

ВИКОРИСТАННЯ МЕТОДІВ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ДАНИХ ДЛЯ МАРКЕТИНГОВИХ ДОСЛІДЖЕНЬ

Марія Репетило², Мирослав Дацко², Андрій Панчишин³

Львівський національний університет імені Івана Франка
79008 м. Львів, проспект Свободи, 18

¹e-mail: mariia.repetylo@lnu.edu.ua; ORCID: 0009-0008-3759-6094

²e-mail: myroslav.datsko@lnu.edu.ua; ORCID: 0000-0002-5181-4632

³e-mail: andriy.panchyshyn@lnu.edu.ua; ORCID: 0009-0004-1652-5810

Анотація. За останні десятиліття зросла здатність підприємств швидко генерувати та збирати дані. Тому аналіз, інтерпретація та максимальне використання даних є процесом, що потребує значної уваги через експоненційне зростання баз даних підприємств. Інтелектуальний аналіз даних передбачає пошук у базах даних потенційно корисної інформації, такої як правила, шаблони, закономірності та інші тенденції, приховані в даних. Стаття присвячена використанню у маркетингових дослідженнях сучасних інформаційних технологій, що дають змогу працювати підприємствам з великими базами даних для сегментування клієнтів, аналізу профілів клієнтів, визначення їх потреб, тощо. В результаті проведеного дослідження були побудовані асоціативні правила для покупців, які можна застосувати для кращого задоволення їх потреб та збільшення прибутку внаслідок перехресних продажів. Отже актуальність дослідження визначається необхідністю застосування методів інтелектуального аналізу даних для прийняття оптимальних рішень у маркетингу.

Ключові слова: маркетингові дослідження, методи інтелектуального аналізу даних, асоціативні правила, кластеризація.

Постановка проблеми. Розвиток інформаційних технологій а з ними і електронної комерції, онлайн-реклами, соціальних медіа змінив розвиток маркетингу. Це створило можливості до широкого застосування методів прямого маркетингу. Підприємства накопичують значні обсяги даних та на їх основі можуть використовувати інтелектуальний аналіз даних.

Аналіз ринкового кошика є одним із методів, який використовується великими роздрібними мережами для збільшення продажів шляхом розуміння купівельної поведінки покупців через отримання детальної інформації про їхню поведінку.

Інтелектуальний аналіз даних дозволяє визначити потенційно корисну інформацію з даних. Результати інтелектуального аналізу даних у вигляді прогнозів можна застосувати до таких заходів, як розробка індивідуальних маркетингових кампаній спрямованих на збільшення шансів утримати наявних клієнтів, а також залучити нових.

Аналіз останніх досліджень і публікацій. Окремі аспекти реалізації етапів маркетингових досліджень знайшли відображення у працях вітчизняних та зарубіжних науковців. Так І. Коростова[8], розглядає технології big data та їх функції, у застосуванні до завдань маркетингу, питання обґрунтування застосування математичного інструментарію методів багатокритеріального аналізу в маркетинговій діяльності набули розвитку у працях Білорус Т.В., Корнілова І.М., Фірсова С.Г. [3], теоретичні аспекти організації та проведення маркетингових досліджень розглянули Мейш А. В. та Матвійчук О. В.[9]. Аналіз сучасних методів проведення маркетингових досліджень на споживчому ринку провели Ільченко Т. В., та Помазан Л. М.[5] Дослідженню методів інтелектуального аналізу даних приділяли увагу A.Sahaya Arthy [1], D. M. Saputra, D. Saputra, L. D. Oswari [4], K. Satyam [6], N. Taher, D. Elzanfaly, S. Salama [10]. Проте, незважаючи на досягнення представників наукової спільноти, багато теоретичних та практичних питань даної проблематики ще недостатньо вирішені.

Постановка завдання. Реалії сьогодення такі, що значна частина покупців переходить на використання онлайн-магазинів. Тому підприємці активно шукають способи збільшити свій прибуток. З огляду на результати вивчення літературних джерел за даною проблематикою метою дослідження є системне застосування методів інтелектуального аналізу даних для оптимізації прийняття маркетингових рішень.

Методи дослідження. Для досягнення мети дослідження та розв'язання поставлених завдань у роботі використовувались різні методи наукового дослідження: системний та систематизування для виокремлення етапів маркетингового дослідження; метод структурно-логічного аналізу – під час проведення попередньої обробки транзакційних даних; методи інтелектуального аналізу даних– для реалізації етапів маркетингового дослідження.

Виклад основного матеріалу дослідження. Для ілюстрації проведення маркетингового дослідження скористаємось загальнодоступним набором даних.[7] Набір даних містить усі транзакції, здійснені за один рік для зареєстрованого у Великій Британії підприємства, що продає різноманітні подарунки.

Набір даних для аналізу містить 541909 записів та понад 25000 транзакцій. Назви полів даних представлені в табл. 1. Для проведення дослідження використовувалася мова програмування Python, зокрема такі бібліотеки pandas, datetime, math, numpy, matplotlib, seaborn, plotly, scipy, sklearn.cluster. Проведений попередній аналіз даних показав, що окремі поля набору даних від'ємні, що може означати наявність транзакції повернення або відсутність ідентифікації клієнтів для окремих записів тому такі записи можна не враховувати. Подальше дослідження вхідних даних показало, що існувала також проблема різних описів для одного товару. Така ситуація ілюструє помилку опису що призводить до низької якості даних і неправильного аналізу: такі вхідні дані було відкореговано.

Таблиця 1

Поля набору даних

№	Назва	Значення
1	InvoiceNo	Унікальний ідентифікатор для кожної окремої транзакції (рахунку)
2	StockCode	Ідентифікатор товарів
3	Description	Опис товарів
4	Quantity	Кількість придбаних товарів
5	InvoiceDate	Дата транзакцій
6	UnitPrice	Ціна за одиницю товару
7	CustomerID	Ідентифікатор клієнтів
8	Country	Країна покупця

Для сегментації клієнтів попередньо застосуємо RFM (Recency, Frequency, and Monetary Value) аналіз, заснований на моделі цінності клієнта, щоб знайти клієнтські сегменти, оскільки даний набір даних обмежується записами про продажі та не містить іншої інформації про клієнтів.

Цінність клієнта є важливим показником для вимірювання активів клієнта, ця концепція лежить в основі CRM (Customer Relationship Management) [12].

Цінність клієнта зустрічається в літературі під різними назвами, такими як пожиттєва цінність клієнта «CLV», «власний капітал клієнта» або «рентабельність клієнта».

Модель RFM є найбільш широко використовуваною моделлю для аналізу CLV (Customer Lifetime Value). Ця модель базується на принципі Парето, згідно з яким «80% вашого бізнесу припадає на 20% ваших клієнтів».

На базі моделі RFM було визначено три важливі інформаційні атрибути про клієнта, а саме: давність останньої покупки, частота покупок, та їх сума.

На наступному етапі застосуємо алгоритм кластеризації K-середніх. Кластеризація – це процес групування набору об'єктів даних у групи або кластери, де об'єкти всередині кластера схожі один на одного та відрізняються від об'єктів в інших кластерах [10].

Базуючись на метриці евклідової відстані, можна описати алгоритм k-середніх як задачу оптимізації яка мінімізує внутрішньокластерну суму квадратичних помилок (SSE), яку іноді також називають інерцією кластера. Отже, метою кластеризації є мінімізація загальної внутрішньокластерної дисперсії, або функції квадрата помилки:

$$F = \sum_{j=1}^k \sum_{i=1}^n \|x_i^{(j)} - c_j\|^2 \rightarrow \min \quad , \quad (1)$$

де k – кількість кластерів, $j=1, k$

n – кількість об'єктів, $i=1, n$

$x_i^{(j)}$ – об'єкт i в кластері j ,

c_j – центроїд кластера j .

Важливою проблемою при кластеризації є широкий діапазон значень, які може мати кожна змінна. Цю проблему можна вирішити, застосувавши логарифмування. Також важливим аспектом попередньої обробки даних є те, що перетворення мають бути зворотними, тобто має зберігатися можливість повернутися до первинних даних для надання практичних рекомендацій щодо проведеного дослідження.

Графічний аналіз атрибутів RFM аналізу показав висхідну тенденцію суми вартості покупок з відповідною тенденцією зростання частоти та зменшення давності покупок, очевидні закономірності, які можна побачити на рисунку 1, полягають у тому, що клієнти, які купують з вищою частотою, як правило, витрачають більше.

Проведемо кластеризацію покупців використавши дані RFM аналізу. Для визначення кількості кластерів застосуємо метод ліктя.

Метод ліктя передбачає, що кожне значення суми квадратів помилок у межах кластера WSS (Within-Cluster-Sum of Squared Errors) відображається на графіку, де вісь Y є значенням WSS, а вісь X є значенням кількості кластерів. Оптимальною є кількість кластерів, що відповідає абсцисі місця перегину графіку.[4]

Згідно методу, у нашому випадку, доцільним було розглянути 3-7 кластерів, тому проведемо кластеризацію з $k = 3, 5$ та 7 кластерів для порівняння результатів. Результати проведеної кластеризації зображено на рис. 2. Кластеризація при $k=3$ показує, що кластери мають значну різницю в грошовій цінності клієнта. Кластер 1

складається з цінних клієнтів, які часто роблять покупки, і це, безперечно, важливий сегмент для кожного бізнесу. Також в кластерах з мітками 0 і 2 отримуємо групи клієнтів з низькими та середніми витратами. Виходячи з тенденції, частота та давність ідеально співвідносяться з грошовою цінністю, що було згадано раніше, тобто висока грошова цінність відповідає низькій давності та високій частоті покупок.

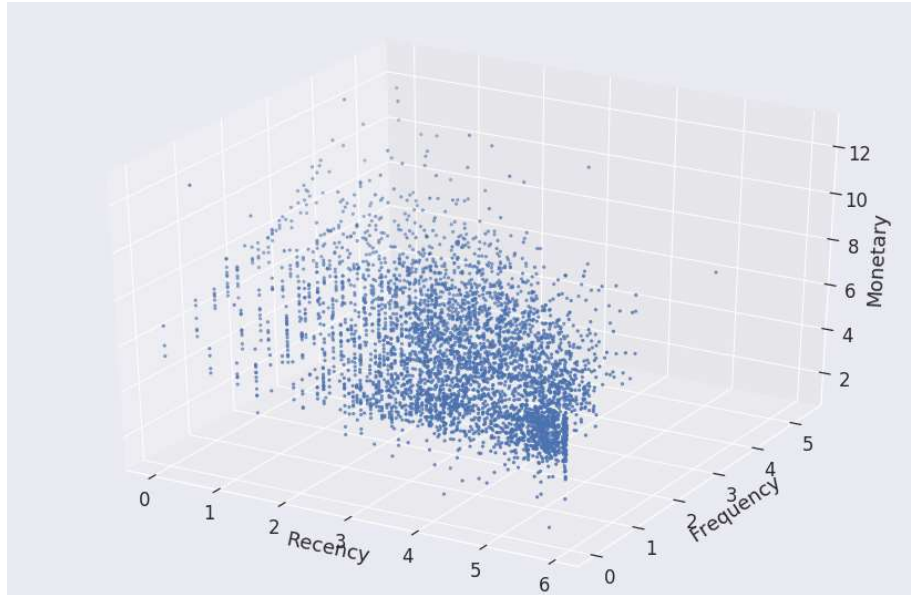


Рис. 1. Графік отриманих атрибутів RFM-аналізу

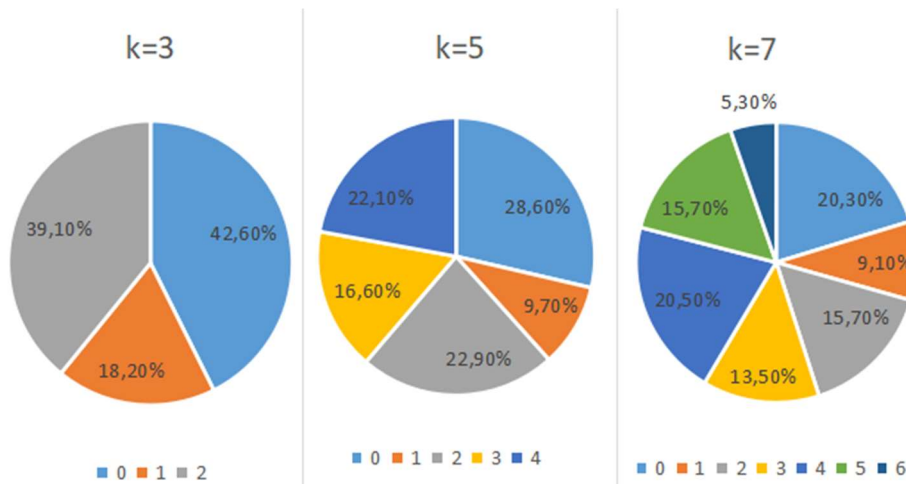


Рис. 2. Структура кластерів

За допомогою кластеризації з $k=5$ отримуємо кластер 0, що схожий на кластер 3, але відзначається більш недавніми покупками. Кластер 1 складається з покупців, що роблять покупки часто та на великі суми. Кластер 2 складається з тих, хто витрачає великі суми, але не так часто, як кластер 1. Кластер 4 робить покупки середнього розміру регулярно, але нечасто. Кластер 3 купує недорогі товари з відносно низькою частотою, але частіше, ніж кластер 1 і остання покупка зроблена нещодавно. Можна припустити, що ця група клієнтів реагує на знижки.

У кластеризації на 7 сегментів визначається кластер 6, що вказує на покупців, що роблять покупки часто і на великі суми. Кластери 1 і 5 показують хороші витрати та частоту, з тією лише різницею, наскільки нещодавні були їхні останні покупки, кластер 1 зробив покупки більш нещодавно. Кластер 0 займає четверте місце за витратами та показує достатню частоту, але з моменту останньої покупки пройшло багато часу. Кластер 5 схожий на кластер 0, покупки у ньому зроблено нещодавно але мають дещо кращу періодичність.

Зауважимо, що покращити якість кластеризації можна включивши, за наявності, іншу інформацію про клієнтів і деталі покупок.

Також для дослідження якості побудованих кластерів застосуємо силуетний аналіз. Силуетний графік може показувати невдалий вибір кількості кластерів для наведених даних через наявність кластерів із нижчими середніми показниками силуету, а також через великі коливання розміру силуетних графіків. На рисунках 3,4,5 наведено силуетний аналіз для кожної групи кластерів.

Проаналізуємо результати силуетного аналізу проведеної кластеризації. Оскільки немає кластерів із нижчими за середні оцінками силуету, усі варіанти кластерів допустимі, проте розміри силуетних ділянок сильно відрізняються в усіх варіантах.

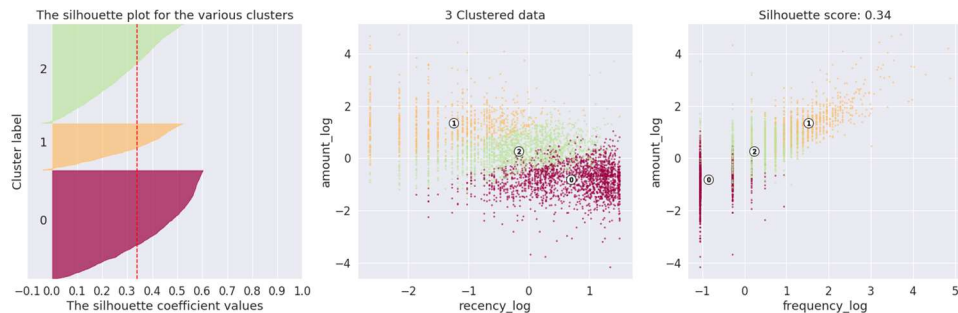


Рис. 3. Силуетний аналіз для 3 кластерів

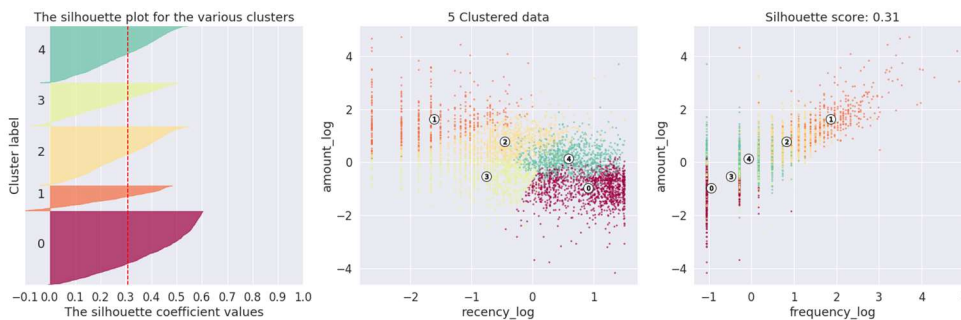


Рис. 4. Силуетний аналіз для 5 кластерів

У результаті найкращим варіантом може бути той, який надає простіше бізнес-пояснення, а також класифікує клієнтів у фокус-групи з розмірами, ближчими до бажаних.



Рис. 5. Силуетний аналіз для 7 кластерів

Центри кластерів після проведення зворотнього перетворення до вхідних даних наведено на рис. 6.

for 3 clusters the silhouette score is 0.34

Centers of each cluster:

	amount	recency	frequency
0	261.952265	116.604917	1.190876
1	3967.994380	7.236580	10.044493
2	1006.914317	33.819966	3.152227

for 5 clusters the silhouette score is 0.31

Centers of each cluster:

	amount	recency	frequency
0	213.876290	159.060239	1.088129
1	5708.668108	4.285608	13.677542
2	1929.872406	22.442129	5.413014
3	372.314665	14.590855	1.665686
4	863.093356	100.092666	2.395562

for 7 clusters the silhouette score is 0.31

Centers of each cluster:

	amount	recency	frequency
0	809.713152	107.590047	2.277095
1	2115.751105	4.436558	6.395614
2	239.805507	36.372861	1.132543
3	667.345658	13.698858	2.663541
4	205.016462	225.462781	1.082459
5	2414.804796	38.026754	6.003854
6	10182.351681	4.961015	20.687947

Рис. 6. Центри кластерів

Продати більше товарів клієнту можливо за допомогою перехресного продажу шляхом аналізу тенденцій покупок клієнтів. Тому потрібно дослідити клієнтські транзакції та знайти потенційні доповнення до початкових потреб клієнта скориставшись правилами асоціації.

Оскільки проведений попередній аналіз даних показав, що лише кілька товарів відповідають за основну частину продажів, то з огляду на це можна скоротити набір даних. Тому було вилучено дані які не містять товари, що входять в список топ-15 товарів в продажах, з якими можна ознайомитись у табл. 2.

Провівши пошук транзакцій, які містять принаймні два елементи отримали 4664 продажі, що містять в собі популярні товари, які відповідальні за 8,73% обсягу продажу і складають близько 5% транзакцій.

Таблиця 2

Топ 15 популярних товарів

№	Назва	Кількість
1	White hanging heart t-light holder	1978
2	Regency cakestand 3 tier	1703
3	Jumbo bag red retrospot	1600
4	Party bunting	1379
5	Assorted colour bird ornament	1375
6	Lunch bag red retrospot	1289
7	Set of 3 cake tins pantry design	1146
8	Postage	1099
9	Jumbo bag vintage doily	1080
10	Lunch bag black skull	1052
11	Lunch bag suki design	1043
12	Popcorn holder	1035
13	Pack of 72 retrospot cake cases	1029
14	Spotty bunting	1009
15	Lunch bag vintage doily	1006

Видобуток асоціативних правил – це метод виявлення частих шаблонів, кореляцій, асоціацій або причинно-наслідкових структур у наборах даних, які містяться в численних базах даних, таких як реляційні бази даних, транзакційні бази даних та інші типи сховищ даних.

Підтримка (Support) правила асоціації – це відсоток груп, які містять усі елементи, перелічені в цьому правилі асоціації. Відсоткове значення обчислюється серед усіх груп, які розглядалися. Це відсоткове значення розраховується за формулою 2 та показує, як часто антецедент та наслідок правила зустрічаються серед усіх груп, які розглядалися [11].

$$Support = \frac{number(x)}{N} \quad (2)$$

де N – кількість всіх транзакцій.

Достовірність, чи Впевненість (Confidence) визначається за формулою 3 та показує як часто правило виявляється істинним або як часто елементи X і Y зустрічаються разом у наборі даних, коли X уже вказано [2]. Це відношення транзакцій, які містять X і Y, до кількості записів, які містять X.

$$Confidence = \frac{number(X, Y)}{number(X)} \quad (3)$$

Для створення асоціативних правил потрібно вказати рівні підтримки та довіри. Варто почати з вищої підтримки, оскільки нижча призведе до більшої кількості частих наборів елементів і, отже, довшого часу обчислень. Якщо вказати підтримку 0,01, то отримуємо 663 273 набори елементів лише для 15 товарів. Це значення зросло б у рази, якщо зменшити підтримку або збільшити кількість елементів у наборі даних.

На наступному кроці встановимо значення довіри 0,6 та згенеруємо правила. Результатом є 25247 правил. Відомо, що значення підйому рівне 1 вказує на те, що ймовірність появи антецеденту та наслідку разом не залежать один від одного. Отже, варто розглядати правила, які мають підйом більше 1. Підрахунок показує, що всі згенеровані правила відповідають цьому критерію, отже всі правила можна використовувати. У таблиці 3 відображено перші 5 правил з найвищою впевненістю.

Побудувавши асоціативні правила підприємство може, наприклад, з використанням цільової реклами пропонувати покупцям інші асоційовані відповідно до отриманих правил товари.

Таблиця 3

Асоціативні правила з найвищою впевненістю

№	№ rule	consequent	antecedent	support	confidence	lift
1	20	jumbo bag vintage doily	jumbo bag red retrospot, lunch bag red retrospot, lunch bag vintage doily	60	0.90	4.52
2	133	lunch bag vintage doily	lunch bag red retrospot, jumbo bag vintage doily, lunch bag black skull., lunch bag suki design	49	0.89	4.59
3	134	lunch bag vintage doily	lunch bag red retrospot, jumbo bag vintage doily, lunch bag suki design	48	0.88	4.58
4	131	lunch bag vintage doily	lunch bag red retrospot, jumbo bag vintage doily, lunch bag black skull.	48	0.87	4.50
5	21	jumbo bag vintage doily	jumbo bag red retrospot, lunch bag suki design, lunch bag vintage doily	48	0.87	4.33

У табл. 4 відображено перші 5 правил з найвищою підтримкою.

Таблиця 4

Асоціативні правила з найвищими показниками підтримки

№	№ rule	Consequent	Antecedent	Support	Confidence	Lift
1	123	lunch bag vintage doily	jumbo bag vintage doily, lunch bag red retrospot	154	0.77	4.01
2	28	jumbo bag vintage doily	lunch bag vintage doily, jumbo bag red retrospot	153	0.80	4.00
3	92	lunch bag red retrospot	lunch bag vintage doily, lunch bag suki design	149	0.66	2.67
4	77	lunch bag red retrospot	lunch bag black skull, lunch bag suki design	142	0.60	2.40
5	4	jumbo bag red retrospot	jumbo bag vintage doily, lunch bag red retrospot	138	0.69	2.68

Побудова асоціативних правил є простим процесом, але дорогим з точки зору обчислень, тому що чим більше в наявних даних є наборів елементів, тим більше можна знайти правил, що збільшує час обчислень. Загалом потрібно прагнути до балансу підтримки та впевненості, що призводить до раціональної кількості правил.

Проте, слід зауважити, що є ситуації, коли потрібно видобути всі можливі правила, щоб знайти серед них рідкісні та, можливо, перспективні.

Висновки та перспективи подальших досліджень. У статті розглянуто етапи маркетингового дослідження покупців із використанням транзакційних даних. Так у результаті сегментації клієнтів, що використовує дані про частоту, давність, та суму покупок споживачів, та кластеризації К-середніх отримано три можливі кількості сегментів, з яких можна вибрати потрібний варіант відповідно до потреб та специфіки маркетингових кампаній підприємства.

Проаналізовано відмінності між отриманими сегментами та проведено силуетний аналіз для оцінки якості побудованих кластерів.

У результаті аналізу ринкового кошика отримано понад 25 тисяч правил для найпопулярніших товарів. Показано, що ряд з них можна використовувати, оскільки значення підйому для кожного правила становить більше одиниці, що вказує на те, що наявність антецедента позитивно впливає на наслідок правила, тобто поява другого товару в чеку є наслідком появи першого. У подальшому отримані результати можуть бути основою для реалізації маркетингових заходів, зважаючи на особливості клієнтів.

1. A. Sahaya Arthy. Prediction Using Data Mining Techniques and Tools. International Journal of Engineering Research in Computer Science and Engineering (IJERCSE) Vol. 6 Issue 1, January 2019. URL: https://www.technoarete.org/common_abstract/pdf/IJERCSE/v6/i1/Ext_20867.pdf (дата звернення: 08.11.2023).
2. Association Rule Learning. Java Tpoint. URL: <https://www.javatpoint.com/association-rule-learning> (дата звернення: 12.10.2023).
3. Білорус Т.В. Організація маркетингових досліджень із використанням методів багатокритеріального аналізу. *Економіка та суспільство*. № 7 (2016). С. 206-215. URL: https://economyandsociety.in.ua/journals/7_ukr/35.pdf (дата звернення: 18.10.2023).
4. D. M. Saputra, D. Saputra, L. D. Oswari. Effect of Distance Metrics in Determining K-Value in KMeans Clustering Using Elbow and Silhouette Method. *Advances in Intelligent Systems Research*, Vol. 172. Sriwijaya International Conference on Information Technology and Its Applications. 2019, p. 341-346.
5. Льченко Т.В., Помазан Л.М. Сучасні методи проведення маркетингових досліджень на споживчому ринку. *Інвестиції: практика та досвід*. 2023. № 8. С. 131-135. URL: <https://dspace.dsau.dp.ua/handle/123456789/8020> (дата звернення: 19.10.2023).
6. K. Satyam, Silhouette Method. Better than Elbow Method to find Optimal Clusters, 2020. *Towards Data Science*. URL: <https://towardsdatascience.com/silhouette-method-better-than-elbow-method-to-find-optimal-clusters-378d62ff6891> (дата звернення: 20.10.2023).
7. Kaggle URL: <https://www.kaggle.com/> URL: (дата звернення: 21.10.2023).
8. Коростова І. О. Big data в маркетингу. *Ефективна економіка*. 2021. № 11. URL: <http://www.economy.nauka.com.ua/?op=1&z=9530> (дата звернення: 19.10.2023).
9. Мейш А. В. Маркетингові дослідження: їх роль та потреба у ринковій економіці. *Вісник Хмельницького національного університету. Серія: «Економічні науки»*. 2021. №1. С.80-84. URL: <http://journals.khnu.km.ua/vestnik/wp-content/uploads/2021/01/17-2.pdf> (дата звернення: 23.10.2023).
10. N.M. Taher, D. Elzanfaly, S. Salama. Investigation in Customer Value Segmentation Quality under Different Preprocessing Types of RFM Attributes. Vol. 4, Issue 4, 2016, pp. 5-10. URL: <https://doi.org/10.3991/ijes.v4i4.6532> (дата звернення: 24.10.2023).

11. Support in an association rule, 2021. IBM Documentation. URL: <https://www.ibm.com/docs/en/db2/9.7?topic=associations-support-in-association-rule> (дата звернення: 26.10.2023).
12. V. Kumar, W. Reinartz. Strategic Customer Relationship Management Today, *Springer*, 2012, pp. 5-6. URL: <https://doi.org/10.1007/978-3-642-20110-3>. 22 (дата звернення: 30.10.2023).

References

1. A. Sahaya Arthy (2019). Prediction Using Data Mining Techniques and Tools. *International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)* Vol. 6 Issue 1. Retrieved from: https://www.technoarete.org/common_abstract/pdf/IJERCSE/v6/i1/Ext_20867.pdf (accessed 8 November 2023).
2. Association Rule Learning. Java Tpoint. Retrieved from: <https://www.javatpoint.com/association-rule-learning> (accessed 12 October 2023).
3. Bilorus T.V. (2016). Organizaciya marketingovih doslidzhen iz vikoristanniam metodiv bagatokriterialnogo analizu. *Ekonomika ta suspilstvo*. No. 7. pp. 206-215. Retrieved from: https://economyandsociety.in.ua/journals/7_ukr/35.pdf (accessed 18 October 2023).
4. D. M. Saputra, D. Saputra, L. D. Oswari (2019). Effect of Distance Metrics in Determining K-Value in KMeans Clustering Using Elbow and Silhouette Method. *Advances in Intelligent Systems Research*, Vol. 172. *Sriwijaya International Conference on Information Technology and Its Applications*, pp. 341-346.
5. Ilchenko T.V. (2023). Suchasni metodi provedennya marketingovih doslidzhen na spozhivchomu rinku. *Investiciyi: praktika ta dosvid*. No. 8. pp. 131-135. Retrieved from: <https://dspace.dsau.dp.ua/handle/123456789/8020> (accessed 12 October 2023).
6. K. Satyam, Silhouette (2020). Method. Better than Elbow Method to find Optimal Clusters, *Towards Data Science*. Retrieved from: <https://towardsdatascience.com/silhouette-method-better-than-elbow-method-to-find-optimal-clusters-378d62ff6891> (accessed 20 October 2023).
7. Kaggle. Retrieved from: <https://www.kaggle.com> (accessed 21 October 2023).
8. Korostova, I. (2021). Big data in marketing. *Efektivna ekonomika*. Vol. 11, Retrieved from: <http://www.economy.nayka.com.ua/?op=1&z=9530> (accessed 19 October 2023).
9. Mejsh A. V. (2021). Marketingovi doslidzhennya: yih rol ta potreba u rinkovij ekonomici. *Visnik Hmelnickogo nacionalnogo universitetu. Seriya: Ekonomichni nauki*. No. 1. pp. 80-84. Retrieved from: <http://journals.khnu.km.ua/vestnik/wp-content/uploads/2021/01/17-2.pdf> (accessed 23 October 2023).
10. N.M. Taher, D. Elzanfaly, S. Salama (2016). Investigation in Customer Value Segmentation Quality under Different Preprocessing Types of RFM Attributes. Vol. 4, Issue 4, pp. 5-10. Retrieved from: <https://doi.org/10.3991/ijes.v4i4.6532> (accessed 24 October 2023).
11. Support in an association rule (2021). IBM Documentation. Retrieved from: <https://www.ibm.com/docs/en/db2/9.7?topic=associations-support-in-association-rule> (accessed 26 October 2023).
12. V. Kumar, W. Reinartz (2012). Strategic Customer Relationship Management Today. *Springer*, pp. 5-6. Retrieved from: <https://doi.org/10.1007/978-3-642-20110-3> (accessed 30 October 2023).

INTELLIGENCE DATA ANALYSIS METHODS FOR MARKETING RESEARCH**Mariia Repetylo¹, Myroslav Datsko², Andriy Panchyshyn³***Ivan Franko National University of Lviv,
79008 Lviv, Prospekt Svobody, 18*¹*e-mail: mariia.repetylo@lnu.edu.ua; ORCID: 0009-0008-3759-6094*²*e-mail: myroslav.datsko@lnu.edu.ua; ORCID: 0000-0002-5181-4632*³*e-mail: andriy.panchyshyn@lnu.edu.ua; ORCID: 0009-0004-1652-5810*

Abstract. Competition in the world of trade business and the progress of information technology are interrelated issues in the market competition to meet the increasingly high needs of customers. Companies need to develop a business strategy to continue to meet customer desires and market demands. Regardless of the type of market, participants compete to provide the best services, considering it from various aspects, which become material for the consumer's evaluation.

The development of information technologies including e-commerce, online advertising, and social media, has changed the marketing domain. It created opportunities for the widespread use of direct marketing methods. Enterprises accumulate significant amounts of data and can use intelligent data analysis on their basis. Market basket analysis is one method that large retail chains use to increase sales by understanding the buying behavior of customers through detailed information about their behavior.

The results of intelligent data analysis in the form of predictions can be applied to activities such as the development of individual marketing campaigns aimed at increasing the chances of retaining existing customers, as well as attracting new ones.

The article discusses the stages of marketing research of buyers using transactional data. The clustering was made with a K-mean method based on the frequency, age, and amount of consumer purchases. The results of the clustering obtained three possible segments and each of them can be chosen according to the needs and specifics of the company's marketing campaigns. The differences between the obtained segments were analyzed and a silhouette analysis was performed to assess the quality of the constructed clusters.

Finally, associative rules were obtained as a result of the analysis of the market basket. It is shown that a number of them can be used since the lift value for each rule is greater than the threshold. It indicates that the presence of an antecedent positively affects the consequence of the rule, that is, the appearance of the second product in the check is a consequence of the appearance of the first. In the future, the obtained results can be the basis for the implementation of marketing measures, taking into account the features of customers.

Keywords: marketing research, data analysis methods, associative rules, clustering.

Стаття надійшла до редколегії 20.10.2023

Прийнята до друку 29.11.2023