



Прийнято 20.04.2026. Прорецензовано 10.05.2026. Опубліковано 30.05.2026.

УДК 681.518:004.85:658.8

DOI: 10.31471/1993-9981-2026-1(56)-195-207

МАТЕМАТИЧНА МОДЕЛЬ АДАПТИВНОЇ СЕГМЕНТАЦІЇ КОРИСТУВАЧІВ В АВТОМАТИЗОВАНИХ СИСТЕМАХ УПРАВЛІННЯ МАРКЕТИНГОВИМИ КАМПАНІЯМИ

Коржик М. В.

Кандидат технічних наук, Доцент

Національний технічний університет України «Київський політехнічний інститут
ім. Ігоря Сікорського»;

03056, Берестейський проспект, 37, м. Київ, Україна;

<https://orcid.org/0000-0003-3453-6258>

e-mail: korzhyk@kpi.ua

Бокушев Е. Н.

Старший викладач кафедри

Національний технічний університет України «Київський політехнічний інститут
ім. Ігоря Сікорського»;

03056, Берестейський проспект, 37, м. Київ, Україна;

<https://orcid.org/0009-0009-6725-3330>

e-mail: bokushev@gmail.com

Дехтярук К. Б.

Аспірант, Науковий співробітник

Національний технічний університет України «Київський політехнічний інститут
ім. Ігоря Сікорського»;

03056, Берестейський проспект, 37, м. Київ, Україна;

<https://orcid.org/0009-0001-3777-6277>

e-mail: kostiadekhthiaruk@gmail.com

Анотація. У статті розроблено математичну модель адаптивної сегментації користувачів для автоматизованих систем управління маркетинговими кампаніями, орієнтованих на роботу в умовах динамічного інформаційного середовища та безперервного надходження даних. Актуальність дослідження зумовлена необхідністю підвищення ефективності маркетингових рішень шляхом оперативного врахування змін у поведінці споживачів та автоматизованого формування релевантних цільових аудиторій. Запропоновано модифікований алгоритм кластеризації, який враховує часову динаміку поведінкових характеристик користувачів та забезпечує автоматичне оновлення сегментів у режимі реального часу без необхідності повної перебудови кластерної структури. Формалізовано багатовимірний простір ознак користувачів, що включає демографічні,

Запропоноване посилання: Коржик, М. В., Бокушев, Е. Н. & Дехтярук, К. Б. (2026). Математична модель адаптивної сегментації користувачів в автоматизованих системах управління маркетинговими кампаніями. *Методи та прилади контролю якості*, 1(56), 195-207. doi: 10.31471/1993-9981-2026-1(56)-195-207

* Відповідальний автор



поведінкові та транзакційні параметри. Для забезпечення коректної обробки різнорідних даних запропоновано процедури нормалізації та масштабування ознак, які дозволяють підвищити стійкість алгоритму до неоднорідності вхідної інформації. У роботі визначено критерії оцінювання якості сегментації на основі індексу Silhouette та модифікованого індексу Девіса–Болдіна, що забезпечують комплексний аналіз внутрішньої однорідності кластерів та рівня їх відокремленості. Розроблено механізм адаптації сегментів із використанням ковзного вікна та адаптивного коефіцієнта згладжування, який дозволяє ефективно відстежувати дрейф даних, своєчасно реагувати на зміни поведінки користувачів та підтримувати актуальність сформованих сегментів у довготривалій перспективі. Проведено порівняльний аналіз запропонованого підходу з поширеними методами кластеризації, зокрема K-Means, DBSCAN та Gaussian Mixture Models. Отримані результати свідчать про переваги розробленого алгоритму за показниками точності сегментації, стабільності кластерної структури та обчислювальної ефективності при роботі з великими потоками маркетингових даних.

Ключові слова: адаптивна сегментація, кластеризація, потокова обробка даних, машинне навчання, автоматизована система управління, маркетингові кампанії, персоналізація.

Вступ

Сучасний стан цифрового маркетингу характеризується стрімким зростанням обсягів даних про поведінку користувачів та необхідністю їх оперативної обробки для прийняття управлінських рішень. За даними досліджень, 71% споживачів очікують персоналізованих взаємодій від компаній, при цьому 81% користувачів ігнорують нерелевантні повідомлення [1]. Це формує потребу в автоматизованих системах, здатних адаптивно сегментувати аудиторію та оптимізувати маркетингові кампанії без значного втручання оператора.

Ключовою задачею автоматизації управління маркетинговими кампаніями є сегментація користувачів — процес розподілу аудиторії на однорідні групи за визначеними ознаками. Традиційні підходи до сегментації, засновані на статичних правилах або ручному аналізі, мають суттєві обмеження: низьку швидкість адаптації до змін поведінки користувачів, обмежену масштабованість та неможливість урахування багатовимірних залежностей у даних [2, 3].

Методи машинного навчання (ML) відкривають принципово нові можливості для вирішення задачі сегментації. Зокрема, алгоритми кластеризації дозволяють автоматично виявляти приховані структури у даних та групувати користувачів за комплексом поведінкових, демографічних та контекстних ознак [4]. Однак класичні алгоритми кластеризації (K-Means, DBSCAN) працюють у статичному режимі та не враховують темпоральну динаміку поведінки користувачів [5]. Поточкові (streaming) алгоритми кластеризації, такі

як Mini-Batch K-Means [6] та BIRCH [7], частково вирішують проблему масштабованості, проте не забезпечують механізму адаптації до дрейфу даних.

Актуальність дослідження зумовлена необхідністю розробки математичних моделей, які забезпечують адаптивне оновлення сегментів у реальному часі з урахуванням дрейфу даних при збереженні обчислювальної ефективності, що є критичним для автоматизованих систем управління. Особливу значущість проблема набуває для українського ринку, де цифрова економіка зростає на 15% щорічно, проте бракує локалізованих інструментів автоматизації маркетингу [8].

Метою статті є розробка математичної моделі адаптивної сегментації користувачів для автоматизованих систем управління маркетинговими кампаніями, що забезпечує динамічне оновлення кластерів на основі поточних даних з контрольованою швидкістю адаптації.

Аналіз літературних джерел та постановка проблеми

Проблема сегментації користувачів у контексті автоматизації маркетингу досліджувалася у працях багатьох вітчизняних та зарубіжних вчених. У роботі [2] запропоновано використання алгоритму K-Means для сегментації клієнтів e-commerce платформ на основі RFM-аналізу (Recency, Frequency, Monetary). Автори продемонстрували підвищення точності таргетування на 15–18%, проте модель не враховує часову динаміку поведінки та потребує повного перенавчання при зміні патернів.

У дослідженні [3] застосовано ієрархічну кластеризацію з урахуванням демографічних та поведінкових ознак. Перевагою підходу є можливість візуалізації дендрограми сегментів, однак обчислювальна складність $O(n^3)$ робить його непридатним для систем реального часу з великими обсягами даних.

Роботи [4, 9] присвячені застосуванню алгоритму DBSCAN для виявлення кластерів довільної форми у просторі поведінкових ознак. Алгоритм демонструє робастність до шуму та не потребує апріорного визначення кількості кластерів, проте чутливий до вибору параметрів ϵ та MinPts, що ускладнює автоматизацію.

Окрему групу становлять потокові алгоритми кластеризації, орієнтовані на обробку даних у реальному часі. Mini-Batch K-Means [6] зменшує обчислювальні витрати за рахунок використання підвибірок, однак не адаптує кількість кластерів. Алгоритм BIRCH [7] будує деревоподібну структуру CF-Tree для інкрементальної кластеризації, але не враховує часову динаміку ознак. DenStream [10] розширює DBSCAN для поточкових даних через механізм мікрокластерів з часовим затуханням, проте потребує значних обчислювальних ресурсів для високовимірних просторів.

Підходи на основі глибокого навчання [11] пропонують використання автоенкодерів для зниження розмірності простору ознак з подальшою кластеризацією у латентному просторі. Ці методи забезпечують високу якість сегментації для складних нелінійних залежностей, але потребують значних обчислювальних ресурсів та великих обсягів навчальних даних.

З позицій теорії автоматичного управління задача адаптивної сегментації може розглядатися як задача адаптивного управління з ідентифікацією [12]. Зворотний зв'язок за якістю кластеризації формує контур управління, де регульованою величиною є розбиття множини користувачів, а керуючим впливом — параметри кластеризації. Цей підхід дозволяє застосувати апарат теорії стійкості для аналізу збіжності алгоритму [13].

Серед вітчизняних досліджень слід відзначити роботи [14], де розглянуто автоматизацію процесів прийняття рішень у маркетингових системах, та [15], де досліджено методи інтелектуального аналізу даних для задач сегментації у контексті українського e-commerce ринку.

Аналіз літературних джерел дозволяє виявити наступні не вирішені задачі: відсутність математичних моделей, що одночасно забезпечують адаптивність сегментації з урахуванням дрейфу даних, обчислювальну ефективність та інтеграцію з автоматизованими системами управління кампаніями; недостатня формалізація критеріїв оптимальності сегментації з урахуванням бізнес-метрик; відсутність механізмів визначення вагових коефіцієнтів ознак та плавного оновлення сегментів без повного перенавчання моделі.

Архітектура автоматизованої системи сегментації

Перед формалізацією математичної моделі визначимо архітектуру автоматизованої системи, в якій функціонує модуль адаптивної сегментації. Система побудована за модульним принципом та включає наступні компоненти (рис. 1), з'єднані у послідовний конвеєр обробки даних: модуль збору та обробки даних, модуль формування простору ознак (Feature Engineering), модуль адаптивної сегментації (кластеризація та механізм адаптації), модуль аналізу якості кластеризації та систему управління кампаніями.

На рис. 1 представлено модульну архітектуру системи, яка реалізує послідовний конвеєр обробки даних із замкненим контуром управління сегментацією. Дані з джерел надходять до модуля збору та ETL-обробки, далі передаються до модуля Feature Engineering для формування вектора ознак, після чого підготовлені ознаки подаються на вхід модуля адаптивної сегментації (кластеризація). Результати кластеризації оцінюються модулем аналізу якості, який через зворотний зв'язок (позначений пунктирною лінією) передає сигнал до механізму адаптації параметрів кластеризації, формуючи замкнений контур

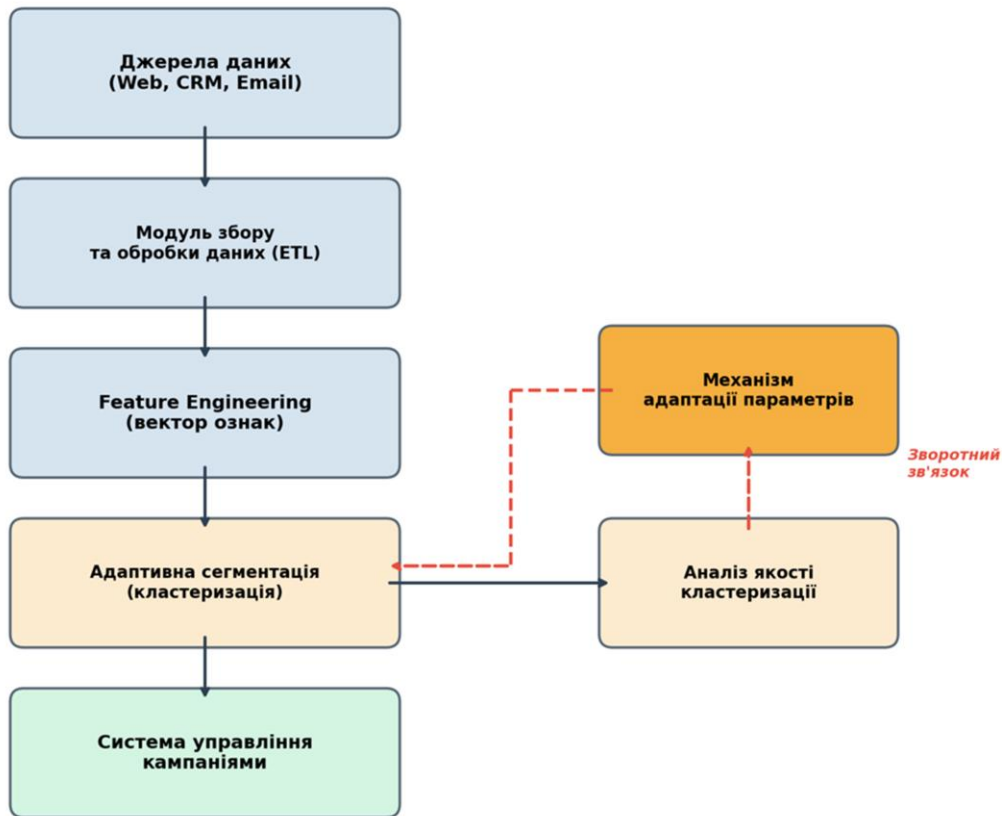


Рисунок 1 – Архітектура автоматизованої системи адаптивної сегментації

управління. Така декомпозиція дозволяє незалежно масштабувати кожен модуль та замінювати окремі компоненти без перебудови всієї системи.

Модуль збору даних агрегує інформацію з різних джерел: систем веб-аналітики, CRM-систем, email-маркетингових платформ та соціальних мереж. Зібрані дані проходять ETL-обробку (Extract, Transform, Load) та нормалізуються для подальшого аналізу. Модуль Feature Engineering формує вектор ознак кожного користувача, який передається до модуля кластеризації. Критичною особливістю архітектури є наявність зворотного зв'язку: модуль аналізу якості оцінює результати кластеризації та передає метрики якості (Silhouette Score, міру дрейфу) до механізму адаптації, який коригує параметри кластеризації (коефіцієнт згладжування, кількість кластерів), формуючи замкнений контур управління.

Математична модель адаптивної сегментації

Формалізація простору ознак користувачів

Нехай $U = \{u_1, u_2, \dots, u_n\}$ - множина користувачів системи, де n - загальна кількість користувачів. Кожен користувач u_i характеризується вектором ознак у момент часу t :

$$x_i(t) = [B_i(t), D_i, C_i(t)]^T \in \mathbb{R}^d, \quad (1)$$

де $x_i(t)$ - вектор ознак i -го користувача у момент часу t ;

$B_i(t)$ - вектор поведінкових ознак (частота візитів, тривалість сесій, кількість покупок);

D_i - вектор демографічних ознак (вік, стать, геолокація);

$C_i(t)$ - вектор контекстних ознак (тип пристрою, час доби, канал залучення).

Простір ознак $X \subseteq \mathbb{R}^d$ є d -вимірним евклідовим простором, де $d = \dim(B) + \dim(D) + \dim(C)$. Для забезпечення порівнянності ознак різної природи застосовується нормалізація Min-Max:

$$\tilde{x}_{ik} = \frac{x_{ik} - x_k^{\min}}{x_k^{\max} - x_k^{\min}}, \quad (2)$$

де \tilde{x}_j^k - нормалізоване значення k -ї ознаки i -го користувача;

x_{jk} - вихідне значення;

x_k^{\min} та x_k^{\max} - мінімальне та максимальне значення k -ї ознаки по всій множині користувачів.

Для урахування відносної важливості різних категорій ознак вводиться матриця вагових коефіцієнтів

$$W = \text{diag}(w_1, w_2, \dots, w_d),$$

де $w_k \geq 0$ та $\sum w_k = 1$. Вагові коефіцієнти визначаються на основі взаємної інформації (mutual information) між ознаками та цільовою змінною конверсії:

$$w_k = \frac{I(x_k; y)}{\sum_{j=1}^d I(x_j; y)}, \quad (3)$$

де $I(x^k; y)$ - взаємна інформація між k -ю ознакою та цільовою змінною конверсії. Такий підхід забезпечує автоматичне визначення ваг без експертного втручання та адаптацію до специфіки конкретного набору даних. Зважена евклідова відстань між користувачами визначається як:

$$d_W(x_i, x_j) = \sqrt{(x_i - x_j)^T W (x_i - x_j)}, \quad (4)$$

де $d_W(x_i, x_j)$ - зважена евклідова відстань між користувачами u_i та u_j ;

W - діагональна матриця вагових коефіцієнтів.

Модифікований алгоритм адаптивної кластеризації

Задача сегментації формалізується як задача оптимізації. Необхідно знайти розбиття множини користувачів $S = \{S_1, S_2, \dots, S_K\}$ на K непустих кластерів, що мінімізує цільову функцію:

$$J(S) = \sum_{j=1}^K \sum_{x_i \in S_j} d_W^2(x_i, \mu_j) \rightarrow \min, \quad (5)$$

де $J(S)$ - цільова функція кластеризації;

μ_j - центроїд j -го кластера;

W - матриця вагових коефіцієнтів ознак.

Оптимальна кількість кластерів K^* визначається автоматично на основі комбінованого нормалізованого критерію:

$$K^* = \arg \max [\alpha \cdot \tilde{SC}(K) - \beta \cdot \tilde{DB}(K)], \quad (6)$$

де $\tilde{SC}(K)$ та $\tilde{DB}(K)$ - нормалізовані до діапазону $[0, 1]$ значення Silhouette Score та індексу Davies-Bouldin відповідно, отримані за формулою Min-Max нормалізації

по діапазону $K \in [K_{\min}, K_{\max}]$; α, β - вагові коефіцієнти ($\alpha = 0,6, \beta = 0,4$ визначені експериментально). Нормалізація забезпечує порівнянність метрик, оскільки $SC \in [-1, 1]$, а $DB \in (0, +\infty)$ мають різні діапазони значень.

Silhouette Score для i -го об'єкта визначається як:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}, s(i) \in [-1, 1], \quad (7)$$

де $a(i)$ - середня відстань від об'єкта i до інших об'єктів його кластера (когезія);

$b(i)$ - мінімальна середня відстань від об'єкта i до об'єктів інших кластерів (сепарація). Значення $s(i) \in [-1, 1]$, де 1 відповідає ідеальній кластеризації.

Механізм адаптивного оновлення сегментів

Ключовою відмінністю запропонованої моделі від існуючих підходів є механізм адаптивного оновлення сегментів, що базується на принципі ковзного вікна. Нехай $T = \{t_1, t_2, \dots, t_m\}$ — множина часових точок спостереження. Для кожного моменту часу t_m визначається вікно спостереження:

$$W(t_m) = \{x_i(t) \mid t \in [t_m - \tau, t_m]\}, \quad (8)$$

де τ - ширина вікна (визначається експериментально). Центроїди кластерів оновлюються за принципом експоненційного згладжування:

$$\mu_j(t) = (1 - \lambda) \cdot \mu_j(t - 1) + \lambda \cdot \hat{\mu}_j(t), \quad (9)$$

де $\mu_j(t)$ - оновлений центроїд j -го кластера у момент t ;

$\mu_j(t - 1)$ - попереднє значення центроїда;

$\hat{\mu}_j(t)$ - центроїд, обчислений на даних вікна $W(t_m)$;

$\lambda \in (0, 1)$ - коефіцієнт згладжування, що контролює швидкість адаптації.

Значення λ визначає баланс між стабільністю сегментів та швидкістю реакції на зміни. Запропоновано адаптивне визначення λ на основі формалізованої міри дрейфу даних. Міра дрейфу $\Delta(t)$ обчислюється як максимальне значення статистики Колмогорова-Смірнова по всіх ознаках:

$$\Delta(t) = \max_{k=1, \dots, d} KS_k(t), \Delta(t) \in [0, 1], \quad (10)$$

де $KS^k(t)$ - значення статистики Колмогорова–Смірнова для k -ї ознаки, обчислене як супремум абсолютної різниці між емпіричними функціями розподілу у поточному та попередньому вікнах;

d - розмірність простору ознак.

Оскільки $KS \in [0, 1]$, гарантується $\Delta(t) \in [0, 1]$. Адаптивний коефіцієнт згладжування визначається як:

$$\lambda(t) = \lambda_0 + (1 - \lambda_0) \cdot \Delta(t), \quad (11)$$

де λ_0 - базове значення коефіцієнта згладжування.

Оскільки $\Delta(t) \in [0, 1]$, гарантується $\lambda(t) \in [\lambda_0, 1] \subset (0, 1)$. За відсутності дрейфу ($\Delta = 0$) маємо $\lambda = \lambda_0$ (повільна адаптація), при максимальному дрейфі ($\Delta = 1$) маємо $\lambda = 1$ (повне оновлення центроїда).

Критерій необхідності повної перебудови сегментації визначається нерівністю, що фіксує деградацію якості кластеризації:

$$\frac{SC(t - 1) - SC(t)}{SC(t - 1)} > \delta, \quad (12)$$

де $SC(t)$ - поточне значення Silhouette Score;

$SC(t - 1)$ - попереднє значення;

δ - порогове значення деградації якості ($\delta = 0,05$).

Формула (12) спрацьовує лише при зниженні SC , на відміну від варіанту з модулем, який реагував би і на різке покращення. При виконанні умови (12) ініціюється повна перебудова кластерів з оновленням K^* .

Умови збіжності алгоритму

Для аналізу збіжності процедури оновлення центроїдів (9) розглянемо стаціонарний випадок, коли дрейф даних відсутній, тобто розподіл ознак не змінюється з часом. У цьому випадку $\Delta(t) = 0$, $\lambda(t) = \lambda_0 = \text{const}$, і формула (9) описує геометрично збіжний процес. Позначимо істинний центроїд кластера як μ_j^* . Тоді похибка $e_j(t) = \mu_j(t) - \mu_j^*$ задовольняє рекурентному співвідношенню:

$$e_j(t) = (1 - \lambda_0) \cdot e_j(t - 1) + \lambda_0 \cdot \varepsilon(t). \quad (13)$$

За умови $0 < \lambda_0 < 1$ маємо $|1 - \lambda_0| < 1$, що гарантує експоненційну збіжність $e_j(t) \rightarrow 0$ при $t \rightarrow \infty$ зі швидкістю $O((1 - \lambda_0)^t)$. Для нестационарного випадку (з дрейфом)

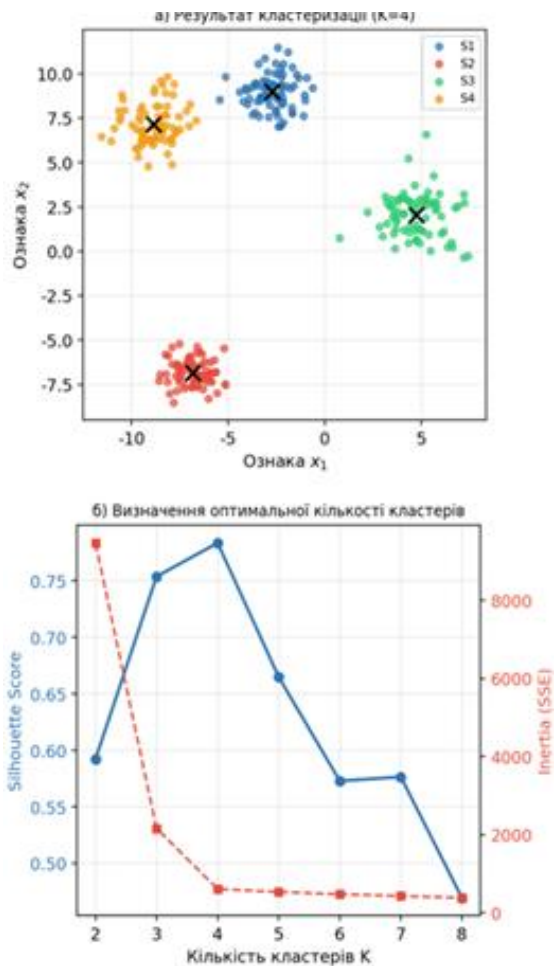


Рисунок 2 – Результат кластеризації та визначення оптимальної кількості кластерів

алгоритм забезпечує обмеженість похибки відстеження, яка залежить від швидкості дрейфу та параметра λ_0 . Детальний аналіз стійкості в нестационарному випадку потребує окремого дослідження та є предметом подальшої роботи.

На рис. 2а представлено результат кластеризації набору даних на $K = 4$ кластери. Кластери мають чітко виражену просторову структуру з мінімальним перекриттям, що підтверджується високим значенням Silhouette Score. Центроїди кластерів (позначені хрестиками) розташовані у геометричних центрах відповідних груп. На рис. 2б наведено залежність Silhouette Score та Inertia (SSE) від кількості кластерів K . Максимум Silhouette Score досягається при $K=4$, що збігається з характерним «ліктем» на кривій Inertia. Комбінований нормалізований критерій (6) автоматично визначає цей оптимум без експертного втручання.

Алгоритм адаптивної сегментації

На основі розробленої математичної моделі сформульовано алгоритм адаптивної сегментації (Алгоритм 1), що реалізується у складі автоматизованої системи управління маркетинговими кампаніями.

Алгоритм 1. Адаптивна сегментація користувачів

Вхід: множина користувачів U , потік даних $D(t)$, параметри $\tau, \delta, \lambda_0, P$.

Вихід: актуальне розбиття

$$S(t) = \{S_1, S_2, \dots, S_k\}.$$

Крок 1. Ініціалізація. Виконати початкову кластеризацію на повному наборі даних. Визначити K^* за критерієм (6). Обчислити початкові центроїди $\mu_j(0)$. Для нових користувачів без поведінкової історії (cold start) виконати призначення до найближчого кластера на основі демографічних ознак D_i .

Крок 2. Збір даних. Для кожного нового моменту t сформувати вікно $W(t_m)$ згідно з (8). Оновити вектори ознак $x_i(t)$. Для нових користувачів ініціалізувати вектор ознак із заповненням поведінкових компонент середніми значеннями найближчого кластера.

Крок 3. Адаптивне оновлення центроїдів. Обчислити $\hat{\mu}_j(t)$ на даних вікна. Обчислити $\Delta(t)$ за формулою (10). Оновити центроїди за формулою (9) з адаптивним λ за (11). Саме на цьому кроці реалізується адаптація: коефіцієнт згладжування визначається автоматично на основі виявленого дрейфу даних (формула (11)), що відрізняє запропонований підхід від простої перекластеризації — центроїди не обчислюються заново, а плавно коригуються з урахуванням історії та інтенсивності змін у даних.

Крок 4. Перепризначення користувачів. Для кожного u_i визначити найближчий центроїд: $S(u_i) = \arg \min_j d^7(x_i, \mu_j)$.

Крок 5. Оцінка якості. Кожні P циклів ($P = 5$ за замовчуванням) обчислити апроксимований $SC(t)$ на випадковій підвибірці розміром $\min(n, 5000)$, що знижує складність з $O(n^2)$ до $O(n'^2)$, де $n' \ll n$. Якщо виконується умова (12), перейти до Кроку 1 з повною перебудовою.

Крок 6. Повторити з Кроку 2 для наступного моменту часу.

Обчислювальна складність одного циклу оновлення (Кроки 2–4) становить $O(n \cdot K \cdot d)$, де n - кількість користувачів у вікні, K - кількість кластерів, d - розмірність простору ознак. Оцінка якості (Крок 5) виконується з періодом P та має складність $O(n'^2)$, де n' - розмір підвибірки. Амортизована складність одного циклу з урахуванням оцінки якості становить $O(n \cdot K \cdot d + n'^2/P)$. Повна перебудова кластерів (Крок 1) виконується рідко та має складність $O(n \cdot K \cdot d \cdot I)$, де I - кількість ітерацій збіжності.

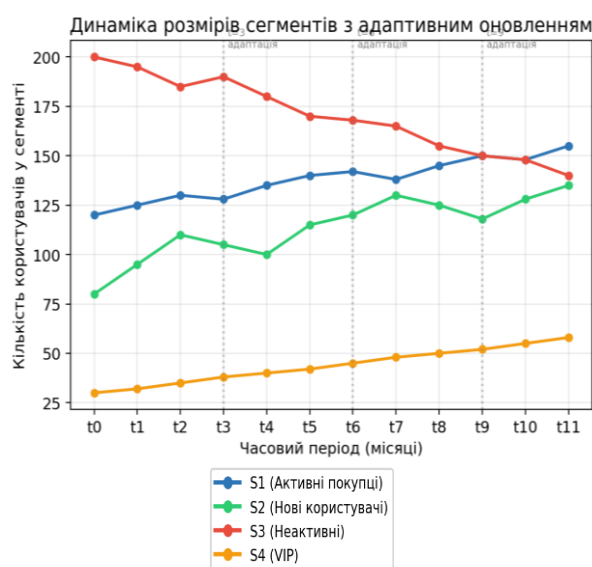


Рисунок 3 – Динаміка розмірів сегментів з адаптивним оновленням

Рисунок 3 ілюструє динаміку розмірів чотирьох сегментів протягом 12 місяців спостереження. Вертикальні пунктирні лінії відповідають моментам адаптивного оновлення центроїдів. Спостерігається поступове зростання сегментів S1 (активні покупці) та S4 (VIP-клієнти), що свідчить про ефективність маркетингових кампаній з утримання. Водночас сегмент S3 (неактивні користувачі) стабільно зменшується. Адаптивний механізм оновлення забезпечує плавну зміну меж сегментів без різких стрибків, що є критичним для стабільності маркетингових кампаній, орієнтованих на конкретні сегменти.

Таблиця 1 – Порівняння методів сегментації (mean ± std, n = 30)

Метод	SC	DB	CH	CTR, %	ROI
К-Means (стат.)	0.58±0.03	1.42±0.08	312±18	3.8±0.4	2.5±0.2
DBSCAN	0.54±0.02	1.58±0.10	285±15	3.5±0.3	2.3±0.2
Ієрархічна	0.56±0.01	1.48±0.05	298±12	3.6±0.3	2.4±0.2
Mini-Batch К-М.	0.55±0.04	1.50±0.09	290±20	3.6±0.4	2.4±0.3
BIRCH	0.57±0.03	1.45±0.07	305±16	3.7±0.3	2.5±0.2
Автоenk. + К-М.	0.62±0.02	1.31±0.06	345±14	4.2±0.3	2.8±0.2
Запропонований	0.72±0.02	1.08±0.05	418±16	5.2±0.4	3.2±0.3

Експериментальне дослідження

Для верифікації запропонованої моделі проведено серію експериментів на реальних даних маркетингових кампаній e-commerce платформи. Набір даних містить 50 000 користувачів з 12-місячною історією взаємодій. Простір ознак включає $d = 18$ параметрів: 8 поведінкових (частота візитів, середній час сесії, кількість переглянутих сторінок, глибина перегляду, кількість покупок, середній чек, частота повернень, активність у e-mail), 5 демографічних (вік, стать, регіон, тип зайнятості, сімейний стан) та 5 контекстних (основний пристрій, переважний час активності, канал залучення, тип браузера, мова інтерфейсу).

Експерименти проведено у середовищі Python 3.11 з використанням бібліотек scikit-learn, NumPy та Pandas. Апаратне забезпечення: персональний комп'ютер з CPU Intel Core i5-10400, 16 GB RAM (обчислювальна складність алгоритму не потребує спеціалізованого серверного обладнання). Для оцінки якості використано метрики: Silhouette Score (SC), Davies-Bouldin Index (DB), Calinski-Harabasz Index (CH). Бізнес-метрики CTR та ROI оцінювалися методом ретроспективного А/В-тестування: для кожного місяця спостережень визначалася сегментація різними методами, після чого на основі історичних даних про відгук користувачів на кампанії обчислювалися показники CTR та ROI для кожного варіанту сегментації. Кожен експеримент повторювався 30 разів з різною ініціалізацією для оцінки статистичної стійкості.

У таблиці 1 наведено порівняння запропонованого методу адаптивної сегментації як із статичними, так і з потоковими (streaming) базовими методами. Результати

представлено у форматі mean ± std за 30 запусків.

Результати експериментів (табл. 1) демонструють статистично значущу перевагу запропонованого методу ($p < 0.001$ за тестом Вілкоксона) над усіма базовими методами, включаючи потокові. Порівняно з найкращим базовим методом (автоенкодер + К-Means) Silhouette Score підвищено на 16.1% (з 0.62 до 0.72), індекс Davies-Bouldin знижено на 17.6% (з 1.31 до 1.08). Порівняно зі статичним К-Means (найпоширенішим у практиці) покращення є ще більш суттєвим: SC зріс на 24.1%, CTR - на 36.8% (з 3.8% до 5.2%), ROI - на 28.0% (з 2.5 до 3.2). Важливо, що запропонований метод перевищує і потокові алгоритми Mini-Batch К-Means та BIRCH, які не враховують дрейф даних.

Рисунок 4 наочно демонструє перевагу запропонованого методу за бізнес-метриками. Запропонований адаптивний підхід забезпечує ROI = 3.2, що на 28% перевищує статичний К-Means (2.5) та на 14.3% - метод на основі автоенкодера (2.8). Аналогічна тенденція спостерігається для CTR: запропонований метод досягає 5.2%, що є найвищим показником серед усіх розглянутих підходів. Варто зазначити, що статичні та потокові методи демонструють близькі результати (ROI 2.3–2.5), тоді як адаптивний підхід забезпечує якісний стрибок за рахунок урахування дрейфу даних.

На рис. 5 представлено динаміку збіжності якості кластеризації, оціненої на даних поточного вікна спостереження. Адаптивна сегментація досягає стаціонарного значення SC ≈ 0.80 після 25 ітерацій оновлення, тоді як статичний метод стабілізу-

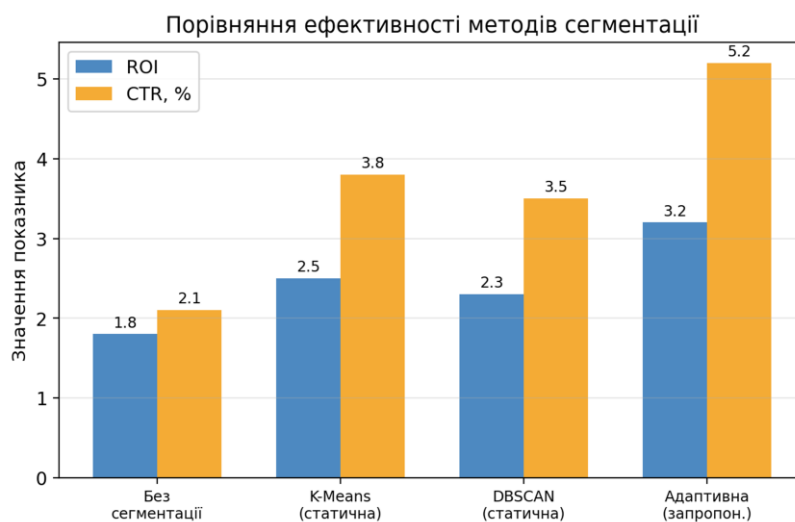


Рисунок 4 – Порівняння ефективності методів сегментації за ROI та CTR

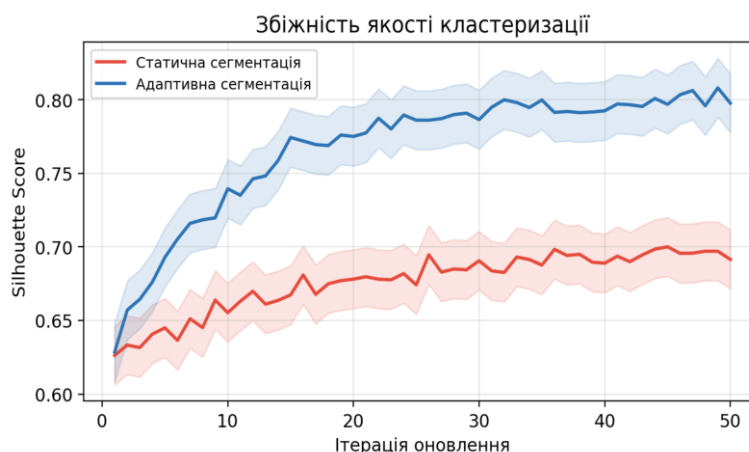


Рисунок 5 – Збіжність якості кластеризації для статичного та адаптивного підходів

ється на рівні $SC \approx 0.70$. Значення SC на рис. 5 є вищими за наведені у табл. 1 ($SC = 0.72$ та 0.58 відповідно), оскільки табл. 1 містить усереднені результати по всьому набору даних за 30 запусків, тоді як рис. 5 відображає SC на даних ковзного вікна для найкращого запуску. Різниця між підходами є статистично значущою ($p < 0.001$ за тестом Вілкоксона).

Рисунок 5 демонструє процес збіжності якості кластеризації для двох підходів. Адаптивна сегментація (синя крива) досягає стаціонарного значення $SC \approx 0.80$ після приблизно 25 ітерацій оновлення. Для статичного підходу (червона крива) на кожній ітерації виконувалася повна перекластеризація K-Means на даних поточного вікна без механізму згладжування центроїдів та без адаптивного коефіцієнта — тобто без урахування попереднього стану кластерів.

Статичний підхід стабілізується на нижчому рівні $SC \approx 0.70$, оскільки кожна перекластеризація не враховує накопичену історію та є чутливою до шуму в поточних даних. Затінені області відповідають довірчим інтервалам ± 0.02 . Різниця між підходами стає статистично значущою ($p < 0.05$) вже після 10-ї ітерації та зростає з часом, що підтверджує накопичувальний ефект адаптації до дрейфу даних.

Аналіз обчислювальної ефективності (табл. 2) підтверджує, що амортизований час одного циклу оновлення для 50 000 користувачів становить 1.2 секунди при використанні підвибірки $n' = 5000$ для оцінки якості з періодом $P = 5$. Повна перебудова кластерів ініціюється в середньому 1 раз на 15–20 циклів, що підтверджує ефективність критерію (12).

Таблиця 2 – Обчислювальна ефективність методів

Метод	Час оновлення, с	Пам'ять, МБ	Складність
K-Means (повне)	38.5	256	$O(nKdI)$
Mini-Batch K-Means	5.8	180	$O(bKdI)$
BIRCH	8.2	210	$O(nBd)$
DBSCAN	125.3	512	$O(n^2d)$
Запропон. (оновл.)	1.2	128	$O(nKd+n^2/P)$

Висновки

У статті розроблено математичну модель адаптивної сегментації користувачів для автоматизованих систем управління маркетинговими кампаніями. Основні результати дослідження полягають у наступному.

Формалізовано простір ознак користувачів як d -вимірний евклідов простір з трьома категоріями ознак (поведінкові, демографічні, контекстні) та запропоновано зважену метрику відстані з автоматичним визначенням вагових коефіцієнтів на основі взаємної інформації.

Розроблено модифікований алгоритм кластеризації з адаптивним оновленням центрів на основі експоненційного згладжування з адаптивним коефіцієнтом, що залежить від формалізованої міри дрейфу даних (статистика Колмогорова–Смірнова). Доведено експоненційну збіжність алгоритму у стаціонарному випадку.

Запропоновано нормалізований комбінований критерій визначення оптимальної кількості кластерів та критерій деградації якості для ініціювання повної перебудови сегментації.

Експериментальне дослідження на реальних даних (50 000 користувачів, 18 ознак, 12 місяців, 30 повторних запусків) підтвердило статистично значущу перевагу запропонованого підходу: підвищення Silhouette Score на 16.1% порівняно з найкращим базовим методом та на 24.1% порівняно зі статичним K-Means; зростання CTR на 36.8% та ROI на 28.0% порівняно зі статичною кластеризацією при часі оновлення 1.2 секунди.

Перспективами подальших досліджень є: строгий аналіз стійкості алгоритму у нестационарному випадку; інтеграція методів глибокого навчання для автоматичного визначення вагових коефіцієнтів ознак; розробка механізму динамічної зміни кількості кластерів без повної перебудови; впровадження федеративного навчання для забезпечення приватності даних та розширення моделі для мультиканальних маркетингових кампаній.

Подяки

Відсутні.

Конфлікт інтересів

Відсутній.

Список використаних джерел

1. McKinsey & Company. The value of personalization — and how to get it right. McKinsey Digital Report. 2023. 18 p.
2. Christy L, Joy, Umamakeswari, Arumugam, Priyatharsini, L., Neyaa, A. RFM Ranking – An Effective Approach to Customer Segmentation. Journal of King Saud University - Computer and Information Sciences. 2021. Vol. 33. № 10. P. 1245-1251. doi: 10.1016/j.jksuci.2018.09.004
3. Perapu, Purushottam. Customer Segmentation Using K-Means Clustering for Personalized Marketing Campaigns. International Journal of Scientific Research in Computer Science Engineering and Information Technology. 2025. Vol. 11. P. 810-815. doi: 10.32628/CSEIT25113344
4. Ram, Anant, Jalal, Sunita, Jalal, Anand, Manoj, Kumar. A Density Based Algorithm for Discovering Density Varied Clusters in Large Spatial Databases. International Journal of Computer Applications. 2010. Vol. 3. doi: 10.5120/739-1038

5. Oti, Eric, Olusola, Michael, Eze, Francis, Enogwe, Samuel. Comprehensive Review of K-Means Clustering Algorithms. *International Journal of Advances in Scientific Research and Engineering*. 2021. Vol. 07. P. 64-69. doi: 10.31695/IJASRE.2021.34050
6. Sculley, D. Web-scale k-means clustering. *Proceedings of the 19th international conference on World wide web*. 2010. P. 1177-1178. doi: 10.1145/1772690.1772862
7. Schikuta, Erich. Grid-Clustering: An efficient hierarchical Clustering method for very large data sets. *Proceedings - International Conference on Pattern Recognition*. 1996. Vol. 2. doi: 10.1109/ICPR.1996.546732
8. Державна служба статистики України. Цифрова економіка та інформаційне суспільство в Україні. Статистичний збірник. Київ, 2024. 98 с.
9. Schubert, Erich, Sander, Jörg, Ester, Martin, Kriegel, Hans, Xu, Xiaowei. DBSCAN revisited, revisited: Why and how you should (still) use DBSCAN. *ACM Transactions on Database Systems*. 2017. Vol. 42. P. 1-21. doi: 10.1145/3068335
10. Cao, Feng, Ester, Martin, Qian, Weining, Zhou, Aoying. Density-Based Clustering over an Evolving Data Stream with Noise. *Proceedings of the 2006 SIAM International Conference on Data Mining*. 2006. doi: 10.1137/1.9781611972764.29
11. Xie, Junyuan, Girshick, Ross, Farhadi, Ali. Unsupervised Deep Embedding for Clustering Analysis. *International Conference on Machine Learning*. 2016. P. 478-487. doi: 10.48550/arXiv.1511.06335
12. Appiah Osei, Derrick, Aliu, Abass. A critical Review of Robust and Adaptive Control Algorithms for Dynamic Environments. *International Journal of Innovative Research in Engineering & Multidisciplinary Physical Sciences*. 2026. Vol. 14. doi: 10.37082/IJIRMP.S.v14.i1.232956
13. Simpkins, Charles. *System Identification: Theory for the User*, 2nd Edition (Ljung, L.; 1999) [On the Shelf]. *Robotics & Automation Magazine, IEEE*. 2012. Vol. 19. P. 95-96. doi: 10.1109/MRA.2012.2192817
14. Биба В. К. Автоматизація прийняття маркетингових рішень. *Журнал стратегічних економічних досліджень*. 2023. № 4 (15). С. 97-110. doi: 10.30857/2786-5398.2023.4.12
15. Попко, Олена, Тивончук, Павло. Стратегічний маркетинговий аналіз ринку IT-послуг в Україні. *Економіка та суспільство*. 2025. № 71. doi: 10.32782/2524-0072/2025-71-58
16. Salvatori, Benedetta, Wegener, Silke, Kotzaeridi, Grammata, Herding, Annika, Eppel, Florian, Dressler-Steinbach, Iris, Henrich, Wolfgang, Piersanti, Agnese, Morettini, Micaela, Tura, Andrea, Göbl, Christian. Identification and validation of gestational diabetes subgroups by data-driven cluster analysis. *Diabetologia*. 2024. Vol. 67. P. 1-15. doi: 10.1007/s00125-024-06184-7
17. Cui, Geng, Wong, Man Leung, Lui, Hon-Kwong. Machine Learning for Direct Marketing Response Models: Bayesian Networks with Evolutionary Programming. *Management Science*. 2006. Vol. 52. P. 597-612. doi: 10.1287/mnsc.1060.0514

References

1. McKinsey & Company. The value of personalization — and how to get it right. *McKinsey Digital Report*. 2023. 18 p.
2. Christy L, Joy, Umamakeswari, Arumugam, Priyatharsini, L., Neyaa, A. RFM Ranking – An Effective Approach to Customer Segmentation. *Journal of King Saud University - Computer and Information Sciences*. 2021. Vol. 33. № 10. P. 1245-1251. doi: 10.1016/j.jksuci.2018.09.004
3. Perapu, Purushottam. Customer Segmentation Using K-Means Clustering for Personalized Marketing Campaigns. *International Journal of Scientific Research in Computer Science Engineering and Information Technology*. 2025. Vol. 11. P. 810-815. doi: 10.32628/CSEIT25113344

4. Ram, Anant, Jalal, Sunita, Jalal, Anand, Manoj, Kumar. A Density Based Algorithm for Discovering Density Varied Clusters in Large Spatial Databases. *International Journal of Computer Applications*. 2010. Vol. 3. doi: 10.5120/739-1038
5. Oti, Eric, Olusola, Michael, Eze, Francis, Enogwe, Samuel. Comprehensive Review of K-Means Clustering Algorithms. *International Journal of Advances in Scientific Research and Engineering*. 2021. Vol. 07. P. 64-69. doi: 10.31695/IJASRE.2021.34050
6. Sculley, D. Web-scale k-means clustering. *Proceedings of the 19th international conference on World wide web*. 2010. P. 1177-1178. doi: 10.1145/1772690.1772862
7. Schikuta, Erich. Grid-Clustering: An efficient hierarchical Clustering method for very large data sets. *Proceedings - International Conference on Pattern Recognition*. 1996. Vol. 2. doi: 10.1109/ICPR.1996.546732
8. Derzhavna sluzhba statystyky Ukrainy. Tsyfrova ekonomika ta informatsijne suspilstvo v Ukraini. *Statystychnyj zbirnyk*. Kyiv, 2024. 98 s. [in Ukrainian]
9. Schubert, Erich, Sander, Jörg, Ester, Martin, Kriegel, Hans, Xu, Xiaowei. DBSCAN revisited, revisited: Why and how you should (still) use DBSCAN. *ACM Transactions on Database Systems*. 2017. Vol. 42. P. 1-21. doi: 10.1145/3068335
10. Cao, Feng, Ester, Martin, Qian, Weining, Zhou, Aoying. Density-Based Clustering over an Evolving Data Stream with Noise. *Proceedings of the 2006 SIAM International Conference on Data Mining*. 2006. doi: 10.1137/1.9781611972764.29
11. Xie, Junyuan, Girshick, Ross, Farhadi, Ali. Unsupervised Deep Embedding for Clustering Analysis. *International Conference on Machine Learning*. 2016. P. 478-487. doi: 10.48550/arXiv.1511.06335
12. Appiah Osei, Derrick, Aliu, Abass. A critical Review of Robust and Adaptive Control Algorithms for Dynamic Environments. *International Journal of Innovative Research in Engineering & Multidisciplinary Physical Sciences*. 2026. Vol. 14. doi: 10.37082/IJIRMPS.v14.i1.232956
13. Simpkins, Charles. System Identification: Theory for the User, 2nd Edition (Ljung, L.; 1999) [On the Shelf]. *Robotics & Automation Magazine, IEEE*. 2012. Vol. 19. P. 95-96. doi: 10.1109/MRA.2012.2192817
14. Byba V. K. Avtomatyzatsiia pryiniattia marketynhovykh rishen. *Zhurnal stratehichnykh ekonomichnykh doslidzen*. 2023. № 4 (15). S. 97-110. doi: 10.30857/2786-5398.2023.4.12 [in Ukrainian]
15. Popko, Olena, Tyvonchuk, Pavlo. Stratehichni marketynhovyi analiz rynku IT-posluh v Ukraini. *Ekonomika ta suspilstvo*. 2025. № 71. doi: 10.32782/2524-0072/2025-71-58 [in Ukrainian]
16. Salvatori, Benedetta, Wegener, Silke, Kotzaeridi, Grammata, Herding, Annika, Eppel, Florian, Dressler-Steinbach, Iris, Henrich, Wolfgang, Piersanti, Agnese, Morettini, Micaela, Tura, Andrea, Göbl, Christian. Identification and validation of gestational diabetes subgroups by data-driven cluster analysis. *Diabetologia*. 2024. Vol. 67. P. 1-15. doi: 10.1007/s00125-024-06184-7
17. Cui, Geng, Wong, Man Leung, Lui, Hon-Kwong. Machine Learning for Direct Marketing Response Models: Bayesian Networks with Evolutionary Programming. *Management Science*. 2006. Vol. 52. P. 597-612. doi: 10.1287/mnsc.1060.0514

MATHEMATICAL MODEL OF ADAPTIVE USER SEGMENTATION IN AUTOMATED MARKETING CAMPAIGN MANAGEMENT SYSTEMS

Korzhyk M. V.

Candidate of Technical Sciences, Associate Professor
National Technical University of Ukraine «Igor Sikorsky Kyiv Polytechnic Institute»
03056, Beresteyskyi Avenue, 37, Kyiv, Ukraine
<https://orcid.org/0000-0003-3453-6258>
e-mail: korzhyk@kpi.ua

Bokushev E. N.

Senior Lecturer of the Department
National Technical University of Ukraine «Igor Sikorsky Kyiv Polytechnic Institute»
03056, Beresteyskyi Avenue, 37, Kyiv, Ukraine
<https://orcid.org/0009-0009-6725-3330>
e-mail: bokushev@gmail.com

Dehtyaruk K. B.

Postgraduate Student, Research Fellow
National Technical University of Ukraine «Igor Sikorsky Kyiv Polytechnic Institute»
03056, Beresteyskyi Avenue, 37, Kyiv, Ukraine
<https://orcid.org/0009-0001-3777-6277>
e-mail: kostiadekhtiaruk@gmail.com

Abstract. The article develops a mathematical model of adaptive user segmentation for automated marketing campaign management systems focused on working in a dynamic information environment and continuous data flow. The relevance of the study is due to the need to increase the effectiveness of marketing decisions by promptly taking into account changes in consumer behavior and automated formation of relevant target audiences. A modified clustering algorithm is proposed that takes into account the temporal dynamics of user behavioral characteristics and provides automatic updating of segments in real time without the need for a complete restructuring of the cluster structure. A multidimensional space of user features is formalized, including demographic, behavioral, and transactional parameters. To ensure correct processing of heterogeneous data, procedures for normalization and scaling of features are proposed that allow increasing the stability of the algorithm to the heterogeneity of input information. The paper defines criteria for assessing the quality of segmentation based on the Silhouette index and the modified Davis–Boldin index, which provide a comprehensive analysis of the internal homogeneity of clusters and the level of their separation. A segment adaptation mechanism using a sliding window and an adaptive smoothing coefficient has been developed, which allows for effective tracking of data drift, timely response to changes in user behavior, and long-term relevance of formed segments. A comparative analysis of the proposed approach with common clustering methods, including K-Means, DBSCAN, and Gaussian Mixture Models, has been conducted. The results obtained indicate the advantages of the developed algorithm in terms of segmentation accuracy, cluster structure stability, and computational efficiency when working with large streams of marketing data.

Keywords: adaptive segmentation, clustering, stream data processing, machine learning, automated management system, marketing campaigns, personalization.