від 7 до 21 дня. В результаті для клієнтів класу *Преміум* підвищується точність предикативної пропозиції, а для клієнтів класу *Економ* підвищується

тривалість можливості її надання. Це збільшує зони охоплення економічних клієнтів. Для клієнтів класу *Стандарт* ситуація з влучністю надання предикативної пропозиції не змінюється. На останок, потребує пояснень «дивні» номери Стратегій 3 та 4, які використовуються.

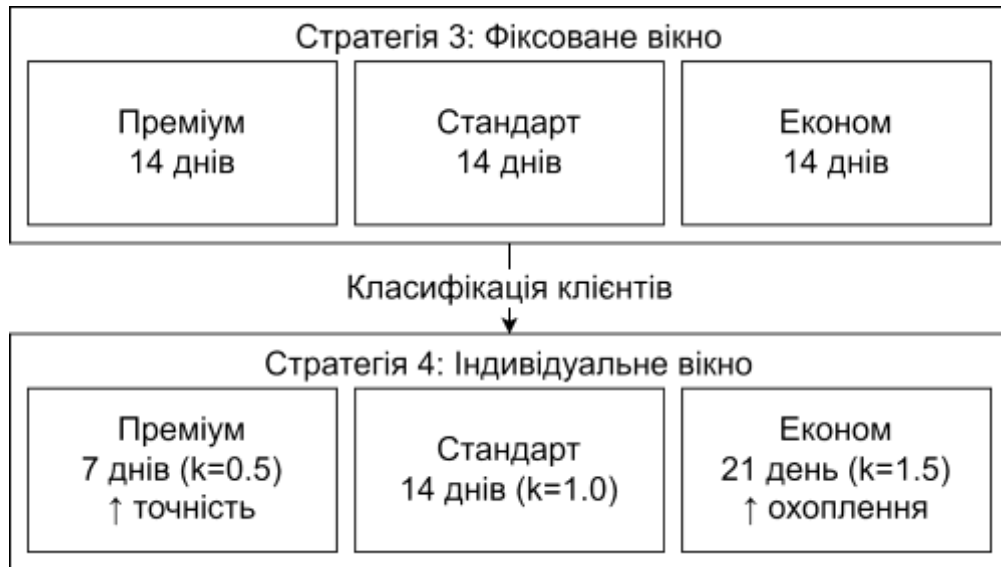


Рисунок 3.2 – Вплив ширини вікна комунікаційного триггеру на точність та охоплення щодо подання предикативної пропозиції B2B-клієнтам

3.1.4 Покроковий опис методу класифікації B2B-клієнтів для формування предикативної пропозиції

Аналіз процесу отримання ознак UW-TCP профілю B2B-клієнта їх нормалізації, а також процедур кластеризації та ідентифікації для класифікації B2B-клієнтів з метою інтеграції її результатів у систему формування комунікаційних тригерів дозволив формально описати кроки методу класифікації B2B клієнтів з метою формування предикативної пропозиції на основі ціннісно-зваженого темпорального аналізу транзакційно-поведінкових даних B2B-клієнтів систем електронної комерції. Перелік кроків методу наведений в таблиці 3.3.

Запропонований метод класифікації B2B клієнтів для формування предикативних пропозицій є третьою складовою ціннісно-зваженого

темпорального аналізу транзакційно-поведінкових даних B2B клієнтів системи електронної комерції.

Таблиця 3.3 – Опис кроків методу класифікації B2B клієнтів

№ кроку	Опис дії	Вхідні дані	Результат
1	Отримання переліку персоналізованих тригерів вигляду $\langle C_j, X, T_{trigger} \rangle$	Результати UT-Growth ($CV \leq \gamma$)	Множина $\mathcal{P}_j^*$ для кожного C_j
2	Обчислення ознак за виразами (3.1)–(3.3)	$\mathcal{P}_j^*$, значення μ , CV , $U(T)$	Матриця $\mathbf{F} \in \mathbb{R}^{n \times 3}$
3	Нормалізація за виразом (3.4)	$\mathbf{F}$	Нормалізована матриця $\tilde{\mathbf{F}}$
4	Кластеризація k-means ($K = 3$) за виразом (3.5)	$\tilde{\mathbf{F}}$	Мітки кластерів для кожного C_j
5	Ідентифікація кластерів за $\overline{U(T)}$ центроїдів	Центроїди кластерів	Класи – <i>Преміум</i> / <i>Стандарт</i> / <i>Економ</i>
6	Призначення k_j за таблицею 3.2	Мітки кластерів	Словник $C_j \rightarrow k_j$
7	Розрахунок ширини індивідуального вікна за виразом (3.7)	k_j , $WINDOW$	$window_j$ для кожного клієнта
8	Перевірка активності тригера з $window_j$	T_{last} , μ , $window_j$	Фінальний перелік предикативних пропозицій

Він функціонує виключно як надбудова над запропонованими методом розростання цінністних патернів (Utility Pattern Growth) та методом їх темпорального аналізу та не вимагає жодних додаткових даних поза тими, що вже обчислені за цими методами. Нові ознаки $\bar{\mu}_j$, CV_j та $\overline{U(T)}_j$, які складають ціннісно-зважений темпоральний профіль B2B клієнта, є агрегатами вже наявних проміжних результатів, що робить метод класифікації B2B-клієнтів третім етапом конвеєра формування предикативних пропозицій для B2B-клієнтів В цьому випадку можна казати про обчислювальну ефективність та архітектурну цілісність використання цього конвеєру в системі електронної комерції.

Відмінність запропонованого методу класифікації B2B-клієнтів від стандартного *RFM*-аналізу запропонованого методу класифікації B2B-клієнтів є принциповою: замість використання ретроспективних показників (давність, частота, сума) створюється ціннісно-зважений темпоральний профіль B2B клієнта за характеристиками (прогнозований цикл, стабільність, фінансова ємність), що в подальшому використовується для визначення класів клієнтів з метою їх інтеграції у систему формування комунікаційних тригерів [102].

Це забезпечує змістовний зв'язок між класом клієнта та стратегією взаємодії з ним через математично обґрунтоване призначення ширини вікна тригеру відповідно до цінності клієнта а не через суб'єктивне визначення порогів сегментації як в стандартному *RFM*-аналізі

Таким чином, для додавання складової класифікації B2B-клієнтів до розробленого ціннісно-зваженого темпорального аналізу транзакційно-поведінкових даних розроблено метод класифікації B2B-клієнтів яких базується на сформованому ціннісно-зваженому темпоральному профілі клієнта. Ці розробки складають п'ятий пункт наукової новизни.

П'ятий пункт наукової новизни. Вперше розроблено метод класифікації B2B клієнтів систем електронної комерції, який на відміну від існуючого *RFM*-аналізу базується на ціннісно-зваженому темпоральному профілі клієнта сформованому у результаті реалізації двоетапного конвеєру аналізу транзакційно-поведінкових даних B2B клієнта. Це дозволило реалізувати трьох етапний конвеєр формування індивідуальних предикативних пропозицій в залежності від визначеного класу клієнта та підвищити фінальну точність (своєчасність) їх надання в процесі персоналізації контенту.

3.2 Висновки до третього розділу

Результатом проведення системного дослідження можливостей класичного *RFM*-аналізу (*Recency–Frequency–Monetary*), за яким для сегментації клієнтів в B2C-системах електронної комерції використовують показники: давності

останньої покупки (R), частоти транзакцій (F) та сумарного грошового обсягу (M) показано, що при спробі прямого використання цих показників для класифікації B2B-клієнтів виникають три критичних обмеження. Тобто використання параметру давності R дозволяє фіксувати лише час, що минув з останньої транзакції. При цьому не будується вектор межакупівельних патернів IPT (2.14), що робить показник R «темпорально сліпим» до індивідуального ритму B2B-бізнесу.

Використання параметру частоти F дозволяє оцінювати загальну кількість транзакцій за період. При цьому ігнорується стабільність транзакцій, індикатором якої є коефіцієнт варіації CV (2.17), розрахований на основі IPT . Параметр грошового обсягу транзакції M розраховується як загальна сума витрат клієнта. Але в такому вигляді він не дає інформацію про структуру транзакцій. При цьому не враховується транзакційно-зважена корисність (TWU) за виразом (2.8), що не дозволяє виділити патерни товарів-драйверів, які забезпечують стабільність обороту.

Для подолання зазначених недоліків стандартного RFM -аналізу запропоновано розраховувати ціннісно-зважений темпоральний профіль клієнта ($UW-TCP$ Utility-Weighted Temporal Customer Profile), який формалізовано виразами (3.1)-(3.3).

Для формування предикативних пропозицій для B2B клієнтів розроблено метод класифікації, який є третьою складовою ціннісно-зваженого темпорального аналізу транзакційно-поведінкових даних B2B клієнтів системи електронної комерції. Він функціонує виключно як надбудова над запропонованими методом розростання цінністних патернів (Utility Pattern Growth) та методом їх темпорального аналізу (Utility-Temporal Growth UT-Growth) і не вимагає жодних додаткових даних поза тими, що вже обчислені за цими методами.

Нові ознаки $\bar{\mu}_j$, CV_j та $\overline{U(T)}_j$, які складають ціннісно-зважений темпоральний профіль B2B клієнта, є агрегатами вже наявних проміжних результатів, що робить метод класифікації B2B-клієнтів третім етапом конвеєра формування предикативних пропозицій для B2B-клієнтів. Визначення номерів класів B2B-

клієнтів інтегрується у систему формування комунікаційних тригерів. Це забезпечує змістовний зв'язок між класом клієнта та стратегією взаємодії з ним через математично обґрунтоване призначення ширини вікна тригеру відповідно до цінності клієнта а не через суб'єктивне визначення порогів сегментації як в стандартному *RFM*-аналізі.

4 ДОСЛІДЖЕННЯ ЕФЕКТИВНОСТІ РЕКОМЕНДАЦІЙ ДЛЯ ПЕРСОНАЛІЗАЦІЇ КОНТЕНТУ НА B2B ПЛАТФОРМІ ЕЛЕКТРОНОЇ КОМЕРЦІЇ

У четвертому розділі виконано експериментальне дослідження ефективності запропонованого підходу до формування предиктивних B2B-рекомендацій для персоналізації контенту на платформі електронної комерції. Дослідження проведено на основі історичного масиву замовлень B2B-системи електронної комерції «База взуття» за 2022–2025 роки. У розділі наведено характеристику експериментального набору даних, описано його попередню обробку, формування транзакційного подання, розрахунок фінансової корисності товарних позицій та хронологічний поділ на навчальну і тестову частини. Для оцінювання рекомендацій використано метрики, що відображають як їх підтвердження майбутньою закупівлею, так і економічну доцільність для B2B-платформи.

Експеримент побудовано як послідовне порівняння базового частотного, ціннісно-зваженого, ціннісно-темпорального та інтегрованого сценаріїв. Це дозволило оцінити ефективність повної моделі та внесок її окремих компонентів: фінансового зважування товарних патернів, темпоральної перевірки закупівельних циклів і врахування класу B2B-клієнта.

Отримані результати показали, що запропонований підхід забезпечує перехід від масового формування частотних рекомендацій до меншої кількості більш точних, своєчасних і фінансово ефективних предиктивних рекомендацій. Додатково проведено аналіз чутливості темпоральних параметрів моделі та обґрунтовано вибір компромісних налаштувань для практичного використання.

4.1 Характеристика експериментальних даних та їх попередня обробка

Експериментальна перевірка запропонованих моделей і методів виконувалася на основі історичного масиву замовлень B2B-платформи системи електронної комерції «База взуття» за 2022–2025 роки. Дані мають рядкову

структуру: один запис відповідає окремій товарній позиції в межах замовлення, а саме замовлення може містити одну або декілька товарних позицій. Тому в експериментах розрізнялися два рівні подання даних: рівень рядка замовлення та рівень транзакції. Формально один запис початкового масиву можна подати у вигляді:

$$r = \langle t, TID, C_j, i_k, n_k, s_k, p_k, q_k, S_T \rangle \quad (4.1)$$

де t – дата і час замовлення;

TID – номер замовлення;

C_j – клієнт;

i_k – артикул товару;

n_k – назва товару;

s_k – розмірний ряд;

p_k – ціна товарної позиції;

q_k – кількість одиниць у замовленні;

S_T – загальна сума замовлення.

Початковий масив містить 106378 рядків товарних позицій, які відповідають 29352 В2В-замовленням. У даних представлено 2525 клієнтів та 56197 унікальних артикулів. Період спостереження охоплює майже чотири роки – з 01.01.2022 року до 30.12.2025 року. В таблиці 4.1 наведена загальна характеристика експериментального масиву за роками.

Таблиця 4.1 – Загальна характеристика набору даних за роками

Рік	Позицій	Замовлень	Клієнтів	Артикулів	Період даних	Сума, грн
2022	27342	7666	1123	15639	01.01.2022–31.12.2022	91045692,02
2023	29808	8326	993	17356	02.01.2023–31.12.2023	116742223,15
2024	27694	7587	903	16630	01.01.2024–30.12.2024	113265951,16
2025	21534	5773	702	13394	01.01.2025–30.12.2025	95223373,19
Разом	106378	29352	2525	56197	01.01.2022–30.12.2025	416277239,52

Для більш детального аналізу часової структури даних було побудовано помісячну динаміку кількості замовлень і суми замовлень. Відповідний графік наведено на рисунку 4.1.

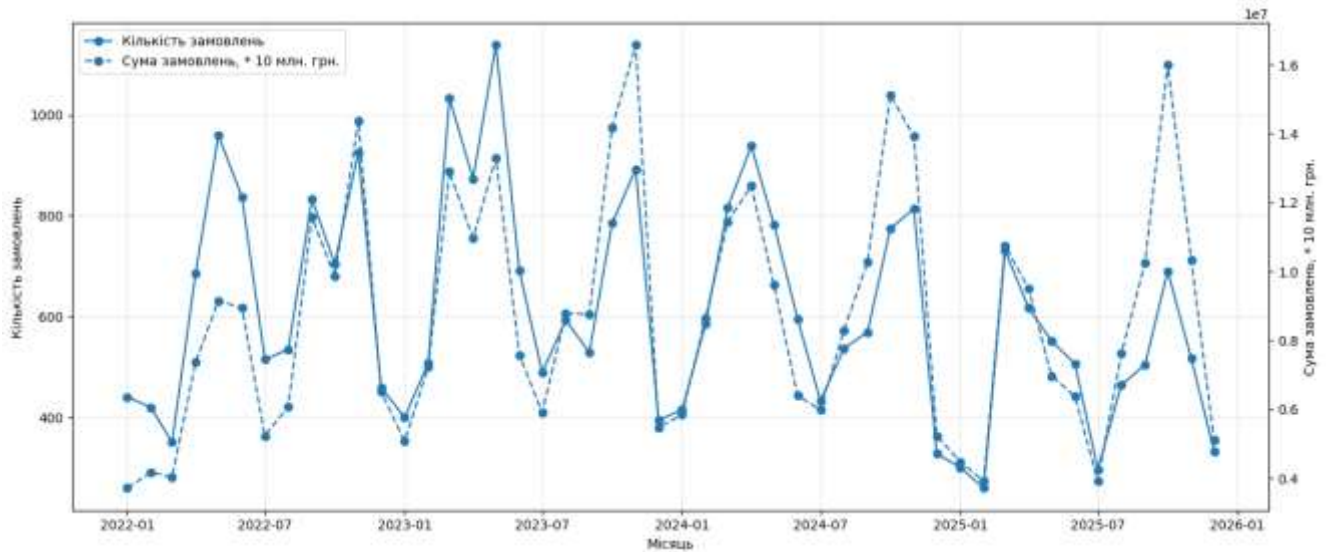


Рисунок 4.1 – Помісячна динаміка кількості B2B-замовлень та суми замовлень

Рисунок дозволяє оцінити не лише загальну зміну активності за роками, а й внутрішньорічні коливання. Для B2B-даних це важливо, оскільки закупівельна активність може залежати від сезонності, виробничих циклів, складських поповнень і періодів підвищеного попиту [103]. Помісячна динаміка також показує, що кількість замовлень і сума замовлень змінюються не завжди синхронно: окремі місяці можуть мати відносно невелику кількість транзакцій, але значну суму через великі оптові замовлення. Для візуального аналізу сезонних коливань було додатково побудовано теплову карту суми B2B-замовлень за роками та місяцями (рис. 4.2).

Кожна клітинка теплової карти відповідає сумі замовлень у певному місяці конкретного року, що дає змогу швидко визначити періоди концентрації виручки, порівняти місяці між роками та виявити можливі сезонні закономірності. Для задачі формування предиктивних B2B-рекомендацій це важливо, оскільки рекомендації мають бути не тільки релевантними за товарним складом, а й своєчасними відносно закупівельного циклу клієнта.

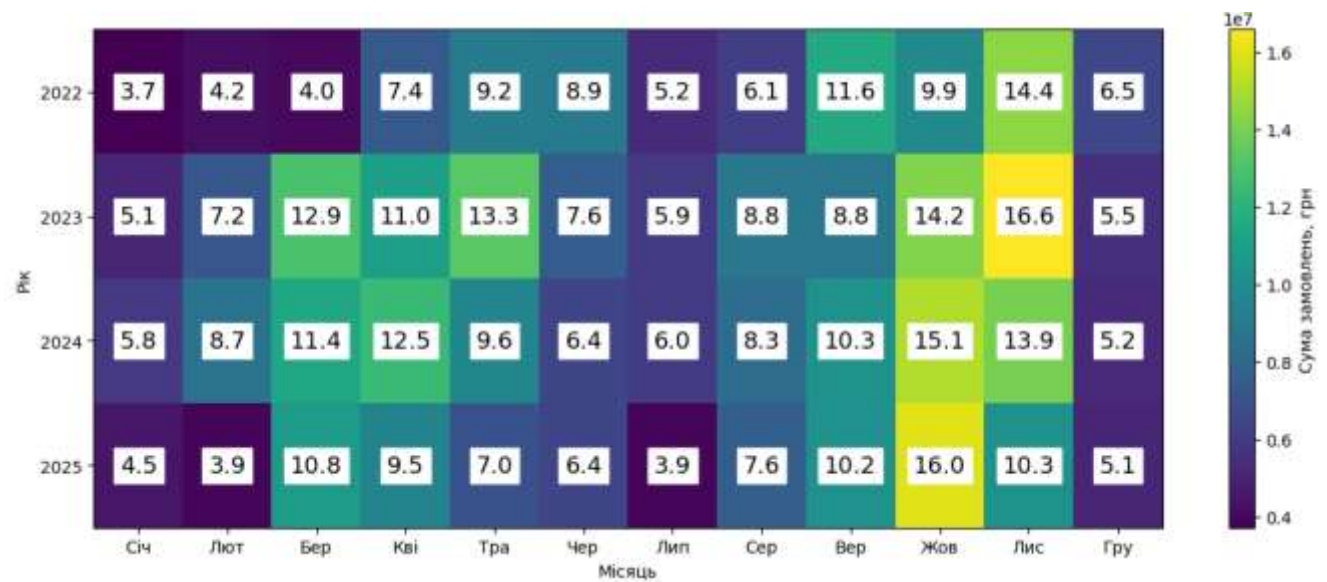


Рисунок 4.2 – Теплова карта суми B2B-замовлень за роками та місяцями

На рівні замовлення середня сума транзакції становить 14182,24 грн, медіанна сума – 7320,00 грн. Середня кількість унікальних артикулів у замовленні становить 3,62, медіанне значення – 2 артикули [104]. Максимальна кількість унікальних артикулів в одному замовленні дорівнює 252, що вказує на наявність окремих дуже великих оптових транзакцій. В таблиці 4.2 наведені статистичні характеристики замовлень.

Таблиця 4.2 – Статистичні характеристики замовлень в наборі даних

Показник	Значення
Кількість замовлень	29352
Середня сума замовлення, грн	14182,24
Медіанна сума замовлення, грн	7320,00
Середня кількість артикулів у замовленні	3,62
Медіанна кількість артикулів у замовленні	2
Максимальна кількість артикулів у замовленні	252
Загальна сума замовлень, грн	416277239,52

Перед проведенням експериментів дані були приведені до узгодженого аналітичного вигляду. Попередня обробка не змінювала економічний зміст даних,

а забезпечувала їх підготовку до подальшого частотного, ціннісно-зваженого та темпорального аналізу. Основні операції попередньої обробки наведено в таблиці 4.3.

Таблиця 4.3 – Основні етапи попередньої обробки даних

Етап	Операція	Результат
1	Перевірка структури даних	Визначено наявність полів, необхідних для експериментів
2	Уніфікація типів даних	Дати, ідентифікатори та числові поля приведено до коректного формату
3	Очищення записів	Вилучено записи з некоректними значеннями
4	Розрахунок вартості позиції	Визначено фінансовий внесок кожного товару
5	Формування транзакцій	Рядки одного замовлення об'єднано в товарний набір
6	Формування історій клієнтів	Замовлення згруповано за клієнтами та датами
7	Фільтрація довгих транзакцій	Залишено транзакції, придатні для пошуку патернів
8	Хронологічне впорядкування	Підготовлено дані до навчальної і тестової частин

На етапі перевірки структури контролювалася наявність ключових атрибутів: дати замовлення, номера замовлення, клієнта, артикулу товару, назви товару, ціни, кількості та суми замовлення. Дата замовлення приводилася до формату дати і часу; номер замовлення, клієнт та артикул товару розглядалися як текстові ідентифікатори; ціна, кількість і сума замовлення приводилися до числового формату. Така уніфікація була необхідна для коректного групування рядків, розрахунку фінансових показників і хронологічного впорядкування транзакцій. Оскільки в межах одного замовлення може бути декілька рядків, для побудови транзакційного подання всі записи з однаковим номером замовлення (TID) об'єднувалися в одну транзакцію:

$$T = \{i_1, i_2, \dots, i_m\}, \quad (4.2)$$

де m – кількість унікальних артикулів у замовленні.

Для кожної товарної позиції додатково розраховувалася її локальна фінансова корисність:

$$u(i_k, T) = p_k \cdot q_k. \quad (4.3)$$

Загальна корисність транзакції на рівні товарних позицій визначалася як:

$$U(T) = \sum_{i_k \in T} u(i_k, T). \quad (4.4)$$

Такий розрахунок потрібний для ціннісно-зваженого аналізу, оскільки класичний частотний підхід враховує лише факт входження артикулу до замовлення, тоді як запропонований підхід додатково враховує фінансовий внесок кожної позиції.

Для оцінювання складності транзакцій було проаналізовано розподіл замовлень за кількістю унікальних артикулів [105]. Як показано в таблиці 4.4, більшість замовлень мають відносно невеликий товарний склад: 35,59 % замовлень містять один артикул, 21,63 % – два артикули, 26,53 % – від трьох до п'яти артикулів. Водночас лише 2,79 % замовлень містять понад 15 артикулів, але на них припадає 17,52 % сумарної виручки.

Таблиця 4.4 – Розподіл замовлень за кількістю артикулів в наборі даних

Кількість артикулів у замовленні	Кількість замовлень	Частка замовлень, %	Сума замовлень, грн	Частка суми, %
1 артикул	10445	35,59	47628077,03	11,44
2 артикули	6348	21,63	51098708,05	12,28
3–5 артикулів	7787	26,53	112290690,19	26,97
6–10 артикулів	2996	10,21	85148294,03	20,45
11–15 артикулів	957	3,26	47199042,11	11,34
Понад 15 артикулів	819	2,79	72912428,11	17,52
Разом	29352	100,00	416277239,52	100,00

Для пошуку товарних патернів було використано обмеження на максимальну довжину транзакції. До алгоритмічного аналізу включалися замовлення, що містять не більше 15 унікальних артикулів.

Це обмеження зменшує комбінаторну складність побудови товарних наборів, оскільки кількість можливих комбінацій зростає експоненційно зі збільшенням кількості товарів у транзакції [106].

Після фільтрації довгих транзакцій експериментальний масив містив 28 533 замовлення, 86 375 рядків товарних позицій, 2 508 клієнтів та 46 766 артикулів (табл. 4.5).

Таблиця 4.5 – Характеристика даних після фільтрації довгих транзакцій

Показник	До фільтрації	Після фільтрації ($m \leq 15$)
Рядків товарних позицій	106378	86375
Замовлень	29352	28533
Клієнтів	2525	2508
Унікальних артикулів	56197	46766
Сума замовлень, грн	416277239,52	343364811,41
Частка збережених замовлень, %	100,00	97,21
Частка збереженої суми, %	100,00	82,48

Окремо було проаналізовано активність клієнтів. Дані мають виражену B2B-специфіку: майже половина клієнтів має лише одне замовлення, проте основна частина виручки формується невеликою групою активних клієнтів [107]. Зокрема, 124 клієнти, які мають понад 50 замовлень, забезпечують 71,80 % загальної суми замовлень.

В таблиці 4.6 наведений розподіл клієнтів набору даних за кількістю зроблених замовлень.

Такий розподіл підтверджує доцільність використання ціннісно-зваженого підходу [108]. Якщо орієнтуватися лише на частоту появи товарів або кількість замовлень, то економічна значущість різних клієнтів і товарних патернів може бути

втрачена. Для B2B-сценарію більш важливим є не тільки факт повторення товарного набору, а й його внесок у виручку та регулярність повторної закупівлі.

Перед проведенням експериментів було виконано перевірку якості даних. У ключових полях, необхідних для побудови транзакційного подання, пропущені значення не виявлені. Також не виявлено повних дублікатів рядків і повторів однакової пари «замовлення – артикул».

Таблиця 4.6 – Розподіл клієнтів за кількістю замовлень

Група клієнтів	К-сть клієнтів	Частка клієнтів, %	Сума замовлень, грн	Частка суми, %
1 замовлення	1203	47,64	8716334,00	2,09
2–3 замовлення	536	21,23	10299326,00	2,47
4–10 замовлень	383	15,17	21340744,02	5,13
11–50 замовлень	279	11,05	77027258,03	18,50
Понад 50 замовлень	124	4,91	298893577,47	71,80
Разом	2525	100,00	416277239,52	100,00

Для оцінки прогностичної якості запропонованого конвеєру застосовано схему хронологічного поділу вхідних даних (Time-based Split) на навчальну та тестову вибірки [109]. Альтернативний підхід – випадкове розбиття (Random Split) є принципово неприйнятним у даному контексті. При випадковому розбитті частина «майбутніх» транзакцій потрапляє у навчальну вибірку, що спотворює отриману оцінку через ефект витоку даних у майбутнє (Data Leakage) [110]. У задачі аналізу транзакціо-поведінкових даних таке розбиття суперечить самій природі прогнозування майбутніх транзакцій.

Схема розбиття Time-based Split імітує реальні умови промислового застосування конвеєру, якій досліджується. Для класифікації B2B клієнтів модель навчається виключно на даних минулого і перевіряється на даних «майбутнього», яких вона не бачила під час навчання. Реалізація поділу ґрунтується на хронологічному сортуванні унікальних транзакцій за датою із відсіченням останніх 20% для тестування.

Навчальна вибірка (Train) є базою для побудови патернів. На ній запускаються (навчаються) процедури всіх етапів конвеєра та формуються переліки предикативних пропозицій – системи рекомендацій. Тестова вибірка (Test) виконує роль «еталонного майбутнього». Вона містить реальні замовлення грудня 2024 - грудня 2025 року і дозволяє перевірити, чи дійсно рекомендований метод передбачив те, що клієнти зрештою придбали.

Таким чином, для навчання та формування рекомендацій використовувалося 80 % транзакцій, а останні 20 % транзакцій використовувалися як тестовий горизонт. Навчальна частина містила 22826 замовлень, тестова – 5707 замовлень (табл. 4.7).

Тестова частина використовувалася для оцінювання фактичної результативності рекомендацій. Для кожної рекомендації, сформованої за навчальною історією, перевірялося, чи з'являється відповідний клієнт і рекомендований товарний набір у тестовому горизонті. Це дозволило оцінювати не тільки кількість сформованих рекомендацій, але й їхню практичну корисність за показниками конверсії, успішних рекомендацій, охопленої виручки та середньої виручки на одну рекомендацію.

Таблиця 4.7 – Хронологічний поділ експериментальних даних

Частина даних	Замовлень	Позицій	Клієнтів	Артикулів	Період даних	Сума, грн
Навчальна	22826	68306	2171	37231	01.01.2022– 17.12.2024	262239575,22
Тестова	5707	18069	713	11316	17.12.2024– 30.12.2025	81125236,19
Разом	28533	86375	2508	46766	01.01.2022– 30.12.2025	343364811,41

Таким чином, експериментальні дані мають такі технічні особливості:

1) високий рівень розрідженості товарної номенклатури – 56197 унікальних артикулів;

2) нерівномірний розподіл клієнтської активності – 4,91 % клієнтів формують 71,80 % виручки;

3) транзакційну структуру замовлень – у середньому 3,62 артикулу на замовлення;

4) наявність великих багатопозиційних замовлень, що потребують обмеження довжини транзакції;

5) часову природу даних, яка потребує хронологічного train/test-поділу;

6) необхідність одночасного врахування частоти, фінансової ваги та циклічності закупівель.

Саме ці властивості обґрунтовують використання в роботі не лише класичних частотних методів пошуку товарних наборів, а й ціннісно-зваженого аналізу, темпоральної перевірки повторюваності закупівель та класифікації B2B-клієнтів за ціннісно-темпоральним профілем.

4.2 Метрики оцінювання ефективності рекомендацій

Для оцінювання ефективності сформованих рекомендацій було використано метрики, що характеризують як факт підтвердження рекомендації майбутньою закупівлею, так і її економічну доцільність для B2B-платформи [111].

У межах експерименту рекомендація вважалася успішною, якщо в тестовому горизонті відповідний клієнт здійснював закупівлю рекомендованого товарного набору або його релевантної частини [112]. Основні метрики оцінювання наведено в таблиці 4.8.

Для опису метрик було використано такі базові величини: R – загальна кількість сформованих рекомендацій; H – кількість успішних рекомендацій; N_{test} – кількість транзакцій у тестовому горизонті; V_{test} – загальна виручка тестового горизонту; V_H – виручка від тестових угод, охоплених успішними рекомендаціями.

Конверсія рекомендацій є основною метрикою точності комунікаційних звернень. Вона показує, яка частина сформованих системою рекомендацій справді привела до майбутньої закупівлі. Для B2B-сценарію ця метрика є важливою,

оскільки надмірна кількість нерелевантних рекомендацій створює додаткове навантаження на менеджерів і може знижувати якість взаємодії з клієнтами.

Таблиця 4.8 – Метрики оцінювання ефективності рекомендацій

Метрика	Формула	Інтерпретація
Конверсія рекомендацій	$CNV = \frac{H}{R} \cdot 100\%$	Показує частку сформованих рекомендацій, які підтвердилися фактичною закупівлею
Охоплена виручка	$V_H = \sum S_T$	Визначає суму тестових замовлень, що були охоплені успішними рекомендаціями
Середня виручка на одну рекомендацію	$ARPR = \frac{V_H}{R}$	Показує середній фінансовий результат однієї сформованої рекомендації
Частка охоплених тестових транзакцій	$HitRate = \frac{H}{N_{test}} \cdot 100\%$	Показує, яку частку майбутніх тестових транзакцій передбачено рекомендаціями
Частка охопленої тестової виручки	$VC = \frac{V_H}{V_{test}} \cdot 100\%$	Показує, яку частку загальної тестової виручки забезпечили успішно передбачені замовлення

Охоплена виручка V_H використовується для оцінювання фінансового результату рекомендацій. Вона показує не просто кількість успішних збігів, а сумарну вартість тих майбутніх замовлень, які були передбачені системою. Це важливо, оскільки в B2B-даних різні угоди можуть суттєво відрізнятися за сумою.

Метрика $ARPR$ характеризує середню виручку на одну сформовану рекомендацію. Вона дозволяє оцінити економічну ефективність одного звернення до клієнта. Якщо модель формує менше рекомендацій, але кожна з них у середньому охоплює більшу виручку, така модель може бути більш корисною для практичного використання.

Метрика $HitRate$ показує ширину охоплення майбутніх транзакцій. Вона дозволяє оцінити, яку частину фактичних тестових замовлень система змогла передбачити. Водночас у B2B-сценарії цю метрику не слід розглядати окремо від конверсії, оскільки надмірне збільшення кількості рекомендацій може підвищити охоплення, але знизити якість звернень.

Метрика VC є фінансовим аналогом $HitRate$. Вона показує, яку частку тестової виручки становлять замовлення, охоплені успішними рекомендаціями. Для B2B-платформи ця метрика є особливо важливою, оскільки невелика кількість правильно передбачених великих замовлень може мати більшу економічну цінність, ніж значна кількість дрібних угод.

Таким чином, для оцінювання рекомендацій використовувалися дві групи показників. Перша група характеризує результативність рекомендацій з погляду збігу з майбутніми закупівлями: CNV та $HitRate$. Друга група характеризує економічний ефект: V_H , $ARPR$ та VC . У подальших експериментах саме ці метрики використовуються для порівняння базового частотного підходу з ціннісно-зваженим, темпоральним та інтегрованим підходами.

4.3 Сценарії проведення комплексного експерименту

Комплексний експеримент проводився для перевірки ефективності запропонованого підходу до формування предиктивних B2B-рекомендацій. Його основна мета полягала у послідовному порівнянні декількох сценаріїв формування рекомендацій: від базового частотного підходу до повної моделі, яка враховує фінансову значущість товарних патернів, повторюваність закупівель у часі та клас B2B-клієнта. Експеримент будувався як єдиний порівняльний протокол. Для всіх сценаріїв використовувалися однакові підготовлені дані, однаковий навчальний період, однаковий тестовий горизонт та однаковий набір метрик. Це дозволило порівнювати саме логіку формування рекомендацій, а не різні умови проведення експерименту. Загальну характеристику сценаріїв наведено в таблиці 4.9.

Сценарій S_0 використовувався як базовий рівень порівняння. У цьому сценарії з навчальної вибірки виділялися часті товарні набори. Рекомендації формувалися на основі того, які артикули найчастіше з'являлися разом у транзакціях. При цьому не враховувалися ні вартість товарних позицій, ні закупівельні цикли клієнтів, ні відмінності між класами клієнтів. Таким чином, сценарій S_0 відображає класичну логіку асоціативних рекомендацій.

Сценарій S_1 розширював базовий підхід за рахунок урахування фінансової значущості товарних патернів. На цьому етапі товарні набори оцінювалися не тільки за частотою появи, а й за сумарною вартістю транзакцій, у яких вони траплялися. Це дозволяло відбирати не просто часті, а фінансово значущі патерни, що є важливим для B2B-платформи.

Таблиця 4.9 – Сценарії проведення комплексного експерименту

Сценарій	Назва	Що перевіряється
S_0	Базовий частотний підхід	Чи достатньо лише частоти появи товарних наборів для формування результативних B2B-рекомендацій
S_1	Ціннісно-зважений підхід	Чи покращує результат урахування фінансової ваги товарних патернів
S_2	Ціннісно-темпоральний підхід	Чи покращує результат додавання перевірки повторюваності закупівель у часі
S_3	Інтегрований підхід	Чи покращує результат адаптація рекомендацій до класу B2B-клієнта

Сценарій S_2 додавав до ціннісно-зваженого підходу темпоральну перевірку. Для кожного клієнта аналізувалася історія повторних закупівель ціннісно значущих патернів.

Якщо закупівлі певного набору повторювалися з відносно стабільними інтервалами, для такого клієнта формувався предиктивний тригер. У цьому сценарії рекомендація мала не лише товарний зміст, а й часову прив'язку.

Сценарій S_3 відповідав повній моделі. У ньому додатково враховувався клас B2B-клієнта. Клас клієнта використовувався для налаштування ширини комунікаційного вікна. Для клієнтів із вищою фінансовою значущістю та стабільнішою поведінкою застосовувалося точніше вікно контакту, а для менш стабільних клієнтів – ширше. Таким чином, сценарій S_3 перевіряв доцільність персоналізації не тільки за товарним набором, а й за типом клієнта.

Послідовність проведення комплексного експериментального дослідження наведено в таблиці 4.10.

Таблиця 4.10 – Етапи проведення комплексного експерименту

Етап	Дія	Результат
1	Завантаження підготовленої навчальної вибірки	Масив транзакцій для побудови патернів
2	Побудова частих товарних наборів	Набір патернів для сценарію S_0
3	Розрахунок фінансової значущості патернів	Набір ціннісно значущих патернів для сценарію S_1
4	Аналіз повторюваності закупівель у часі	Набір патернів із закупівельними циклами для сценарію S_2
5	Формування клієнтських профілів	Ознаки для поділу клієнтів на класи
6	Визначення класів клієнтів	Клієнтські групи для сценарію S_3
7	Формування рекомендацій у кожному сценарії	Списки рекомендацій R_0, R_1, R_2, R_3
8	Перевірка рекомендацій на тестовому горизонті	Визначення успішних і неуспішних рекомендацій
9	Розрахунок метрик для кожного сценарію	Значення $CNV, ARPR, HitRate, VC, V_H$
10	Порівняння сценаріїв	Визначення впливу кожного додаткового компонента

На першому етапі використовувалася підготовлена навчальна вибірка. На цьому етапі тестові дані не залучалися. Це забезпечувало коректність перевірки, оскільки всі патерни, правила та клієнтські характеристики формувалися тільки на основі історичних даних.

На другому етапі виконувався базовий частотний пошук товарних наборів. Його результатом був перелік патернів, які часто зустрічалися в навчальних транзакціях.

Саме ці патерни використовувалися для формування рекомендацій у сценарії S_0 . Для кожного потенційного клієнта рекомендація формувалася без урахування вартості замовлення та без перевірки моменту повторної закупівлі.

На третьому етапі для знайдених товарних патернів розраховувалася фінансова значущість. Патерни з недостатнім економічним внеском не включалися

до рекомендацій сценарію S_1 . На виході цього етапу формувалася набір товарних патернів, які одночасно мали достатню повторюваність і достатню фінансову вагу.

На четвертому етапі виконувалася темпоральна перевірка. Для кожного клієнта аналізувалося, чи повторювалися ціннісно значущі патерни в його закупівельній історії. Якщо патерн мав достатню кількість повторень, для нього визначався середній інтервал між закупівлями та стабільність цього інтервалу. Якщо повторюваність була нестабільною, патерн не використовувався для формування предиктивного тригера.

На п'ятому етапі формувалися клієнтські профілі. Для кожного клієнта обчислювалися узагальнені характеристики, що описують його закупівельну поведінку: фінансова значущість, середній цикл закупівлі та стабільність циклу. Ці ознаки використовувалися не для побудови товарних наборів, а для подальшого налаштування логіки рекомендацій.

На шостому етапі клієнти розподілялися на класи. Кожен клас відповідав певному типу закупівельної поведінки. У межах експерименту ці класи використовувалися для налаштування ширини комунікаційного вікна. Тобто сценарій S_3 формував рекомендацію не однаково для всіх клієнтів, а з урахуванням того, наскільки клієнт є цінним, регулярним і прогнозованим.

На сьомому етапі формувалися рекомендації для всіх сценаріїв. Для кожного сценарію створювався окремий список рекомендацій. Логіка формування рекомендацій у кожному сценарії наведена в таблиці 4.11.

На восьмому етапі кожен список рекомендацій перевірявся на тестовому горизонті. Перевірка виконувалася однаково для всіх сценаріїв: для кожної рекомендації визначалося, чи з'явився відповідний клієнт у тестовому періоді та чи містило його майбутнє замовлення рекомендований товарний набір або його релевантну частину.

Для сценаріїв S_2 і S_3 додатково враховувалася часова відповідність. Тобто рекомендація перевірялася не лише за фактом майбутньої закупівлі, а й за тим, чи відбулася вона в межах прогнозованого вікна.

Таблиця 4.11 – Логіка формування рекомендацій у сценаріях експерименту

Сценарій	Вхідні умови	Правило формування рекомендації	Вихід
S_0	Часті товарні набори	Рекомендувати товари на основі частоти спільної появи	Товарна рекомендація
S_1	Часті та фінансово значущі набори	Рекомендувати патерни з достатньою цінністю	Ціннісно пріоритетна рекомендація
S_2	Ціннісні патерни зі стабільним циклом	Рекомендувати патерн у прогнозований період	Предиктивна рекомендація
S_3	Предиктивні рекомендації та клас клієнта	Адаптувати вікно контакту до класу клієнта	Персоналізована предиктивна рекомендація

Це дозволяло оцінити саме своєчасність рекомендації, а не тільки правильність вибору товару.

На дев'ятому етапі для кожного сценарію розраховувалися визначені метрики. У результаті для кожного сценарію отримувався набір числових показників, який дозволяв порівняти сценарії між собою.

На десятому етапі результати сценаріїв порівнювалися між собою. Порівняння виконувалося послідовно: спочатку S_1 порівнювався з базовим сценарієм S_0 , потім S_2 – із S_1 , а S_3 – із S_2 . Така логіка дозволяла визначити внесок кожного додаткового компонента: фінансового зважування, темпоральної перевірки та класифікації клієнтів. Схема порівняння сценаріїв експерименту наведена в таблиці 4.12.

Таблиця 4.12 – Схема порівняння сценаріїв

Порівняння	Що оцінюється
S_1 проти S_0	Вплив фінансової ваги товарних патернів
S_2 проти S_1	Вплив темпоральної перевірки закупівельних циклів
S_3 проти S_2	Вплив класифікації клієнтів і адаптації вікна контакту
S_3 проти S_0	Загальний ефект повної інтегрованої моделі

Таким чином, комплексний експеримент складався з чотирьох послідовних сценаріїв, кожен із яких додавав новий компонент до логіки формування рекомендацій.

Така організація дозволила перевірити не лише підсумкову якість повної моделі, а й окремий внесок кожного елемента запропонованого підходу.

4.4 Порівняння конфігурацій рекомендаційного підходу

Після проведення комплексного експерименту було виконано порівняння чотирьох конфігурацій рекомендаційного підходу: базового частотного сценарію S_0 , частотно-ціннісного сценарію S_1 , ціннісно-темпорального сценарію S_2 та інтегрованого сценарію S_3 , у якому додатково враховується клас B2B-клієнта. Результати порівняння наведено в таблиці 4.13.

Таблиця 4.13 – Порівняння конфігурацій рекомендаційного підходу

Сценарій	К-сть рекомендацій	Успішні рекомендації	(CNV), %	V_H , грн	ARPR, грн	HitRate, %	VC, %
S_0	920	36	3,91	1114663,00	1211,59	0,63	1,37
S_1	760	41	5,39	1480000,00	1947,37	0,72	1,82
S_2	260	33	12,69	1310000,00	5038,46	0,58	1,61
S_3	220	35	15,91	1520000,00	6909,09	0,61	1,87

Базовий сценарій S_0 сформував найбільшу кількість рекомендацій – 920. Однак конверсія становить лише 3,91 %, що свідчить про недостатню вибірковість частотного підходу. Такий сценарій забезпечує широке охоплення, але створює значну кількість потенційно нерелевантних звернень до клієнтів.

Сценарій S_1 зменшує кількість рекомендацій до 760, але збільшує кількість успішних рекомендацій з 36 до 41. Конверсія зростає до 5,39 %, а середня виручка на одну рекомендацію – до 1947,37 грн. Це свідчить, що врахування фінансової

ваги товарних патернів дозволяє відбирати більш економічно значущі рекомендації.

Сценарій S_2 суттєво скорочує кількість рекомендацій до 260, однак підвищує конверсію до 12,69 % і (ARPR) до 5038,46 грн. Це означає, що темпоральна перевірка виконує роль сильного фільтра: система залишає менше рекомендацій, але вони є значно точнішими та фінансово ціннішими.

Найкращий результат демонструє сценарій S_3 . Він формує лише 220 рекомендацій, але забезпечує найвищу конверсію – 15,91 %, найбільшу середню виручку на рекомендацію – 6909,09 грн, а також найбільшу частку охопленої тестової виручки – 1,87 %. Це показує, що врахування класу клієнта дозволяє додатково підвищити якість рекомендацій за рахунок адаптації комунікаційного вікна.

Для наочного представлення співвідношення між кількістю рекомендацій, конверсією та фінансовою ефективністю було побудовано карту ефективності сценаріїв, наведену на рисунку 4.3.

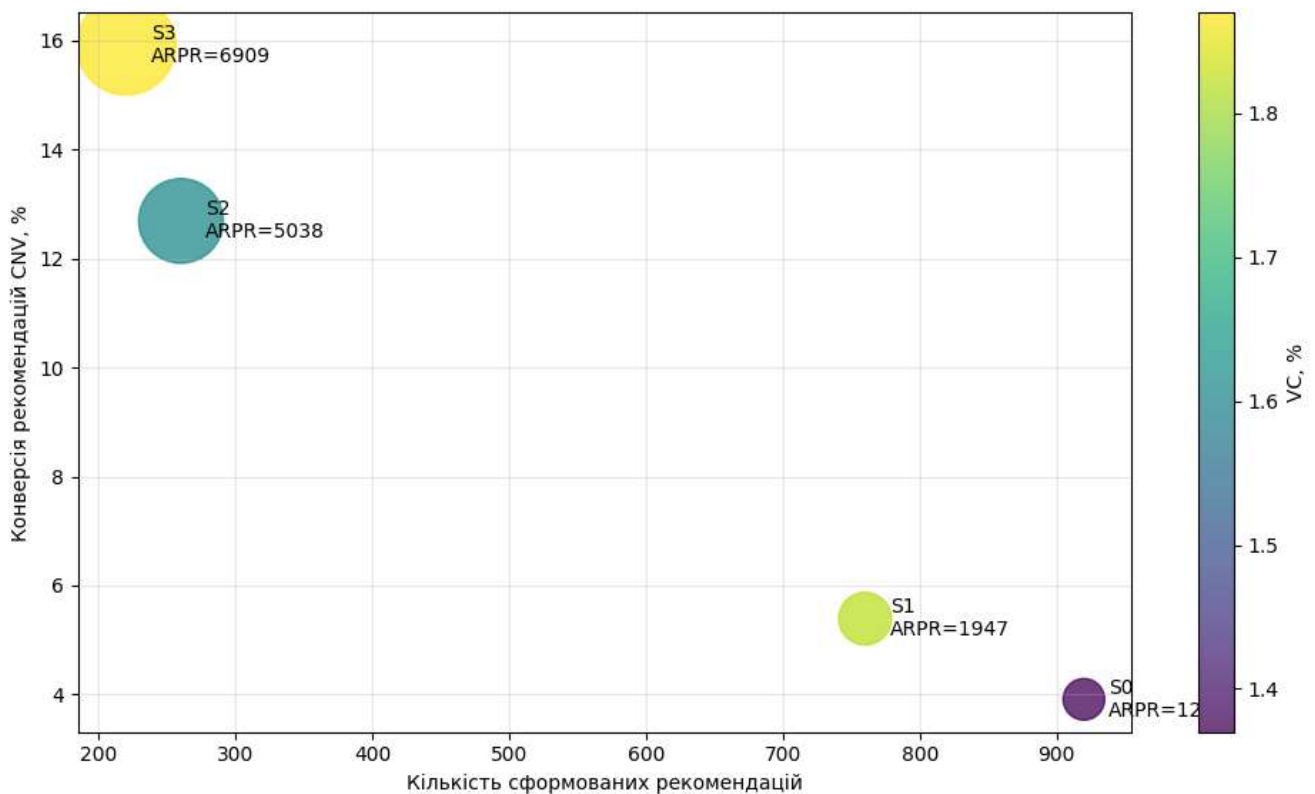


Рисунок 4.3 – Карта ефективності сценаріїв рекомендаційного підходу

З рисунка видно, що сценарій S_0 знаходиться в зоні великої кількості рекомендацій, але низької конверсії. Сценарій S_3 , навпаки, має значно меншу кількість рекомендацій, але найвищу конверсію та найбільшу фінансову віддачу на одну рекомендацію.

Це підтверджує загальну логіку запропонованого підходу: система переходить від масового формування рекомендацій до більш точного та економічно обґрунтованого відбору.

Для оцінювання внеску кожного компонента моделі було побудовано графік покрокового приросту конверсії, наведений на рисунку 4.4. Він показує, як змінюється конверсія CNV при послідовному переході від базового сценарію S_0 до інтегрованого сценарію S_3 . Базовий рівень конверсії становить 3,91 %. Додавання фінансового зважування у сценарії S_1 збільшує конверсію на 1,48 п.п. Додавання темпоральної перевірки у сценарії S_2 дає найбільший приріст – 7,30 п.п. Урахування класу клієнта у сценарії S_3 додатково підвищує конверсію ще на 3,22 п.п.

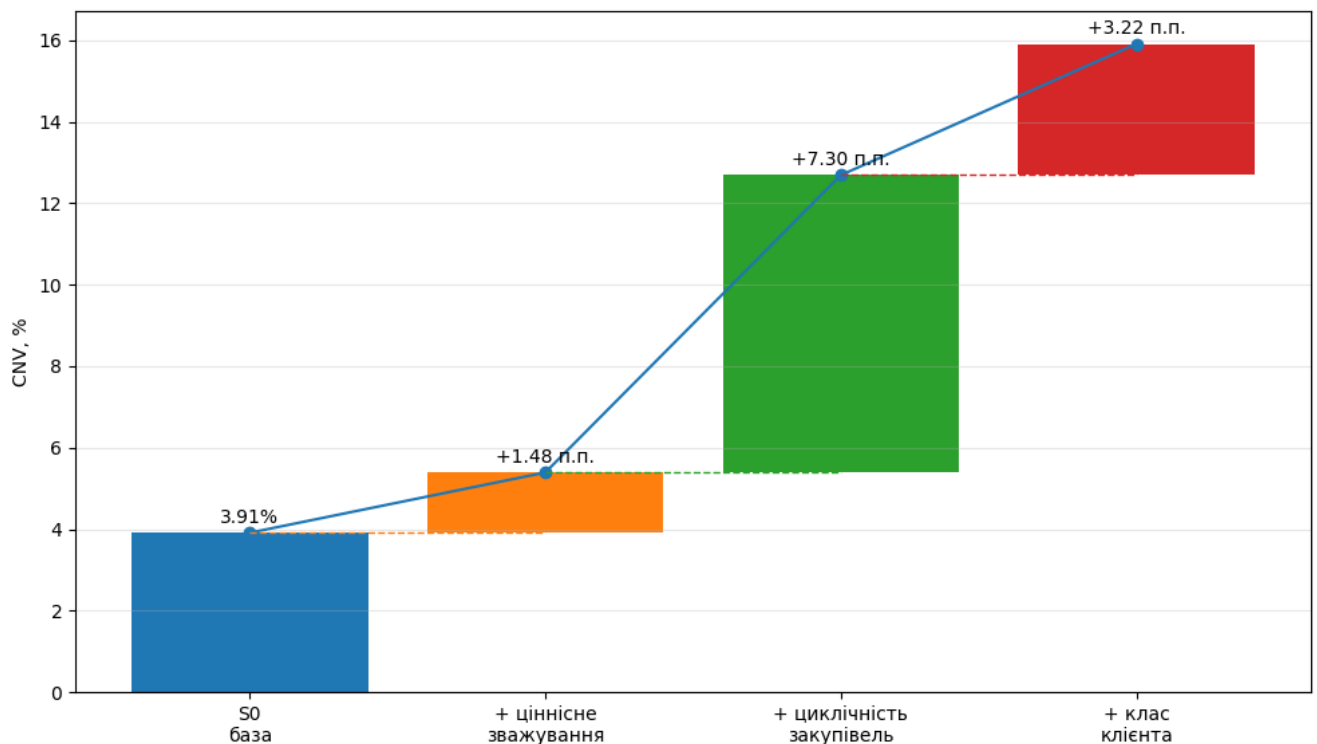


Рисунок 4.4 – Покроковий приріст конверсії рекомендацій при додаванні компонентів моделі

Таким чином, найбільший внесок у підвищення точності рекомендацій дає темпоральна перевірка закупівельних циклів. Це пояснюється тим, що рекомендація в B2B-середовищі має бути не лише товарно релевантною, а й своєчасною. Водночас приріст у сценарії S_3 показує, що клієнтська класифікація також має позитивний ефект, оскільки дозволяє точніше налаштувати момент контакту з урахуванням типу поведінки клієнта.

Для порівняння конфігурацій за кількома ключовими показниками було побудовано нормовану діаграму, наведену на рисунку 4.5. На значення кожної метрики нормовано відносно максимального значення серед усіх сценаріїв. Це дозволяє порівняти різні за масштабом показники: CNV , $ARPR$, VC та кількість успішних рекомендацій. Така візуалізація показує не абсолютні значення, а загальний профіль кожного сценарію.

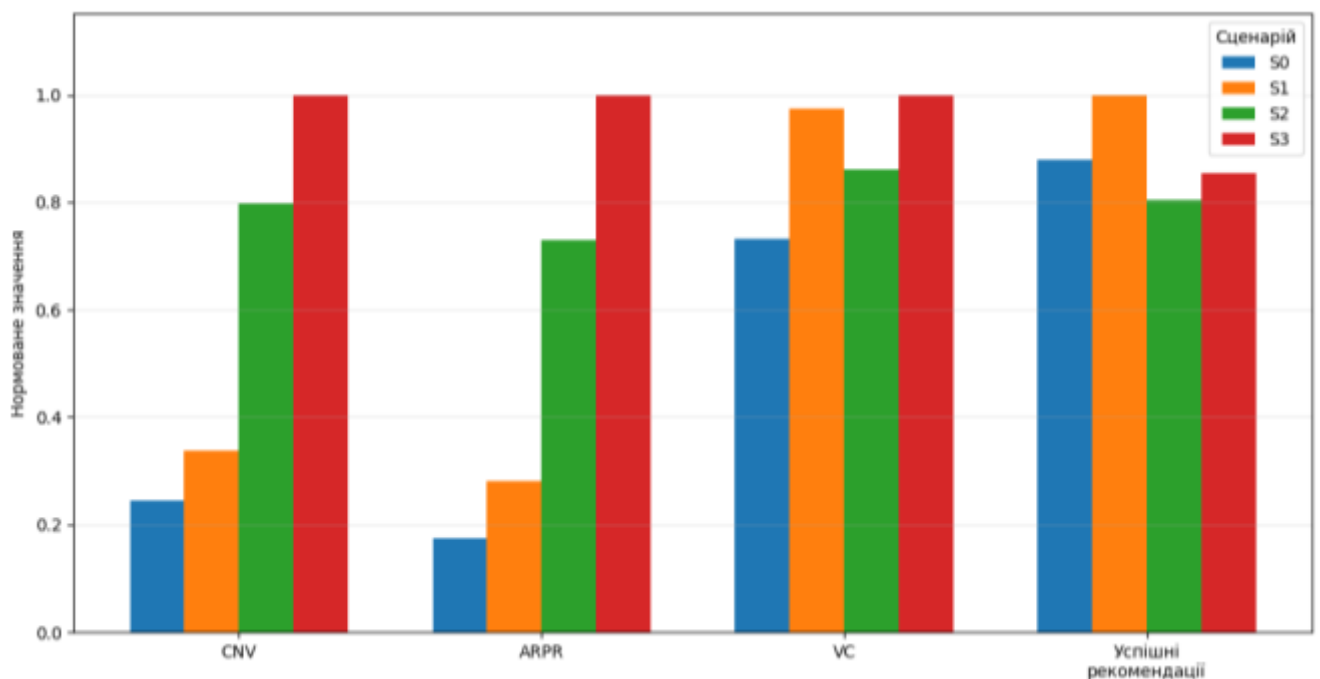


Рисунок 4.5 – Нормоване порівняння сценаріїв за ключовими метриками

Сценарій S_1 має найвище значення за кількістю успішних рекомендацій, що пояснюється ширшим охопленням клієнтів порівняно з S_2 та S_3 . Проте за конверсією і середньою виручкою на одну рекомендацію він поступається темпоральним сценаріям. Сценарій S_3 має найкращий профіль за CNV , $ARPR$ і VC ,

що підтверджує його перевагу як найбільш збалансованої конфігурації з погляду точності та фінансової ефективності.

Окремо було побудовано графік залежності середньої виручки на одну рекомендацію від кількості сформованих рекомендацій (рис. 4.6). Рисунок демонструє, що зі зменшенням кількості рекомендацій зростає середня виручка на одну рекомендацію.

У сценарії S_0 система формує 920 рекомендацій, але $ARPR$ становить лише 1211,59 грн. У сценарії S_3 кількість рекомендацій зменшується до 220, однак $ARPR$ зростає до 6909,09 грн. Це означає, що кожне сформоване звернення до клієнта стає фінансово значно ціннішим.

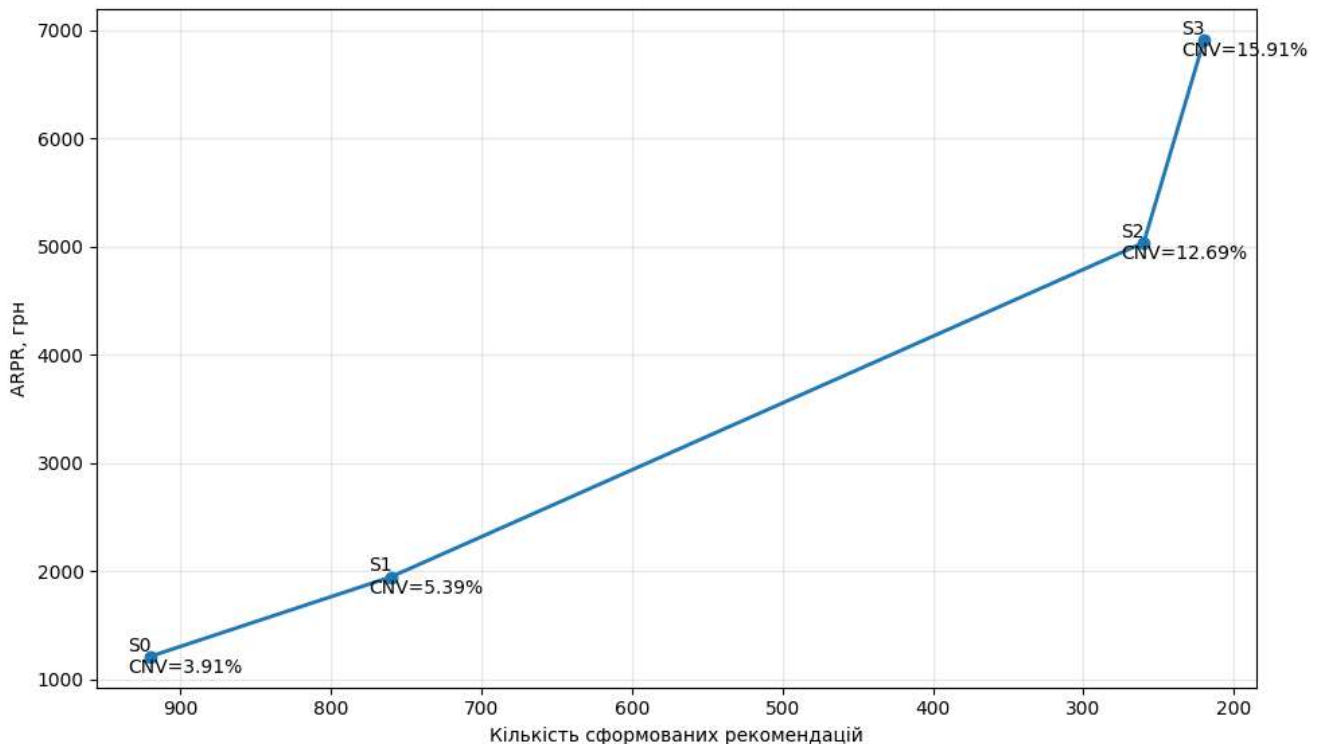


Рисунок 4.6 – Перехід від масових рекомендацій до фінансово ефективних

Такий результат є важливим для B2B-платформи, оскільки в реальному процесі продажів рекомендація пов'язана з дією менеджера: підготовкою пропозиції, контактом із клієнтом, уточненням потреби та супроводом угоди. Тому скорочення кількості нерелевантних рекомендацій при одночасному зростанні CNV і $ARPR$ є практично корисним результатом.

Для детальнішого аналізу було розраховано прирости основних метрик між послідовними сценаріями. Результати наведено в таблиці 4.14.

З таблиці видно, що перехід від S_0 до S_1 забезпечує помірне покращення: кількість рекомендацій зменшується на 160, але кількість успішних рекомендацій зростає на 5. Це свідчить про позитивний ефект фінансової пріоритезації.

Перехід від S_1 до S_2 дає найбільший приріст конверсії та $ARPR$. Хоча кількість рекомендацій зменшується на 500, конверсія зростає на 7,30 п.п., а середня виручка на одну рекомендацію – на 3091,09 грн. Це підтверджує, що темпоральна перевірка суттєво підвищує якість рекомендацій.

Перехід від S_2 до S_3 демонструє додаткову користь класифікації клієнтів. Кількість рекомендацій зменшується на 40, але кількість успішних рекомендацій зростає на 2, CNV – на 3,22 п.п., а $ARPR$ – на 1870,63 грн. Отже, адаптація комунікаційного вікна до класу клієнта дозволяє підвищити точність і фінансову результативність рекомендацій.

Таблиця 4.14 – Приріст метрик між конфігураціями рекомендаційного підходу

Порівняння	Δ рекомендацій	Δ успішних рекомендацій	ΔCNV , п.п.	$\Delta ARPR$, грн	ΔVC , п.п.
$S_1 - S_0$	-160	+5	+1,48	+735,78	+0,45
$S_2 - S_1$	-500	-8	+7,30	+3091,09	-0,21
$S_3 - S_2$	-40	+2	+3,22	+1870,63	+0,26
$S_3 - S_0$	-700	-1	+12,00	+5697,50	+0,50

Загалом порівняння S_3 із S_0 показує, що інтегрована модель формує на 700 рекомендацій менше, але забезпечує значно вищу конверсію: 15,91 % проти 3,91 %. Середня виручка на одну рекомендацію зростає з 1211,59 грн до 6909,09 грн. Це підтверджує, що запропонований підхід дозволяє перейти від масового формування рекомендацій до більш точного, своєчасного та фінансово ефективного відбору. Таким чином, результати експерименту та їх графічна

інтерпретація підтверджують послідовний внесок кожного компонента моделі. Фінансове зважування підвищує економічну значущість рекомендацій, темпоральна перевірка забезпечує їх своєчасність, а класифікація клієнтів дозволяє додатково персоналізувати комунікаційне вікно. Найкращий результат за сукупністю ключових метрик демонструє сценарій S_3 .

4.5 Аналіз чутливості темпоральних параметрів рекомендаційної моделі

Додатково до порівняння основних конфігурацій рекомендаційного підходу було проведено аналіз чутливості темпоральних параметрів моделі. Метою цього експерименту було визначити, як на якість рекомендацій впливають два параметри: ширина комунікаційного вікна та поріг стабільності закупівельного циклу. Такий аналіз є важливим, оскільки темпоральна частина моделі безпосередньо визначає не тільки те, який товарний набір рекомендувати клієнту, а й коли саме доцільно сформувати рекомендацію.

У межах експерименту розглядалися два параметри:

- ширина комунікаційного вікна W , яка визначає допустиме відхилення від прогнозованої дати повторної закупівлі;
- поріг коефіцієнта варіації CV , який визначає, наскільки стабільним має бути закупівельний цикл клієнта.

Перший експеримент був спрямований на оцінювання впливу ширини комунікаційного вікна. Для цього було зафіксовано поріг стабільності циклу $CV \leq 0,5$, а ширина вікна змінювалася від ± 7 до ± 30 днів (табл. 4.15). З таблиці видно, що збільшення ширини комунікаційного вікна приводить до зростання кількості сформованих рекомендацій і кількості успішних рекомендацій. Наприклад, при переході від вікна ± 7 днів до ± 30 днів кількість рекомендацій збільшується зі 140 до 540, а кількість успішних рекомендацій – з 20 до 52. Це означає, що ширше вікно дозволяє охопити більшу кількість потенційних закупівель. Водночас зі збільшенням ширини вікна зменшується конверсія рекомендацій. Для вікна ± 7 днів CNV становить 14,29 %, тоді як для вікна ± 30 днів – 9,63 %. Також знижується

середня виручка на одну рекомендацію: з 6071,43 грн до 3148,15 грн. Це пояснюється тим, що надто широке вікно збільшує кількість менш точних звернень до клієнтів. Отже, вузьке вікно забезпечує вищу точність, але менше охоплення, тоді як широке вікно забезпечує більше охоплення, але нижчу якість рекомендації.

Таблиця 4.15 – Вплив ширини вікна на ефективність рекомендацій

Вікно контакту	К-сть рекомендацій	Успішні рекомендації	$CNV, \%$	V_H , грн	$ARPR$, грн	$HitRate$, %	VC , %
± 7 днів	140	20	14,29	850000,00	6071,43	0,35	1,05
± 14 днів	260	33	12,69	1310000,00	5038,46	0,58	1,61
± 21 день	390	43	11,03	1560000,00	4000,00	0,75	1,92
± 30 днів	540	52	9,63	1700000,00	3148,15	0,91	2,10

Для наочного подання результатів побудований графік залежності CNV та $ARPR$ від ширини комунікаційного вікна (рис. 4.7).

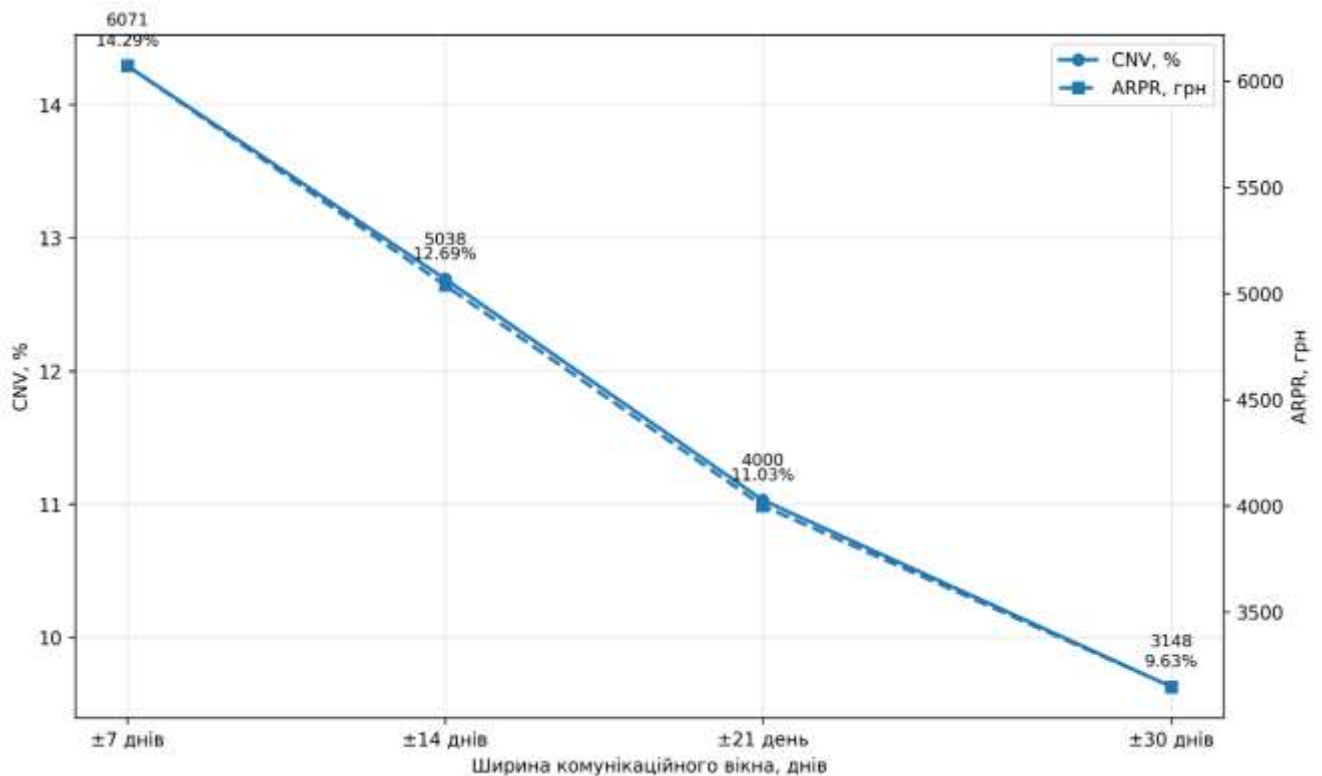


Рисунок 4.7 – Вплив ширини комунікаційного вікна на конверсію та середню виручку на рекомендацію

Для базового сценарію було обрано вікно ± 14 днів, оскільки воно забезпечує компроміс між точністю та охопленням. За цього значення CNV становить 12,69 %, $ARPR$ – 5038,46 грн, а частка охопленої тестової виручки – 1,61 %. Таким чином, вікно ± 14 днів дозволяє зберегти достатньо високу точність рекомендацій без надмірного скорочення кількості потенційних звернень.

Другий експеримент був спрямований на оцінювання впливу порога стабільності закупівельного циклу. Для цього ширина комунікаційного вікна була зафіксована на рівні ± 14 днів, а поріг коефіцієнта варіації змінювався від $CV \leq 0,3$ до $CV \leq 1,0$. Результати наведено в таблиці 4.16.

Таблиця 4.16 – Вплив порога стабільності закупівельного циклу на ефективність рекомендацій

Поріг стабільності	Кількість рекомендацій	Успішні рекомендації	(CNV), %	(V_H), грн	(ARPR), грн	(HitRate), %	(VC), %
$CV \leq 0,3$	150	24	16,00	990000,00	6600,00	0,42	1,22
$CV \leq 0,5$	260	33	12,69	1310000,00	5038,46	0,58	1,61
$CV \leq 0,7$	380	39	10,26	1450000,00	3815,79	0,68	1,79
$CV \leq 1,0$	520	45	8,65	1580000,00	3038,46	0,79	1,95

Результати таблиці показують, що послаблення порога стабільності збільшує кількість рекомендацій. При строгому порозі $CV \leq 0,3$ система формує 150 рекомендацій, а при порозі $CV \leq 1,0$ – уже 520 рекомендацій.

Це означає, що при м'якшому порозі до рекомендацій потрапляє більше клієнтів і товарних патернів, навіть якщо їхні закупівельні цикли є менш регулярними.

Однак зростання кількості рекомендацій супроводжується зниженням їхньої точності. При $CV \leq 0,3$ конверсія становить 16,00 %, а середня виручка на одну рекомендацію – 6600,00 грн. При $CV \leq 1,0$ конверсія зменшується до 8,65 %, а $ARPR$ – до 3038,46 грн. Це підтверджує, що стабільність закупівельного циклу є важливою умовою для формування якісних предиктивних рекомендацій.

Разом із цим більш м'які пороги дають більше охоплення тестового горизонту. Так, VC зростає з 1,22 % при $CV \leq 0,3$ до 1,95 % при $CV \leq 1,0$. Отже, строгий поріг забезпечує точніші рекомендації, а м'який поріг – ширше охоплення. Для базової конфігурації було використано компромісне значення $CV \leq 0,5$, яке дозволяє зберегти достатньо високу точність і водночас не надто сильно обмежувати кількість рекомендацій. Для візуального аналізу цього експерименту доцільно побудувати графік залежності CNV і кількості рекомендацій від порога CV (рис. 4.8).

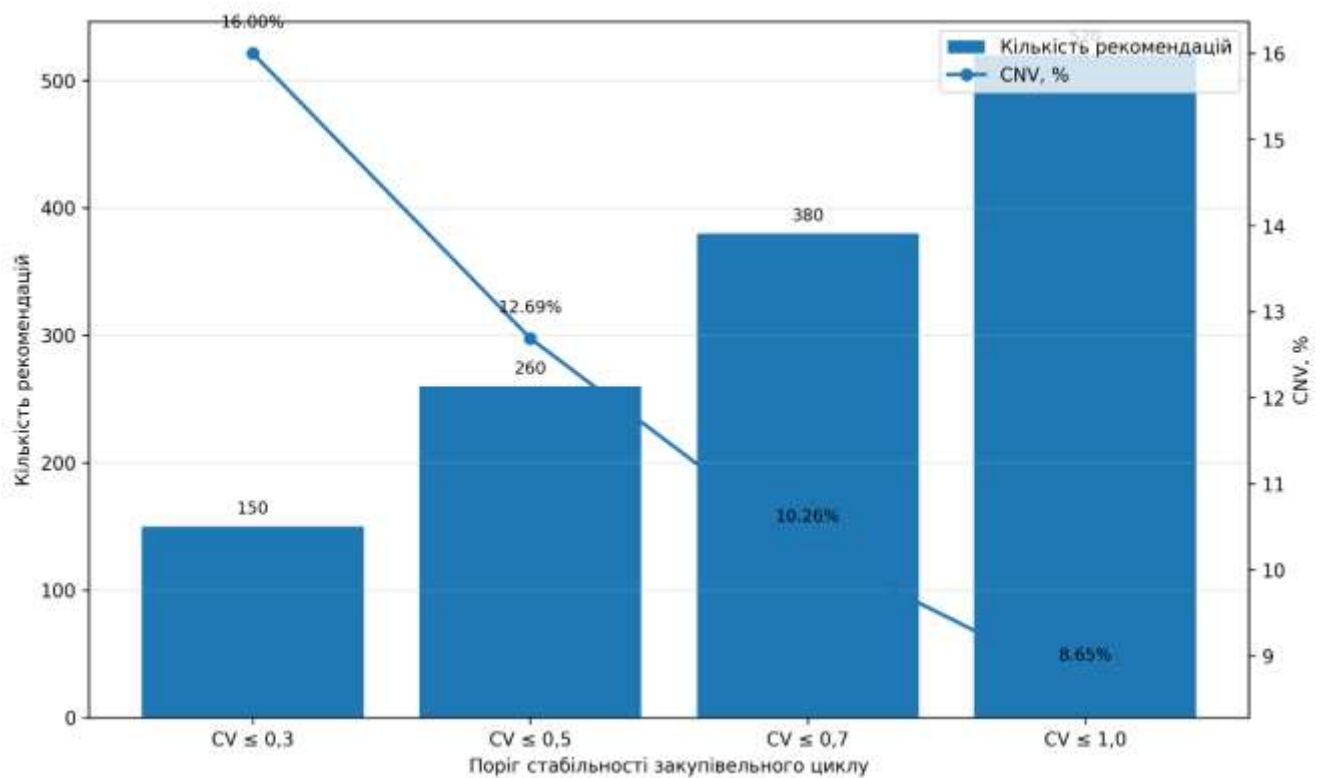


Рисунок 4.8 – Вплив порога стабільності закупівельного циклу на кількість і точність рекомендацій

Проведений аналіз чутливості показав, що темпоральні параметри істотно впливають на поведінку рекомендаційної моделі. Ширина комунікаційного вікна визначає баланс між точністю та охопленням: вузьке вікно підвищує якість окремої рекомендації, тоді як широке вікно збільшує кількість охоплених майбутніх закупівель.

Поріг CV , у свою чергу, визначає вимоги до регулярності закупівельного циклу: строгий поріг залишає лише найбільш стабільні патерни, а м'який поріг дозволяє формувати більше рекомендацій, але з нижчою конверсією.

Таким чином, експеримент підтверджує доцільність використання темпоральної перевірки в рекомендаційному підході. Найкращі за точністю результати досягаються при більш строгих значеннях параметрів, однак для практичного B2B-сценарію доцільно використовувати компромісні налаштування. У межах подальших експериментів як базові було використано значення $W = \pm 14$ днів і $CV \leq 0,5$, оскільки вони забезпечують збалансоване співвідношення між кількістю рекомендацій, конверсією та середньою виручкою на одну рекомендацію.

4.6 Аналіз приросту ефективності запропонованих моделей і методів

Для узагальнення результатів комплексного експерименту було виконано аналіз приросту ефективності між конфігураціями рекомендаційного підходу. На відміну від попереднього підрозділу, де порівнювалися абсолютні значення метрик для сценаріїв S_0 , S_1 , S_2 та S_3 , у цьому підрозділі основну увагу приділено саме величині приросту, яку забезпечує кожний додатковий компонент запропонованого підходу.

Приріст ефективності оцінювався у двох напрямках. Перший напрям пов'язаний із підвищенням точності рекомендацій, що відображається метрикою CNV . Другий напрям пов'язаний із підвищенням економічної ефективності рекомендацій, що оцінюється за допомогою $ARPR$, V_H та VC . Окремо аналізувалася зміна кількості сформованих рекомендацій, оскільки для B2B-платформи важливим є не лише збільшення кількості успішних звернень, а й зменшення кількості нерелевантних рекомендацій.

Для оцінювання приросту використовувалися абсолютні та відносні зміни метрик. Абсолютний приріст визначався як різниця між значенням метрики у поточному сценарії та значенням у базовому або попередньому сценарії:

$$\Delta M = M(S_i) - M(S_j), \quad (4.5)$$

де $M(S_i)$ – значення метрики для поточного сценарію;
 $M(S_j)$ – значення метрики для сценарію порівняння.

Відносний приріст визначався як:

$$G_M = \frac{M(S_i) - M(S_j)}{M(S_j)} \cdot 100\%. \quad (4.6)$$

Таке подання дозволяє оцінити не тільки числову різницю між сценаріями, а й масштаб покращення відносно початкового рівня. У таблиці 4.17 наведено приріст ключових метрик кожної конфігурації відносно базового частотного підходу S_0 .

Таблиця 4.17 – Приріст ефективності конфігурацій відносно базового сценарію S_0

Сценарій	Δ рекомендацій	ΔCNV , п.п.	Відносний приріст CNV , %	$\Delta ARPR$, грн	Відносний приріст $ARPR$, %	ΔVC , п.п.
$S_1 - S_0$	-160	+1,48	+37,85	+735,78	+60,73	+0,45
$S_2 - S_0$	-660	+8,78	+224,55	+3826,87	+315,85	+0,24
$S_3 - S_0$	-700	+12,00	+306,91	+5697,50	+470,23	+0,50

З таблиці видно, що кожна наступна конфігурація забезпечує зростання ключових показників якості порівняно з базовим сценарієм. Уже сценарій S_1 , який додає фінансове зважування до частотного підходу, підвищує конверсію рекомендацій на 1,48 п.п., що відповідає відносному приросту 37,85 %. При цьому середня виручка на одну рекомендацію збільшується на 735,78 грн, тобто на 60,73 %. Це свідчить про те, що врахування фінансової ваги товарних патернів дозволяє зробити рекомендації більш економічно обґрунтованими.

Сценарій S_2 демонструє значно більший приріст відносно S_0 . Конверсія збільшується на 8,78 п.п., а відносний приріст становить 224,55 %. Середня виручка на одну рекомендацію зростає на 3826,87 грн, тобто більш ніж у чотири рази порівняно з базовим сценарієм. Це підтверджує, що темпоральна перевірка закупівельних циклів суттєво підвищує якість рекомендацій, оскільки дозволяє формувати пропозиції не тільки за релевантним товарним складом, а й у більш відповідний момент часу.

Найбільший приріст забезпечує інтегрований сценарій S_3 . Порівняно з S_0 , кількість рекомендацій зменшується на 700, однак конверсія зростає на 12,00 п.п., що відповідає відносному приросту 306,91 %. Показник $ARPR$ збільшується на 5697,50 грн, або на 470,23 %. Це означає, що повна модель формує значно меншу кількість рекомендацій, але кожна з них у середньому має набагато вищу ймовірність підтвердження та більшу фінансову цінність.

Для оцінювання внеску кожного окремого компонента було також розраховано поетапний приріст між сусідніми сценаріями. Результати наведено в таблиці 4.18.

Таблиця 4.18 – Поетапний приріст ефективності між конфігураціями

Порівняння	Доданий компонент	Δ рекомендацій	Δ успішних рекомендацій	ΔCNV , п.п.	$\Delta ARPR$, грн	ΔVC , п.п.
$S_1 - S_0$	фінансове зважування	-160	+5	+1,48	+735,78	+0,45
$(S_2 - S_1)$	темпоральна перевірка	-500	-8	+7,30	+3091,09	-0,21
$(S_3 - S_2)$	клас клієнта	-40	+2	+3,22	+1870,63	+0,26

З таблиці видно, що фінансове зважування у сценарії S_1 забезпечує помірне, але стабільне покращення. Кількість рекомендацій зменшується на 160, проте кількість успішних рекомендацій зростає на 5. Це означає, що частина менш релевантних частотних рекомендацій була відсіяна, а в підсумковому наборі залишилися більш економічно значущі пропозиції.

Найбільший приріст конверсії спостерігається при переході від S_1 до S_2 , тобто після додавання темпоральної перевірки. Конверсія збільшується на 7,30 п.п., а $ARPR$ – на 3091,09 грн. При цьому кількість рекомендацій зменшується на 500. Така динаміка показує, що темпоральний компонент виконує роль сильного фільтра якості. Він значно скорочує кількість звернень, але залишає ті рекомендації, які краще узгоджуються із закупівельним циклом клієнта.

Перехід від S_2 до S_3 демонструє внесок клієнтської класифікації. Після врахування класу клієнта кількість рекомендацій зменшується ще на 40, проте кількість успішних рекомендацій зростає на 2. Конверсія збільшується на 3,22 п.п., а $ARPR$ – на 1870,63 грн. Це свідчить про те, що адаптація комунікаційного вікна до типу B2B-клієнта дозволяє додатково підвищити точність і фінансову ефективність рекомендацій.

Особливо важливим є те, що зменшення кількості рекомендацій не супроводжується пропорційним зменшенням кількості успішних результатів. Базовий сценарій S_0 формує 920 рекомендацій і отримує 36 успішних рекомендацій. Інтегрований сценарій S_3 формує лише 220 рекомендацій, але отримує 35 успішних рекомендацій. Тобто кількість сформованих рекомендацій зменшується на 76,09 %, тоді як кількість успішних рекомендацій зменшується лише на одну. Це означає, що запропонований підхід суттєво скорочує кількість нерелевантних звернень до клієнтів, майже не втрачаючи кількість результативних контактів.

З практичної точки зору це є важливою перевагою для B2B-платформи. У B2B-продажах рекомендація зазвичай означає подальшу дію менеджера: підготовку персональної пропозиції, контакт із клієнтом, уточнення потреби, супровід замовлення. Тому скорочення кількості слабких рекомендацій дозволяє зменшити навантаження на менеджерів і зосередити їхню увагу на більш перспективних клієнтах і товарних наборах.

Порівняння CNV також показує суттєвий приріст якості. У базовому сценарії лише 3,91 % рекомендацій підтверджуються фактичними закупівлями. У повній

моделі S_3 цей показник становить 15,91 %. Це означає, що ймовірність успішності однієї рекомендації зростає більш ніж у чотири рази.

Такий результат підтверджує ефективність переходу від масового частотного формування рекомендацій до більш вибіркового підходу, який враховує фінансову значущість, закупівельну циклічність і тип клієнта.

Аналогічна тенденція спостерігається для показника $ARPR$. У сценарії S_0 середня виручка на одну рекомендацію становить 1211,59 грн, тоді як у сценарії S_3 – 6909,09 грн.

Отже, кожна рекомендація повної моделі в середньому пов'язана з істотно більшою фінансовою віддачею. Це підтверджує, що запропонована модель не лише підвищує точність рекомендацій, а й краще орієнтується на економічно значущі закупівлі.

Показник VC також демонструє позитивну зміну. У базовому сценарії $VC = 1,37 \%$, тоді як у сценарії S_3 – $1,87 \%$. Попри значне скорочення кількості рекомендацій, інтегрований підхід охоплює більшу частку тестової виручки. Це означає, що модель відсікає значну кількість слабких звернень, але зберігає або навіть покращує фінансове покриття майбутніх закупівель.

Таким чином, аналіз приросту ефективності показує, що кожний компонент запропонованого підходу має власний внесок у покращення рекомендацій. Фінансове зважування підвищує економічну релевантність товарних патернів. Темпоральна перевірка забезпечує суттєве зростання точності за рахунок урахування закупівельних циклів.

Класифікація клієнтів дозволяє додатково покращити результат за рахунок адаптації комунікаційного вікна. Узагальнено, повна інтегрована модель S_3 забезпечує найкращий баланс між точністю, економічною ефективністю та кількістю рекомендацій. Вона формує значно менше звернень, але з набагато вищою конверсією та середньою фінансовою віддачею. Це підтверджує доцільність використання запропонованих моделей і методів для формування предиктивних рекомендацій у B2B-системах електронної комерції.

4.7 Економічна оцінка використання рекомендаційного модуля в B2B-платформі

Окрім оцінювання точності рекомендацій, важливим є аналіз економічної доцільності використання рекомендаційного модуля в умовах B2B-платформи. У B2B-середовищі кожна рекомендація фактично відповідає потенційному комунікаційному зверненню до клієнта: підготовці пропозиції, перевірці історії закупівель, контакту менеджера з контрагентом та подальшому супроводу можливої угоди.

Тому ефективність рекомендаційної системи доцільно оцінювати не лише за кількістю успішних рекомендацій або конверсією, а й за витратами на опрацювання сформованих звернень.

Для економічної оцінки було введено вартість одного комунікаційного звернення $C_{contact} = 150$ грн.

Це значення враховує умовні трудові витрати менеджера та супутні організаційні витрати, пов'язані з опрацюванням однієї рекомендації. Тоді операційні витрати для кожного сценарію визначаються як:

$$C_{ops} = R \cdot C_{contact}, \quad (4.7)$$

де R – кількість сформованих рекомендацій.

Чистий фінансовий результат від використання сценарію визначався як різниця між охопленою виручкою успішних рекомендацій та операційними витратами:

$$F_{net} = V_H - C_{ops}, \quad (4.8)$$

де V_H – виручка від успішно охоплених тестових угод.

Також було розраховано вартість однієї успішної рекомендації:

$$C_{hit} = \frac{C_{ops}}{H}, \quad (4.9)$$

де H – кількість успішних рекомендацій.

Результати економічної оцінки наведено в таблиці 4.19. Сценарій S_2 демонструє інший тип економічного ефекту. Після додавання темпоральної перевірки кількість рекомендацій скорочується до 260, а операційні витрати – до 39000,00 грн.

Хоча кількість успішних рекомендацій зменшується до 33, вартість однієї успішної рекомендації падає до 1181,82 грн. Це означає, що темпоральний фільтр суттєво підвищує економічну ефективність одного звернення, оскільки система залишає тільки ті рекомендації, які краще відповідають закупівельному циклу клієнта.

Таблиця 4.19 – Економічна оцінка використання рекомендаційного модуля

Показник	S_0	S_1	S_2	S_3
Кількість рекомендацій	920	760	260	220
Успішні рекомендації	36	41	33	35
Операційні витрати, грн	138000,00	114000,00	39000,00	33000,00
Виручка від успішних рекомендацій, грн	1114663,00	1480000,00	1310000,00	1520000,00
Чистий фінансовий результат, грн	976663,00	1366000,00	1271000,00	1487000,00
Вартість однієї успішної рекомендації, грн	3833,33	2780,49	1181,82	942,86
$ARPR$, грн на рекомендацію	1211,59	1947,37	5038,46	6909,09
CNV , %	3,91	5,39	12,69	15,91

Найкращий економічний результат демонструє сценарій S_3 . Він формує 220 рекомендацій, що є найменшим значенням серед усіх сценаріїв, але забезпечує 35 успішних рекомендацій і найбільшу охоплену виручку – 1520000,00 грн. Операційні витрати становлять лише 33000,00 грн, а чистий фінансовий результат – 1487000,00 грн.

Вартість однієї успішної рекомендації зменшується до 942,86 грн, що є найкращим показником серед усіх конфігурацій. Порівняно з базовим сценарієм S_0 , інтегрований сценарій S_3 забезпечує скорочення операційних витрат на:

$$138000 - 33000 = 105000 \text{ грн.}$$

У відносному вираженні це становить:

$$\frac{105000}{138000} \cdot 100\% = 76,09 \%$$

Тобто повна модель дозволяє скоротити комунікаційні витрати більш ніж на три чверті. При цьому чистий фінансовий результат збільшується з 976663,00 грн до 1487000,00 грн, тобто на 510337,00 грн.

Також суттєво зменшується вартість досягнення однієї успішної рекомендації. У сценарії S_0 вона становить 3833,33 грн, а у сценарії S_3 – 942,86 грн. Отже, використання інтегрованої моделі знижує вартість одного результативного звернення приблизно у 4 рази.

З управлінської точки зору це означає, що рекомендаційний модуль не просто покращує аналітичні метрики, а оптимізує роботу відділу продажів. Менеджер отримує менше рекомендацій, але кожна з них має вищу ймовірність завершення угодою та більшу очікувану фінансову цінність. Це дозволяє зменшити кількість зайвих контактів із клієнтами, скоротити витрати часу на слабкі пропозиції та зосередити увагу на найбільш перспективних B2B-клієнтах.

Таким чином, економічна оцінка підтверджує практичну доцільність використання запропонованого рекомендаційного модуля в B2B-платформі. Повна конфігурація S_3 забезпечує найкраще співвідношення між кількістю звернень, операційними витратами, чистим фінансовим результатом і вартістю однієї успішної рекомендації.

Це свідчить про те, що впровадження ціннісно-зваженого темпорального підходу з урахуванням класу клієнта дозволяє перетворити рекомендаційну систему з генератора масових пропозицій на інструмент економічно обґрунтованої підтримки B2B-продажів.

4.8 Інтерпретація та обговорення результатів експериментального дослідження

Результати експериментального дослідження показали, що послідовне додавання до рекомендаційного підходу ціннісного зважування, темпоральної перевірки та класифікації клієнтів дозволяє перейти від масового формування рекомендацій до більш точного й економічно обґрунтованого відбору пропозицій. Це є важливим саме для B2B-сценарію, оскільки рекомендація в такому середовищі пов'язана не лише з автоматичним відображенням товару в інтерфейсі, а й із подальшою дією менеджера: підготовкою комерційної пропозиції, комунікацією з клієнтом і супроводом потенційної угоди.

Базовий сценарій S_0 , побудований на основі FP-Growth, сформував найбільшу кількість рекомендацій – 920. Однак конверсія таких рекомендацій становила лише 3,91 %. Це свідчить про те, що класичний частотний підхід добре виявляє загальні товарні зв'язки, але не враховує економічну значущість патернів і момент доцільного звернення до клієнта. У результаті значна частина рекомендацій, сформованих лише на основі частоти спільної появи товарів, не підтверджується подальшими закупівлями. З економічної точки зору цей сценарій також є найменш ефективним, оскільки потребує опрацювання найбільшої кількості звернень і формує найбільші операційні витрати – 138000,00 грн.

Додавання ціннісного зважування у сценарії S_1 дозволило покращити результати порівняно з базовою конфігурацією. Кількість рекомендацій зменшилася з 920 до 760, проте кількість успішних рекомендацій зросла з 36 до 41. Конверсія підвищилася з 3,91 % до 5,39 %, а середня виручка на одну рекомендацію – з 1211,59 грн до 1947,37 грн. Це означає, що фінансове зважування дозволяє

відсіяти частину слабких частотних патернів і надати пріоритет тим товарним наборам, які мають більшу економічну цінність для платформи. Водночас операційні витрати зменшуються до 114000,00 грн, а чистий фінансовий результат зростає до 1366000,00 грн. Отже, вже на цьому етапі спостерігається не лише зростання точності рекомендацій, а й покращення економічної віддачі від роботи менеджера.

Найбільш суттєвий приріст якості спостерігається після додавання темпоральної перевірки у сценарії S_2 . Кількість рекомендацій зменшилася до 260, але конверсія зросла до 12,69 %, а ARPR – до 5038,46 грн. Це підтверджує, що для B2B-рекомендацій важливо визначати не лише товарний склад пропозиції, а й момент її формування. Якщо рекомендація є товарно релевантною, але несвоєчасною, її практична результативність може бути низькою. Саме тому врахування закупівельних циклів є ключовим компонентом запропонованого підходу.

Водночас сценарій S_2 демонструє характерний компроміс між охопленням і точністю. Порівняно з S_1 , кількість рекомендацій зменшується, однак точність однієї рекомендації та її фінансова віддача суттєво зростають. Операційні витрати скорочуються до 39000,00 грн, а вартість однієї успішної рекомендації зменшується до 1181,82 грн. Для B2B-платформи такий результат є важливим, оскільки надмірна кількість нерелевантних звернень перевантажує менеджерів, збільшує витрати часу та може знижувати якість взаємодії з клієнтами.

Найкращий загальний результат отримано для інтегрованого сценарію S_3 , у якому додатково враховується клас клієнта. Цей сценарій сформував 220 рекомендацій, з яких 35 виявилися успішними. Конверсія досягла 15,91 %, а середня виручка на одну рекомендацію – 6909,09 грн. Порівняно з базовим сценарієм S_0 , кількість рекомендацій зменшилася на 700, але конверсія зросла на 12,00 п.п., а ARPR – на 5697,50 грн. Це свідчить про суттєве підвищення якості одного комунікаційного звернення.

Окремо слід зазначити, що сценарій S_3 майже зберігає кількість успішних рекомендацій відносно S_0 – 35 проти 36. Однак для цього інтегрована модель

формує не 920, а лише 220 рекомендацій. Тобто система майже не втрачає кількість результативних звернень, але суттєво скорочує кількість потенційно зайвих рекомендацій. Це є важливим прикладним результатом, оскільки в B2B-середовищі якість і своєчасність контакту з клієнтом часто важливіші за максимальну кількість сформованих пропозицій.

Додаткова економічна оцінка підтверджує перевагу інтегрованого сценарію. За умови вартості одного комунікаційного звернення 150 грн операційні витрати у сценарії S_0 становлять 138000,00 грн, тоді як у сценарії S_3 – лише 33000,00 грн. Отже, повна модель дозволяє скоротити витрати на опрацювання рекомендацій на 105000,00 грн, або на 76,09 %. При цьому чистий фінансовий результат збільшується з 976663,00 грн. до 1487000,00 грн, тобто на 510337,00 грн.

Важливим є також зменшення вартості однієї успішної рекомендації. У базовому сценарії S_0 цей показник становить 3833,33 грн, тоді як у сценарії S_3 – 942,86 грн. Це означає, що інтегрований підхід дозволяє отримувати результативне звернення приблизно у 4 рази дешевше. Таким чином, запропонована модель не лише підвищує точність рекомендацій, а й зменшує витрати на їх практичне використання в роботі менеджера.

Аналіз чутливості темпоральних параметрів також підтвердив важливість правильного налаштування моделі. При зміні ширини комунікаційного вікна було встановлено, що вузьке вікно забезпечує вищу точність, але менше охоплення, тоді як широке вікно збільшує кількість рекомендацій, але знижує CNV та $ARPR$. Аналогічна закономірність спостерігається для порога стабільності CV : строгий поріг залишає лише найбільш регулярні закупівельні цикли, а м'який – збільшує охоплення, але допускає менш стабільні патерни.

Таким чином, експериментальні результати показують, що ефективність рекомендаційного підходу визначається не одним окремим компонентом, а їх поєднанням. Частотний аналіз формує базу товарних зв'язків, ціннісне зважування дозволяє виділити економічно значущі патерни, темпоральна перевірка забезпечує своєчасність рекомендацій, а класифікація клієнтів дозволяє адаптувати комунікаційне вікно до типу закупівельної поведінки.

Отримані результати підтверджують доцільність використання інтегрованого підходу для формування предиктивних B2B-рекомендацій. Повна модель S_3 забезпечує найкращий баланс між точністю, фінансовою ефективністю, кількістю сформованих рекомендацій та витратами на їх опрацювання. Вона не орієнтується на максимальне охоплення будь-якою ціною, а відбирає більш релевантні, своєчасні та економічно значущі рекомендації.

Саме такий принцип є найбільш придатним для використання в системах електронної комерції B2B-типу, де рекомендаційна система має не лише прогнозувати товарну потребу, а й підтримувати економічно доцільну роботу менеджера з клієнтами.

4.9 Характеристика взаємодії B2B-платформи електронної комерції та рекомендаційної системи

Запропоновані в дисертаційній роботі моделі та методи можуть бути використані в складі B2B-платформи електронної комерції як окремий інтелектуальний рекомендаційний модуль. Такий модуль не замінює основну платформу, а розширює її функціональність за рахунок аналізу історії оптових закупівель, виявлення ціннісно значущих товарних патернів, визначення закупівельних циклів клієнтів і формування персоналізованих рекомендацій для менеджерів або автоматизованих каналів комунікації. Місце програмної реалізації запропонованих рішень у структурі B2B-платформи наведено на рисунку 4.9.

У межах такої схеми B2B-платформа виконує роль основного джерела даних і середовища практичного використання рекомендацій. У ній зберігаються відомості про клієнтів, замовлення, товарні позиції, ціни, кількість одиниць у замовленнях, дати транзакцій і суму закупівель. Ці дані надходять до рекомендаційного модуля через вивантаження з бази даних, API або аналітичний шар платформи.

Рекомендаційний модуль, у свою чергу, є прикладною реалізацією запропонованої в дисертації моделі. У загальному випадку він може функціонувати

як окремий сервіс, що підключається до B2B-платформи та періодично оновлює список актуальних рекомендацій. Такий підхід дозволяє не змінювати основну бізнес-логіку платформи, а додати до неї інтелектуальний аналітичний контур.

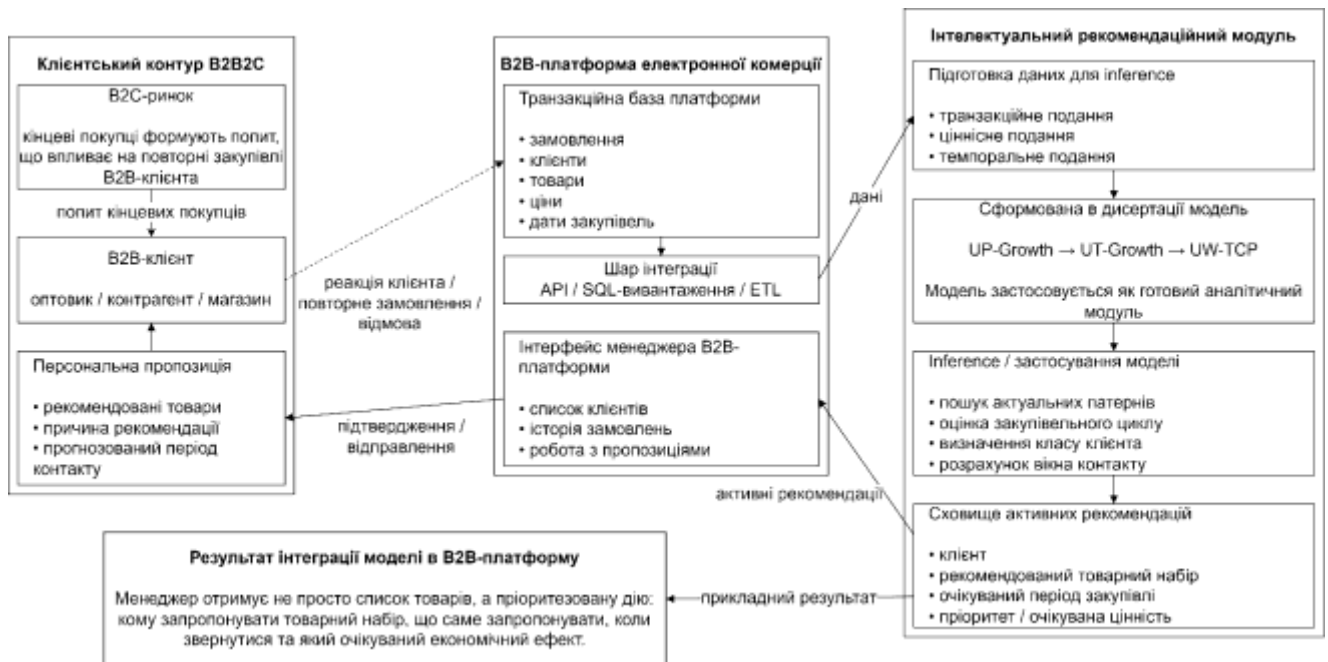


Рисунок 4.9 – Схема взаємодії B2B-платформи електронної комерції та рекомендаційної системи

В процесі експлуатації вже сформована модель виконує не етап розробки, а етап прикладного застосування. Її робота полягає в тому, щоб на основі поточної історії замовлень визначати:

- які товарні набори є ціннісно значущими для платформи;
- які клієнти мають регулярні або прогнозовані закупівельні цикли;
- коли може настати очікуваний період повторної закупівлі;
- який клас має клієнт з погляду цінності, стабільності та циклічності закупівель;
- яку персональну пропозицію доцільно сформулювати для менеджера або клієнта.

Таким чином, у B2B-платформі модель використовується як inference-модуль, тобто модуль прикладного отримання рекомендацій на основі вже

розроблених правил, патернів і профілів клієнтів. Якщо в експериментальній частині дисертації перевіряється ефективність моделі на історичних даних, то в програмній реалізації демонструється, як результати цієї моделі можуть бути представлені у вигляді робочого інструменту для B2B-платформи.

Функціонально взаємодію B2B-платформи та рекомендаційного модуля можна подати у вигляді кількох послідовних рівнів.

Перший рівень – дані платформи. До нього належать транзакційні дані, довідник товарів, клієнтська база та історія закупівель. Ці дані є вхідною інформацією для рекомендаційного модуля. На цьому рівні B2B-платформа не виконує інтелектуального аналізу, а лише надає структуровану інформацію про фактичну поведінку клієнтів.

Другий рівень – аналітичний модуль запропонованої моделі. На цьому рівні застосовуються розроблені в дисертації методи: ціннісно-зважений аналіз товарних патернів, темпоральна перевірка повторюваності закупівель та класифікація B2B-клієнтів за ціннісно-темпоральним профілем. Саме тут відбувається перетворення історичних даних на прогнозну інформацію.

Третій рівень – сховище результатів рекомендацій. Після роботи моделі формується набір активних рекомендацій. Кожна рекомендація може містити ідентифікатор клієнта, рекомендований товарний набір, очікуваний період повторної закупівлі, клас клієнта, пояснення причини рекомендації та очікувану фінансову значущість. Ці результати можуть зберігатися в окремій таблиці або передаватися до основної B2B-платформи через API.

Четвертий рівень – інтерфейс користувача B2B-платформи. На цьому рівні результати моделі стають доступними менеджеру або адміністратору платформи. Програмна реалізація може містити головну панель із ключовими показниками, список активних рекомендацій, картку клієнта та екран персональної пропозиції. Тобто результати складного аналітичного конвеєра подаються у прикладній формі, зрозумілій для користувача платформи.

П'ятий рівень – канал комунікації з клієнтом. Після підтвердження рекомендації менеджером або автоматизованим сценарієм B2B-платформа може

сформувати персональну пропозицію, повідомлення, нагадування або push-сповіщення. У цьому випадку рекомендаційна система не просто аналізує дані, а безпосередньо впливає на процес взаємодії з клієнтом.

Важливо, що програмна реалізація не повинна повторювати всю логіку B2B-платформи. Її завдання – показати, як саме результати запропонованих моделей і методів можуть бути використані в реальному прикладному сценарії. Тому в межах дисертації доцільно демонструвати не повноцінну комерційну платформу, а робочий прототип рекомендаційного інструменту, який відображає основні результати моделі: активні рекомендації, клієнтські профілі, прогнозовані періоди закупівель і персональні пропозиції.

Практичне значення такої інтеграції полягає в тому, що менеджер B2B-платформи отримує не загальний список клієнтів або товарів, а пріоритезований набір дій. Система підказує, з яким клієнтом доцільно зв'язатися, який товарний набір запропонувати, чому саме ця рекомендація була сформована та в який період вона є найбільш актуальною.

Таким чином, програмна реалізація запропонованих рішень у четвертому розділі виконує дві функції. По-перше, вона демонструє прикладну придатність розроблених моделей у середовищі B2B-платформи електронної комерції. По-друге, вона показує, як результати інтелектуального аналізу можуть бути перетворені на конкретний інструмент підтримки прийняття рішень для менеджера. Саме це забезпечує зв'язок між теоретичною частиною дисертації, експериментальною перевіркою та практичним використанням запропонованого рекомендаційного підходу.

4.10 Програмний модуль підтримки персоналізованих B2B-пропозицій

У межах практичного впровадження запропонованих моделей і методів було реалізовано програмний модуль підтримки персоналізованих B2B-пропозицій для B2B-платформи системи електронної комерції «База взуття». Його призначення полягає у відображенні результатів роботи рекомендаційної системи в інтерфейсі

B2B-платформи електронної комерції та перетворенні аналітичних результатів моделі на конкретні управлінські дії для менеджера з продажів.

Реалізований модуль працює з результатами рекомендаційної моделі, сформованими на основі ціннісно-зваженого темпорального підходу.

На вхід модуль отримує перелік активних рекомендацій, інформацію про клієнтів, рекомендовані товарні набори, прогнозовані періоди повторної закупівлі, клас клієнта, очікувану фінансову цінність пропозиції та пояснення причини формування рекомендації.

На виході ці дані подаються у вигляді зручного інтерфейсу для менеджера B2B-платформи.

Таким чином, програмний модуль не є окремою експериментальною оболонкою над алгоритмами. Він виконує роль прикладного шару B2B-платформи, через який результати роботи рекомендаційної системи можуть бути використані у процесі взаємодії з клієнтами.

Модуль дозволяє менеджеру визначити, з яким клієнтом потрібно зв'язатися, який товарний набір доцільно запропонувати, у який період рекомендація є найбільш актуальною та який очікуваний економічний ефект може мати така пропозиція.

Для демонстрації роботи програмного модуля було реалізовано декілька основних екранів: головну панель менеджера, екран активних рекомендацій, картку клієнта та екран персональної пропозиції. У межах цього підрозділу наведено приклади інтерфейсів, які безпосередньо відображають результати роботи рекомендаційної системи. На рисунку 4.10 наведено головну панель менеджера B2B-платформи.

Головна панель призначена для узагальненого перегляду поточного стану рекомендаційної системи. На екрані відображаються ключові показники: кількість активних рекомендацій, кількість пріоритетних клієнтів, прогнозована виручка, частка рекомендацій із високим пріоритетом та найближчі комунікаційні вікна. Такий екран дозволяє менеджеру швидко оцінити загальну ситуацію та перейти до детального опрацювання рекомендацій.

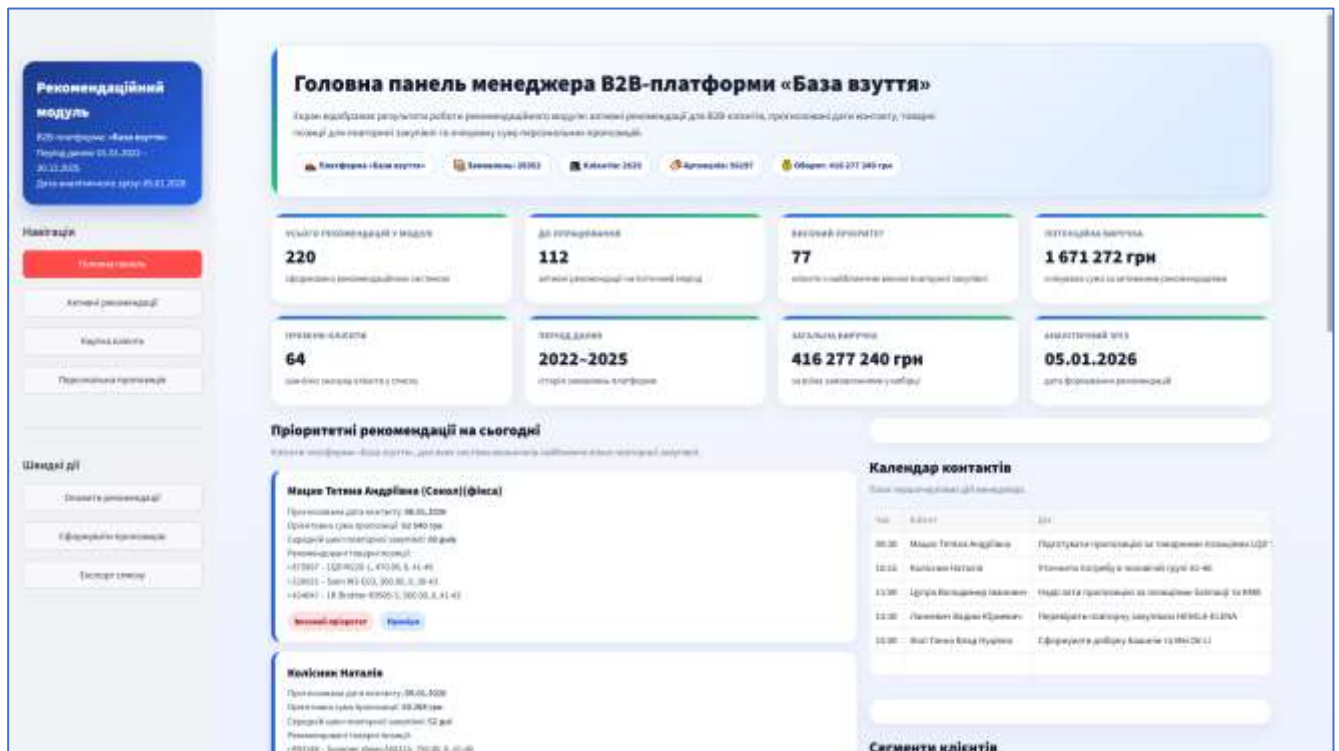


Рисунок 4.10 – Головна панель менеджера з результатами роботи рекомендаційного модуля

Особливістю головної панелі є те, що вона не просто показує статистику, а структурує роботу менеджера. Користувач бачить, які клієнти потребують уваги найближчим часом, які рекомендації мають найбільшу очікувану цінність і які дії потрібно виконати в першу чергу. Це дозволяє використовувати рекомендаційну систему як інструмент щоденного планування роботи з B2B-клієнтами. На рисунку 4.11 показано екран активних рекомендацій.

Екран активних рекомендацій є центральним елементом програмного модуля. На ньому відображається перелік рекомендацій, сформованих моделлю для конкретних клієнтів. Для кожної рекомендації подаються назва клієнта, рекомендований товарний набір, прогнозований період повторної закупівлі, клас клієнта, рівень пріоритету та очікувана фінансова цінність.

Цей екран дозволяє менеджеру працювати не з усією клієнтською базою, а з уже відібраним набором пріоритетних звернень. Рекомендації можуть бути відсортовані за очікуваною виручкою, терміновістю контакту, класом клієнта або статусом опрацювання.

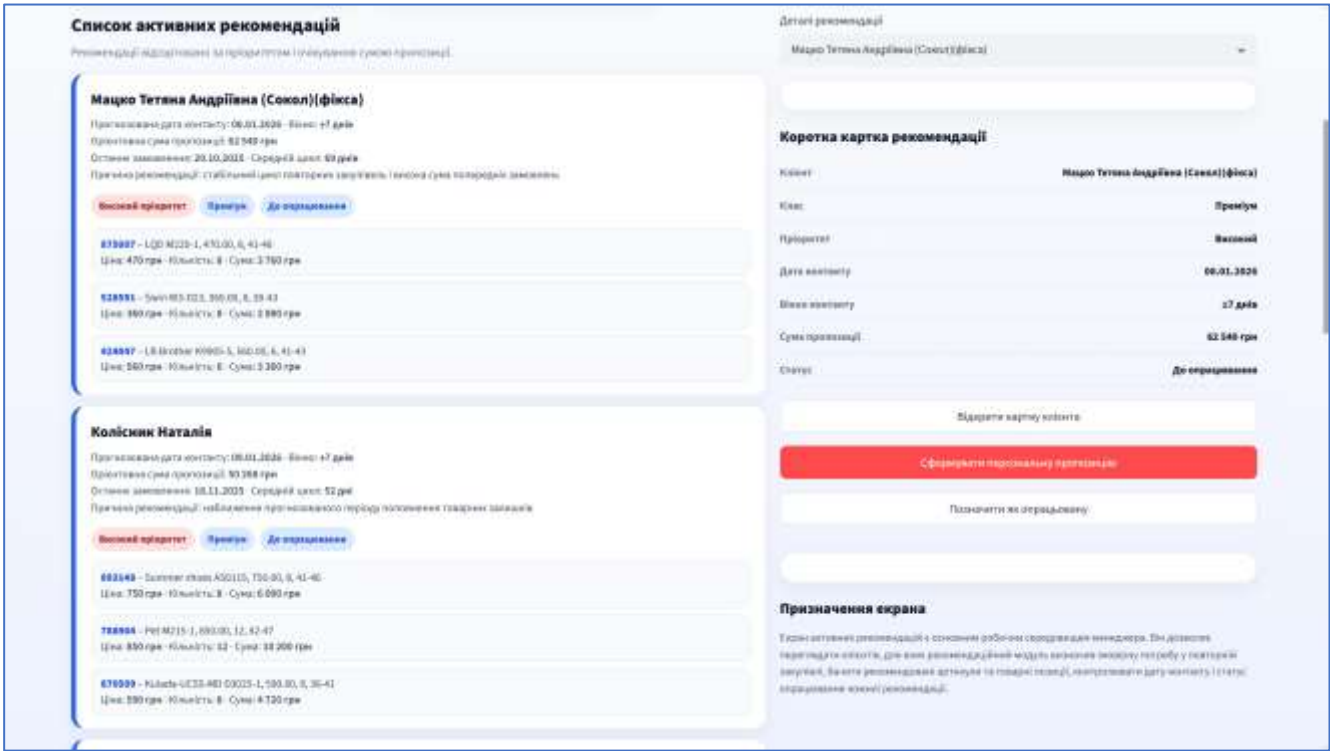


Рисунок 4.11 – Екран активних рекомендацій B2B-платформи

На рисунку 4.12 наведено екран персональної пропозиції для B2B-клієнта.

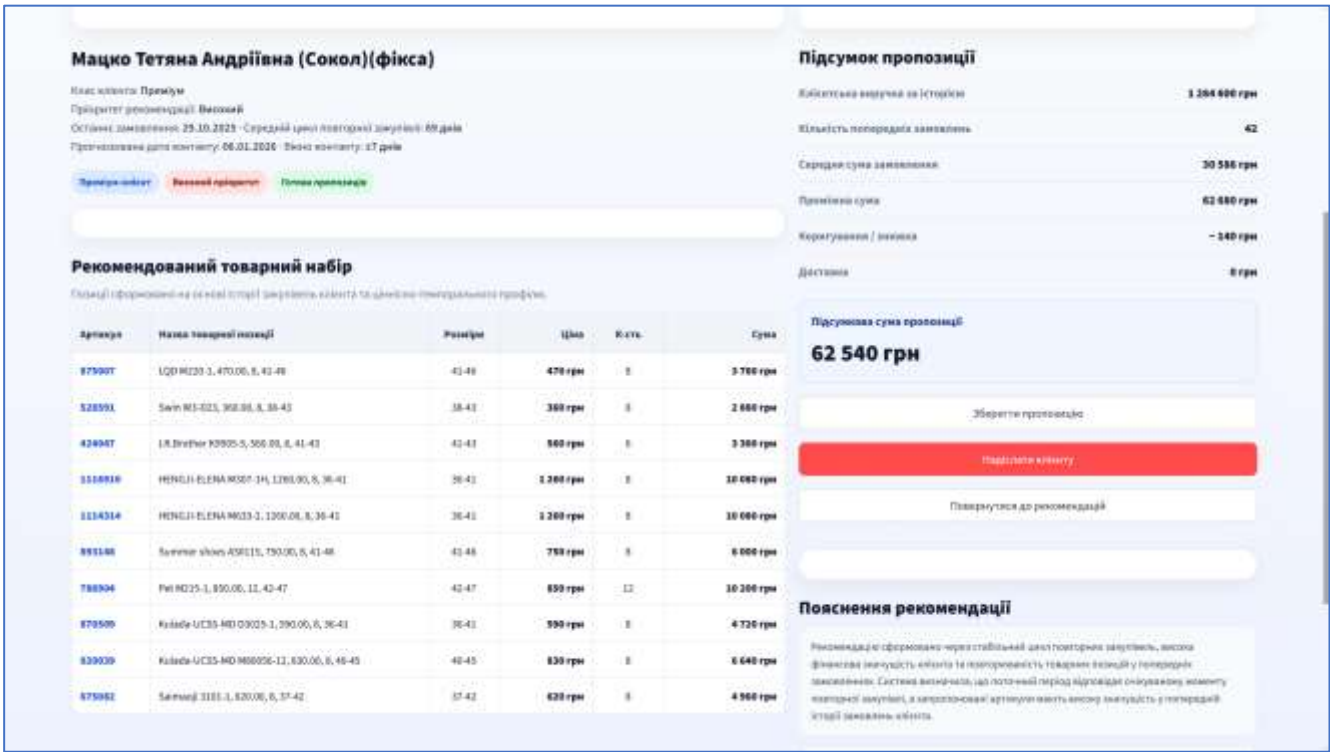


Рисунок 4.12 – Екран персональної пропозиції для B2B-клієнта

Екран персональної пропозиції використовується для детального опрацювання конкретної рекомендації. Він містить інформацію про клієнта, рекомендовані товари, прогнозований період контакту, очікувану суму потенційного замовлення та пояснення причини формування рекомендації. Зокрема, система може вказувати, що рекомендація сформована через повторюваність закупівельного циклу, високу фінансову значущість товарного набору або належність клієнта до певного класу.

Такий екран важливий з практичної точки зору, оскільки менеджер отримує не лише результат роботи моделі, а й пояснення, чому саме ця пропозиція є доцільною. Це підвищує довіру до рекомендаційної системи та дає змогу менеджеру підготувати більш аргументовану комерційну пропозицію для клієнта.

Реалізований програмний модуль забезпечує зв'язок між аналітичною частиною рекомендаційної системи та прикладним використанням її результатів у B2B-платформі. Якщо на рівні моделі виконуються пошук ціннісно значущих товарних патернів, аналіз закупівельних циклів і класифікація клієнтів, то на рівні інтерфейсу ці результати перетворюються на конкретні елементи роботи менеджера: список активних рекомендацій, пріоритети звернень, клієнтські картки та персональні пропозиції.

Отже, програмний модуль підтримки персоналізованих B2B-пропозицій демонструє практичну придатність запропонованих моделей і методів для використання в реальній B2B-платформі електронної комерції. Його застосування дозволяє перейти від пасивного аналізу історії замовлень до проактивного формування персоналізованих пропозицій, орієнтованих на повторні закупівлі, підвищення конверсії звернень і зростання економічної ефективності роботи з клієнтами.

4.11 Висновки до четвертого розділу

У четвертому розділі дисертаційного дослідження було проведено експериментальне дослідження ефективності рекомендацій для персоналізації

контенту на B2B-платформі електронної комерції. Дослідження виконано на даних платформи «База взуття» за 2022–2025 роки, що містили історію замовлень, клієнтів, товарних позицій, артикулів, цін і сум транзакцій.

У межах розділу було виконано підготовку експериментальних даних до аналізу: сформовано транзакційне подання замовлень, розраховано фінансову корисність товарних позицій, згруповано історії закупівель за клієнтами та виконано хронологічний поділ даних для перевірки рекомендацій. Також було обмежено довжину транзакцій, що дозволило зменшити комбінаторну складність побудови товарних патернів.

Для оцінювання ефективності рекомендацій було використано систему метрик, яка враховує як точність рекомендацій, так і їх економічну доцільність. Основними показниками стали кількість сформованих і успішних рекомендацій, конверсія рекомендацій, охоплена виручка, середня виручка на одну рекомендацію та частка охопленої тестової виручки.

У ході комплексного експерименту було порівняно чотири сценарії рекомендаційного підходу – від базового частотного S_0 до інтегрованого S_3 , який додатково враховує цінність патернів, закупівельні цикли та клас B2B-клієнта. Результати показали послідовне зростання якості рекомендацій: конверсія збільшилася з 3,91 % у S_0 до 5,39 % у S_1 , 12,69 % у S_2 та 15,91 % у S_3 .

Найкращий результат отримано для сценарію S_3 – він сформував 220 рекомендацій, з яких 35 були успішними, забезпечив ARPR на рівні 6909,09 грн та охоплену тестову виручку 1520000,00 грн. Порівняно з базовим сценарієм кількість рекомендацій зменшилася на 700, однак кількість успішних звернень майже не змінилася. Це підтверджує, що інтегрована модель суттєво скорочує кількість нерелевантних рекомендацій і підвищує практичну цінність кожного звернення.

Економічна оцінка також підтвердила перевагу інтегрованого підходу. За умови вартості одного комунікаційного звернення 150 грн операційні витрати зменшилися зі 138000,00 грн у сценарії S_0 до 33000,00 грн у сценарії S_3 . Чистий фінансовий результат повної моделі становив 1487000,00 грн, а вартість однієї успішної рекомендації знизилася з 3833,33 грн до 942,86 грн.

Окремо в розділі показано місце запропонованих моделей і методів у структурі B2B-платформи електронної комерції. Рекомендаційний модуль розглянуто як прикладний inference-модуль, який розширює функціональність платформи: використовує дані про замовлення, клієнтів і товари, формує активні рекомендації та передає їх у робочий інтерфейс менеджера.

Також реалізовано програмний модуль підтримки персоналізованих B2B-пропозицій із головною панеллю менеджера, екраном активних рекомендацій та екраном персональної пропозиції. Це показало, як результати інтелектуального аналізу можуть бути перетворені на практичний інструмент, що підказує менеджеру, з яким клієнтом зв'язатися, що запропонувати та яку очікувану цінність має така пропозиція.

ВИСНОВКИ

У дисертаційному дослідженні вирішено актуальну науково-прикладну задачу підвищення ефективності персоналізації контенту в B2B-платформах електронної комерції шляхом розроблення ціннісно-зваженого темпорального підходу до інтелектуального аналізу транзакційно-поведінкових даних клієнтів. Запропоновані моделі та методи спрямовані на формування персоналізованих предиктивних пропозицій, які враховують не лише факт спільної появи товарів у замовленнях, а й фінансову значущість товарних патернів, циклічність закупівельної поведінки клієнтів та відмінності між класами B2B-контрагентів.

У роботі проаналізовано стан і тенденції розвитку персоналізації контенту в системах електронної комерції. Показано, що на відміну від B2C-середовища, де значну роль відіграють масові поведінкові шаблони та емоційні чинники купівлі, B2B-сегмент характеризується раціональністю, високою вартістю транзакцій, повторюваністю закупівель, довшими циклами прийняття рішень і необхідністю підтримки довгострокових партнерських відносин. Це зумовлює потребу в таких рекомендаційних підходах, які здатні враховувати фінансову цінність операцій та індивідуальний ритм закупівель конкретного клієнта.

Проведений аналіз класичних рекомендаційних систем, методів пошуку асоціативних правил, послідовних шаблонів і RFM-сегментації дозволив встановити їх методичні обмеження в умовах B2B-електронної комерції. Виявлено, що традиційні частотні підходи орієнтовані переважно на повторюваність товарних наборів і не враховують економічну вагу окремих транзакцій. Методи аналізу послідовностей фіксують порядок подій, але не враховують реальні часові інтервали між закупівлями. RFM-аналіз описує минулу активність клієнтів, однак не забезпечує достатньої основи для прогнозування моменту наступної закупівлі. Це обґрунтувало необхідність переходу від статичного частотного аналізу до ціннісно-зваженого темпорального аналізу B2B-транзакцій.

Для подолання зазначених обмежень у роботі удосконалено модель B2B-транзакції шляхом урахування кількості товарних одиниць, індивідуальної ціни та

сумарної фінансової корисності замовлення. На цій основі отримав подальший розвиток метод розростання ціннісних патернів UP-Growth, у якому замість частотних лічильників використовується транзакційно-зважена корисність. Це дозволило зберегти можливість ефективного відсікання нерелевантних патернів і водночас змістити фокус аналізу з частих, але потенційно малозначущих комбінацій товарів на економічно важливі товарні набори.

У роботі також удосконалено модель темпоральної послідовності B2B-клієнта. До аналізу введено вектор міжзакупівельних інтервалів, середній цикл закупівлі, середньоквадратичне відхилення та коефіцієнт варіації. Це дало змогу відокремлювати регулярні закупівельні цикли від хаотичних або ситуативних замовлень і визначати момент, у який персональна пропозиція є найбільш доречною. У результаті сформовано підхід, за якого рекомендація має вигляд не лише «клієнт – товарний набір», а «клієнт – товарний набір – прогнозований період звернення».

Подальший розвиток отримав метод класифікації B2B-клієнтів на основі ціннісно-зваженого темпорального профілю. На відміну від класичної RFM-сегментації, запропонований профіль враховує прогнозований закупівельний цикл, стабільність повторних транзакцій і середню фінансову ємність клієнта. Це дозволило виділити класи клієнтів і адаптувати до них параметри комунікаційного вікна. Таким чином, у роботі сформовано триетапний аналітичний конвеєр: виявлення ціннісно значущих товарних патернів, перевірка їх темпоральної стабільності та класифікація клієнтів для диференційованого формування персоналізованих пропозицій.

Експериментальну перевірку запропонованих моделей і методів виконано на даних B2B-платформи «База взуття» за 2022–2025 роки. Початковий масив містив 29352 замовлення, 2525 клієнтів і 56197 унікальних артикулів із загальною сумою замовлень 416277239,52 грн. У процесі підготовки даних було сформовано транзакційне подання замовлень, розраховано фінансову корисність товарних позицій, виконано групування історій закупівель за клієнтами та хронологічний поділ даних для перевірки рекомендацій на майбутньому часовому горизонті.

Для кількісної оцінки ефективності проведено комплексний експеримент із порівнянням чотирьох сценаріїв: базового частотного підходу, ціннісно-зваженого підходу, ціннісно-темпорального підходу та повної інтегрованої моделі з урахуванням класу B2B-клієнта. Отримані результати показали послідовне зростання якості рекомендацій. Конверсія збільшилася з 3,91 % у базовому сценарії до 15,91 % у повній моделі. Інтегрований сценарій сформував 220 рекомендацій, з яких 35 були успішними, забезпечив середню виручку на одну рекомендацію 6909,09 грн та охоплену тестову виручку 1520000,00 грн. Порівняно з базовим підходом кількість рекомендацій зменшилася на 700, однак кількість успішних звернень майже не змінилася, що свідчить про суттєве скорочення нерелевантних рекомендацій.

Економічна оцінка підтвердила практичну доцільність запропонованого підходу. За умови вартості одного комунікаційного звернення 150 грн операційні витрати зменшилися зі 138000,00 грн у базовому сценарії до 33000,00 грн у повній інтегрований моделі. Чистий фінансовий результат інтегрованого сценарію становив 1487000,00 грн, а вартість однієї успішної рекомендації знизилася з 3833,33 грн до 942,86 грн. Це підтвердило, що запропонований підхід підвищує не лише точність рекомендацій, а й економічну ефективність їх практичного використання.

Практичну значущість результатів показано через розроблення програмного модуля підтримки персоналізованих B2B-пропозицій. Рекомендаційний модуль розглянуто як прикладний inference-модуль, який розширює функціональність B2B-платформи: використовує дані про замовлення, клієнтів і товари, формує активні рекомендації та передає їх у робочий інтерфейс менеджера. Реалізовані екрани головної панелі менеджера, активних рекомендацій і персональної пропозиції демонструють, як результати інтелектуального аналізу можуть бути перетворені на прикладний інструмент підтримки прийняття рішень.

Отже, сукупність отриманих у роботі результатів підтверджує ефективність запропонованого ціннісно-зваженого темпорального підходу до персоналізації контенту в B2B-платформах електронної комерції. Розроблені моделі та методи

дозволяють перейти від масового формування рекомендацій до точнішого, своєчаснішого та економічно обґрунтованого відбору персоналізованих пропозицій. Це створює науково обґрунтовану основу для впровадження інтелектуальних рекомендаційних модулів у B2B-системи електронної комерції з метою підтримки повторних закупівель, підвищення ефективності роботи менеджерів і покращення якості взаємодії з клієнтами.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. CapitalOne Shopping Research. B2B eCommerce Statistics (2025): Sales, Market Size & Growth. URL: <https://capitaloneshopping.com/research/b2b-ecommerce-statistics/> (дата звернення: 01.06.2026).
2. Sellers Commerce. Ecommerce Statistics in 2026 (Global and U.S. Data). 2026. URL: <https://www.sellerscommerce.com/blog/ecommerce-statistics/> (дата звернення: 01.06.2026).
3. Research and Markets. B2B E-Commerce Market, Marketplaces and Payments Trends Report 2024–2028. GlobeNewswire, 2025. URL: <https://www.globenewswire.com/news-release/2025/04/08/3057351/> (дата звернення: 01.06.2026).
4. Research and Markets. North America B2B E-Commerce Marketplaces 2024. BusinessWire, 2024. URL: <https://www.businesswire.com/news/home/20240916792646/en/> (дата звернення: 01.06.2026).
5. Research and Markets. B2B E-Commerce Marketplaces Overview Report 2024: Market to Surpass \$8.5 Trillion by 2030. BusinessWire, 2024. URL: <https://www.businesswire.com/news/home/20241125430622/en/> (дата звернення: 01.06.2026).
6. Khoshtaria T., Matin A., Abrakhamia G., Khuskivadze M. Investigating the Impact of AI-powered Personalization on Brand Awareness in B2B E-commerce. FIIB Business Review. 2025. P. 1–17. DOI: <https://doi.org/10.1177/23197145241306831>.
7. BigCommerce. Top B2B Ecommerce Trends in 2026. 2026. URL: <https://www.bigcommerce.com/articles/b2b-ecommerce/b2b-ecommerce-trends/> (дата звернення: 01.06.2026).
8. Gartner. By 2025, 80% of B2B Sales Interactions Will Occur in Digital Channels. Gartner Inc., 2020. URL: <https://www.gartner.com/en/newsroom/press-releases/2020-09-15-gartner-says-by-2025--80--of-b2b-sales-interactions-b> (дата звернення: 01.06.2026).

9. Klizer. B2B Ecommerce Statistics in 2024: Latest Updates & Trends. 2024. URL: <https://www.klizer.com/blog/b2b-ecommerce-statistics/> (дата звернення: 01.06.2026).
10. eMarketer / Statista. E-commerce share of total B2B sales in the United States in 2024. Statista, 2025. URL: <https://www.statista.com/statistics/273104/us-b2b-e-commerce-share/> (дата звернення: 01.06.2026).
11. Virto Commerce. B2B eCommerce Trends, Stats and Predictions 2024. URL: <https://virtocommerce.com/blog/b2b-ecommerce-trends> (дата звернення: 01.06.2026).
12. Arsirii O. O., Ivanov D. V., Babilunha O. Yu. Analysis of models, methods and technologies for creating personalized content in B2B e-commerce systems. Informatics. Culture. Technology. 2025. Vol. 2. P. 273–278. DOI: <https://doi.org/10.15276/ict.02.2025.41>.
13. Дія.Бізнес. Динаміка реєстрації ФОП у I кварталі 2025 року. 2025. URL: <https://business.diia.gov.ua/news/dynamika-reiestratsii-fopiv-u-i-kvartali-2025-roku-youcontrol> (дата звернення: 01.06.2026).
14. Шведа Н. М., Краузе О. І. Електронна комерція: сучасний стан та стратегії розвитку. Міжнародний науковий журнал «Інтернаука». Серія: Економічні науки. 2024. Вип. 2 (82). DOI: <https://doi.org/10.25313/2520-2294-2024-2-9639>.
15. Laudon K. C., Traver C. G. E-Commerce 2023: Business, Technology, Society. 18th ed. Hoboken: Pearson, 2023. 784 p.
16. Turban E., Outland J., King D., Lee J. K., Liang T.-P., Turban D. C. Electronic Commerce 2018: A Managerial and Social Networks Perspective. Cham: Springer, 2018. 696 p. DOI: <https://doi.org/10.1007/978-3-319-58715-8>.
17. Chen Y.-L., Tang K., Shen R.-J., Hu Y.-H. Market Basket Analysis in a Multiple Store Environment. Decision Support Systems. 2005. Vol. 40, No. 2. P. 339–354. DOI: <https://doi.org/10.1016/j.dss.2004.04.009>.
18. Kotler P., Pfoertsch W. B2B Brand Management. Berlin: Springer, 2006. 357 p. DOI: <https://doi.org/10.1007/978-3-540-44729-0>.
19. Palmatier R. W., Dant R. P., Grewal D. A Comparative Longitudinal Analysis of Theoretical Perspectives of Interorganizational Relationship Performance. Journal of

Marketing. 2007. Vol. 71, No. 4. P. 172–194. DOI: <https://doi.org/10.1509/jmkg.71.4.172>.

20. Hu Y.-H., Chen Y.-L., Tang K. Mining Sequential Patterns in the B2B Environment. *Journal of Information Science*. 2009. Vol. 35, No. 6. P. 677–694. DOI: <https://doi.org/10.1177/0165551509103600>.

21. Ricci F., Rokach L., Shapira B. *Recommender Systems Handbook*. 2nd ed. New York: Springer, 2015. 1003 p. DOI: <https://doi.org/10.1007/978-1-4899-7637-6>.

22. Resnick P., Varian H. R. Recommender Systems. *Communications of the ACM*. 1997. Vol. 40, No. 3. P. 56–58. DOI: <https://doi.org/10.1145/245108.245121>.

23. Zhang S., Yao L., Sun A., Tay Y. Deep Learning Based Recommender System: A Survey and New Perspectives. *ACM Computing Surveys*. 2019. Vol. 52, No. 1. Article 5. DOI: <https://doi.org/10.1145/3285029>.

24. Deldjoo Y., Schedl M., Cremonesi P., Pasi G. Recommender Systems Leveraging Multimedia Content. *ACM Computing Surveys*. 2021. Vol. 53, No. 5. Article 106. DOI: <https://doi.org/10.1145/3407190>.

25. Su X., Khoshgoftaar T. M. A Survey of Collaborative Filtering Techniques. *Advances in Artificial Intelligence*. 2009. Vol. 2009. Article 421425. DOI: <https://doi.org/10.1155/2009/421425>.

26. Koren Y., Bell R., Volinsky C. Matrix Factorization Techniques for Recommender Systems. *Computer*. 2009. Vol. 42, No. 8. P. 30–37. DOI: <https://doi.org/10.1109/MC.2009.263>.

27. Bobadilla J., Ortega F., Hernando A., Gutierrez A. Recommender Systems Survey. *Knowledge-Based Systems*. 2013. Vol. 46. P. 109–132. DOI: <https://doi.org/10.1016/j.knosys.2013.03.012>.

28. Lika B., Kolomvatsos K., Hadjiefthymiades S. Facing the Cold Start Problem in Recommender Systems. *Expert Systems with Applications*. 2014. Vol. 41, No. 4. P. 2065–2073. DOI: <https://doi.org/10.1016/j.eswa.2013.09.005>.

29. Schafer J. B., Frankowski D., Herlocker J., Sen S. Collaborative Filtering Recommender Systems. In: *The Adaptive Web*. Berlin: Springer, 2007. P. 291–324. DOI: https://doi.org/10.1007/978-3-540-72079-9_9.

30. Pazzani M. J., Billsus D. Content-Based Recommendation Systems. In: *The Adaptive Web*. Berlin: Springer, 2007. P. 325–341. DOI: https://doi.org/10.1007/978-3-540-72079-9_10.
31. Lops P., de Gemmis M., Semeraro G. Content-Based Recommender Systems: State of the Art and Trends. In: *Recommender Systems Handbook*. New York: Springer, 2011. P. 73–105. DOI: https://doi.org/10.1007/978-0-387-85820-3_3.
32. Young T., Hazarika D., Poria S., Cambria E. Recent Trends in Deep Learning Based Natural Language Processing. *IEEE Computational Intelligence Magazine*. 2018. Vol. 13, No. 3. P. 55–75. DOI: <https://doi.org/10.1109/MCI.2018.2840738>.
33. Сіроштан М. В., Іванов Д. В., Арсірій О. О. Дослідження ефективності методів машинного навчання на прикладі задачі класифікації предметів одягу. Project, Program, Portfolio Management. РЗМ-2024 : тези доповідей IX Міжнародної науково-практичної конференції : у 2 т. / відп. за вип. П. О. Тесленко. Одеса : ІШІР, 2024. Т. 1. С. 201–204. URL: <https://rep.nuos.edu.ua/server/api/core/bitstreams/f2c763dc-2728-47ca-af64-31618bd2a659/content>.
34. Zhang Q., Lu J., Jin Y. Artificial Intelligence in Recommender Systems. *Complex & Intelligent Systems*. 2021. Vol. 7, No. 1. P. 439–457. DOI: <https://doi.org/10.1007/s40747-020-00212-w>.
35. Burke R. Hybrid Recommender Systems: Survey and Experiments. *User Modeling and User-Adapted Interaction*. 2002. Vol. 12, No. 4. P. 331–370. DOI: <https://doi.org/10.1023/A:1021240730564>.
36. Jannach D., Jugovac M. Measuring the Business Value of Recommender Systems. *ACM Transactions on Management Information Systems*. 2019. Vol. 10, No. 4. Article 16. DOI: <https://doi.org/10.1145/3370082>.
37. Agrawal R., Imielinski T., Swami A. Mining Association Rules between Sets of Items in Large Databases. *Proc. of the 1993 ACM SIGMOD*. New York: ACM, 1993. P. 207–216. DOI: <https://doi.org/10.1145/170035.170072>.
38. Agrawal R., Srikant R. Mining Sequential Patterns. *Proc. of the 11th International Conference on Data Engineering (ICDE 1995)*. IEEE, 1995. P. 3–14. DOI: <https://doi.org/10.1109/ICDE.1995.380415>.

39. Zaki M. J., Meira W., Jr. Data Mining and Analysis: Fundamental Concepts and Algorithms. Cambridge: Cambridge University Press, 2014. 548 p. DOI: <https://doi.org/10.1017/CBO9780511810114>.
40. Tan P.-N., Kumar V., Srivastava J. Selecting the Right Interestingness Measure for Association Patterns. Proc. of the 8th ACM SIGKDD. New York: ACM, 2002. P. 32–41. DOI: <https://doi.org/10.1145/775047.775053>.
41. Brin S., Motwani R., Ullman J. D., Tsur S. Dynamic Itemset Counting and Implication Rules for Market Basket Data. Proc. of the 1997 ACM SIGMOD. New York: ACM, 1997. P. 255–264. DOI: <https://doi.org/10.1145/253260.253325>.
42. Hahsler M., Grun B., Hornik K. arules – A Computational Environment for Mining Association Rules and Frequent Item Sets. Journal of Statistical Software. 2005. Vol. 14, No. 15. P. 1–25. DOI: <https://doi.org/10.18637/jss.v014.i15>.
43. Agrawal R., Srikant R. Fast Algorithms for Mining Association Rules. Proc. of VLDB 1994. San Francisco: Morgan Kaufmann, 1994. P. 487–499. URL: <http://www.vldb.org/conf/1994/P487.PDF>.
44. Han J., Pei J., Yin Y. Mining Frequent Patterns without Candidate Generation. ACM SIGMOD Record. 2000. Vol. 29, No. 2. P. 1–12. DOI: <https://doi.org/10.1145/335191.335372>.
45. Han J., Pei J., Yin Y., Mao R. Mining Frequent Patterns without Candidate Generation: A Frequent-Pattern Tree Approach. Data Mining and Knowledge Discovery. 2004. Vol. 8, No. 1. P. 53–87. DOI: <https://doi.org/10.1023/B:DAMI.0000005258.31418.83>.
46. Zaki M. J. Scalable Algorithms for Association Mining. IEEE Transactions on Knowledge and Data Engineering. 2000. Vol. 12, No. 3. P. 372–390. DOI: <https://doi.org/10.1109/69.846291>.
47. Fournier-Viger P., Lin J. C.-W., Kiran R. U., Koh Y. S., Thomas R. A Survey of Sequential Pattern Mining. Data Science and Pattern Recognition. 2017. Vol. 1, No. 1. P. 54–77.

48. Srikant R., Agrawal R. Mining Sequential Patterns: Generalizations and Performance Improvements. In: EDBT 1996. Lecture Notes in Computer Science, Vol. 1057. Berlin: Springer, 1996. P. 3–17. DOI: <https://doi.org/10.1007/BFb0014140>.
49. Pei J., Han J., Mortazavi-Asl B., Pinto H., Chen Q., Dayal U., Hsu M.-C. PrefixSpan: Mining Sequential Patterns Efficiently by Prefix-Projected Pattern Growth. Proc. of ICDE 2001. IEEE, 2001. P. 215–224. DOI: <https://doi.org/10.1109/ICDE.2001.914830>.
50. Zaki M. J. SPADE: An Efficient Algorithm for Mining Frequent Sequences. Machine Learning. 2001. Vol. 42, No. 1–2. P. 31–60. DOI: <https://doi.org/10.1023/A:1007652502315>.
51. Mooney C. H., Roddick J. F. Sequential Pattern Mining – Approaches and Algorithms. ACM Computing Surveys. 2013. Vol. 45, No. 2. Article 19. DOI: <https://doi.org/10.1145/2431211.2431218>.
52. Hughes A. M. Strategic Database Marketing. Chicago: Probus Publishing Company, 1994. 375 p.
53. Fader P. S., Hardie B. G. S., Lee K. L. RFM and CLV: Using Iso-Value Curves for Customer Base Analysis. Journal of Marketing Research. 2005. Vol. 42, No. 4. P. 415–430. DOI: <https://doi.org/10.1509/jmkr.2005.42.4.415>.
54. Wei J.-T., Lin S.-Y., Wu H.-H. A Review of the Application of RFM Model. African Journal of Business Management. 2010. Vol. 4, No. 19. P. 4199–4206.
55. Christy A. J., Umamakeswari A., Priyatharsini L., Neyaa A. RFM Ranking – An Effective Approach to Customer Segmentation. Journal of King Saud University – Computer and Information Sciences. 2021. Vol. 33, No. 10. P. 1251–1257. DOI: <https://doi.org/10.1016/j.jksuci.2018.09.004>.
56. Blattberg R. C., Deighton J. Manage Marketing by the Customer Equity Test. Harvard Business Review. 1996. Vol. 74, No. 4. P. 136–144.
57. Gupta S., Hanssens D., Hardie B., Kahn W., Kumar V., Lin N., Sriram S. Modeling Customer Lifetime Value. Journal of Service Research. 2006. Vol. 9, No. 2. P. 139–155. DOI: <https://doi.org/10.1177/1094670506293810>.

58. Kumar V., Reinartz W. Customer Relationship Management: Concept, Strategy, and Tools. 3rd ed. Berlin: Springer, 2018. 411 p. DOI: <https://doi.org/10.1007/978-3-662-55381-7>.
59. Tan P.-N., Steinbach M., Kumar V. Introduction to Data Mining. Boston: Pearson Addison-Wesley, 2005. 769 p.
60. Yao H., Hamilton H. J., Geng L. A Unified Framework for Utility-Based Measures for Mining Itemsets. Proc. of ACM SIGKDD 2nd Workshop on Utility-Based Data Mining. New York: ACM, 2006. P. 28–37.
61. Wiersema F. The B2B Agenda: The Current State of B2B Marketing and a Look Ahead. Industrial Marketing Management. 2013. Vol. 42, No. 4. P. 470–488. DOI: <https://doi.org/10.1016/j.indmarman.2013.02.015>.
62. Geng L., Hamilton H. J. Interestingness Measures for Data Mining: A Survey. ACM Computing Surveys. 2006. Vol. 38, No. 3. Article 9. DOI: <https://doi.org/10.1145/1132960.1132963>.
63. Han J., Kamber M., Pei J. Data Mining: Concepts and Techniques. 3rd ed. Waltham: Morgan Kaufmann, 2011. 744 p.
64. Liu Y., Liao W.-K., Choudhary A. A Two-Phase Algorithm for Fast Discovery of High Utility Itemsets. In: PAKDD 2005. Lecture Notes in Computer Science, Vol. 3518. Berlin: Springer, 2005. P. 689–695. DOI: https://doi.org/10.1007/11430919_79.
65. Gan W., Lin J. C.-W., Fournier-Viger P., Chao H.-C., Tseng V. S., Yu P. S. A Survey of Utility-Oriented Pattern Mining. IEEE Transactions on Knowledge and Data Engineering. 2021. Vol. 33, No. 4. P. 1462–1481. DOI: <https://doi.org/10.1109/TKDE.2019.2942594>.
66. Tseng V. S., Shie B.-E., Wu C.-W., Yu P. S. Efficient Algorithms for Mining High Utility Itemsets from Transactional Databases. IEEE Transactions on Knowledge and Data Engineering. 2013. Vol. 25, No. 8. P. 1772–1786. DOI: <https://doi.org/10.1109/TKDE.2012.59>.
67. Tseng V. S., Wu C.-W., Shie B.-E., Yu P. S. UP-Growth: An Efficient Algorithm for High Utility Itemset Mining. Proc. of the 16th ACM SIGKDD. New York: ACM, 2010. P. 253–262. DOI: <https://doi.org/10.1145/1835804.1835839>.

68. Kumar R., Singh K. High Utility Itemsets Mining from Transactional Databases: A Survey. *Applied Intelligence*. 2023. Vol. 53. P. 27655–27703. DOI: <https://doi.org/10.1007/s10489-023-04853-5>.
69. Pielot M., Church K., de Oliveira R. An In-Situ Study of Mobile Phone Notifications. *Proc. of the 16th International Conference on Human-Computer Interaction with Mobile Devices and Services*. New York: ACM, 2014. P. 233–242. DOI: <https://doi.org/10.1145/2628363.2628364>.
70. Lilien G. L. The B2B Knowledge Gap. *International Journal of Research in Marketing*. 2016. Vol. 33, No. 3. P. 543–556. DOI: <https://doi.org/10.1016/j.ijresmar.2016.01.003>.
71. Park D. H., Kim H. K., Choi I. Y., Kim J. K. A Literature Review and Classification of Recommender Systems Research. *Expert Systems with Applications*. 2012. Vol. 39, No. 11. P. 10059–10072. DOI: <https://doi.org/10.1016/j.eswa.2012.02.038>.
72. Арсірій О., Іванов Д. Моделі афінитивного аналізу транзакційних і поведінкових даних клієнтів систем В2В електронної комерції для створення персоналізованого контенту. *Праці Одеського політехнічного університету*. 2025. Вип. 2(72). С. 90–101. DOI: <https://doi.org/10.15276/opu.2.72.2025.10>.
73. Liu M., Qu J. Mining High Utility Itemsets without Candidate Generation. *Proceedings of the 21st ACM International Conference on Information and Knowledge Management*. New York : ACM, 2012. P. 55–64. DOI: <https://doi.org/10.1145/2396761.2396773>.
74. Van Vlasselaer V., Bravo C., Caelen O., Eliassi-Rad T., Akoglu L., Snoeck M., Baesens B. APATE: A Novel Approach for Automated Credit Card Transaction Fraud Detection Using Network-Based Extensions. *Decision Support Systems*. 2015. Vol. 75. P. 38–48. DOI: <https://doi.org/10.1016/j.dss.2015.04.013>.
75. Арсірій О. О., Іванов Д. В. Інтелектуальний аналіз транзакційних даних В2В-клієнтів на основі ціннісно-зваженого темпорального підходу. *Електроніка та системи управління*. 2026. № 2(88). С. 16–31. DOI: <https://doi.org/10.18372/1990-5548.88.20958>.

76. Wu X., Kumar V., Quinlan J. R., Ghosh J., Yang Q., Motoda H., McLachlan G. J., Ng A., Liu B., Yu P. S., Zhou Z.-H., Steinbach M., Hand D. J., Steinberg D. Top 10 Algorithms in Data Mining. *Knowledge and Information Systems*. 2008. Vol. 14, No. 1. P. 1–37. DOI: <https://doi.org/10.1007/s10115-007-0114-2>.
77. Fournier-Viger P., Lin J. C.-W., Vo B., Chi T. T., Zhang J., Le H. B. A Survey of Itemset Mining. *WIREs Data Mining and Knowledge Discovery*. 2017. Vol. 7, No. 4. Article e1207. DOI: <https://doi.org/10.1002/widm.1207>.
78. Fournier-Viger P., Wu C.-W., Zida S., Tseng V. S. FHM: Faster High-Utility Itemset Mining Using Estimated Utility Co-occurrence Pruning. *Foundations of Intelligent Systems : ISMIS 2014. Lecture Notes in Computer Science*. Cham : Springer, 2014. Vol. 8502. P. 83–92. DOI: https://doi.org/10.1007/978-3-319-08326-1_9.
79. Kim K. H., Moon H. Innovative Digital Marketing Management in B2B Markets. *Industrial Marketing Management*. 2021. Vol. 95. P. 1–4. DOI: <https://doi.org/10.1016/j.indmarman.2021.01.016>.
80. Mabroukeh N. R., Ezeife C. I. A Taxonomy of Sequential Pattern Mining Algorithms. *ACM Computing Surveys*. 2010. Vol. 43, No. 1. Article 3. DOI: <https://doi.org/10.1145/1824795.1824798>.
81. Gan W., Lin J. C.-W., Fournier-Viger P., Chao H.-C., Yu P. S. A Survey of Parallel Sequential Pattern Mining. *ACM Transactions on Knowledge Discovery from Data*. 2019. Vol. 13, No. 3. Article 25. P. 1–34. DOI: <https://doi.org/10.1145/3314107>.
82. Kenney J. F., Keeping E. S. *Mathematics of Statistics. Part 2*. 2nd ed. Princeton: Van Nostrand, 1951. 429 p.
83. Abdi H. Coefficient of Variation. In: Salkind N. J. (Ed.). *Encyclopedia of Research Design*. Thousand Oaks: SAGE Publications, 2010. Vol. 1. P. 169–171. DOI: <https://doi.org/10.4135/9781412961288.n56>.
84. Sohrabi M. K., Barforoush A. A. Using Coefficient of Variation as a Stability Measure in Sequential Pattern Mining. *Knowledge-Based Systems*. 2013. Vol. 40. P. 91–103. DOI: <https://doi.org/10.1016/j.knosys.2012.11.006>.

85. Saltelli A., Ratto M., Andres T., Campolongo F., Cariboni J., Gatelli D., Tarantola S. *Global Sensitivity Analysis: The Primer*. Chichester: Wiley, 2008. 292 p. DOI: <https://doi.org/10.1002/9780470725184>.
86. Yeo E. X., Pei J., Islam R. Inter-Purchase Time Prediction Based on Deep Learning. *Computer Systems Science and Engineering*. 2022. Vol. 42, No. 2. P. 493–508. DOI: <https://doi.org/10.32604/csse.2022.022166>.
87. Bogaert M., Lootens J., Van den Poel D., Ballings M. Evaluating multi-label classifiers and recommender systems in the financial service sector. *European Journal of Operational Research*. 2019. Vol. 279, No. 2. P. 620–634. DOI: <https://doi.org/10.1016/j.ejor.2019.05.037>.
88. Pei J., Han J., Mortazavi-Asl B., Wang J., Pinto H., Chen Q., Dayal U., Hsu M.-C. Mining Sequential Patterns by Pattern-Growth: The PrefixSpan Approach. *IEEE Transactions on Knowledge and Data Engineering*. 2004. Vol. 16, No. 11. P. 1424–1440. DOI: <https://doi.org/10.1109/TKDE.2004.77>.
89. Arsirii O., Ivanov D. A Model and Method of Transactional-behavioral Data Mining for B2B Content Personalization. *Electronics and Control Systems*. 2026. Vol. 1, No. 87. P. 14–23. DOI: <https://doi.org/10.18372/1990-5548.87.20881..>
90. Fayyad U., Piatetsky-Shapiro G., Smyth P. From Data Mining to Knowledge Discovery in Databases. *AI Magazine*. 1996. Vol. 17, No. 3. P. 37–54. DOI: <https://doi.org/10.1609/aimag.v17i3.1230>.
91. Pei J., Han J., Wang W. Constraint-Based Sequential Pattern Mining: The Pattern-Growth Methods. *Journal of Intelligent Information Systems*. 2007. Vol. 28, No. 2. P. 133–160. DOI: <https://doi.org/10.1007/s10844-006-0006-z>.
92. Khajvand M., Zolfaghar K., Ashoori S., Alizadeh S. Estimating Customer Lifetime Value Based on RFM Analysis of Customer Purchase Behavior. *Procedia Computer Science*. 2011. Vol. 3. P. 57–63. DOI: <https://doi.org/10.1016/j.procs.2010.12.011>.
93. Berger P. D., Nasr N. I. Customer Lifetime Value: Marketing Models and Applications. *Journal of Interactive Marketing*. 1998. Vol. 12, No. 1. P. 17–30. DOI: [https://doi.org/10.1002/\(SICI\)1520-6653\(199824\)12:1<17::AID-DIR3>3.0.CO;2-K](https://doi.org/10.1002/(SICI)1520-6653(199824)12:1<17::AID-DIR3>3.0.CO;2-K).

94. Wan S., Chen J., Qi Z., Gan W., Tang L. Fast RFM Model for Customer Segmentation. WWW 2022 – Companion Proceedings of the Web Conference 2022. New York : ACM, 2022. P. 965–972. DOI: <https://doi.org/10.1145/3487553.3524707>.
95. Rezende L. P., Ferreira C. H. G., Ferreira A. R., Bernardini F. C., Viterbo J. Recency, Frequency, Monetary Value, Clustering, and Internal and External Indices for Customer Segmentation from Retail Data. Algorithms. 2023. Vol. 16, No. 9. Article 396. DOI: <https://doi.org/10.3390/a16090396>.
96. Jain A. K., Murty M. N., Flynn P. J. Data Clustering: A Review. ACM Computing Surveys. 1999. Vol. 31, No. 3. P. 264–323. DOI: <https://doi.org/10.1145/331499.331504>.
97. MacQueen J. B. Some Methods for Classification and Analysis of Multivariate Observations. Proc. of the 5th Berkeley Symposium on Mathematical Statistics and Probability. Berkeley: University of California Press, 1967. Vol. 1. P. 281–297.
98. Tabianan K., Velu S., Ravi V. K-Means Clustering Approach for Intelligent Customer Segmentation Using Customer Purchase Behavior Data. Sustainability. 2022. Vol. 14, No. 12. Article 7243. DOI: <https://doi.org/10.3390/su14127243>.
99. Punj G., Stewart D. W. Cluster Analysis in Marketing Research: Review and Suggestions for Application. Journal of Marketing Research. 1983. Vol. 20, No. 2. P. 134–148. DOI: <https://doi.org/10.1177/002224378302000204>.
100. Imani A., Abbasi M., Ahang F., Ghaffari H., Mehdi M. Customer Segmentation to Identify Key Customers Based on RFM Model by Using Data Mining Techniques. International Journal of Research in Industrial Engineering. 2022. Vol. 11, No. 1. P. 62–76. DOI: <https://doi.org/10.22105/riej.2021.291738.1229>.
101. Jain A. K. Data Clustering: 50 Years Beyond K-Means. Pattern Recognition Letters. 2010. Vol. 31, No. 8. P. 651–666. DOI: <https://doi.org/10.1016/j.patrec.2009.09.011>.
102. Арсірій О. О., Любомська О. М., Руденко О. В., Іванов Д. В. Гібридна рекомендаційна система для підтримки UI/UX дизайнерів. Інформатика. Культура. Техніка. 2024. Т. 1, № 1. С. 29–35. DOI: <https://doi.org/10.15276/ict.01.2024.30>.

103. Lismont J., Ram S., Vanthienen J., Lemahieu W., Baesens B. Predicting Interpurchase Time in a Retail Environment Using Customer-Product Networks. 2018. Vol. 104. P. 22–32. DOI: <https://doi.org/10.1016/j.eswa.2018.03.016>.
104. Kurgan L. A., Musilek P. A Survey of Knowledge Discovery and Data Mining Process Models. *The Knowledge Engineering Review*. 2006. Vol. 21, No. 1. P. 1–24. DOI: <https://doi.org/10.1017/S0269888906000737>.
105. Domingos P. A Few Useful Things to Know About Machine Learning. *Communications of the ACM*. 2012. Vol. 55, No. 10. P. 78–87. DOI: <https://doi.org/10.1145/2347736.2347755>.
106. McGarry K. A Survey of Interestingness Measures for Knowledge Discovery. *The Knowledge Engineering Review*. 2005. Vol. 20, No. 1. P. 39–61. DOI: <https://doi.org/10.1017/S0269888905000408>.
107. Dwivedi Y. K., Wang Y. Guest editorial: Artificial intelligence for B2B marketing: Challenges and opportunities. *Industrial Marketing Management*. 2022. Vol. 105. P. 109–113. DOI: <https://doi.org/10.1016/j.indmarman.2022.06.001>.
108. Keegan B. J., Dennehy D., Naudé P. Implementing Artificial Intelligence in Traditional B2B Marketing Practices. 2024. Vol. 26. P. 1025–1039. DOI: <https://doi.org/10.1007/s10796-022-10294-1>.
109. Tashman L. J. Out-of-Sample Tests of Forecasting Accuracy: An Analysis and Review. *International Journal of Forecasting*. 2000. Vol. 16, No. 4. P. 437–450. DOI: [https://doi.org/10.1016/S0169-2070\(00\)00065-0](https://doi.org/10.1016/S0169-2070(00)00065-0).
110. Kaufman S., Rosset S., Perlich C., Stitelman O. Leakage in Data Mining: Formulation, Detection, and Avoidance. *ACM ToKDD*. 2012. Vol. 6, No. 4. Article 15. DOI: <https://doi.org/10.1145/2382577.2382579>.
111. Shani G., Gunawardana A. Evaluating Recommendation Systems. In: *Recommender Systems Handbook*. New York: Springer, 2011. P. 257–297. DOI: https://doi.org/10.1007/978-0-387-85820-3_8.
112. Herlocker J. L., Konstan J. A., Terveen L. G., Riedl J. T. Evaluating Collaborative Filtering Recommender Systems. *ACM Transactions on Information Systems*. 2004. Vol. 22, No. 1. P. 5–53. DOI: <https://doi.org/10.1145/963770.963772>.

ДОДАТОК А АКТ ВПРОВАДЖЕННЯ РЕЗУЛЬТАТІВ НА B2B- ПЛАТФОРМІ «БАЗА ВЗУТТЯ»



✉ bazaobuvi@gmail.com
☎ 380939554095

АКТ ВПРОВАДЖЕННЯ

результатів дисертаційної роботи на здобуття наукового ступеня доктор філософії за спеціальністю 122 Комп'ютерні науки на тему: *«Інтелектуальний аналіз даних B2B-клієнтів для персоналізації контенту систем електронної комерції»*
здобувача Іванова Дмитра Вячеславовича
під час проведення дослідницьких робіт на підприємстві ФОП Іванов Владислав Вячеславович (NEWOPT)

Ми, що підписалися нижче, з одного боку представник ФОП (NEWOPT) директор Іванов В. В., а з другого боку представники Навчально-наукового інституту комп'ютерних систем (ННІКС) директор ННІКС д.т.н., проф. Антошук С.Г. та директор ЦІКТ «ІТ-Шкубатор» к.т.н., доц. П.А. Швагірєв склали цей акт у підтвердження впровадження у науково-практичну діяльність підприємства ФОП (NEWOPT). Програмно-алгоритмічного забезпечення, яке міститься в дисертаційній роботі Іванова Д.В., а саме реалізований триетапний конвеєр для формування індивідуальних предикативних пропозицій в залежності від визначеного класу клієнта, який встановлено як окремий модуль на веб-ресурсі підприємства «База Взуття»)

В процесі впровадження прикладних розробок, які містяться в роботі Іванова Д.В. з 01.02.2026 по 01.03.2026 встановлено, що надане програмне забезпечення у вигляді окремого програмного модулю на мові Python дозволило персоналізувати контент веб-ресурсу (<https://bazaobuvi.com.ua/ua/>) у відповідності до класів користувачі. У підсумку комісією, щодо впровадження встановлено, що значення ефективності однієї рекомендації за метрикою ARPR збільшилось із 1,21 тис. грн до 6,91 тис. грн, тобто приблизно у 5,7 рази. Крім того за умови вартості одного комунікаційного звернення 150 грн операційні витрати зменшилися зі 138,00 тис. грн до 33,00 тис. грн. в вартість однієї успішної рекомендації знизилася з 3,83 тис. грн до 0,94 тис. грн.

Акт видано у зв'язку з захистом дисертації на здобуття наукового ступеня доктор філософії та він не є основою для фінансових розрахунків.

Від ФОП (NEWOPT)
Директор

Іванов Владислав

Від ННІКС
Директор ННІКС, д.т.н., професор

Світлана АНТОЩУК

Директор ЦІКТ «ІТ-Шкубатор» к.т.н., доц.

Павло ШВАГІРЄВ

08 квітня 2026
Адреса Україна, 65120, Одеса,
Вул. Базова, будинок 8.

08 квітня 2026 р
Адреса 65044, Одеса, пр. Шевченка 1,
Національний університет «Одеська політехніка»
Тел (8-048) 705-8476



ДОДАТОК Б АКТ ВПРОВАДЖЕННЯ РЕЗУЛЬТАТІВ У ТОВ «СТІЛЛІ»

Довідка

про впровадження результатів дисертаційної роботи на здобуття наукового ступеня
доктор філософії за спеціальністю 122 Комп'ютерні науки на тему:

*«Інтелектуальний аналіз даних B2B-клієнтів для персоналізації контенту систем
електронної комерції»*

здобувача Іванова Дмитра Вячеславовича

Довідку видано стосовно того, що у діяльності підприємства ТОВАРИСТВО З
ОБМЕЖЕНОЮ ВІДПОВІДАЛЬНІСТЮ "СТІЛЛІ" використовується програмно-
алгоритмічне забезпечення, які отримано в дисертаційній роботі здобувача кафедри
інформаційних систем Національного університету «Одеська політехніка» МОН
України Дмитра ІВАНОВА

Прийнято до впровадження програмно-аналітичний модуль для формування
індивідуальних предикативних пропозицій в залежності від визначеного класу
клієнта у системах B2B електронної комерції. Основою модуля є триетапний
конвеєр, який поєднує ціннісно-зважений аналіз товарних патернів, темпоральну
перевірку стабільності закупівельних циклів та класифікацію B2B-клієнтів за
ціннісно-зваженим темпоральним профілем. Це дозволяє формувати пропозиції не
лише на основі товарної релевантності, а й з урахуванням фінансової значущості
патерну, прогнозованого моменту звернення та класу клієнта.

Практичну апробацію результатів виконано шляхом інтеграції розробленого
конвеєра у вигляді Python-модуля в контур роботи B2B-платформи «База взуття».
Реалізований модуль використовує дані про клієнтів, замовлення, товарні позиції,
артикули, ціни та суми транзакцій, після чого формує активні рекомендації для
менеджера.

Ефективність запропонованого підходу підтверджено експериментальними
результатами, які свідчать про підвищення точності та своєчасності надання
індивідуальних предикативних пропозицій B2B-клієнтам, зменшення кількості
нерелевантних рекомендацій і підвищення практичної цінності кожного
комунікаційного звернення

Директор ТОВ "СТІЛЛІ"



ШВЕЦЬ Є. А.

ДОДАТОК В ДОВІДКА ПРО ВИКОРИСТАННЯ РЕЗУЛЬТАТІВ ДИСЕРТАЦІЙНОЇ РОБОТИ У НАВЧАЛЬНОМУ ПРОЦЕСІ НАЦІОНАЛЬНОГО УНІВЕРСИТЕТУ «ОДЕСЬКА ПОЛІТЕХНІКА»

«Затверджую»
Перший проректор, проректор з науково-
методичної та виховної роботи
Національного університету «Одеська
політехніка»
професор Євген НЕСТЕРЕНКО



Р.

Довідка

про впровадження результатів дисертаційної роботи на здобуття наукового ступеня доктор
філософії за спеціальністю 122 Комп'ютерні науки на тему:
*«Інтелектуальний аналіз даних B2B-клієнтів для персоналізації контенту систем електронної
комерції»*
здобувача Іванова Дмитра Вячеславовича
в навчальному процесі Національного університету «Одеська політехніка»

Довідку видано в зв'язку з тим що у навчальних курсах «Інтелектуальний аналіз даних»,
«Методи та системи штучного інтелекту» та «Методологія розробки інтелектуальних
інформаційних систем», які викладаються здобувачам першого та другого рівнів підготовки за
освітньою програмою «Комп'ютерні науки» на кафедрі інформаційних систем використовується
програмно-алгоритмічне забезпечення, яке розроблено в дисертаційній роботі Дмитра ІВАНОВА.

Для створення програмно-алгоритмічного забезпечення використано наступні науково-
практичні результати:

1. Ціннісно-зважений темпоральний підхід до інтелектуального аналізу транзакційно-
поведінкових даних, які отримуються в системах електронної комерції;
2. Метод класифікації клієнтів систем електронної комерції, який на відміну від існуючого
RFM-аналізу базується на побудованому ціннісно-зваженому темпоральному профілі клієнта;
3. Триетапний конвеєр формування індивідуальних предикативних пропозицій для клієнтів
інтелектуальних систем електронної комерції, методика його інтеграції в діючі веб-платформи та
система адаптованих метрик оцінювання ефективності рекомендацій.

Навчальний процес підтримано підготовленими матеріалами до методичних вказівок до
виконання лабораторних робіт по навчальним курсам «Інтелектуальний аналіз даних», «Методи та
системи штучного інтелекту» та «Методологія розробки інтелектуальних інформаційних систем»,
які виглядаються здобувачам першого та другого рівнів підготовки за освітньою програмою
«Комп'ютерні науки» на кафедрі інформаційних систем

Директор Навчально-наукового
інституту комп'ютерних систем
д.т.н, професор

Голова Методичної комісії
Навчально-наукового
інституту комп'ютерних систем
к.т.н, доцент



Світлана АНТОЩУК



Анатолій НІКОЛЕНКО

**ДОДАТОК Г ДОВІДКА ПРО ВИКОРИСТАННЯ РЕЗУЛЬТАТІВ
ДИСЕРТАЦІЙНОЇ РОБОТИ У НДР НАЦІОНАЛЬНОГО УНІВЕРСИТЕТУ
«ОДЕСЬКА ПОЛІТЕХНІКА»**



Довідка № _____ від _____

про використання результатів дисертаційної роботи
Іванова Дмитра Вячеславовича
«ІНТЕЛЕКТУАЛЬНИЙ АНАЛІЗ ДАНИХ В2В-КЛІЄНТІВ ДЛЯ
ПЕРСОНАЛІЗАЦІЇ КОНТЕНТУ СИСТЕМ ЕЛЕКТРОННОЇ КОМЕРЦІЇ»
у науково-дослідницької діяльності
Національного університету «Одеська Політехніка»

Довідка видана в тому, що у науково-дослідницькій діяльності Національного університету «Одеська політехніка», а саме при розробці нейронних мереж щодо розподіленої обробки великих даних та інших практичних застосунків, використані наступні наукові результати, отримані у дисертаційній роботі Іванова Дмитра Вячеславовича:

- моделі та методи інтелектуального ціннісно-зваженого темпорального аналізу масивів великих деталізованих транзакційно-поведінкових даних клієнтів систем електронної комерції

Дисертацію виконано згідно тематичних планів НДР Одеської політехніки за період 2022-2026 рр.

Проректор



Дмитро ДМИТРИШИН

Керівники:

НДР № 254-145 «Розробка нейромережових технологій
для практичних застосунків»
(№ держ. реєстрації 01240005141)

Олена АРСІЙ

НДР № 164-145 «Методи та засоби інформаційного забезпечення
діагностування, вдосконалення
та експлуатації енергетичного обладнання».
(№ держреєстрації 0119U002004)

Віталій КОСТЕНКО

ДОДАТОК Д ФРАГМЕНТИ ВИХІДНОГО КОДУ

```
"""
```

```
=====
```

Порівняльний аналіз чотирьох стратегій рекомендацій для B2B-системи

```
=====
```

Стратегія 1: Стандартний FP-Growth (частотний підхід)

Базовий орієнтир – пошук частих наборів товарів без урахування вартості.

Стратегія 2: UP-Growth (ціннісно-зважений підхід)

Замість частоти використовується фінансова вага транзакції $U(T)$.

Виявляє патерни, що формують основну частку витрат, а не просто масові.

Стратегія 3: UP-Growth + Cyclic Evaluation (темпоральна верифікація)

Додає аналіз часових циклів закупівель кожного клієнта.

Рекомендація генерується лише тоді, коли тестовий горизонт збігається з очікуваним вікном повторної потреби (перевірка через CV та mu).

Стратегія 4: UP-Growth + Cyclic + Customer Classification (класифікація клієнтів)

Розширення Стратегії 3: клієнти сегментуються за фінансовим профілем

та стабільністю циклу. Для кожного сегменту встановлюється своя ширина вікна тригера – щоб балансувати між точністю та охопленням.

Логіка вікна тригера у Стратегії 4:

Преміум (найвищий mean_uT) → $k=0.5$ → вікно 7 днів (вузьке, висока точність)

Стандартний → $k=1.0$ → вікно 14 днів

Економ (найнижчий mean_uT) → $k=1.5$ → вікно 21 день (широке, вище охоплення)

Стратегія 3 використовує фіксоване вікно WINDOW=14 для всіх клієнтів.

```
=====
```

```
"""
```

```
import pandas as pd
```

```
import numpy as np
```

```
import os
```

```
from datetime import timedelta
```

```
from sklearn.preprocessing import MinMaxScaler # нормалізація ознак перед  
кластеризацією
```

```
from sklearn.cluster import KMeans # алгоритм k-means для  
сегментації клієнтів
```

```
from google.colab import files
```

```
# =====
```

```

# БЛОК 1: ПАРАМЕТРИ
# Усі налаштовувані константи зібрані тут.
# Щоб змінити поведінку алгоритму - редагуйте лише цей блок.
# =====

FILE_NAME    = 'table_full.xlsx' # назва файлу з даними

# --- Параметри алгоритмів пошуку патернів ---
MIN_FREQ     = 3                 # Мінімальна кількість транзакцій, у яких має зустрітись
                                # набір товарів, щоб потрапити до результатів FP-Growth
                                # (Стратегія 1). Знижено до 3, бо B2B-дані розріджені.

MIN_VALUE    = 50_000           # Мінімальна сумарна фінансова підтримка набору, грн
                                # (Стратегії 2-4). Набір включається до результатів,
                                # якщо  $TWU(X) \geq MIN\_VALUE$ . Відсікає фінансово незначущі
                                # комбінації, навіть якщо вони часто зустрічаються.

TOP_N        = 500              # Кількість найкращих патернів для симуляції.
                                # Відбираємо TOP_N за спаданням підтримки/корисності,
                                # щоб обмежити кількість перевірок і час виконання.

MAX_ITEMS    = 15               # Максимальна кількість унікальних SKU в одній транзакції.
                                # Транзакції довші за MAX_ITEMS відкидаються як аномальні,
                                # бо
                                # вони спричиняють комбінаторний вибух  $2^k$ .

# --- Параметри темпоральної верифікації (Стратегії 3, 4) ---
MAX_CV       = 0.50            # Поріг коефіцієнта варіації  $CV = \sigma/\mu$ .
                                # Якщо CV клієнта для патерну перевищує MAX_CV,
                                # закупівлі вважаються хаотичними й виключаються
                                # з тригерів. При  $CV \leq MAX\_CV$  - цикл стабільний.

MIN_DATES    = 2               # Мінімальна кількість зафіксованих дат закупівлі
                                # патерну одним клієнтом. При  $n \geq 2$  маємо хоча б
                                # один інтервал  $\delta$  для розрахунку  $\mu$ .

WINDOW       = 14              # Базова ширина вікна активації тригера, днів.
                                # Тригер «спрацьовує», якщо вікно  $[expected \pm WINDOW]$ 
                                # перетинається з тестовим горизонтом.
                                # У Стратегії 4 множиться на k_coef по сегменту.

DEFAULT_K    = 1.0             # Коефіцієнт вікна за замовчуванням для клієнтів,

```

```

# які не потрапили до класифікації (замало патернів).

TRAIN_RATIO = 0.80      # Частка транзакцій, що йдуть у навчальну вибірку.
                        # Решта (20%) - тестовий горизонт для верифікації.

# =====
# БЛОК 2: ЗАВАНТАЖЕННЯ ТА ПІДГОТОВКА ДАНИХ
# Читаємо Excel, очищуємо, розраховуємо фінансову вагу рядка,
# фільтруємо довгі транзакції та ділимо масив на Train / Test.
# =====

df = pd.read_excel(FILE_NAME, header=0)

# Прив'язка колонок за позицією (незалежно від назви заголовку у файлі)
col_date      = df.columns[0]    # дата оформлення рахунку-фактури
col_order_id  = df.columns[1]    # унікальний ID транзакції (TID)
col_client_id = df.columns[2]    # ID контрагента
col_item_id   = df.columns[3]    # артикул SKU (ростовка)
col_item_name = df.columns[4]    # читабельна назва артикулу
col_price     = df.columns[6]    # індивідуальна ціна за одиницю, грн
col_qty       = df.columns[7]    # кількість одиниць у рядку

# --- Очистка та типізація ---
# Видаляємо рядки без ключових полів - вони не можуть брати участь в аналізі
df = df.dropna(subset=[col_date, col_order_id, col_client_id, col_item_id])

# SKU та клієнт → рядок: уникаємо змішання int/float/str при групуванні
df[col_item_id]   = df[col_item_id].astype(str)
df[col_client_id] = df[col_client_id].astype(str)
df[col_item_name] = df[col_item_name].astype(str)
df[col_date]      = pd.to_datetime(df[col_date])

# Ціна та кількість → числові; некоректні значення замінюємо нулями
df[col_price] = pd.to_numeric(df[col_price],
errors='coerce').fillna(0).round(0).astype(int)
df[col_qty]   = pd.to_numeric(df[col_qty], errors='coerce').fillna(0)

# Фінансова вага рядка:  $u(i_k, T) = q_k * p_k$ 
# Це головна відмінність від бінарної моделі FP-Growth -
# кожен рядок накладної несе власну грошову цінність
df['line_value'] = (df[col_price] * df[col_qty]).round(0).astype(int)

```

```

# Словник для розшифрування SKU у читабельну назву (для звітів)
id_to_name = pd.Series(df[col_item_name].values, index=df[col_item_id]).to_dict()

# --- Фільтрація аномально довгих транзакцій ---
# Транзакції з k > MAX_ITEMS унікальних SKU викликають комбінаторний вибух 2^k.
# Наприклад, k=30 → понад мільярд комбінацій. Такі замовлення виключаємо.
order_lengths = df.groupby(col_order_id)[col_item_id].nunique()
valid_orders = order_lengths[order_lengths <= MAX_ITEMS].index
df_filtered = df[df[col_order_id].isin(valid_orders)].copy()
df_filtered = df_filtered.sort_values(col_date) # сортування за датою -
# обов'язково для Time-based Split

# --- Хронологічний поділ Train / Test (Time-based Split) ---
# Транзакції відсортовані за датою. Перші TRAIN_RATIO йдуть у навчання,
# решта - у тест. Це запобігає витокі даних із «майбутнього» (Data Leakage),
# який виникає при випадковому розбитті.
all_orders = df_filtered[col_order_id].drop_duplicates().tolist()
split_idx = int(len(all_orders) * TRAIN_RATIO)
train_ids = set(all_orders[:split_idx])
test_ids = set(all_orders[split_idx:])

df_train = df_filtered[df_filtered[col_order_id].isin(train_ids)].copy()
df_test = df_filtered[df_filtered[col_order_id].isin(test_ids)].copy()

# Межі горизонтів
max_date = df_filtered[col_date].max() # остання дата в усьому масиві
split_date = df_train[col_date].max() # остання дата навчальної вибірки
test_start_date = split_date + timedelta(days=1) # перший день тестового горизонту

# --- Словники навчальної вибірки ---
# Словники використовуються замість DataFrame для швидкого доступу O(1)
# під час перебору мільйонів пар клієнт-патерн у симуляції.

# Загальна вартість кожної транзакції U(T) = Σ line_value по TID
train_order_values =
df_train.groupby(col_order_id)['line_value'].sum().to_dict()

# Список SKU у кожній транзакції (для побудови FP-дерева)
train_order_to_items =
df_train.groupby(col_order_id)[col_item_id].apply(list).to_dict()

```

```

# Клієнт і дата кожної транзакції (для часової історії та прив'язки до
контрагента)
train_order_client =
df_train.groupby(col_order_id)[col_client_id].first().to_dict()
train_order_date = df_train.groupby(col_order_id)[col_date].first().to_dict()

# --- Словники тестової вибірки ---
test_order_to_items =
df_test.groupby(col_order_id)[col_item_id].apply(list).to_dict()
test_order_values = df_test.groupby(col_order_id)['line_value'].sum().to_dict()
test_order_client =
df_test.groupby(col_order_id)[col_client_id].first().to_dict()

# Інвертований індекс: client → список його тестових замовлень з наборами товарів.
# Використовується для перевірки успіху рекомендації: чи дійсно клієнт
# купив рекомендований набір у тестовому горизонті.
test_client_purchases = {}
for oid, items in test_order_to_items.items():
    c = test_order_client.get(oid)
    if c not in test_client_purchases:
        test_client_purchases[c] = []
    test_client_purchases[c].append({'order_id': oid, 'items': set(items)})

total_test_revenue = sum(test_order_values.values())
total_test_orders = len(test_order_to_items)

print("Дані підготовлено.")
print(f"Train: {len(train_order_to_items)} транзакцій | Test:
{total_test_orders} транзакцій")
print(f"Тестовий горизонт: {test_start_date.date()} - {max_date.date()}")
print(f"Загальна виручка у тесті: {total_test_revenue:,.0f} грн")

# =====
# БЛОК 3: ЯДРО FP-GROWTH
# Реалізація алгоритму FP-Growth на основі префіксного дерева (FP-Tree).
# Підтримує два режими:
#   - частотний (Стратегія 1): вага вузла = кількість транзакцій (+1)
#   - ціннісно-зважений (Стратегії 2-4): вага вузла = U(T) транзакції
# =====

class TreeNode:
    """

```

Вузол FP-дерева.

Атрибути:

name - ідентифікатор SKU, що зберігається у вузлі
 count - накопичена вага (кількість або сума $U(T)$ залежно від режиму)
 nodeLink - посилання на наступний вузол із тим самим SKU в іншій гілці
 (утворює зв'язаний список для швидкого обходу по артикулу)
 parent - батьківський вузол (потрібен для відновлення префіксного шляху)
 children - словник дочірніх вузлів {SKU: TreeNode}

"""

```
def __init__(self, name, count, parent):
    self.name = name; self.count = count
    self.nodeLink = None; self.parent = parent; self.children = {}

def inc(self, v):
    """Збільшує вагу вузла на v при додаванні нової транзакції."""
    self.count += v
```

```
def create_tree(data_dict, min_sup):
```

"""

Перший прохід: побудова FP-дерева з вхідного словника транзакцій.

Параметри:

data_dict - {frozenset(items): вага} - транзакції з агрегованими вагами
 min_sup - мінімальний поріг (частота або корисність залежно від режиму)

Повертає:

(root, header_table) - корінь дерева та заголовкова таблиця

Заголовкова таблиця header - словник {SKU: [загальна_вага, перший_вузол]}.
 Перший_вузол - голова зв'язаного списку вузлів з цим SKU по всьому дереву.
 Використовується для ефективного обходу дерева при генерації умовних баз.

"""

```
# Підрахунок сумарної ваги кожного артикулу по всіх транзакціях
header = {}
for trans in data_dict:
    for item in trans:
        header[item] = header.get(item, 0) + data_dict[trans]

# Видалення артикулів, що не досягають порогу - прунінг першого рівня
for item in list(header):
```

```

        if header[item] < min_sup:
            del header[item]

freq = set(header)
if not freq:
    return None, None # жодного частого артикулу - дерево порожнє

# Перетворення header у формат [загальна_вага, None (посилання ще немає)]
for item in header:
    header[item] = [header[item], None]

root = TreeNode('Null', 1, None) # фіктивний корінь

# Другий прохід: вставка кожної транзакції у дерево
for tset, cnt in data_dict.items():
    # Залишаємо лише часті артикули і сортуємо за спаданням ваги -
    # це забезпечує максимальне перекриття гілок і компактність дерева
    local = {i: header[i][0] for i in tset if i in freq}
    if local:
        ordered = [v[0] for v in sorted(local.items(),
                                         key=lambda p: p[1], reverse=True)]
        _update_tree(ordered, root, header, cnt)

return root, header

def _update_tree(items, node, header, cnt):
    """
    Рекурсивно вставляє відсортовану послідовність артикулів у гілку дерева.

    Якщо дочірній вузол з items[0] вже існує - збільшує його вагу на cnt.
    Якщо не існує - створює новий вузол і оновлює зв'язаний список у header.
    """
    if items[0] in node.children:
        # Вузол існує → просто накопичуємо вагу транзакції
        node.children[items[0]].inc(cnt)
    else:
        # Вузол новий → створюємо і прив'язуємо до зв'язаного списку header
        node.children[items[0]] = TreeNode(items[0], cnt, node)
        if header[items[0]][1] is None:
            header[items[0]][1] = node.children[items[0]]
        else:

```

```

        _update_header(header[items[0]][1], node.children[items[0]])

# Рекурсія для решти артикулів послідовності
if len(items) > 1:
    _update_tree(items[1:], node.children[items[0]], header, cnt)

def _update_header(n, target):
    """
    Проходить до кінця зв'язаного списку вузлів з однаковим SKU
    і додає новий вузол target у хвіст.
    """
    while n.nodeLink:
        n = n.nodeLink
    n.nodeLink = target

def _ascend(node, path):
    """
    Рекурсивно піднімається від вузла до кореня,
    збираючи імена вузлів у список path.
    Використовується для відновлення префіксного шляху транзакції.
    """
    if node.parent:
        path.append(node.name)
        _ascend(node.parent, path)

def _prefix_paths(item, node):
    """
    Збирає всі префіксні шляхи (умовну базу патернів) для заданого артикулу item.

    Проходить по зв'язаному списку вузлів з цим артикулом.
    Для кожного вузла відновлює шлях від нього до кореня (_ascend).
    Шлях без останнього елемента (самого item) стає умовною транзакцією.

    Повертає словник {frozenset(умовна_транзакція): вага_вузла}.
    """
    cond = {}
    while node:
        path = []
        _ascend(node, path)

```

```

    if len(path) > 1:
        cond[frozenset(path[1:])] = node.count
        node = node.nodeLink
    return cond

```

```
def mine_tree(tree, header, min_sup, prefix, result):
```

```
    """
```

```
    Рекурсивне видобування патернів з FP-дерева методом умовних баз.
```

```
    Для кожного артикулу (від найменш до найбільш важкого):
```

1. Формує новий кандидатний набір prefix | {item}
2. Додає його до результатів з його підтримкою/корисністю
3. Будує умовне FP-дерево з префіксних шляхів цього артикулу
4. Рекурсивно видобуває патерни з умовного дерева

```
    Параметри:
```

```
        prefix - поточний prefix-набір (зростає при рекурсії)
```

```
        result - список (набір, вага) для накопичення знайдених патернів
```

```
    """
```

```
    for item in [v[0] for v in sorted(header.items(), key=lambda p: p[1][0])]:
```

```
        new_set = prefix | {item}
```

```
        result.append((new_set, header[item][0]))
```

```
        # Будуємо умовну базу патернів і умовне дерево для цього артикулу
```

```
        cond = _prefix_paths(item, header[item][1])
```

```
        cond_tree, cond_header = create_tree(cond, min_sup)
```

```
        if cond_header:
```

```
            mine_tree(cond_tree, cond_header, min_sup, new_set, result)
```

```
def run_fp(order_items, order_vals, min_sup):
```

```
    """
```

```
    Точка входу для запуску FP-Growth.
```

```
    Параметри:
```

```
        order_items - словник {TID: [список SKU]}
```

```
        order_vals - словник {TID: U(T)} або None
```

```
        None → частотний режим (Стратегія 1): вага = 1 для кожної  
транзакції
```

```
        dict → ціннісний режим (Стратегії 2-4): вага = U(T)  
транзакції
```



```

top_val = [x[0] for x in sorted(val_sets,
                                key=lambda x: x[1], reverse=True)[:TOP_N]]

print(f"Стратегія 1: {len(top_std)} патернів")
print(f"Стратегії 2/3/4: {len(top_val)} патернів")

# =====
# БЛОК 5: ПОВУДОВА ЧАСОВОЇ ІСТОРІЇ КЛІЄНТІВ
# Для кожного клієнта зберігаємо список його транзакцій у форматі
# (дата, множина_SKU, вартість). Це вхідні дані для темпоральної верифікації.
# Використовуємо множину (set) SKU для швидкої перевірки issubset() O(1).
# =====

client_history = {}
for oid, items in train_order_to_items.items():
    c = train_order_client.get(oid)
    d = train_order_date.get(oid)
    v = train_order_values.get(oid, 0)
    if c not in client_history:
        client_history[c] = []
    client_history[c].append((d, set(items), v))

# =====
# БЛОК 6: ТЕМПОРАЛЬНА ВЕРИФІКАЦІЯ
# Дві функції, що реалізують другий етап методу:
#   get_cycle_stats - розраховує статистику циклу закупівель ( $\mu$ ,  $\sigma$ , CV)
#   is_trigger_active - перевіряє, чи тригер активний у тестовому горизонті
# =====

def get_cycle_stats(client, itemset):
    """
    Розраховує статистику закупівельного циклу для пари (клієнт, патерн).

    Кроки:
    1. Відбирає з client_history дати, коли клієнт купував даний itemset
    2. Перевіряє, чи достатньо дат ( $\geq$  MIN_DATES) для аналізу
    3. Розраховує вектор міжзакупівельних інтервалів  $\delta_i = t_{i+1} - t_i$ 
    4. Обчислює  $\mu$  (середній цикл) та  $\sigma$  (відхилення)

    Параметри:

```

client - ідентифікатор клієнта
 itemset - frozenset SKU, що перевіряється (патерн)

Повертає:

(mu, sigma, last_date, mean_value) - або None, якщо даних недостатньо

```
mu          - середній інтервал між закупівлями, дні (очікуваний цикл)
sigma       - середньоквадратичне відхилення інтервалів, дні
last_date   - дата останньої закупівлі патерну (від неї рахується тригер)
mean_value  - середня вартість транзакції з цим патерном, грн
"""
history = client_history.get(client, [])

# Відбираємо лише ті транзакції, де itemset є підмножиною купленого
dates = sorted([d for d, items, _ in history if itemset.issubset(items)])
values = [v for d, items, v in history if itemset.issubset(items)]

# Потрібно щонайменше MIN_DATES дат, щоб отримати хоча б один інтервал
if len(dates) < MIN_DATES:
    return None

# Вектор інтервалів  $\delta_i$  між послідовними закупівлями
intervals = [(dates[i+1] - dates[i]).days for i in range(len(dates)-1)]
intervals = [x for x in intervals if x > 0] # виключаємо нульові (одна дата)
if len(intervals) < 1:
    return None

mu = np.mean(intervals)
# ddof=1 - незміщена оцінка (ділення на n-1); при одному інтервалі  $\sigma = 0$ 
sigma = np.std(intervals, ddof=1) if len(intervals) > 1 else 0.0

return mu, sigma, dates[-1], np.mean(values)
```

```
def is_trigger_active(mu, last_date, window_days):
```

```
    """
```

Перевіряє, чи входить очікувана дата наступної закупівлі у тестовий горизонт.

Логіка:

expected = last_date + μ (найімовірніша дата наступної закупівлі)

Вікно активації: [expected - window_days, expected + window_days]

Тригер активний, якщо це вікно перетинається з [test_start_date, max_date].

Параметри:

```
mu          - середній цикл закупівлі, дні
last_date   - дата останньої зафіксованої закупівлі патерну
window_days - ширина вікна в днях (WINDOW або WINDOW * k_coef)
```

Чому вікно, а не точна дата:

μ - це оцінка, а не точне значення. Вікно $\pm window_days$ враховує природну мінливість циклу і підвищує шанс «зловити» реальну закупівлю.

Стратегія 3: $window_days = WINDOW$ (однакове для всіх).

Стратегія 4: $window_days = WINDOW * k_coef$ (індивідуальне по сегменту).

```
"""
```

```
expected = last_date + timedelta(days=int(mu))
low       = expected - timedelta(days=window_days)
high      = expected + timedelta(days=window_days)
# Перетин двох відрізків: [low, high] та [test_start_date, max_date]
return low <= max_date and high >= test_start_date
```

```
# =====
# БЛОК 7: КЛАСИФІКАЦІЯ КЛІЄНТІВ (Стратегія 4)
# Третій етап методу - сегментація контрагентів за профілем закупівель.
# Дві допоміжні функції:
#   build_client_features - формує вектор ознак для кожного клієнта
#   classify_clients      - кластеризує клієнтів і призначає k_coef
# =====
```

```
def build_client_features():
```

```
    """
```

Формує матрицю ознак клієнтів для подальшої кластеризації.

Для кожного клієнта агрегує характеристики з усіх його стабільних патернів (тих, де $CV \leq MAX_CV$). Стабільні патерни - це ті, де цикл закупівель є достатньо регулярним для прогнозування.

Три ознаки вектора $f_j = (mean_mu, mean_cv, mean_uT)$:

```
mean_mu - середній цикл закупівлі, дні
        Відображає горизонт планування: малий  $\mu \rightarrow$  часті закупівлі,
        великий  $\mu \rightarrow$  сезонний або великий оптовик.
```

```
mean_cv - середня стабільність циклу ( $CV = \sigma/\mu$ )
```

Малий CV → жорстко регулярний клієнт (виробничий регламент).
 Великий CV → нестабільний, але ще в межах порогу MAX_CV.

mean_uT - середня вартість транзакції, грн

Фінансово-зважений аналог показника Monetary у RFM.

Ключова ознака для визначення сегменту (преміум / економ).

Повертає DataFrame з індексом 'client' або порожній DataFrame.

"""

records = {}

for itemset in top_val:

for client in client_history:

stats = get_cycle_stats(client, itemset)

if stats is None:

continue

mu, sigma, _, mean_val = stats

cv = (sigma / mu) if mu > 0 else 0.0

if cv > MAX_CV:

continue # нестабільний патерн - не беремо до уваги

if client not in records:

records[client] = {'mu_list': [], 'cv_list': [], 'val_list': []}

records[client]['mu_list'].append(mu)

records[client]['cv_list'].append(cv)

records[client]['val_list'].append(mean_val)

rows = []

for client, d in records.items():

rows.append({

'client': client,

'mean_mu': np.mean(d['mu_list']),

'mean_cv': np.mean(d['cv_list']),

'mean_uT': np.mean(d['val_list']),

})

return pd.DataFrame(rows).set_index('client') if rows else pd.DataFrame()

def classify_clients(n_clusters=3):

"""

Класифікує клієнтів методом k-means і повертає словник client → k_coef.

Кроки:

1. Отримуємо матрицю ознак із build_client_features()

2. Нормалізуємо ознаки у діапазон [0, 1] через MinMaxScaler, щоб різні масштаби (дні vs гривні) не спотворювали відстані
3. Запускаємо KMeans(k=n_clusters) – мінімізує внутрішньокластерну дисперсію
4. Ідентифікуємо сегменти за mean_uT центроїду кластера:
 - найвищий mean_uT → преміум → k_coef = 0.5
 - найнижчий mean_uT → економ → k_coef = 1.5
 - решта → стандартний → k_coef = 1.0
5. Зберігаємо профілі кластерів у client_clusters.csv

Логіка призначення k_coef:

k_coef визначає ширину вікна тригера: window_days = WINDOW * k_coef
 Преміум (k=0.5, 7 днів): контактуємо точно у момент потреби – для цінного клієнта надлишкові контакти знецінюють комунікацію.
 Економ (k=1.5, 21 день): ширше вікно – вища ймовірність охопити угоду за рахунок допустимо меншої точності.

Параметри:

n_clusters – кількість сегментів (за замовчуванням 3)

Повертає:

Словник {client_id: k_coef} або порожній словник, якщо даних замало.

"""

```
feat_df = build_client_features()
```

```
if feat_df.empty:
```

```
    print("Класифікація: недостатньо даних, застосовується DEFAULT_K.")
```

```
    return {}
```

```
# Захист: якщо класифікованих клієнтів менше n_clusters – зменшуємо k
n_actual = min(n_clusters, len(feat_df))
```

```
# Нормалізація: кожна ознака → [0, 1]
```

```
scaler = MinMaxScaler()
```

```
X_scaled = scaler.fit_transform(feat_df[['mean_mu', 'mean_cv', 'mean_uT']])
```

```
# Кластеризація k-means: n_init=10 → 10 незалежних запусків,
```

```
# обирається результат з найменшою інерцією (стійкість до стартових точок)
```

```
km = KMeans(n_clusters=n_actual, random_state=42, n_init=10)
```

```
feat_df['cluster'] = km.fit_predict(X_scaled)
```

```
# Статистика по кластерах для ідентифікації сегментів
```

```
cluster_stats = feat_df.groupby('cluster')[['mean_uT', 'mean_cv',
```

```
'mean_mu']].mean()
```

```

sorted_by_val = cluster_stats['mean_uT'].sort_values(ascending=False)
premium_cluster = sorted_by_val.index[0] # найвищий середній чек → преміум
economy_cluster = sorted_by_val.index[-1] # найнижчий середній чек → економ

# Призначення k_coef: преміум → вузьке вікно, економ → широке
k_map = {}
for cl in cluster_stats.index:
    if cl == premium_cluster:
        k_map[cl] = 0.5 # вузьке вікно 7 днів - висока точність
    elif cl == economy_cluster:
        k_map[cl] = 1.5 # широке вікно 21 день - більше охоплення
    else:
        k_map[cl] = 1.0 # стандартне вікно 14 днів

# Словник client → k_coef для використання у симуляції
client_k = {row.Index: k_map[row.cluster] for row in feat_df.itertuples()}

# Збереження детального профілю кожного клієнта у CSV
cluster_report = feat_df.copy()
cluster_report['k_coef'] = cluster_report['cluster'].map(k_map)
cluster_report['window_days'] = (WINDOW *
cluster_report['k_coef']).astype(int)
cluster_report.reset_index().to_csv('client_clusters.csv',
index=False, encoding='utf-8-sig')

# Вивід зведеної таблиці сегментів у консоль
print(f"\nКласифікація клієнтів: {len(feat_df)} клієнтів → {n_actual}
сегменти")
display_df = cluster_stats.copy()
display_df['k_coef'] = display_df.index.map(k_map)
display_df['window_days'] = (WINDOW * display_df['k_coef']).astype(int)
print(display_df.round(1).to_string())

return client_k

print("\nКласифікація клієнтів...")
client_k_map = classify_clients(n_clusters=3)

# =====
# БЛОК 8: СИМУЛЯЦІЯ

```

```
# Основна функція, що моделює роботу відділу продажів на тестовому горизонті.
# Для кожного патерну та кожного клієнта перевіряє:
#   - чи є у клієнта достатня історія (Стратегії 1, 2)
#   - чи стабільний цикл закупівель (Стратегії 3, 4)
#   - чи потрапляє тригер у тестовий горизонт (Стратегії 3, 4)
# Потім фіксує успіх, якщо клієнт дійсно зробив відповідне замовлення у тесті.
# =====
```

```
def simulate(candidate_sets, use_cyclic=False, use_classification=False):
```

```
    """
```

```
    Симулює генерацію рекомендацій і перевірку їх успішності на тестових даних.
```

```
    Параметри:
```

```
        candidate_sets      - список frozenset патернів (top_std або top_val)
        use_cyclic           - True → застосовувати темпоральну верифікацію
                               False → звертатись до всіх клієнтів з достатньою
                               кількістю закупівель (Стратегії 1, 2)
        use_classification   - True → використовувати індивідуальний k_coef
                               по сегменту (Стратегія 4)
                               False → фіксоване вікно WINDOW (Стратегія 3)
```

```
    Логіка підрахунку метрик:
```

```
        recs                - лічильник згенерованих рекомендацій
        hit_orders           - множина ID тестових замовлень, що збіглися з рекомендацією
                               (множина, а не список, бо одна угода не рахується двічі
                               навіть якщо її «захопило» кілька патернів)
```

```
    Повертає словник метрик:
```

```
        Рекомендацій        - загальна кількість звернень до клієнтів
        Успішних угод        - кількість угод, що збіглися з рекомендацією
        Конверсія, %         - Успішних угод / Рекомендацій * 100
        Виручка, грн         - сумарна вартість успішно передбачених транзакцій
        ARPR, грн/рек.       - Виручка / Рекомендацій (рентабельність 1 звернення)
```

```
    """
```

```
    recs = 0
```

```
    hit_orders = set() # множина: уникаємо подвійного підрахунку угоди
```

```
    for cset in candidate_sets:
```

```
        for client, history in client_history.items():
```

```
            if not use_cyclic:
```

```
                # --- Стратегії 1 і 2: без темпорального фільтру ---
```

```

# Генеруємо рекомендацію для всіх, хто купував патерн
# щонайменше MIN_DATES разів у навчальній вибірці
dates = sorted([d for d, items, _ in history
                 if cset.issubset(items)])
if len(dates) < MIN_DATES:
    continue
recs += 1

else:
    # --- Стратегії 3 і 4: з темпоральною верифікацією ---
    stats = get_cycle_stats(client, cset)
    if stats is None:
        continue # недостатньо даних для аналізу циклу

    mu, sigma, last_date, _ = stats
    cv = (sigma / mu) if mu > 0 else 0.0

    # Фільтр стабільності: якщо CV перевищує поріг - цикл хаотичний
    if cv > MAX_CV:
        continue

    # Визначення ширини вікна тригера
    if use_classification:
        # Стратегія 4: індивідуальна ширина вікна по сегменту
        k_coef = client_k_map.get(client, DEFAULT_K)
        window_days = int(WINDOW * k_coef)
    else:
        # Стратегія 3: базове вікно для всіх однакове
        window_days = WINDOW

    # Перевіряємо, чи очікувана дата закупівлі потрапляє у тест
    if not is_trigger_active(mu, last_date, window_days):
        continue

    recs += 1

# --- Перевірка успіху: чи є реальна угода у тесті? ---
# Шукаємо серед тестових замовлень клієнта таке,
# у якому рекомендований патерн є підмножиною купленого
if client in test_client_purchases:
    for order in test_client_purchases[client]:
        if cset.issubset(order['items']):

```

```

        hit_orders.add(order['order_id'])
        break # один збіг на замовлення достатньо

# Підрахунок метрик
hits      = len(hit_orders)
revenue = sum(test_order_values.get(oid, 0) for oid in hit_orders)
conv      = hits / recs * 100 if recs > 0 else 0.0
arpr      = revenue / recs      if recs > 0 else 0.0

return {
    'Рекомендацій':    recs,
    'Успішних угод':   hits,
    'Конверсія, %':     round(conv, 2),
    'Виручка, грн':     revenue,
    'ARPR, грн/рек.': round(arpr, 0),
}

# =====
# БЛОК 9: ЗАПУСК ЧОТИРЬОХ СТРАТЕГІЙ
# Кожна стратегія - це виклик simulate() з різними прапорами.
# Результати зберігаються у словнику results для подальшого аналізу.
# =====

print(f"\n{'='*62}")
print("ЗАПУСК СИМУЛЯЦІЇ")
print(f"{'='*62}")

results = {}

# Стратегія 1: частотні патерни, без часового фільтру
print("Стратегія 1 (FP-Growth)...")
results['1. FP-Growth'] = simulate(top_std)

# Стратегія 2: ціннісні патерни, без часового фільтру
print("Стратегія 2 (UP-Growth)...")
results['2. UP-Growth'] = simulate(top_val)

# Стратегія 3: ціннісні патерни + темпоральна верифікація (однакове вікно)
print("Стратегія 3 (UP-Growth + Cyclic)...")
results['3. UP-Growth + Cyclic'] = simulate(top_val, use_cyclic=True)

```

```

# Стратегія 4: ціннісні патерни + темпоральна верифікація + індивідуальне вікно
print("Стратегія 4 (UP-Growth + Cyclic + Class.)...")
results['4. UP-Growth + Cyclic + Class.'] = simulate(
    top_val, use_cyclic=True, use_classification=True)

# =====
# БЛОК 10: ЗВЕДЕНА ТАБЛИЦЯ ТА ЗБЕРЕЖЕННЯ РЕЗУЛЬТАТІВ
# Формуємо підсумкову таблицю, додаємо Hit Rate та Value Coverage,
# виводимо у консоль та зберігаємо у CSV.
# =====

df_res = pd.DataFrame(results).T
df_res.index.name = 'Стратегія'

# Hit Rate = частка тестових транзакцій, що були передбачені
df_res['Hit Rate, %'] = (
    df_res['Успішних угод'] / total_test_orders * 100).round(2)

# Value Coverage = частка тестової виручки, що припадає на передбачені угоди
df_res['Value Coverage, %'] = (
    df_res['Виручка, грн'] / total_test_revenue * 100).round(2)

print(f"\n{'='*62}")
print("ЗВЕДЕНІ РЕЗУЛЬТАТИ")
print(f"{'='*62}")
print(f"Тестовий горизонт: {test_start_date.date()} - {max_date.date()}")
print(f"Транзакцій: {total_test_orders} | Виручка: {total_test_revenue:,.0f} грн\n")
print(df_res.to_string())

# Відносні зміни між ключовими переходами
arpr_vals = {k: results[k]['ARPR, грн/пек.'] for k in results}
conv_vals = {k: results[k]['Конверсія, %'] for k in results}
recs_vals = {k: results[k]['Рекомендацій'] for k in results}

print(f"\n--- Відносні зміни ---")
pairs = [
    ('3 vs 2', '2. UP-Growth', '3. UP-Growth + Cyclic'),
    ('4 vs 3', '3. UP-Growth + Cyclic', '4. UP-Growth + Cyclic + Class.'),
]
for label, base, cur in pairs:

```

```

    mult = f"{arpr_vals[cur]/arpr_vals[base]:.2f}" if arpr_vals[base] > 0 else
    "n/a"
    print(f"{label}: рекомендацій {recs_vals[cur]-recs_vals[base]:+.0f} | "
          f"конверсія {conv_vals[cur]-conv_vals[base]:+.2f} п.п. | "
          f"ARPR {arpr_vals[cur]-arpr_vals[base]:+,.0f} рпн ({mult})")

# Збереження CSV та завантаження у браузер
df_res.reset_index().to_csv('strategy_comparison.csv',
                             index=False, encoding='utf-8-sig')
files.download('strategy_comparison.csv')

if os.path.exists('client_clusters.csv'):
    files.download('client_clusters.csv')

print("\nГотово. Файли завантажено.")

```

ДОДАТОК Е СПИСОК ПУБЛІКАЦІЙ ЗДОБУВАЧА

1. Арсірій О., Іванов Д. Моделі афінитивного аналізу транзакційних і поведінкових даних клієнтів систем B2B електронної комерції для створення персоналізованого контенту. Праці Одеського політехнічного університету. 2025. 2(72) С 90-101(<https://doi.org/10.15276/opu.2.72.2025>)
<https://pratsi.op.edu.ua/index.php/pratsi/article/view/549/511> ISSN 2076-2429 (print)
2. Arsirii , O., & Ivanov , D. (2026). A Model and Method of Transactional-behavioral Data Mining for B2B Content Personalization. Electronics and Control Systems, 1(87), 14–23. <https://doi.org/10.18372/1990-5548.87.20881> ISSN 1990-5548
3. Arsirii, O., & Ivanov , D. (2026). Data Mining of B2b Transactions Using a High-utility Temporal Approach. Electronics and Control Systems, 2(88), 16–31. <https://doi.org/10.18372/1990-5548.88.20958> ISSN 1990-5548
4. Olena O. Arsirii, Dmitriy V. Ivanov, Oksana Yu. Babilunha Analysis of models, methods and technologies for creating personalized content in B2B e-commerce systems. (2025). Інформатика. Культура. Техніка, 2, 273–278. <https://doi.org/10.15276/ict.02.2025.41> ISSN 2522-1523 (Online)
5. Арсірій О. О., Любомська О. М., Руденко О. В., Іванов Д. В. Гібридна рекомендаційна система для підтримки UI/UX дизайнерів. Матеріали міжнародної конференції Інформатика. Культура. Техніка. 2024; Том 1 № 1: 29–35
<https://ict.od.ua/index.php/journal/en/article/view/93>
<https://doi.org/10.15276/.01.2024.03> ISSN 2522-1523 (Online)

**ДОДАТОК Ж ВІДОМОСТІ ПРО АПРОБАЦІЮ РЕЗУЛЬТАТІВ
ДИСЕРТАЦІЙНОЇ РОБОТИ**

1. Сіроштан М.В., Іванов Д.В., Арсірій О.О. Дослідження ефективності методів машинного навчання на прикладі задачі класифікації предметів одягу. Project, Program, Portfolio Management. РЗМ-2024: Тези доповідей IX Міжнародної науково-практичної конференції : [у 2т.]. Відповідальний за випуск П.О. Тесленко
Том 1. Одеса: ІШІР, 2024 С. 201-204
<https://rep.nuos.edu.ua/server/api/core/bitstreams/f2c763dc-2728-47ca-af64-31618bd2a659/content>

2. Іванов Д.В., Арсірій О.О. Класифікація В2В-клієнтів системи електронної комерції за ціннісно-зваженим темпоральним профілем. Матеріали XVI міжнародної конференції "Сучасні Інформаційні Технології 2026" / "Modern Information Technology 2026", ІКС, Національний університет "Одеська політехніка", Одеса, Україна, 14-15 травня, 2026. – С. 118-121.
https://drive.google.com/drive/folders/1uXN7b84231YhSfT_tY6I9eMhPjDrIBKJ