УДК 338.242.2:631
DOI https://doi.org/10.32782/pdau.eco.2025.3.8

МАРКЕТИНГОВА КОНЦЕПЦІЯ КЛАСИФІКАЦІЇ КЛІЄНТІВ

Пархоменко Дмитро Володимирович
здобувач вищої освіти,
Сумський державний університет
ORCID ID: 0009-0003-5279-1879

Бондаренко Алла Федорівна
кандидат технічних наук, доцент,
доцент кафедри маркетингу,
Сумський державний університет
ORCID ID: 0000-0002-8439-1787

Зось-Кіор Микола Валерійович
доктор економічних наук, професор,
професор кафедри менеджменту ім. І.А. Маркіної,
Полтавський державний аграрний університет
ORCID ID: 0000-0001-8330-2909

Класифікація клієнтів є ключовим інструментом сучасного маркетингу, спрямованим на оптимізацію бізнес-стратегій через сегментацію та прогнозування поведінки споживачів. Мета дослідження – проаналізувати теоретичні засади класифікації клієнтів, практичні алгоритми машинного навчання та їхнє застосування для побудови прогнозних моделей. У роботі розглянуто такі методи, як дерева рішень (C4.5, ID3), алгоритми 1R, k-найближчих сусідів (KNN), Random Forest, а також кластеризацію для виявлення груп клієнтів із схожими характеристиками. На прикладі туристичного агентства продемонстровано, як аналіз віку та доходу дозволяє визначити оптимальний тип рекламних матеріалів для різних сегментів. Дослідження акцентує увагу на оцінці якості моделей за допомогою метрик (Accuracy, Precision, Recall, F1, AUC-ROC) та аналізі матриці помилок, що критично важливо для мінімізації фінансових втрат від хибних прогнозів. Окремо розглянуто маркетингові стратегії (масовий, концентрований, диференційований маркетинг) і їхню взаємодію з результатами класифікації. Висновки підкреслюють необхідність інтеграції класифікаційних моделей у системи прийняття рішень для підвищення лояльності клієнтів, зменшення відтоку та максимізації доходів. Робота має практичну цінність для маркетологів, аналітиків даних та менеджерів, які прагнуть трансформувати дані в конкурентні переваги.

Ключові слова: *маркетинг, класифікація клієнтів, машинне навчання, кластеризація, дерево рішень, матриця помилок, маркетингові стратегії, менеджмент, адаптація, ефективність, конкурентоспроможність.*

Dmytro Parkhomenko, Alla Bondarenko
Sumy State University

Mykola Zos-Kior
Poltava State Agrarian University

MARKETING FRAMEWORK FOR CUSTOMERS CLASSIFICATION

Customer classification is a pivotal tool in modern marketing, aimed at optimizing business strategies through consumer segmentation and behavior prediction. The purpose of this study is to analyze the theoretical foundations of customer classification, practical machine learning algorithms, and their application in building predictive models. The paper explores methods such as decision trees (C4.5, ID3), the 1R algorithm, k-nearest neighbors (KNN), Random Forest, and clustering techniques for identifying customer groups with similar characteristics. Using a case study of a travel agency, the research demonstrates how analyzing age and income data enables the identification of optimal advertising materials for different customer segments. The study emphasizes evaluating

model quality through metrics (Accuracy, Precision, Recall, F1-score, AUC-ROC) and analyzing confusion matrices, which are critical for minimizing financial losses caused by prediction errors. Marketing strategies (mass, concentrated, differentiated marketing) and their alignment with classification results are discussed in detail. The conclusions highlight the necessity of integrating classification models into decision-making systems to enhance customer loyalty, reduce churn, and maximize revenue. This work holds practical value for marketers, data analysts, and managers seeking to transform data into competitive advantages.

Keywords: *marketing, customer classification, machine learning, clustering, decision tree, error matrix, marketing strategies, management, adaptation, efficiency, competitiveness.*

Вступ. У сучасних умовах жорсткої конкуренції та стрімкого зростання обсягів даних класифікація клієнтів набуває ключового значення для ефективного управління маркетинговими стратегіями. Традиційні методи сегментації, засновані на ручній обробці інформації, вже не відповідають вимогам ери цифрових технологій, де дані про поведінку, демографію та потреби клієнтів формуються в реальному часі. Зростання очікувань споживачів щодо персоналізації пропозицій, необхідність боротьби з відтоком клієнтів та оптимізації маркетингових бюджетів вимагають використання точних і масштабованих рішень на основі машинного навчання. Алгоритми, такі як дерева рішень (C4.5, ID3), кластеризація або Random Forest, дозволяють не лише виявляти складні залежності у великих масивах даних, але й прогнозувати ризики відтоку, ідентифікувати перспективні сегменти та адаптувати рекламні кампанії під конкретні групи клієнтів. Технологічний прогрес, зокрема доступність інструментів аналітики, робить ці методи доступними навіть для малого бізнесу, що раніше не міг собі дозволити складні IT-рішення. Наукова новизна дослідження полягає в інтеграції теоретичних підходів до класифікації з практичними маркетинговими кейсами, що заповнює прогалину між академічними розробками та реальними потребами бізнесу. Практична значимість підкреслюється можливістю зменшити витрати на маркетинг за рахунок таргетованих рішень, підвищити лояльність клієнтів через індивідуалізовані пропозиції та уникнути фінансових втрат від помилкових прогнозів завдяки використанню метрик якості (F1-міра, AUC-ROC). У контексті цифрової трансформації, де дані стають основним конкурентним активом, це дослідження пропонує структурований підхід до перетворення інформації в стратегічні рішення, що робить його особливо актуальним для компаній, які прагнуть закріпитися на динамічному ринку.

Матеріали та методи. Класифікація – систематичний процес розподілу об'єктів, явищ або понять на класи чи групи на основі спільних ознак або властивостей. Це дозволяє впорядкувати інформацію, полегшуючи її аналіз та розуміння [1]. У межах клієнтської аналітики класифікація відноситься до одного із заздалегідь визначених класів. Наприклад, до класу «перейде за посиланням» чи «не перейде за посиланням». Для проведення класифікації повинні бути наявні ознаки, що характеризують групу до якої належить той чи інший клієнт. Серед таких ознак можуть бути різні характеристики клієнтів, наприклад: стать, вік, освіта, дохід, частота покупок, давність останньої покупки. Класифікація може бути одновимірною (за однією ознакою) та багатовимірною (за двома й більше ознаками) [2]. Задача класифікації полягає в побудові алгоритму, який на основі наявної вибірки об'єктів, для яких відомо, до яких класів вони належать, зможе визначити клас нових, невідомих об'єктів. Тобто, необхідно класифікувати об'єкти, поділені на класи, на основі їх характеристик [3].

У машинному навчанні та статистиці задачі класифікації поділяються на кілька основних типів залежно від кількості класів, до яких може належати об'єкт:

1. Бінарна класифікація: Об'єкти поділяються на два класи. Наприклад, визначення, чи є електронний лист спамом (так/ні) [4].

2. Многокласова (багатовимірна) класифікація: Об'єкти поділяються на більше ніж два класи. Наприклад, розпізнавання рукописних цифр, де кожна цифра від 0 до 9 є окремим класом [5].

Мета статті – дослідити маркетингові концепції класифікації клієнтів.

У цій статті для класифікації клієнтів і аналізу даних використано такі методи: дерева рішень (Decision Trees); алгоритм 1R; метод k-найближчих сусідів (KNN); Random Forest; кластеризація (Clustering); матриця помилок класифікації (Confusion Matrix).

Результати. Розглянемо задачу класифікації на простому прикладі. Припустимо, є база даних про клієнтів туристичного агентства з інформацією про вік і доходи за місяць. Є рекламний матеріал двох видів: більш дорогий і комфортний відпочинок та дешевший, молодіжний відпочинок. Відповідно, визначені два класи клієнтів: клас 1 і клас 2. База даних наведена в табл. 1.

Для наочності представимо нашу базу даних у двомірному просторі (вік і дохід), у вигляді множини об'єктів, що належать класам 1 (помаранчева мітка) і 2 (сіра мітка) [6] – рис. 1.

Таблиця 1

Завдання. **Визначити, до якого класу належить новий клієнт і який здвох видів рекламних матеріалів йому варто відсилати**

Код клієнта	Вік	Дохід	Клас
1	18	25	1
2	22	100	1
3	30	70	1
4	32	120	1
5	24	15	2
6	25	22	1
7	32	50	2
8	19	45	2
9	22	75	1
10	40	90	2

Джерело: складено авторами

Загальне формулювання задачі класифікації в цьому разі може звучати так: «Прогнозування схильності клієнтів до відтоку» або «Розробка моделі схильності відтоку». Конкретне формулювання задачі класифікації: «Визначити, до якого класу належить клієнт, що позначений на графіку синім хрестиком» [2]. Завдяки цьому можна прогнозувати чи стане клієнт відтоком чи ні. Розв'язати задачі дозволять деякі алгоритми класифікації:

1R – це простий алгоритм класифікації, який розробляє набір правил на основі одного вхідного атрибута (відповідно "1" для кількості вхідних атрибутів і "R" для "правил"). Алгоритм працює шляхом аналізу атрибута даних і створення правила на основі цього атрибута для прогнозування результату. Цей процес повторюється для кожного атрибута, щоб створити повний набір правил [7]. Алгоритм може бути використаний для простих моделей, де один фактор (наприклад вік та дохід) є основним для прогнозування покупки.

C4.5 – це алгоритм, розроблений Джоном Россом Квінланом, який створює дерева рішень. Дерево рішень – це інструмент, що використовується для класифікації в машинному навчанні, який використовує деревоподібну структуру, де внутрішні вузли представляють тести, а листя – рішення. C4.5 використовує інформаційно-теоретичні концепції, такі як ентропія, для класифікації даних [8]. Алгоритм може бути використаний для створення складних моделей, які дозволяють класифікувати клієнтів за кількома ознаками (наприклад, поєднання віку та доходу)

Алгоритм k найближчих сусідів (KNN) – це непараметричний, керований алгоритм класифікації, який використовує близькість для прийняття рішень про належність окремої точки даних до пев-

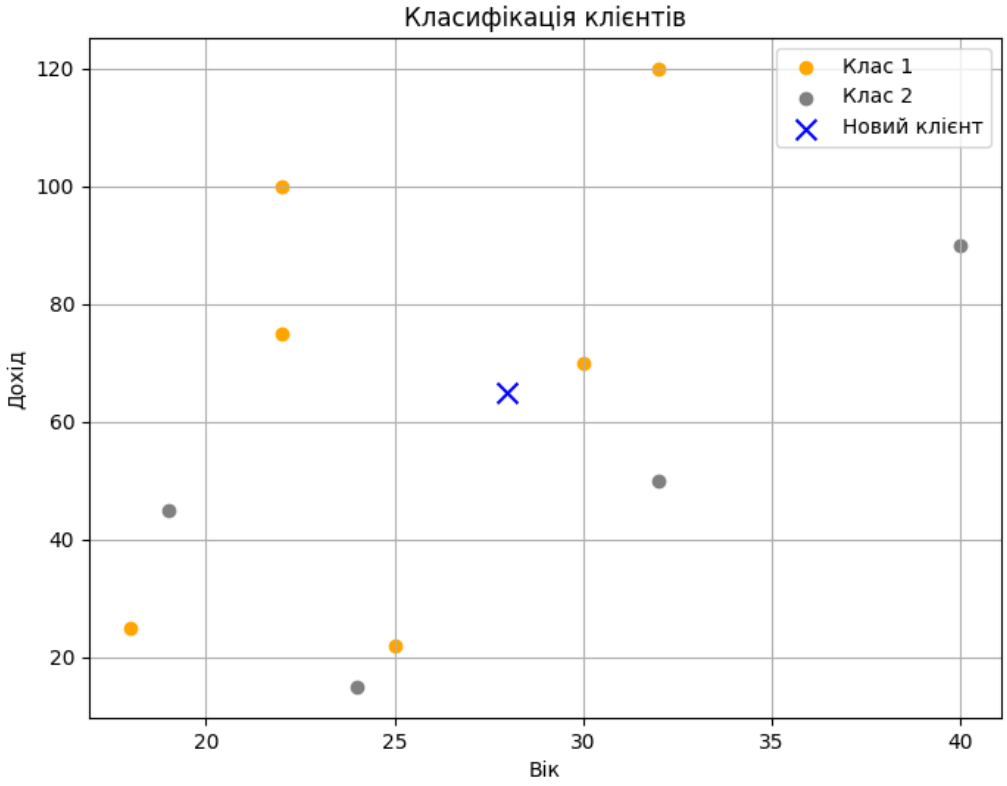

Рис. 1. Двовимірний простір класифікації клієнтів за віком та доходом

Джерело: складено авторами за допомогою мови програмування Python

ної групи. Це один з найпопулярніших і простих алгоритмів класифікації та регресії [9]. Цей алгоритм може допомогти у визначенні групи клієнтів, які найближче схожі на нових користувачів (наприклад, якщо новий клієнт має схожі характеристики з іншими клієнтами)

Random Forest – це популярний алгоритм машинного навчання, розроблений Лео Брейманом та Адель Катлер, який поєднує результати кількох дерев рішень для отримання одного результату. Його простота у використанні та гнучкість сприяли широкому впровадженню, оскільки він вирішує як задачі класифікації, так і регресії [10].

ID3 – це популярний алгоритм побудови дерев рішень, який використовується в машинному навчанні. Його мета – поступово створювати дерево рішень, вибираючи на кожному етапі найкращий атрибут для поділу даних на основі інформаційного виграшу [11].

На рисунку 2 зображено дерево рішень, створене за допомогою алгоритму C4.5. Воно використовується для класифікації на основі трьох атрибутів: Outlook (Прогноз погоди), Humidity (Вологість), і Windy (Вітряність). Внутрішні вузли дерева відповідають тестам на значення певних атрибутів, а листові вузли – кінцевим рішенням (грати чи не грати). Наприклад, якщо прогноз Overcast, то рішення – Play, якщо Sunny, то вибір залежить від вологості, а якщо Rainy – від вітряності. Це демонструє, як алгоритм C4.5 використовує інформаційну ентропію для вибору найбільш значущих атрибутів і побудови дерева, яке може автоматично класифікувати нові випадки.

Цей підхід застосовується в алгоритмах типу дерев рішень (C4.5, ID3, CART), оскільки вони використовують деревоподібну структуру для прийняття рішень. Однак він не підходить для 1R, KNN або SVM. 1R використовує лише один атрибут для класифікації, тоді як KNN базується на близькості до інших точок у просторі ознак, а SVM створює гіперплощини для поділу даних. У цих методах немає ієрархічної структури прийняття рішень, характерної для дерев рішень.

Зважаючи на важливість класифікації клієнтів для маркетингових стратегій, можна зробити наступний крок і перейти до більш складних моделей, які дозволяють передбачити поведінку клієнтів у майбутньому. Одним із таких підходів є моделі схильності клієнтів. Ці моделі намагаються оцінити ймовірність того, що клієнт проявить певну поведінку, наприклад, здійснить покупку, залишить компанію чи скористається певною послугою. Використовуючи техніки класифікації, такі як дерева рішень або алгоритм KNN, можна побудувати прогнози, що допоможуть компаніям приймати стратегічні рішення, оптимізуючи маркетингові кампанії та підвищуючи лояльність клієнтів. Тепер можна більш детально розглянути, як саме алгоритми машинного

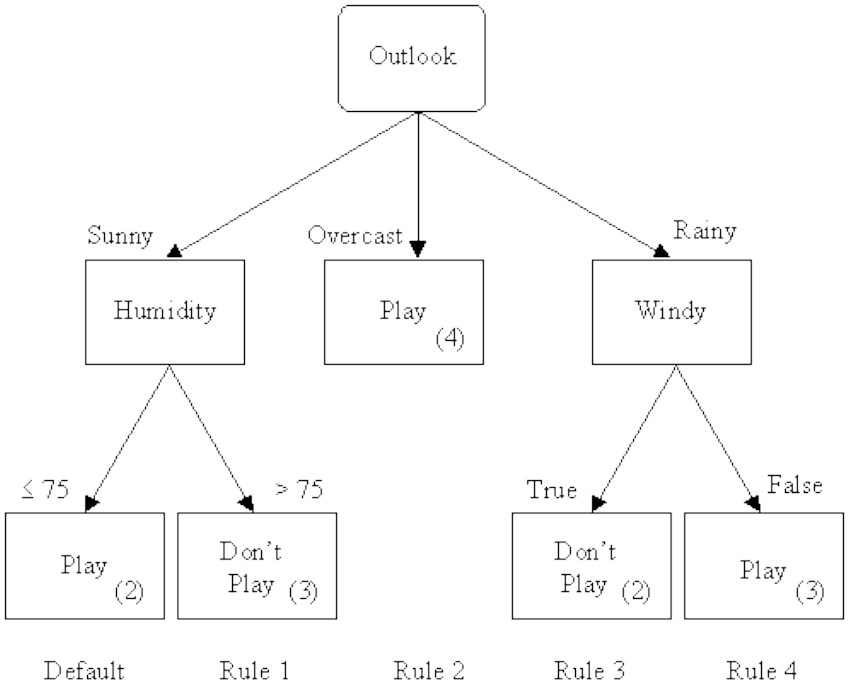

Рис. 2. Дерево рішень, створене за алгоритмом C4.5 для прийняття рішень щодо гри в залежності від погодних умов

Джерело: [10]

навчання допомагають у створенні моделей схильності та прогнозуванні поведінки клієнтів [12]. Виходячи з цього, вхідними даними для таких моделей є дані про клієнта, наприклад, місце проживання, відгуки на сайтах, відгуки на пропозиції за різними каналами комунікації. Є декілька популярних моделей:

Коефіцієнт відтоку клієнтів – коефіцієнт відтоку клієнтів – це відсоток клієнтів, які припиняють вести бізнес з організацією протягом певного періоду часу. Підприємства можуть оцінювати цей показник щорічно, щомісяця, щотижня або щодня [13].

Модель SABONE – ця модель класифікує клієнтів за п'ятьма основними категоріями: безпека, прив'язаність, комфорт, гордість та новизна. Розуміння цих категорій дозволяє компаніям налаштувати свої пропозиції та комунікацію відповідно до потреб кожного сегмента клієнтів [14].

Модель типів особистості (Process commutation model (PCM)) – допомагає визначити тип особистості клієнта, що сприяє кращому розумінню його потреб та способів комунікації. Вона виділяє шість типів особистості: бунтар, фантазер, мислитель, промоутер, наполегливий та гармонійний. Знаючи тип особистості клієнта, компанія може адаптувати свій підхід для більш ефективної взаємодії [15].

Цінність клієнта протягом його життя (Прогнозування CLV) – це загальна вартість клієнта для бізнесу за весь період його взаємовідносин з брендом. Замість того, щоб оцінювати вартість окремих транзакцій, ця цінність враховує всі потенційні транзакції, які можуть бути здійснені протягом періоду взаємодії з клієнтом, і розраховує конкретний дохід від цього клієнта [16].

Одним із важливіших понять класифікації клієнтів – є кластеризація клієнтів. Розуміння своїх клієнтів, а також вивчення структури клієнтського активу є однією з найважливіших задач у сучасному висококонкурентному світі, який бореться за клієнтів [17]. Є різні підходи до вирішення цієї задачі. Одним із них, що набирає популярність останнім часом, – це кластеризація, за допомогою якої компанія може отримати чітке уявлення про структуру свого клієнтського активу. Задача кластеризації клієнтів полягає в поділі всіх клієнтів компанії на групи з подібними характеристиками. Розглянемо приклад зі сфери роздрібної торгівлі, зокрема клієнтів мережі супермаркетів. Відвідувачі магазину мають різні потреби та цілі покупок, що значно впливає на їхню поведінку. Одні клієнти регулярно купують продукти першої необхідності, інші віддають перевагу товарам преміум-сегмента, хтось шукає акційні пропозиції, а дехто використовує магазин як точку швидких покупок по дорозі додому. На основі придбаних товарів можна розподілити клієнтів на певні групи, кожна з яких має власні особливості та потреби.

Розуміння структури клієнтської бази дозволяє компанії ефективніше вибудовувати стратегію взаємодії з покупцями, впроваджувати персоналізовані CRM-рішення та розробляти програми лояльності. Маркетологи, маючи уявлення про поведінкові особливості різних сегментів, можуть створювати таргетовані пропозиції, що краще відповідають очікуванням споживачів і сприяють підвищенню їхньої задоволеності. У сучасних умовах бізнесу особлива увага приділяється персоналізації підходів до клієнтів, і саме аналітичні технології відіграють ключову роль у розробці релевантних пропозицій. Використання методів кластеризації допомагає сфокусувати маркетингові зусилля на перспективних групах споживачів, оптимізуючи ресурси компанії та підвищуючи ефективність її роботи.

У бізнесі завжди є альтернатива: він може орієнтуватись на всіх клієнтів або на деякі групи. Розглянемо детальніше, які варіанти має компанія при розробці маркетингової стратегії і як клієнтська аналітика може в цьому допомогти. Існують три підходи до маркетингових стратегій у компанії, які визначають варіанти роботи з клієнтами: масовий, концентрований і диференційований маркетинг. Ці підходи різняться у способі, яким компанія вибирає цільову аудиторію. Концентрований і диференційований маркетинг відзначаються усвідомленням важливості відходу від масовості та побудовою клієнтської стратегії, яка враховує особливості та відмінності клієнтів. Розглянемо особливості детальніше [18].

Масовий маркетинг передбачає пропозицію одного стандартного продукту для всіх споживачів без урахування їхніх індивідуальних особливостей та потреб (рис. 3). Основна ідея цієї стратегії полягає у досягненні максимальної кількості клієнтів завдяки єдиному маркетинговому підходу. Основні характеристики масового маркетингу включають орієнтацію на широкий ринок без сегментації, використання однакових рекламних повідомлень для всіх споживачів та масове виробництво, що дозволяє знижувати собівартість продукції. Приклади таких компаній – великі FMCG-бренди, як Coca-Cola чи McDonald's, де продукти доступні для всіх споживачів. Переваги масового маркетингу полягають у економії на масштабі виробництва, зниженні витрат на маркетингові дослідження та персоналізацію, а також високому рівні впізнаваності бренду. Однак існують і недоліки, серед яких неможливість задовольнити специфічні потреби окремих груп клієнтів та висока конкуренція через уніфікованість пропозиції.

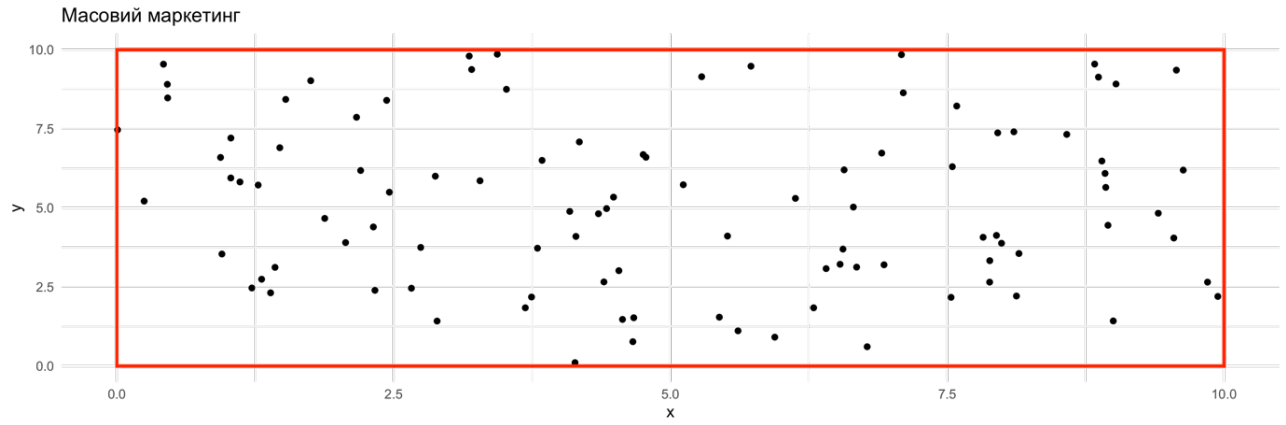

Рис. 3. Приклад масового маркетингу

Джерело: створено авторами за допомогою мови програмування R

Концентрований маркетинг передбачає орієнтацію на один конкретний сегмент ринку (рис. 4), де компанія намагається зайняти домінуючу позицію. Така стратегія є особливо ефективною для малого та середнього бізнесу, який не має ресурсів для роботи з усім ринком. Основні характеристики концентрованого маркетингу включають орієнтацію на одну вузьку групу споживачів, глибоке розуміння їхніх потреб та високу лояльність клієнтів завдяки адаптованій пропозиції. Прикладом можуть служити преміальні бренди, такі як Tesla чи Rolex, які орієнтовані на конкретний прошарок покупців. Переваги цієї стратегії полягають у меншій кількості конкурентів у вибраному сегменті, можливості встановлення високої ціни через ексклюзивність пропозиції та кращій адаптації до потреб клієнтів. Проте існують і недоліки, серед яких ризик втрати прибутку через обмежену кількість клієнтів, а також залежність від одного сегмента ринку, що підвищує ризики у випадку зміни ринкової ситуації.

Диференційований маркетинг передбачає поділ ринку на кілька сегментів і розробку окремих маркетингових стратегій для кожного з них (рис. 5). Компанія створює кілька продуктів або варіацій одного продукту, що відповідають потребам різних категорій споживачів. Основні характеристики цього підходу включають розподіл клієнтів на кілька цільових груп, використання різних маркетингових стратегій для кожного сегмента, а також високі витрати на розробку та просування продуктів. Як приклад можна навести автомобільні компанії, такі як Toyota та BMW, що пропонують моделі для різних сегментів – економ-класу, преміум та спорт. Переваги диференційованого маркетингу полягають у можливості отримання доходів з кількох ринкових ніш, зменшенні залежності від одного сегмента та гнучкості в зміні маркетингової стратегії. Однак існують і недоліки, серед яких високі витрати на дослідження, рекламу та розробку нових продуктів, а також необхідність утримувати кілька маркетингових команд.

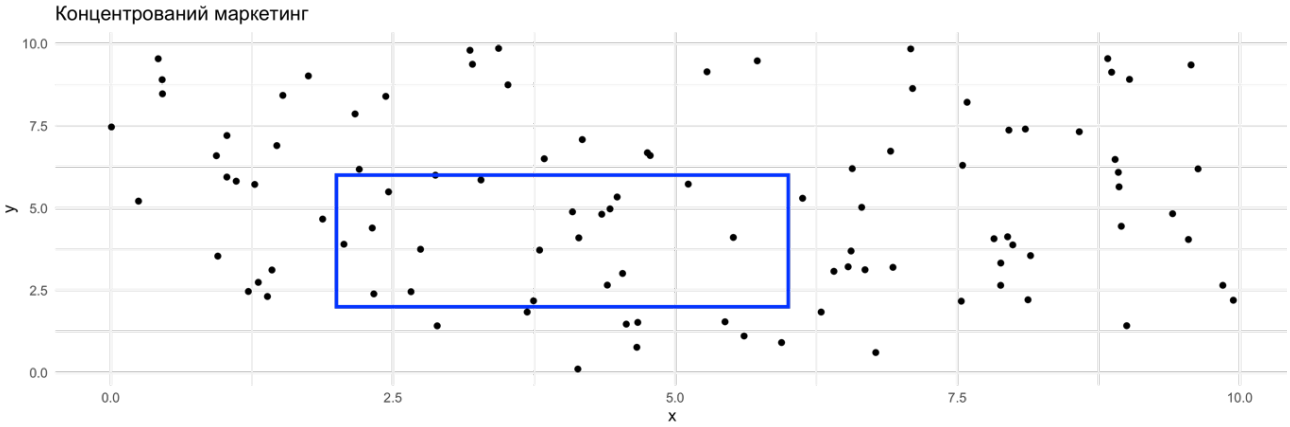

Рис. 4. Приклад концентрованого маркетингу

Джерело: створено авторами за допомогою мови програмування R

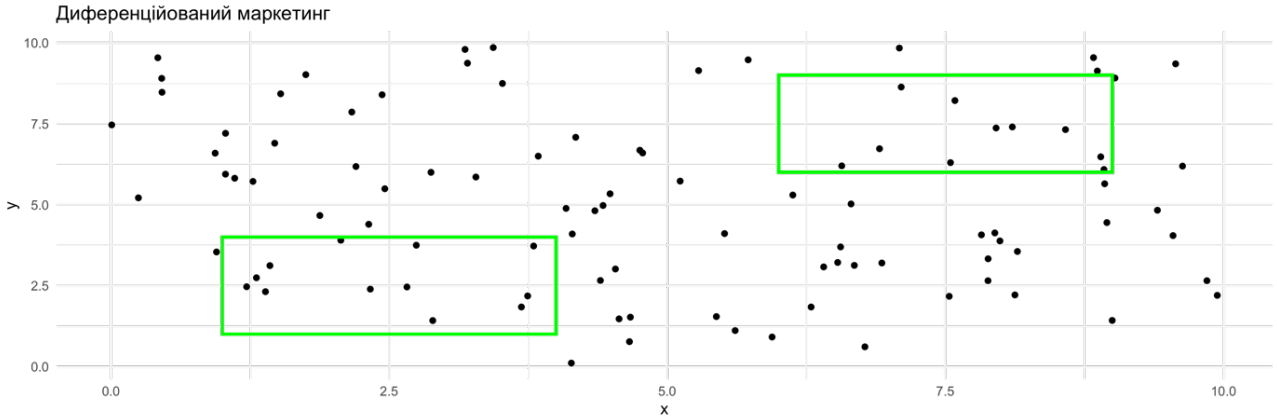

Рис. 5. Приклад диференційованого маркетингу

Джерело: створено авторами за допомогою мови програмування R

Технічний бік кластеризації. Кластеризація є частиною Data Mining – технології та машинного навчання даних. Кластеризація – це автоматичний розподіл клієнтів на однорідні групи залежно від їхньої подібності одразу за кількома характеристиками [19]. Кластерний аналіз виник у антропології завдяки Драйверу та Кроеберу у 1932 році [20], а в психології його вперше застосували Джозеф Зубін у 1938 році [21] та Роберт Тріон у 1939 році [22]. Найбільш відомим застосуванням кластерного аналізу стало використання його Кеттеллом починаючи з 1943 року [23] для класифікації рис у теорії особистості.

Кластеризація є методологічним підходом, що передбачає використання математичних моделей для виявлення груп схожих клієнтів. Цей процес базується на мінімізації внутрішніх відмінностей між об'єктами всередині кожної групи. Основна мета кластерного аналізу – виявлення закономірностей у даних без необхідності формулювання попередніх гіпотез. Застосування кластеризації дозволяє провести попереднє дослідження структури даних та визначити природні кластери. Важливо зазначити, що цей метод дає змогу здійснювати групування об'єктів на основі кількох параметрів, що фактично передбачає роботу у багатовимірному просторі. Кластерний аналіз клієнтської бази ґрунтується на двох основних припущеннях:

можливість поділу клієнтів на групи відповідно до їхніх характеристик;

правильність вибору масштабу та одиниць вимірювання параметрів.

Уявляючи клієнтів як точки у просторі ознак, завдання кластеризації можна розглядати як пошук скупчень цих точок (рис. 6). Під час кластерного аналізу об'єкти поділяються на групи, кожна з яких отримує відповідну мітку (кластер).

Кластер можна описати як групу об'єктів, що мають спільні характеристики. У межах клієнтської аналітики кластер – це стійка група клієнтів, які мають подібні характеристики. Стійкість передбачає стабільність групи в клієнтському активі часу.

У процесі класифікації клієнтів важливо не тільки побудувати ефективні моделі, а й оцінити їхню якість для подальшого використання у практичних задачах. Для цього застосовуються різноманітні інструменти та метрики, що дозволяють визначити, наскільки добре модель справляється з передбаченнями. Одним із основних інструментів для оцінки якості класифікаційної моделі є матриця помилок класифікації, яка також відома як матриця невідповідностей, є таблицею, яка відображає співвідношення між фактичними та прогнозованими класами в результатах роботи класифікаційного алгоритму. Вона допомагає оцінити ефективність моделі, показуючи, скільки разів кожен клас був правильно або неправильно передбачений[24]. У межах завдання прогнозування відтоку є два класи «відтік» та «не відтік». Алгоритм спрогнозував приналежність кожного клієнта одного з класів.

Припустимо, що ми знаємо правильні класи «0» і «1» наших клієнтів, бо попередньо передбачили для них клас. У матриці помилок маємо два класи:

0 – вказує, що клієнт не стане відтоком;

1 – вказує, що клієнт стане відтоком.

Позитивним класом у нашому прикладі є «1». У рамках матриці (табл. 2) описано фактичний клас, тобто прогнозований клас, тобто передбачувані значення. Порівнявши прогнози моделі з фактичними даним, ми отримаємо чотири групи прогнозів матриці помилок класифікації [25].

Розглянемо поняття TP, FP, TN, FN на прикладі прогнозування відтоку (табл. 3).

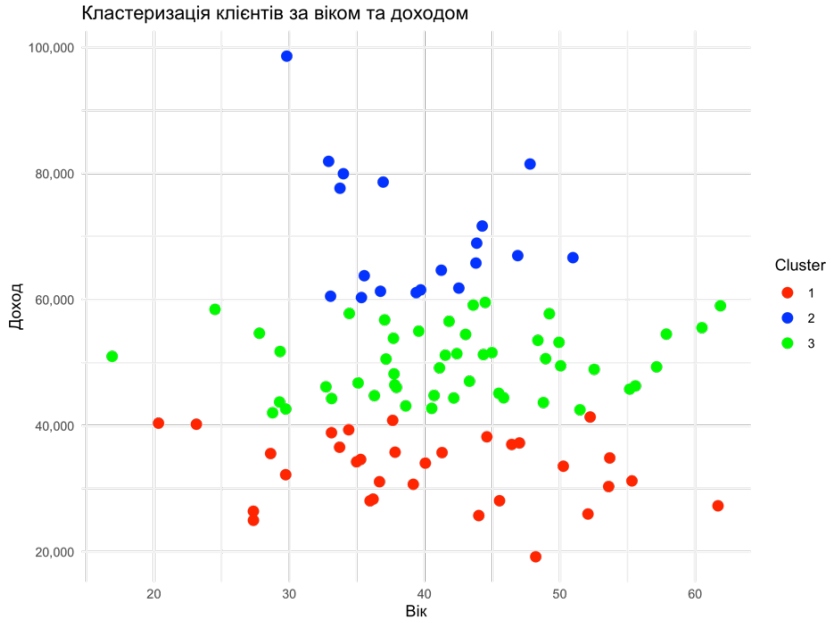

Рис. 6. Приклад візуалізації поділу клієнтів на 3 кластери поділяючи за доходом та віком

Джерело: реалізовано за допомогою мови програмування R

Таблиця 2

Групи прогнозів матриці помилок класифікації

Фактичні значення / прогнозні значення	**x = 1 (Позитивне)**	**x = 0 (Негативне)**
y = 1 (Позитивний прогноз)	Істинно позитивні (True Positive, TP)	Хибно негативні (False Negative, FN)
y = 0 (Негативний прогноз)	Хибно позитивні (False Positive, FP)	Істинно негативні (True Negative, TN)

Джерело: складено авторами, x – фактичні значення, y – прогнозні значення.

Таблиця 3

Поняття TP, FP, TN, FN на прикладі прогнозування відтоку

Фактичні значення \ Прогнозні значення	**Клієнт залишив (x = 1)**	**Клієнт не залишив (x = 0)**
Прогноз: Клієнт залишив (y = 1)	Істинно позитивні (TP): Клієнт дійсно залишив компанію і модель правильно передбачила відтік.	Хибно негативні (FN): Клієнт залишив компанію, але модель передбачила, що він залишиться.
Прогноз: Клієнт не залишив (y = 0)	Хибно позитивні (FP): Клієнт не залишив компанію, але модель помилково передбачила відтік.	Істинно негативні (TN): Клієнт не залишив компанію і модель правильно передбачила, що він залишиться.

Джерело: складено авторами

Істинно позитивні (TP) – це випадки, коли модель правильно передбачила відтік клієнта, який дійсно залишив компанію. Хибно негативні (FN) трапляються, коли модель не змогла передбачити відтік, хоча клієнт насправді залишив компанію. Хибно позитивні (FP) виникають, коли модель помилково передбачила відтік, хоча клієнт насправді залишився. І, нарешті, істинно негативні (TN) – це ситуації, коли модель правильно передбачила, що клієнт залишиться, і він дійсно не залишив компанію. Ці категорії допомагають оцінити точність моделі прогнозування відтоку клієнтів, вказуючи на типи помилок, яких вона може припуститися (табл. 4).

Тепер розглянемо метрики з погляду компонентів матриці.

Правильність / Точність (Accuracy) – це одна з основних метрик для оцінки якості класифікаційної моделі. Вона вимірює частку правильних передбачень серед усіх можливих передбачень. Іншими словами, точність показує, наскільки часто модель дає правильний результат, незалежно від того, чи це позитивний, чи негативний клас.

Таблиця 4

Матриця плутанини з простим прикладом прогнозування відтоку тисячі клієнтів

Фактичні значення \ Прогнозні значення	Клієнт залишив (x = 1)	Клієнт не залишив (x = 0)
Прогноз: Клієнт залишив (y = 1)	150 (Істинно позитивні – TP)	50 (Хибно негативні – FN)
Прогноз: Клієнт не залишив (y = 0)	50 (Хибно позитивні – FP)	750 (Істинно негативні – TN)

Джерело: складено авторами

$$Accuracy = \frac{correct\ classification\ (правильна\ класифікація)}{total\ classification\ (загальна\ кількість\ класифікацій)} = \frac{TP+TN}{TP+TN+FP+FN}$$

Використовуючи дані таблиці ми можемо обрахувати точність:

$$Accuracy = \frac{150+750}{1000} = 0.9 = 90\%$$

класів, не враховуючи їх специфіку. У випадку двійкового класифікатора, що працює з класами «відтік» та «не відтік», можуть виникати два типи помилок: False Positive (FP) та False Negative (FN). False Positive – це випадок, коли модель передбачає відтік клієнта (значення 1), але насправді клієнт залишився (значення 0). Така помилка може призвести до непотрібних витрат, таких як надання бонусів або спеціальних умов клієнтам, які не мають наміру залишати компанію. Наприклад, якщо модель передбачає відтік клієнта, який насправді залишився, компанія може витратити ресурси на надання бонусів або спеціальних пропозицій, які не були необхідні. У нашому прикладі, де FP = 50, це означає, що 5% клієнтів були помилково класифіковані як ті, що йдуть, хоча насправді вони залишилися. У той час як False Negative виникає, коли модель передбачає, що клієнт залишиться (значення 0), але він насправді відходить (значення 1). Ця помилка є більш критичною, оскільки компанія втрачає клієнта, що може спричинити значні фінансові втрати через втрату потенційного доходу. Наприклад, коли модель передбачає, що клієнт залишиться, хоча він фактично йде, компанія не вживає необхідних заходів для утримання клієнта. У нашому прикладі, де FN = 50, це означає, що 5% клієнтів, які насправді пішли, були помилково класифіковані як ті, що залишаються. Для більш точного оцінювання якості моделі використовуються метрики, які враховують помилки для кожного класу окремо. Однією з таких метрик є точність (precision), яка визначає частку правильно класифікованих позитивних випадків серед усіх передбачених позитивних результатів. Ця метрика дозволяє більш детально оцінити надійність позитивних прогнозів, що є важливим для задач, де необхідно зменшити ймовірність помилкових позитивних передбачень [27]. Математичний розрахунок точності має такий вигляд:

$$Precision = \frac{TP}{TP+FP}.$$

У разі прогнозування відтоку запитання може бути сформульоване так: «Яка частка правильно передбачених випадків відтоку серед усіх передбачених відтоків клієнтів?».

$$Precision = \frac{150}{150+50} = 0.75 = 75\%.$$

Що вище значення точності, то краща модель.

Повнота (recall) – це метрика, яка вимірює здатність моделі правильно виявляти позитивні випадки. Вона показує, яка частка реальних позитивних випадків була правильно передбачена [28]. Математична формула повноти має вигляд:

$$Recall = \frac{TP}{TP+FN}.$$

У нашому випадку запитання можна сформулювати так: «Яка частка правильно передбачених випадків відтоку серед усіх клієнтів, які фактично залишили компанію?».

$$Recall = \frac{150}{150+50} = 0.75 = 75\%.$$

Що вище значення повноти, то краща модель.

F1-міра (F1-measure) – це метрика, яка використовується для оцінки ефективності класифікаційної моделі, особливо коли важливо збалансувати точність (precision) і повноту (recall). Вона є гармонійним середнім між цими двома метриками, даючи їм однакову вагу [29].

$$F1measure = \frac{2 * Recall * Precision}{Recall + Precision}.$$

F1-міра допомагає використовувати точність та повноту в одній формулі. В нашому прикладі розрахунок буде такий:

$$F1measure = \frac{2 * 0.75 * 0.75}{0.75 + 0.75} = 0.75 = 75\%.$$

Що вища міра, то точніша модель.

Є загальна формула F-measure, яка поєднує точність (precision) та повноту (recall) в єдину оцінку, при цьому дозволяючи налаштувати важливість кожної з цих метрик за допомогою параметра β.

Загальна формула для F-measure виглядає таким чином:

$$F_{\beta} = \left(1+\beta^2\right) * \frac{Precision * Recal}{\beta^2 * Precision + Recal},$$

де β – це коефіцієнт, що визначає, наскільки важливішою є повнота (recall) порівняно з точністю (precision). Якщо β = 1, то F-measure перетворюється на F1-measure, яка є спеціальним випадком цієї метрики і надає рівну вагу точності та повноті. У такому випадку формула для F1-measure спрощується до:

$$F1measure = \frac{2 * Recall * Precision}{Recall + Precision}.$$

Ця спеціальна форма F-measure, відома як F1-measure, широко використовується в багатьох задачах машинного навчання, коли точність і повнота мають однакову важливість. Однак F-measure у загальному вигляді дозволяє варіювати баланс між точністю та повнотою через зміну параметра β. Наприклад, для задач, де важливо мінімізувати False Positives, значення β може бути менше за одиницю, а для задач, де критично важлива Recall, β буде більшим за одиницю [30].

Існують додатково графічні інструменти, які використовуються для оцінювання моделей класифікації, зокрема ROC-криві (Receiver Operating Characteristic curve) та PR-криві (Precision-Recall curve). ROC-крива дозволяє аналізувати відносини між істинними позитивними (True Positive Rate, TPR) та хибно позитивними (False Positive Rate, FPR) для різних порогів класифікації, допомагаючи оцінити ефективність моделі в умовах зміни порогу. Одним із важливих показників, який можна отримати з ROC-кривої, є AUC (Area Under the Curve), що вимірює площу під цією кривою. AUC варіюється від 0 до 1, де значення, близьке до 1, свідчить про високу якість моделі, яка ефективно розрізняє класи, а значення, близьке до 0.5, вказує на модель, яка не має кращих результатів, ніж випадковий вибір [31].

Зі зміною значення AUC, відповідно, змінюється форма ROC-кривої. Наприклад, у разі ідеальної моделі (AUC = 1) (рис. 7) криві будуть наближатися до верхнього лівого кута (рис. 8), що вказує на те, що модель безпомилково класифікує позитивні та негативні класи. У свою чергу, модель із значенням AUC, що наближається до 0.5, відображатиме випадкові прогнози, що призведе до майже прямої лінії від нижнього лівого до верхнього правого кута (рис. 9). Відповідно, між цими двома крайніми значеннями AUC буде спостерігатися прогресивне покращення моделі, що означатиме зростання її здатності до точного розрізнення класів.

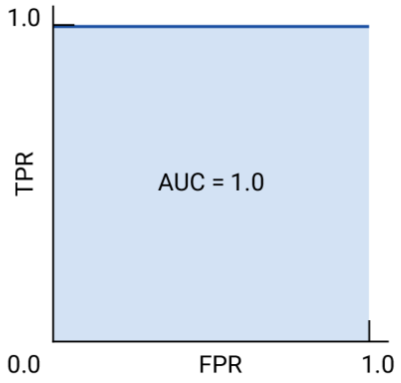

Рис. 7. Приклад ROC-кривою при AUC = 1

Джерело: [31]

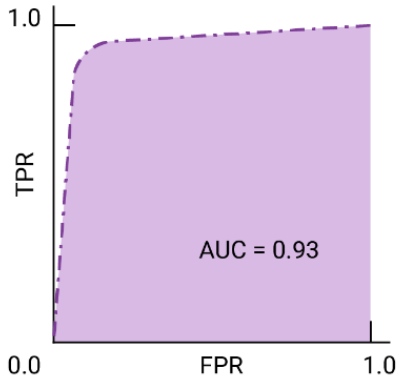

Рис. 8. Приклад ROC-кривою при AUC = 0.93

Джерело: [31]

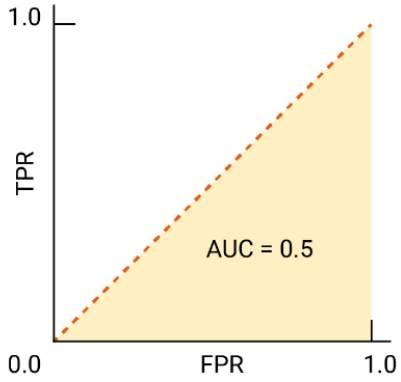

Рис. 9. Приклад ROC-кривою при AUC = 0.5

Джерело: [31]

Класифікація клієнтів становить основу сучасного маркетингу, трансформуючи масиви даних у стратегічні інсайти. Дослідження доводить, що застосування алгоритмів машинного навчання, таких як дерева рішень, KNN або Random Forest, дозволяє виявляти складні закономірності в поведінці клієнтів, враховуючи їхні демографічні,

фінансові та поведінкові характеристики. На прикладі туристичного агентства продемонстровано, як аналіз віку та доходу допомагає сегментувати аудиторію для персоналізації рекламних пропозицій, що безпосередньо впливає на ефективність маркетингових кампаній.

Ключовим аспектом роботи є комплексна оцінка якості моделей за допомогою метрик (Accuracy, F1, AUC-ROC), які визначають баланс між помилками прогнозу та точністю. Особливу увагу приділено кластеризації як інструменту для виявлення природних груп клієнтів, що відкриває можливості для диференційованого підходу та оптимізації витрат. Проте існують обмеження: реальні дані часто містять шум, а дисбаланс класів чи відсутність історичних даних може знижувати точність прогнозів.

Практична цінність дослідження полягає в тому, що компанії, використовуючи класифікацію, здатні не лише запобігати відтоку клієнтів, але й формувати індивідуальні пропозиції, підвищуючи лояльність та довгострокову цінність клієнтів (LTV). Наприклад, прогнозуючи схильність до дорогих турів, агентства можуть таргетувати аудиторію з високим доходом, зменшуючи витрати на масову рекламу.

Для подальшого розвитку теми перспективними напрямками є інтеграція глибокого навчання для аналізу текстових відгуків, врахування часових факторів (сезонність, тривалість взаємин) та вдосконалення алгоритмів для роботи з нестандартними даними. Таким чином, класифікація клієнтів залишається не лище теоретичною концепцією, але й потужним інструментом, який, поєднуючи наукові методи з бізнес-логікою, забезпечує конкурентні переваги в умовах динамічного ринку.

Висновок. Дослідження маркетингової концепції класифікації клієнтів підтвердило її ключову роль у сучасному бізнес-середовищі, де ефективне управління клієнтською базою стає вирішальним чинником конкурентоспроможності. Використання алгоритмів машинного навчання, таких як дерева рішень (C4.5, ID3), KNN, Random Forest та кластеризація, дозволяє не лише сегментувати клієнтів за демографічними, фінансовими та поведінковими характеристиками, але й прогнозувати їхню поведінку, зокрема ризик відтоку. На прикладі туристичного агентства продемонстровано, як аналіз віку та доходу клієнтів дозволяє оптимізувати маркетингові стратегії, персоналізуючи рекламні пропозиції. Це підтверджує, що класифікація є потужним інструментом для підвищення ефективності маркетингових кампаній, зменшення витрат на масову рекламу та зростання лояльності клієнтів. Важливим аспектом роботи є оцінка якості класифікаційних моделей за допомогою метрик (Accuracy, Precision, Recall, F1, AUC-ROC) та аналізу матриці помилок. Це дозволяє мінімізувати фінансові втрати, пов'язані з помилковими прогнозами, та оптимізувати бізнес-процеси. Наприклад, зниження частки хибно негативних прогнозів (FN) критично для запобігання відтоку клієнтів, тоді як контроль хибно позитивних (FP) допомагає уникнути марних витрат на утримання вже лояльних клієнтів. Маркетингові стратегії (масовий, концентрований, диференційований маркетинг) отримують новий рівень ефективності завдяки інтеграції результатів класифікації. Зокрема, диференційований підхід, заснований на кластеризації, дозволяє компаніям точніше визначати потреби різних груп клієнтів і пропонувати їм індивідуальні рішення. Перспективи подальших досліджень полягають у вдосконаленні алгоритмів для роботи з нестандартними даними (наприклад, текстовими відгуками), інтеграції глибокого навчання та врахуванні часових факторів (сезонність, тривалість взаємин із клієнтом). Також актуальним залишається розвиток інструментів візуалізації результатів класифікації для полегшення прийняття рішень маркетологами. У висновку можна стверджувати, що класифікація клієнтів – це не лише теоретичний інструмент, але й практичний механізм перетворення даних на конкурентні переваги. Її впровадження дозволяє компаніям не тільки реагувати на зміни ринку, але й проактивно формувати маркетингові стратегії, спрямовані на довгостроковий розвиток та максимізацію прибутку.

ЛІТЕРАТУРА:

1. Класифікація. Словник української мови. URL: https://slovnyk.ua/index.php?swrd=класифікація
2. Mykhailichenko M., Lozhachevska O., Smagin V., Krasnoshtan O., Zos-Kior M., Hnatenko I. Competitive strategies of personnel management in business processes of agricultural enterprises focused on digitalization. *Management Theory and Studies for Rural Business and Infrastructure Development.* 2021. Vol. 43(3). P. 403–414.
3. Задача класифікації. URL: https://uk.wikipedia.org/wiki/Задача_класифікації
4. Binary Classification. LearnDataSci. URL: https://www.learndatasci.com/glossary/binary-classification/
5. Шість порад для фахівців-початківців з машинного навчання. UAspectr. URL: https://uaspectr.com/2021/05/31/shist-porad-dlya-fahivtsiv-pochatkivtsiv-z-mashynnogo-navchannya/

6. Zhyvko Z., Nikolashyn A., Semenets I., Karpenko Y., Zos-Kior M., Hnatenko I., Klymenchukova N., Krakhmalova N. Secure aspects of digitalization in management accounting and finances of the subject of the national economy in the context of globalization. *Journal of Hygienic Engineering and Design*. 2022. Vol. 39. P. 259–269.

7. 1R Algorithm. James Tharpe. URL: https://www.jamestharpe.com/1r-algorithm/

8. C4.5. GitHub. URL: https://github.com/barisesmer/C4.5?tab=readme-ov-file#c45

9. k-Nearest Neighbors (KNN). IBM. URL: https://www.ibm.com/think/topics/knn

10. Random Forest. IBM. URL: https://www.ibm.com/think/topics/random-forest

11. Sklearn Iterative Dichotomiser 3 (ID3) Algorithms. GeeksforGeeks. URL: https://www.geeksforgeeks.org/sklearn-iterative-dichotomiser-3-id3-algorithms/

12. Прогностична модель схильності: Дізнайтеся про поведінку клієнтів. Zfort. URL: https://www.zfort.com.ua/blog/prognostichna-model-skhilnosti-diznaitesya-pro-povedinku-kliyentiv-za

13. Customer Churn Rate. Zendesk. URL: https://www.zendesk.com/blog/customer-churn-rate/

14. Метод SABONE: Як продавати новий продукт. ABest Agency. URL: https://abest.agency/tpost/3muasco981-metod-sabone-yak-prodavati-novii-produkt

15. Process Communication Model (PCM) Introductions. LemanSkills. ULR: https://lemanskills.com/process-communication-model-pcm-introductions/

16. Customer Lifetime Value (CLV). Qualtrics. URL: https://www.qualtrics.com/experience-management/customer/customer-lifetime-value/

17. Customer Insight. TechTarget. URL: https://www.techtarget.com/searchcustomerexperience/definition/customer-insight-consumer-insight

18. Олійник А. С., Толочій О. Р., Щербина Ю. В. Ефективність проєктного менеджменту сучасних підприємств в умовах диджиталізації. *Вісник Полтавського державного аграрного університету (Серія «Економіка, управління та фінанси»).* 2024. Випуск 2. С. 61–66.

19. Clustering. Solix. URL: https://www.solix.com/uk/kb/clustering/

20. Driver and Kroeber Quantitative Expression of Cultural Relationships. University of California Publications in American Archaeology and Ethnology. Berkeley, CA: University of California Press, 1932. P. 211–256.

21. Zubin J. A technique for measuring like-mindedness. *The Journal of Abnormal and Social Psychology.* 1938. №33 (4). P. 508–516.

22. Tryon R. C. Cluster Analysis: Correlation Profile and Orthometric (factor) Analysis for the Isolation of Unities in Mind and Personality. Edwards Brothers. 1939. P. 212–215.

23. Cattell R. B. The description of personality: Basic traits resolved into clusters. *Journal of Abnormal and Social Psychology*. 1943. № 38 (4). P. 476–506.

24. Какурінов К. В., Чернікова Н. М., Долина Р. М. Цифрові технології в управлінні маркетинговою діяльністю підприємства. *Вісник Полтавського державного аграрного університету (Серія «Економіка, управління та фінанси»).* 2024. Випуск 2. С. 36–43.

25. Терещенко І.О., Буряк О.М. Оцінка та удосконалення маркетингової збутової політики підприємств аграрної сфери. *Вісник Полтавського державного аграрного університету (Серія «Економіка, управління та фінанси»).* 2024. Випуск 1. С. 28–35.

26. Accuracy, Precision, and Recall. Google Developers. URL: https://developers.google.com/machine-learning/crash-course/classification/accuracy-precision-recall?hl=uk

27. Матриця невідповідностей і метрики. URL: https://dumnyj.eu/blog/matrytsia-nevidpovidnostej-i-metryky/

28. Recall. Kolena Docs. URL: https://docs.kolena.com/metrics/recall/

29. F1-Score Guide. V7 Labs. URL: https://www.v7labs.com/blog/f1-score-guide

30. F1-Score. Flowhunt. URL: https://www.flowhunt.io/glossary/f1-score/

31. ROC Curve. DataTab. URL: https://datatab.net/tutorial/roc-curve

REFERENCES:

1. Classification. *Slovnyk ukrainskoi movy* [Ukrainian Language Dictionary]. URL: https://slovnyk.ua/index.php?swrd=класифікація

2. Mykhailichenko, M., Lozhachevska, O., Smagin, V., Krasnoshtan, O., Zos-Kior, M., Hnatenko, I. (2021). Competitive strategies of personnel management in business processes of agricultural enterprises focused on digitalization. *Management Theory and Studies for Rural Business and Infrastructure Development,* 43 (3), 403–414.

3. Classification problem. URL: https://uk.wikipedia.org/wiki/Задача_класифікації

4. Binary Classification. *LearnDataSci*. URL: https://www.learndatasci.com/glossary/binary-classification/

5. Six tips for aspiring machine learning specialists. *UAspectr*. URL: https://uaspectr.com/2021/05/31/shist-porad-dlya-fahivtsiv-pochatkivtsiv-z-mashynnogo-navchannya/

6. Zhyvko, Z., Nikolashyn, A., Semenets, I., Karpenko, Y., Zos-Kior, M., Hnatenko, I., Klymenchukova, N., Krakhmalova, N. (2022). Secure aspects of digitalization in management accounting and finances of the subject of the national economy in the context of globalization. *Journal of Hygienic Engineering and Design,* 39, 259–269.

7. 1R Algorithm. *James Tharpe*. URL: https://www.jamestharpe.com/1r-algorithm/

8. C4.5. *GitHub*. URL: https://github.com/barisesmer/C4.5?tab=readme-ov-file#c45

9. k-Nearest Neighbors (KNN). *IBM*. URL: https://www.ibm.com/think/topics/knn

10. Random Forest. *IBM*. URL: https://www.ibm.com/think/topics/random-forest

11. Sklearn Iterative Dichotomiser 3 (ID3) Algorithms. *GeeksforGeeks*. URL: https://www.geeksforgeeks.org/sklearn-iterative-dichotomiser-3-id3-algorithms/

12. Propensity Model: Learn about customer behavior. *Zfort*. URL: https://www.zfort.com.ua/blog/prognostichna-model-skhilnosti-diznaitesya-pro-povedinku-kliyentiv-za

13. Customer Churn Rate. *Zendesk*. URL: https://www.zendesk.com/blog/customer-churn-rate/

14. SABONE Method: How to sell a new product. *ABest Agency*. URL: https://abest.agency/tpost/3muasco981-metod-sabone-yak-prodavati-novii-produkt

15. Process Communication Model (PCM) Introductions. *LemanSkills*. URL: https://lemanskills.com/process-communication-model-pcm-introductions/

16. Customer Lifetime Value (CLV). *Qualtrics*. URL: https://www.qualtrics.com/experience-management/customer/customer-lifetime-value/

17. Customer Insight. *TechTarget*. URL: https://www.techtarget.com/searchcustomerexperience/definition/customer-insight-consumer-insight

18. Oliinyk, A. S., Tolochii, O. R., Shcherbyna, Y. V. (2024). Efficiency of project management of modern enterprises in the conditions of digitalization. *Bulletin of Poltava State Agrarian University (Series "Economics, Management and Finance"),* 2, 61–66.

19. Clustering. *Solix*. URL: https://www.solix.com/uk/kb/clustering/

20. Driver and Kroeber (1932). Quantitative Expression of Cultural Relationships. *University of California Publications in American Archaeology and Ethnology*. Berkeley, CA: University of California Press. P. 211–256.

21. Zubin, J. (1938). A technique for measuring like-mindedness. *The Journal of Abnormal and Social Psychology,* 33 (4), 508–516.

22. Tryon, R. C. (1939). *Cluster Analysis: Correlation Profile and Orthometric (factor) Analysis for the Isolation of Unities in Mind and Personality*. Edwards Brothers. P. 212–215.

23. Cattell, R. B. (1943). The description of personality: Basic traits resolved into clusters. *Journal of Abnormal and Social Psychology,* 38 (4), 476–506.

24. Kakurinov, K. V., Chernikova, N. M., Dolyna, R. M. (2024). Digital technologies in enterprise marketing management. *Bulletin of Poltava State Agrarian University (Series "Economics, Management and Finance"),* 2, 36–43.

25. Tereshchenko, I. O., Buriak, O. M. (2024). Evaluation and improvement of marketing sales policy of agricultural enterprises. *Bulletin of Poltava State Agrarian University (Series "Economics, Management and Finance"),* 1, 28–35.

26. Accuracy, Precision, and Recall. *Google Developers*. URL: https://developers.google.com/machine-learning/crash-course/classification/accuracy-precision-recall?hl=uk

27. Confusion Matrix and Metrics. URL: https://dumnyj.eu/blog/matrytsia-nevidpovidnostej-i-metryky/

28. Recall. *Kolena Docs*. URL: https://docs.kolena.com/metrics/recall/

29. F1-Score Guide. *V7 Labs*. URL: https://www.v7labs.com/blog/f1-score-guide

30. F1-Score. *Flowhunt*. URL: https://www.flowhunt.io/glossary/f1-score/

31. ROC Curve. *DataTab*. URL: https://datatab.net/tutorial/roc-curve

Стаття надійшла до редакції 15.04.2025