

DOI: <https://doi.org/10.15276/ict.02.2025.41>
УДК 004.4'24

Аналіз моделей, методів та технологій створення персоналізованого контенту в системах B2B електронної комерції

Арсирій Олена Олександрівна¹⁾

Д-р техніч. наук, професор, завідувачка каф. Інформаційних систем
ORCID: <https://orcid.org/0000-0001-8130-9613>; e.arsiriy@gmail.com. Scopus Author ID: 54419480900

Іванов Дмитро Вячеславович¹⁾

Аспірант каф. Інформаційних систем
ORCID: orcid.org/0009-0009-3958-5310; ivanovdima9988@gmail.com

Бабілунга Оксана Юрївна¹⁾

Канд. техніч. наук, доцент каф. Інформаційних систем
ORCID: <https://orcid.org/0000-0001-6431-3557>; babilunga.onpu@gmail.com. Scopus Author ID: 35316966900

¹⁾ Національний університет “Одеська Політехніка”, пр. Шевченка, 1. Одеса, 65044, Україна

АНОТАЦІЯ

Зростання фінансової значущості та структурної складності сегмента ринку систем B2B електронної комерції, а також необхідністю підвищення його ефективності зумовило актуальність проведення системного аналізу моделей, методів та технологій створення персоналізованого контенту в таких системах. Визначено особливості B2B комерції та проведено порівняння систем B2B та B2C за основними показниками. Проведено аналізу існуючих моделей, методів та технологій створення персоналізованого контенту в системах B2B електронної комерції, а саме алгоритмів пошуку частих наборів товарів Apriori, FP-Growth, Eclat, що дозволяють виявляти комбінації товарів, які часто замовляються разом, метрик якості асоціативних правил (підтримка, достовірність, lift, conviction) для оцінювання значущості та корисності знайдених закономірностей в замовленнях товарів. Обґрунтовано необхідність розробки моделей афінитивного аналізу транзакційних і поведінкових даних у поєднанні з методами класифікації клієнтів для підвищення персоналізації товарного, інформаційного та рекомендаційного контенту системи B2B електронної комерції

Ключові слова: електронна комерція; системи B2B; транзакційні та поведінкові дані; афінитивний аналіз; асоціативні правила; Apriori, FP-Growth; Eclat; метрики якості

Актуальність. Актуальність проведення аналізу моделей, методів та технологій створення персоналізованого контенту в системах B2B електронної комерції зумовлена фінансовою значущістю та структурною складністю цього сегмента ринку, а також необхідністю підвищення його ефективності. Критична важливість ефективної B2B-персоналізації підтверджується величезними фінансовими обсягами цього сегмента. Світовий ринок B2B електронної комерції, оцінений у \$11.54 трильйона доларів США у 2024 році, значно перевершує ринок B2C (\$6.55 трлн). Прогноз його зростання до \$60.62 трильйона доларів США до 2034 року вказує на те, що навіть незначне підвищення ефективності взаємодії з клієнтами за допомогою персоналізації матиме мультитрильйонний економічний ефект [1]. Але встановлено, що традиційні методи взаємодії з клієнтами, що розроблені для B2C-сегмента та орієнтовані на емоції, імпульсивні покупки та простий користувацький досвід, є недостатньо ефективними для B2B [2]. Оскільки B2B-закупівлі є раціональними, мають високу вартість та керуються потребами в ROI (повернення інвестицій), виникає нагальна потреба в аналізі та адаптації існуючих моделей для вирішення таких B2B-викликів як: складність закупівельного центру (Buying Center) коли рішення про купівлю товару чи послуги приймається не однією людиною, а групою людей або відділів всередині організації-покупця, індивідуальні цінові умови, розрідженість даних (рідкі, але великі замовлення) та необхідність довгострокового партнерства [3]. Це створює прогалину в дослідженні методів, здатних ефективно працювати у B2B-середовищі. З іншого боку відомо, що для створення справді релевантного та раціонально обґрунтованого контенту (товарного, інформаційного, рекомендаційного) життєво необхідним є глибокий аналіз даних [4]. Це вимагає проведення систематичного аналізу існуючих моделей, які використовують транзакційні дані для виявлення афінитивних правил та типових закупівельних патернів та поведінкові дані для розуміння закупівельних сценаріїв та потреб різних ролей у компанії-

клієнті. Проведення такого аналізу дозволить виявити прогалини в наявних методах, обґрунтувати необхідність інтеграції аналітичних моделей афінитивний аналізу та класифікації із сучасними технологіями великих мовних моделей та закласти основу для розробки адаптованих підходів, здатних підвищити середній чек, лояльність та автоматизацію закупівель в B2B-системах.

Мета та задачі дослідження. Метою даного дослідження є проведення системного аналізу існуючих моделей, методів та технологій, які використовуються для створення персоналізованого контенту в B2B електронній комерції, та ідентифікувати недоліки для подальшого дослідження. До задач дослідження відносимо: визначення особливостей B2B комерції та порівняння систем B2B та B2C за основними показниками; проведення аналізу існуючих моделей, методи та технології створення персоналізованого контенту в системах B2B електронної комерції та також обґрунтування необхідності розробки моделей афінитивного аналізу транзакційних і поведінкових даних у поєднанні з методами класифікації клієнтів для підвищення персоналізації товарного, інформаційного та рекомендаційного контенту системи B2B електронної комерції

Основна частина роботи. Система B2B електронної комерції (B2B E-commerce System) це комплексна онлайн-платформа або набір взаємопов'язаних технологічних рішень, призначених для здійснення комерційних операцій, обміну інформацією та автоматизації бізнес-процесів між двома або більше юридичними особами (підприємствами) [3]. Ця система є цифровим середовищем, що забезпечує організаціям-продавцям можливість представляти свій товарний, інформаційний та рекомендаційний контент (каталоги, прайс-листи, умови співпраці) іншим організаціям-покупцям для їхніх виробничих, комерційних або перепродажних потреб. В таблиці 1 показано результати порівняльного аналізу систем B2B та B2C електронної комерції за основними показниками/

Таблиця 1. Порівняння систем B2B та B2C за основними показниками

Назва показника	B2B (Business-to-Business)	B2C (Business-to-Consumer)
Цільова аудиторія	Інші компанії (стартапи, підприємства, великі корпорації, державні установи, неприбуткові організації). Кількість покупців менша	Індивідуальні споживачі. Кількість покупців значно більша
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	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

Таким чином сучасні системи B2B електронної комерції функціонують у висококонкурентному інтернет середовищі, де від якості контенту залежить швидкість і точність задоволення потреб оптових клієнтів, що є вирішальним для B2B компанії.

Технології створення контенту для B2B e-commerce передбачають генерацію декількох типів контенту:

- товарного (опис товарів, супутні продукти, комерційні пропозиції);
- інформаційного (аналітичні звіти, поради, документація);
- рекомендаційного (персоналізовані добірки товарів та пропозицій).

На сьогодні більшість платформ використовують статичні методи формування контенту на основі фіксованих правил, або базуються на універсальних рекомендаціях для всіх клієнтів. Такий підхід часто не враховує індивідуальні закупівельні патерни, специфіку бізнесу клієнта, сезонність або залежності між товарами. Персоналізація можлива лише за умови детального аналізу транзакційних і поведінкових даних клієнтів. Зокрема:

– *аналіз транзакцій* дозволяє виявляти часті комбінації товарів, що закуповуються разом, та формувати набори рекомендацій;

– *аналіз поведінкових послідовностей* дає змогу визначати типові сценарії взаємодії клієнтів із системою, прогнозувати ймовірність замовлення, виявляти «гарячі» товари.

Об'єднання результатів цих видів аналізу з методами класифікації дозволяє сегментувати клієнтів за закупівельними характеристиками, а відтак формувати персоналізований контент, а саме *товарний* у вигляді адаптованих прайсів та добірок, *інформаційний* у вигляді рекомендації по трендах у категоріях, акційних пропозиціях, а також *рекомендаційний* у вигляді прогнозованих товарних наборів для конкретного клієнта чи групи.

Далі наведено більш детальну інформацію про особливості проведення афінитивного аналізу транзакційних та поведінкових даних клієнтів на прикладі системи B2B електронної комерції яка пропонує широкий спектр взуття для дрібних магазинів взуття (*Wholesale B2B Portal*) з метою створення персоналізованого контенту [5]. Мета афінитивного аналізу в B2B-системі – це виявити асоціації між різними комерційними подіями, транзакціями або поведінковими діями оптових клієнтів. Це дозволяє знайти асоціативні правила (association rules) для кількісного опису взаємного зв'язку між двома чи більше комерційними подіями [6,7]. У теорії аналізу асоціацій базовим визначенням є транзакція, але в B2B вона набуває більшої складності. 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.

Тоді формально якщо I множина всіх товарів $I=\{i_1, i_2, \dots, i_n\}$, то транзакція T – це підмножина I , тобто $T \subseteq I$. Наприклад: $T1=\{\text{Чоловічі черевики Art. A-15, Засіб для догляду №3, Шнурки преміум-класу}\}$ $T2=\{\text{Жіночі босоніжки Art. B-22 (усі розміри), Логістична послуга «Експрес-доставка»}\}$

Асоціативне правило складається з двох наборів предметів, званих умовою (antecedent) і наслідком (consequent), що записуються у вигляді: $A \rightarrow B$ «якщо A тоді B ».

Таким чином, асоціативне правило сформульовано у вигляді: «Якщо умова, то наслідок». Умова і наслідок часто називаються відповідно лівостороннім (left – hand – side – LHS) та правостороннім (right – hand – side – RHS) компонентами асоціативного правила. Формально задачу асоціативних правил (Association Rules) представимо як: вихідні дані: множина товарів (або об'єктів) $I=\{i_1, i_2, \dots, i_n\}$ та база транзакцій D (наприклад, покупки в магазині) $D=\{T1, T2, \dots, Tm\}$, де кожна транзакція $Tj \subseteq I$, мета. Знайти асоціативні правила вигляду: $A \rightarrow B$ Формальна задача пошуку послідовних шаблонів (Sequential Pattern Mining) представимо як: вихідні дані: множина об'єктів $I=\{i_1, i_2, \dots, i_n\}$ та множина послідовностей $S=\{S1, S2, \dots, Sm\}$, де кожна послідовність кортежем $Sj=\langle X1, 2, \dots, Xk \rangle$, а $Xi \subseteq I$, мета: знайти усі послідовні шаблони

вигляду $\langle A \rangle \rightarrow \langle B \rangle$. При цьому *послідовний шаблон* – це впорядкована послідовність подій або товарів, які часто з'являються в такому ж порядку в різних послідовностях.

Приклади асоціативних правил для *Wholesale B2B Portal* [5].

Асоціативне правило (Товар): $\{\text{Жіночі кросівки Art. 403, Чоловічі кросівки Art.510}\} \rightarrow \{\text{Набір устілок «Комфорт»}\}$.

Асоціативне правило (Товар + Послуга): $\{\text{Замовлення більше 1000 пар}\} \rightarrow \{\text{Надання кредитної лінії 60 днів}\}$

Приклади послідовних шаблонів для *Wholesale B2B Portal* [5].

Послідовний шаблон (Сезонність): $\langle \text{Замовлення зимових чобіт} \rangle \rightarrow \langle \text{Замовлення засобів для захисту від солі (через 2 місяці)} \rangle$.

Послідовний шаблон (Прогнозування): $\langle \text{Запит на індивідуальну ціну для нової колекції, Отримання технічної документації} \rangle \rightarrow \langle \text{Оформлення основного оптового замовлення (протягом 14 днів)} \rangle$

Таким чином, за допомогою асоціативних правил шукають одночасну присутність елементів у транзакції, за допомогою послідовних шаблонів – послідовну появу елементів у часі або порядку.

Існують показники, за допомогою яких можна охарактеризувати зв'язок між наборами товарів, встановлений за допомогою АП. *Підтримка асоціативного правила (support)* – це кількість транзакцій, які містять як умову, так і наслідок. Наприклад, для асоціації $A \rightarrow B$ підтримка $support\ S(A \rightarrow B)$ виглядає як (1):

$$support\ S(A \rightarrow B) = P(A \cup B) = \frac{\text{кількість транзакцій, які містять A і B}}{\text{загальна кількість транзакцій}} \quad (1)$$

Достовірність асоціативного правила (confidence) $A \rightarrow B$ – це міра точності правила і визначається як відношення кількості транзакцій, що містять умову і наслідок, до кількості транзакцій, що містять лише умову. Достовірність АП розраховується за формулою 2:

$$confidence\ C(A \rightarrow B) = P(A|B) = \frac{P(A \cup B)}{P(A)} = \frac{\text{кількість транзакцій, які містять A і B}}{\text{кількість транзакцій які містять A}}, \quad (2)$$

$$confidence(A \rightarrow B) = \frac{support(A \cup B)}{support(A)}.$$

Значущість АП обчислюється як різниця між підтримкою правила загалом і добутком підтримки лише умови та лише наслідку.

Асоціативне правило для послідовних шаблонів має вигляд $\langle A \rangle \rightarrow \langle B \rangle$.

Підтримка асоціативного правила (support) – це кількість послідовностей, які містять шаблон, визначається за формулою (3)

$$support\ S(\langle A, B \rangle) = P(\langle A \rangle \cup \langle B \rangle) = \frac{\text{кількість послідовностей, які містять } \langle A, B \rangle}{\text{загальна кількість послідовностей}}. \quad (3)$$

Достовірність виникнення послідовності (confidence), вказує на ймовірність того, що після A буде B у послідовності (4)

$$confidence\ C(\langle A \rangle \rightarrow \langle B \rangle) = \frac{support(\langle A, B \rangle)}{support(\langle A \rangle)}. \quad (4)$$

У процесі пошуку АП може проводитися виявлення всіх асоціацій, підтримка та достовірність для яких перевищують заданий мінімум. Алгоритми для асоціативних правил (Association Rule Mining) шукають часто повторювані комбінації товарів або подій у транзакціях. Тобто в основі таких алгоритмів поняття популярних наборів (frequent itemset), які також можна назвати частими предметними наборами, які часто зустрічаються множинами (відповідно, вони пов'язані з поняттям частоти). Під частотою розуміємо проста кількість транзакцій, у яких міститься даний предметний набір. Тоді популярними наборами будуть ті, які зустрічаються частіше, ніж у заданому числі транзакцій. Методика пошуку асоціативних правил (АП) з використанням популярних наборів складається з двох кроків:

а) знайти найпопулярніші набори;

б) на їх основі згенерувати АП, які відповідають умовам мінімальної підтримки та достовірності.

Алгоритм Apriori (Apriori Property) – це класичний та один із найвпливовіших алгоритмів, який використовується в галузі Data Mining (добування даних) для пошуку частих наборів елементів (frequent itemsets) у базах даних транзакцій [8]. Основна ідея полягає в тому, що будь-яка підмножина частого набору елементів також повинна бути частою. Це дає змогу алгоритму використовувати вже знайдені часті набори меншого розміру для побудови (генерації) потенційних частих наборів більшого розміру (кандидатів) і ефективно відсіювати (прунінг) ті набори-кандидати, які не можуть бути частими.

FP-Growth (Frequent Pattern Growth) – алгоритм для ефективного пошуку частих наборів у транзакційних базах даних без генерації кандидатів (на відміну від Apriori). Основна ідея алгоритму полягає в тому, що замість повного перебору всіх можливих комбінацій товарів (Apriori), FP-Growth будує компакту структуру даних FP-дерево (Frequent Pattern Tree) та на його основі рекурсивно шукає часті шаблони [9].

Основні етапи роботи FP-Growth:

1. Обчислення підтримки кожного товару. Проходимо всі транзакції. Визначаємо підтримку для кожного товару. Відкидаємо товари, які не проходять $\min \text{support}$.

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 компактне 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 – 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 $\geq \min \text{support}$.

Eclat (Equivalence CLAss Transformation) це алгоритм для пошуку частих наборів товарів у транзакційних даних, який працює через перетини списків ідентифікаторів транзакцій (TID-списки), на відміну від перебору кандидатів у Apriori чи побудови дерев у FP-Growth. [10].

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:

1) *Побудова TID-списків*. Для кожного товару визначаємо всі транзакції, де він присутній.

2) *Рекурсивне злиття TID-списків*. Перетинаємо списки для комбінацій товарів. Підраховуємо підтримку для кожного перетину. Якщо підтримка $\geq \min \text{support}$ то перетин зберігаємо

3) *Формування частих наборів*. Рекурсивно об'єднуємо товари та їх TID-списки, поки є часті комбінації

Наведемо результати порівнянь розглянутих алгоритмів пошуку асоціативних правил в таблиці 2.

Таблиця 2. Порівняння алгоритмів пошуку асоціативних правил

Алгоритм	Принцип роботи	Переваги	Недоліки
Apriori	Генерація кандидатів	Простий, класичний	Повільний на великих наборах
FP-Growth	FP-дерево	Швидкий, без кандидатів	Складніший у реалізації
Eclat	Перетин TID-списків	Простий, швидкий на малих даних	Потребує багато пам'яті на великих даних

Висновки. Проведення системного аналізу існуючих моделей, методів та технологій, які використовуються для створення персоналізованого контенту в B2B електронній комерції дозволило ідентифікувати їх недоліки для подальшого дослідження. Визначено особливості B2B комерції та показано порівняння систем B2B та B2C за основними показниками. Проаналізовано основні алгоритми пошуку частих наборів товарів Apriori, FP-Growth, Eclat, що дозволяють виявляти комбінації товарів, які часто замовляються разом. Визначено метрики якості асоціативних правил такі як підтримка та достовірність для оцінювання значущості та корисності знайдених закономірностей в замовленнях товарів. Обґрунтовано необхідність розробки моделей афінівного аналізу транзакційних і поведінкових даних у

поєднанні з методами класифікації клієнтів для підвищення персоналізації товарного, інформаційного та рекомендаційного контенту системи B2B електронної комерції.

СПИСОК ЛІТЕРАТУРИ

1. Панасюк Т. С., Зубенко І. Р. «Перспективи розвитку b2c електронної комерції в світі». 2024. DOI: <https://doi.org/10.5281/zenodo.14603373>.
2. “Global Ecommerce Statistics: Trends to Guide Your Store in 2025”. – URL: <https://www.shopify.com/enterprise/blog/global-ecommerce-statistics>.
3. Шведа Н. М., Краузе О. І. «Електронна комерція: сучасний стан та стратегії розвитку». *Міжнародний науковий журнал «Інтернаука»*. Серія: Економічні науки. 2024. 2 (82). DOI: <https://doi.org/10.25313/2520-2294-2024-2-9639>.
4. Арсірій О. О., Любомська О. М., Руденко О. В., Іванов Д. В. «Гібридна рекомендаційна система для підтримки UI/UX дизайнерів». *Одеська політехніка*. 2024. р. 29–35. DOI: <https://doi.org/10.15276/ict.01.2024.03>.
5. «Система електронної комерції B2B База взуття». – URL: <https://bazaobuvi.com.ua/ua/>
6. Hu Y. H., Chen Y. L., Tang K. “Mining sequential patterns in the B2B environment”. *Journal of Information Science*. 2009; 35 (6): 677–694. DOI: <https://doi.org/10.1177/0165551509103600>
7. Cramer-Flood E. “Global Ecommerce Forecast & Growth Projections”. 2022. – URL: <https://www.insiderintelligence.com/content/globalecommerce-forecast-2022>.
8. Agrawal R., Srikant R. “Fast Algorithms for Mining Association Rules in Large Databases”. In *Proceedings of the 20th International Conference on Very Large Data Bases (VLDB '94)*. Morgan Kaufmann Publishers Inc. San Francisco, CA, USA. 1994. p. 487–499.
9. Han J., Pei J., Yin Y., et al. “Mining Frequent Patterns without Candidate Generation: A Frequent-Pattern Tree Approach”. *Data Mining and Knowledge Discovery*. 2004; 8: 53–87. DOI: <https://doi.org/10.1023/B:DAMI.0000005258.31418.83>.
10. Zaki M. J. “Scalable algorithms for association mining”. In *IEEE Transactions on Knowledge and Data Engineering*. 2000; 12 (3): 372–390. DOI: <https://doi.org/10.1109/69>.

DOI: <https://doi.org/10.15276/ict.02.2025.41>

UDC 004.4'24

Analysis of models, methods and technologies for creating personalized content in B2B e-commerce systems

Olena O. Arsirii¹⁾

Doctor of Engineering Sciences, Professor, Head of the Department of Information Systems
ORCID: <https://orcid.org/0000-0001-8130-9613>; e.arsirii@gmail.com. Scopus Author ID: 54419480900

Dmitriy V. Ivanov¹⁾

Postgraduate student, Department of Information Systems
ORCID: orcid.org/0009-0009-3958-5310; vanovdima9988@gmail.com

Oksana Yu. Babilunha¹⁾

PhD, Associate Professor of the Department of Information Systems
ORCID: <https://orcid.org/0000-0001-6431-3557>; babilunga.onpu@gmail.com. Scopus Author ID: 35316966900

¹⁾ Odesa Polytechnic National University, 1, Shevchenko Ave. Odesa, 65044, Ukraine

ABSTRACT

The increasing financial significance and structural complexity of the B2B e-commerce market segment, along with the imperative to enhance its efficiency, necessitate a systematic analysis of models, methods, and technologies used for personalized content creation within such systems. The study has identified the unique characteristics of B2B commerce and compared B2B and B2C systems based on key performance indicators. An analysis of existing models, methods, and technologies for personalized content creation in B2B e-commerce systems has been conducted, including: algorithms for finding frequent item sets (e.g., Apriority, FP-Growth, Eclat), which enable the detection of product combinations frequently ordered together. Metrics for evaluating the quality of association rules (Support, Confidence, Lift, Conviction), which are essential for assessing the significance and utility of the discovered patterns in product orders. The need to develop affinity analysis models of transactional and behavioral data, in combination with customer classification methods, is substantiated to enhance the personalization of product, informational, and recommendation content within B2B e-commerce systems.

Keywords: e-commerce; B2B systems; transactional and behavioral data; affinity analysis; association rules; Apriori; FP-Growth; Eclat; quality metrics