



✉ Correspondence author

O. O. Narushynska
narushynska@gmail.com

Article received 23.10.2023 p.

Article accepted 26.10.2023 p.

UDK 004.75.62

УДК 004.75.62

О. О. Нарушинська, В. І. Мотрунич, М. В. Арзубов, В. М. Теслюк

Національний університет "Львівська політехніка", м. Львів, Україна

ІНФОРМАЦІЙНА ТЕХНОЛОГІЯ ДЛЯ АНАЛІЗУ ПУНКТІВ ПРОДАЖУ МОБІЛЬНОГО ОПЕРАТОРА НА ОСНОВІ МЕТОДІВ КЛАСТЕРИЗАЦІЇ

Висвітлено дослідження, спрямоване на розроблення та впровадження інформаційної технології моніторингу та аналізу сегментів пунктів продажу мобільного оператора з використанням методів кластеризації. Дослідження вирішує актуальну проблему в сфері маркетингу та бізнес-оптимізації, а саме: покращення стратегій мережі пунктів продажу мобільного зв'язку.

У дослідницькій роботі запропоновано новітній підхід до сегментації пунктів продажу мобільного оператора з використанням алгоритмів кластеризації. Розроблено програмний продукт, що містить алгоритми машинного навчання для кластеризації пунктів продажу за ключовими параметрами. Здійснено комплексний аналіз бази даних мобільного оператора для визначення критичних характеристик пунктів продажу, таких як прибутковість, патерни використання послуг мобільного оператора, кількість нових та втрачених клієнтів, географічне розташування та інші важливі індикатори.

Особливу увагу звернено на розроблення інструментарію для підготовки та опрацювання вхідних даних, що забезпечує точність подальшої кластеризації. За допомогою створеного продукту мобільний оператор зможе ідентифікувати найприбутковіші пункти продажу, виявити можливості для зростання та розробити цільові стратегії для кожного сегмента.

Застосовуючи розроблену технологію, мобільний оператор отримує можливість не тільки виявити ключові та прибуткові точки продажів, але й розробити цілеспрямовані стратегії для різних груп пунктів продажу з урахуванням їх унікальних характеристик. Такий підхід підсилює позиції компанії на ринку, сприяючи підвищенню задоволеності споживачів та збільшенню прибутків.

Крім того, розроблена система має потенціал для аналізу динаміки зміни пунктів продажу з часом, що дає змогу мобільному оператору підтримувати оптимальну стратегію розвитку та задовольняти потреби клієнтів, що може сприяти підвищенню прибутковості та конкурентоспроможності бізнесу.

Ключові слова: машинне навчання; навчання без вчителя; опрацювання даних, К-середніх; патерни поведінки користувачів.

Вступ / Introduction

У сучасному світі маркетинг і бізнес-оптимізація є надзвичайно важливими сферами, особливо в галузі мобільного зв'язку. Зважаючи на постійне зростання конкуренції та зміну уподобань клієнтів, операторам мобільного зв'язку доводиться шукати нові способи для поліпшення якості своїх послуг та оптимізації мережі пунктів продажу [6]. Однією із ключових проблем, які постають перед мобільними операторами, є ефективна сегментація пунктів продажу на основі їх характеристик та значень.

Сучасні інформаційні технології, зокрема методи кластеризації, надають можливість вирішити цю проблему та створити інструмент, який допоможе операторам мобільного зв'язку аналізувати та оптимізувати свою мережу пунктів продажу. Проте, незважаючи на активний розвиток цих технологій, деякі питання залишаються відкритими та потребують подальших досліджень та розроблення [8].

Актуальність цього дослідження полягає у тому, що в сучасних умовах важливо знати, як класифікувати та групувати пункти продажу для реалізації стратегій,

спрямованих на збільшення прибутковості та задоволення потреб клієнтів. Застосування кластерного аналізу в цьому контексті є кроком до оптимізації мережі пунктів продажу мобільного зв'язку та поліпшення обслуговування клієнтів.

Об'єкт дослідження – процеси моніторингу та аналізу сегментів пунктів продажу мобільного оператора на основі методів кластеризації.

Предмет дослідження – методи, моделі та засоби моніторингу та аналізу сегментів пунктів продажу мобільного оператора на основі методів кластеризації.

Мета роботи – підвищення ефективності моніторингу та аналізу сегментів пунктів продажу мобільного оператора на основі методів кластеризації.

Для досягнення зазначеної мети поставлено такі основні завдання дослідження:

- аналіз актуальних досліджень та публікацій у галузі маркетингу та бізнес-оптимізації, зокрема в контексті мобільного зв'язку;
- дослідження впливу методів кластеризації для моніторингу та аналізу сегментів пунктів продажу мобільного оператора. Важливо визна-

чити, які методи є найефективнішими для досягнення поставленої мети;

- вибір оптимальних підходів та методів для розроблення системи моніторингу та аналізу. Це передбачає розгляд різних підходів до сегментації пунктів продажу та вибір найефективніших для поставленого завдання;
- розроблення базової структури системи моніторингу та аналізу сегментів пунктів продажу мобільного оператора. Цей етап передбачає створення основної архітектури програмного продукту та визначення його ключових функціональних елементів.

Отже, завдання цього дослідження охоплює комплексний аналіз та розроблення інструментів для ефективного моніторингу та аналізу сегментів пунктів продажу мобільного оператора з використанням методів кластеризації.

Розроблені засоби інформаційних технологій дадуть змогу оператору мобільного зв'язку розподіляти пункти продажу на кластери за спільними ознаками та аналізувати їхню продуктивність. Розроблена система дасть змогу відстежувати зміну пунктів продажу з часом і адаптувати стратегію розвитку мережі для забезпечення задоволеності клієнтів та підвищення конкурентоспроможності.

Отже, відповідне дослідження має велике значення для сфери маркетингу та розвитку мобільних операторів, оскільки результати дослідження можуть сприяти підвищенню ефективності та прибутковості бізнесу в цій галузі.

Матеріали і методи дослідження. У роботі використано аналітичні методи та інструменти для ведення дослідження та аналізування пунктів продажу мобільного оператора. Для аналізу опрацьовано дані, отримані від мобільного оператора, що містять інформацію про споживацьку активність клієнтів, їхні платежі, використання послуг, а також географічні дані про розташування пунктів продажу.

Для кластеризації пунктів продажу застосовано метод кластерного аналізу, зокрема, алгоритм К-середніх (K-means) [1]. Цей метод дає змогу групувати пункти продажу на основі активності споживачів та інших характеристик.

Для оцінювання результатів кластеризації та моделювання використано метрику індексу Данна та низку статистичних показників [10].

Аналіз останніх досліджень та публікацій. Вивчення споживацької поведінки та кластеризація роздрібних пунктів продажу – актуальні напрями в сучасному дослідженні та бізнес-аналітиці [6]. Попередні дослідження доводять, що споживацька поведінка відіграє критичну роль у визначенні успіху роздрібної торгівлі [7]. Аналіз цін на товари та їх вплив на прибуток пунктів продажу стає ключовим елементом стратегічного планування. Високий рівень еластичності цін на окремі товари свідчить про необхідність диференціації стратегій ціноутворення залежно від пріоритетів споживачів [8].

Застосування кластерного аналізу К-середніх [10] у контексті сегментації клієнтів мобільних операторів дає підставу для висновку, що цей метод можна ефективно

застосовувати для групування клієнтів за їхніми звичками використання, уподобаннями та поведінкою, щоб краще розуміти їхні потреби і налагодити відповідні маркетингові стратегії [5].

У контексті мобільних операторів виділено важливі ознаки та змінні, які необхідно використовувати для сегментації клієнтів, такі як тривалість дзвінків, використання даних, місцезнаходження та демографічні дані клієнтів [4]. Обговорюються кроки попереднього опрацювання, необхідні для підготовки даних до кластерного аналізу, включаючи нормалізацію або стандартизацію [1], [10].

Стаття А. Т. Rosario та співавторів пропонує новий підхід до кластеризації пунктів продажу, із основним акцентом на споживацькій поведінці, а не лише на місцевій конкуренції [8]. Цей підхід враховує еластичність цін на товари та дохід за товар на пункт продажу для кластеризації “схожих” пунктів продажу. Результати дослідження підкреслюють вплив конкуренції, характеристик пунктів продажу та даних про торговельну зону на поведінку споживача.

Важливим аспектом кластерного аналізу є інтерпретованість результатів [2], [4]. Методологією SHAP можна скористатись для визначення важливості ознак у кожному кластері, що допоможе розуміти особливості кластерів та забезпечити краще розуміння даних.

Проаналізувавши дослідження стосовно моніторингу та аналізу сегментів пунктів продажу мобільних операторів, можна стверджувати, що попередні дослідження вказують на важливість аналізу споживацької поведінки та інтерпретованості результатів кластеризації для оптимізації роздрібної торгівлі [9]. І навіть більше, розуміння цих аспектів стає критичним для розвитку стратегій ціноутворення та управління пунктами продажу в сучасному бізнесі.

Результати дослідження та їх обговорення / Research results and their discussion

Під час розроблення програмних засобів кластеризації пунктів продажу мобільного оператора виникла необхідність у виборі оптимального алгоритму кластеризації, який би не лише відповідав специфіці досліджуваної задачі з дослідницького погляду, але й задовольняв вимоги бізнес-використання. З огляду на різноманітність алгоритмів кластеризації, ми врахували декілька можливих підходів до групування пунктів продажу за спільними характеристиками.

1. Алгоритм К-середніх є одним із найпоширеніших методів кластеризації. Основною перевагою алгоритму К-середніх є його ефективність для оброблення великих наборів даних та легке визначення кількості кластерів. Однак важливо враховувати його чутливість до початкових умов, можливість отримання різних результатів під час кожного запуску і обмеженість в роботі з несферичними кластерами [7].
2. Ієрархічна кластеризація використовує ієрархічну структуру для групування об'єктів. Вона має перевагу щодо можливості візуалізації результатів у вигляді дерева та гнучкості вибору кількості кластерів. Проте вона менш ефективна для великих наборів даних та потребує значних обчислювальних ресурсів, якщо об'єктів багато [3].

3. DBSCAN(Density-Based Spatial Clustering of Applications with Noise) ґрунтується на густоті об'єктів у просторі даних. Він відрізняється від інших алгоритмів здатністю виявляти кластери будь-якої форми та виділяти шумові точки. Однак його недоліком є необхідність коректно налаштувати параметри та чутливість до масштабування даних [6].

Кожен із цих алгоритмів має певні переваги та обмеження у контексті кластеризації пунктів продажу мобільного оператора. Вибір оптимального алгоритму визначається характеристиками даних та бізнес-потребами дослідження. Наприклад, алгоритм К-середніх може бути ефективним для швидкого групування великих обсягів даних, тоді як ієрархічна кластеризація може бути корисною для виявлення ієрархічної структури пунктів продажу, а DBSCAN – для виділення різних форм кластерів та ідентифікації шуму.

Вхідні дані програми містять інформацію про щоденну активність користувачів, таку як кількість виконаних дзвінків, відправлених повідомлень, використаних даних тощо. Крім того, в програмі використовуються дані про надання додаткових послуг, які можуть придбати користувачі. Також враховуються дані про тарифні плани, обсяги передавання даних, кількість хвилин розмов, кількість повідомлень та інші відповідні параметри.

Ці вхідні дані важливі для аналізу та кластеризації пунктів продажу мобільного оператора. Вони дають змогу встановити зв'язок між патерном поведінки користувача і популярністю пунктів продажу, а також зрозуміти, які фактори впливають на вибір тарифних планів або використання додаткових послуг. Загальна інформація про користувачів і їх активність дає змогу виявити закономірності та тренди у використанні послуг мобільного оператора, що може бути корисним під час планування маркетингових стратегій та вдосконалення пропозицій для клієнтів.

Щоб отримати необхідні дані про пункти продажу, потрібно правильно опрацювати вхідні дані, а

саме здійснити агрегацію цих даних. Спочатку потрібно зробити користувачів унікальними, додавши характеристики за кожен день. Після того потрібно згрупувати дані за характеристикою, яка являє собою назву пункту продажу, в якому відбулась операція. Відповідно, у результаті можна отримати необхідні дані для кластеризації пунктів продажу.

Одним із важливих етапів аналізування даних є їх попереднє опрацювання для видалення шуму, пропущених значень та аномалій. Це допомагає підготувати дані для подальшого аналізу та забезпечує високу якість моделей і висновків.

Перший крок – визначення шуму та аномалій у наборі даних. Це можна здійснити за допомогою аналізування розподілу даних, виявлення відхилень від типових значень або використання статистичних методів. Також варто звернути увагу на логічні аспекти бізнесу, оскільки аномалія може не бути виявленою з використанням статистики та все ж не задовольнятиме вимоги бізнес-процесів.

Після виявлення шуму, пропущених значень і аномалій необхідно видалити їх або замінити в наборі даних. Це може передбачати видалення окремих спостережень або використання методів згладжування, таких як фільтри або методи згортки, для зменшення впливу шуму. Ми використали заміну на моду та медіану під час аналізу неагрегованих даних.

Результатом роботи алгоритму є дані про кожен пункт продажу та визначений для нього номер кластера, назва та його короткий опис. Для результатів створюється окрема таблиця з чітко визначеними характеристиками для запису, тому потрібно мати всі характеристики для кластеризації.

На рис. 1 зображено структуру розробленої бази даних. Вона містить три таблиці, в яких вказано дані про користувачів, їх платежі та дані про пункти продажу, в яких здійснено платіж.

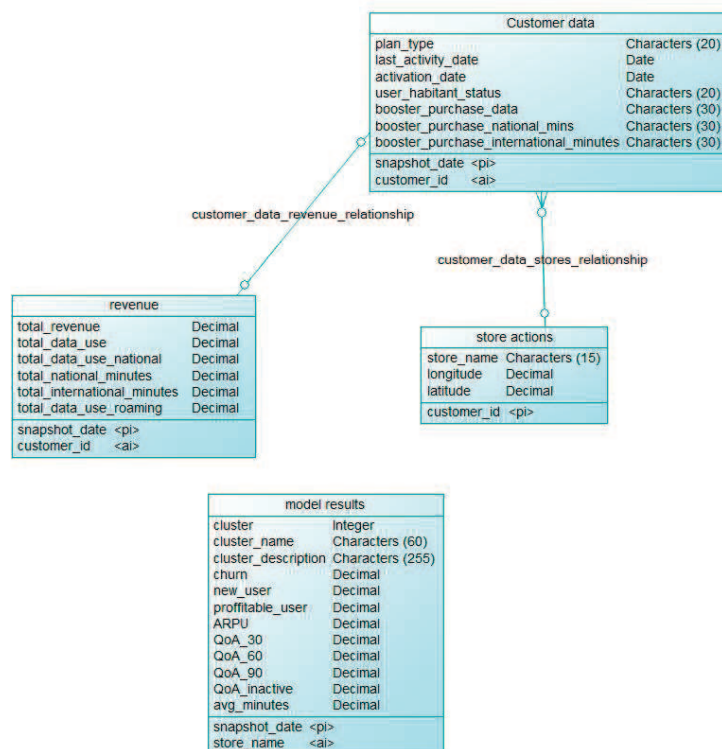


Рис. 1. Схема розробленої бази даних / The schema of the developed database

У таблиці Customer data містяться загальні дані про користувачів, а саме дата активації, дата останньої активності, тип підключеного плану тощо. Дані в цій таблиці оновлюються щодня, тому, щоб досягти унікальності за ідентифікаторами користувача, треба агрегувати дані з кожного дня місяця.

Подібна структура таблиці revenue, яка містить дані про прибутковість користувачів, використання хвилини та мобільного інтернету, а у таблиці store actions вказано назви пунктів продажу, в яких була здійснена операція, та координати цих пунктів продажу.

Розроблено програмне забезпечення, зокрема діаграму класів (рис. 2), яка охоплює:

- клас Тренування, який надасть змогу тренувати та за потреби перетреноувати модель;
- клас Передбачення, що виконує кластеризацію даних використовуючи вже серіалізовану модель;
- клас Підготовка даних, створений для перетворень та підготовки даних до операції кластеризації;

- клас Трансформер – клас-обгортка класу нормалізації, для можливості серіалізувати її.

Для кращого розуміння структури програмного продукту розроблено схему взаємодії моделі з іншими середовищами для її коректного застосування. На рис. 3 зображено схему роботи моделі в хмарному середовищі. Docker контейнер містить натреновану модель сегментації. Кожного першого дня місяця запускається AWS (Amazon Web Services) CloudWatch тригер, він передає на виконання скрипт у Docker контейнері для сегментації даних. Вхідні дані автоматично вибираються з бази даних Vertica, опрацьовуються та трансформуються у набір даних з відповідними ознаками для використання моделлю. Якщо під час використання скрипта сегментації відбувається збій, система надсилає повідомлення відповідальним особам з використанням AWS SES (Simple Email Service). Якщо сегментація відбулась успішно, результати будуть збережені у базу даних Vertica та додатково, у форматі csv, у файлому сховищі S3.

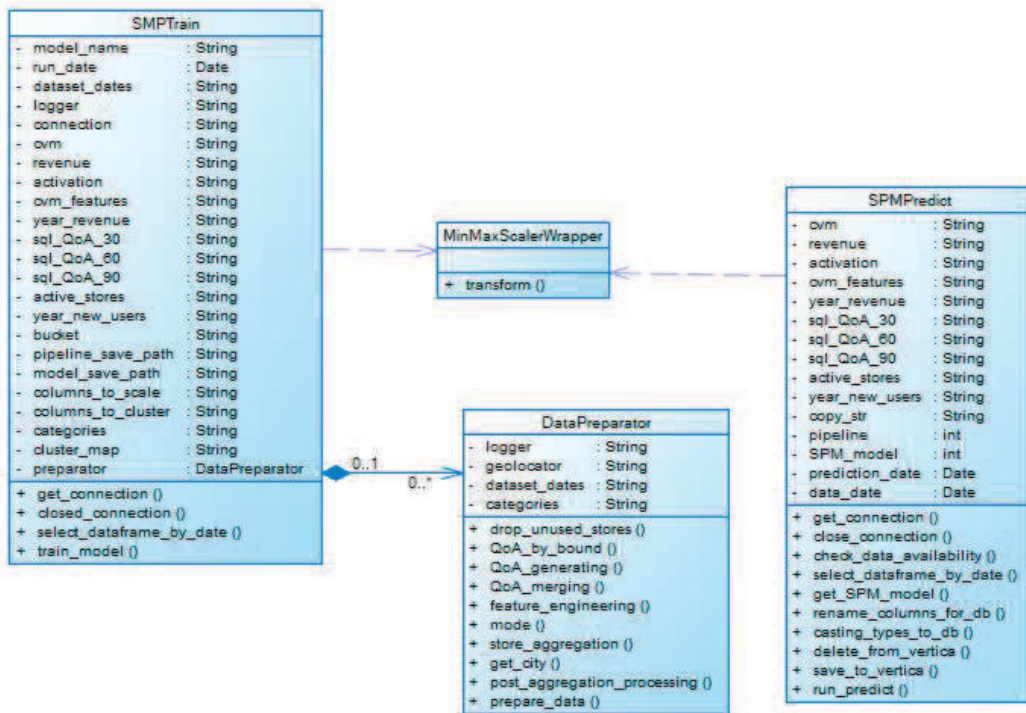


Рис. 2. Діаграма класів розробленого програмного засобу / Class diagram of the developed software

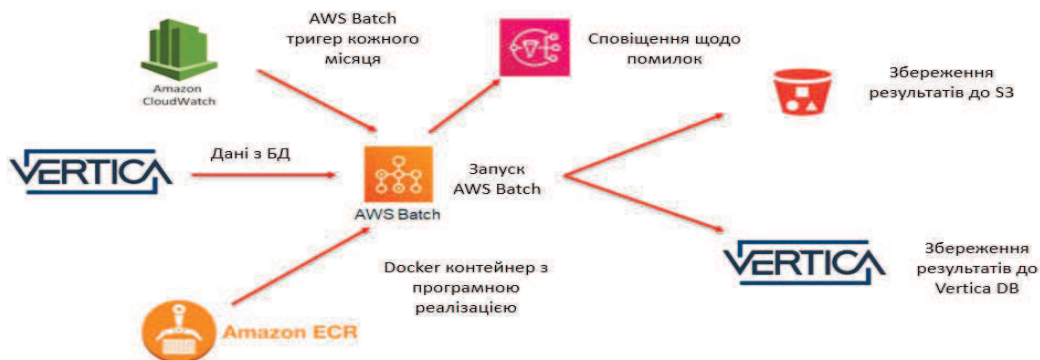


Рис. 3. Схема функціонування моделі в хмарі / Functioning diagram of the cloud-based model

Під час дослідження ми вирішили розробити декілька варіантів кластеризації, щоб можна було порівняти результати та вибрати той, який найкраще підходить для вирішення бізнес-проблеми.

Основні характеристики пунктів продажу, які згенеровано на основі даних про користувачів:

- Середня кількість покупців на пункт продажу.
- Відсоток постійних жителів серед користувачів.
- Відсоток туристів серед користувачів.
- Середній вік (стаж) у мережі в днях.
- Середній місячний дохід.
- Нові платежі (середньомісячні).
- Середня кількість вхідних хвилин за місяць для користувачів пункту продажу.
- Середня кількість вихідних хвилин за місяць для користувачів пункту продажу.
- Середнє використання даних у гігабайтах за місяць для користувачів пункту продажу.
- Відсоток прибуткових користувачів.

- Відсоток прибуткових постійних жителів.
- Середній дохід на одного користувача.
- Відсоток відтоку клієнтів.
- Відсоток відтоку постійних жителів.
- Відсоток відтоку туристів.

Ці характеристики використовують для аналізу різних аспектів продуктивності пунктів продажу, в контексті мобільних операторів або компаній, які надають телекомунікаційні послуги.

Розглянемо результати кластеризації на семи кластерах. Алгоритм поділив дані на три великі кластери, один набагато менший та три найменші. Це свідчить, що алгоритм виділив тільки три великих кластери, а інші було виділено як найвіддаленіші від цих кластерів входження. Така кластеризація не забезпечить практичного і корисного застосування, оскільки більшість пунктів продажу входять у великі кластери й об'єднані на загальному рівні (табл. 1).

Табл. 1. Опис сегментації на сім кластерів / Description of segmentation into 7 clusters

filter	0	1	2	3	4	5	6
Clusters	0	1	2	3	4	5	6
base	83042	9006	87464	92190	70088	74228	66768
shop count	8	23	3	2	12	4	29
churn_mean	7%	5%	6%	18%	5%	12%	4%
churn_resident_mean	2%	3%	2%	1%	3%	1%	3%
churn_tourist_mean	5%	2%	4%	17%	2%	11%	2%
new_user_mean	10%	23%	9%	23%	8%	18%	8%
new_user_Resident_mean	3%	22%	3%	0%	5%	2%	6%
new_user_tourist_mean	6%	1%	6%	22%	3%	17%	3%
activation_day_mode	10.25	7.04	11.33	6.00	10.33	5.00	10.76
network_age_mean	578.85	386.54	661.83	228.84	673.49	347.60	580.96
total_count_sms_mean	6.71	7.28	5.94	1.88	7.22	3.07	7.94
total_mins_incoming_mean	298.24	320.68	229.21	44.97	351.14	101.53	350.28
total_data_usage_transfer_mean	0.30	0.32	0.31	0.27	0.30	0.23	0.36
total_data_usage_mean	8.43	7.13	10.00	5.15	7.45	7.14	8.60
total_minutes_outgoing_mean	456.91	540.38	295.49	53.74	547.69	124.20	556.12
total_revenue_mean	240.19	198.27	280.45	205.22	234.85	209.78	242.34

Ця таблиця використовує колірне кодування для відображення даних, кольори представляють різні діапазони значень для кожного показника. Ось опис колірного кодування:

- Червоний колір ідентифікує значення або відсотки, що можуть вказувати на негативні тенденції (наприклад, високий відсоток відмови).
- Зелений колір відображає значення або відсотки, що можуть бути позитивним індикатором (наприклад, високе значення прибутковості).
- Жовтий колір зазвичай використовується для позначення середніх значень, які можуть потребувати уваги, але не є критичними.

Ці кольори зазвичай вибирають для того, щоб швидко звернути увагу на показники, які вимагають втручання або подальшого аналізу.

Результати кластеризації на п'ять кластерів подано в табл. 2. Пункти продажу розділено на два великі сегменти, один середній та два менші, але все ще достатньо великі, щоб бути повноцінними кластерами та містити корисну інформацію. З опису отриманих кластерів чітко видно розподіл на найприбутковіший кластер, кластер середньої прибутковості, кластери нижчої

від середньої прибутковості та кластер з найнижчою прибутковістю. Саме такого розподілу очікував клієнт. Це означає, що такий результат підходить найкраще, тому в кінцевому випадку ми вирішили брати за основу кластеризацію на п'ять кластерів.

Індекс Данна цього варіанта кластеризації дорівнює 0,685, а це набагато краще порівняно з попереднім варіантом. Тому вибрано саме цей варіант для продовження розроблення та дослідження.

Наступним експериментом є тестування системи на нових даних. Для тестування вибрано найактуальніші дані на момент розроблення системи. Результатом такої кластеризації був розподіл, схожий на початковий розподіл, на якому було навчено модель (див. табл. 3).

Отримані результати дають підставу стверджувати, що результат навчання моделі успішно застосовується на раніше не відомих даних та зберігає свою загальну структуру. Кластери мають приблизно таку саму кількість пунктів продажу, за винятками нових, які могли з'явитися в проміжок часу між старими та новими даними.

Індекс Данна (I) цієї кластеризації майже не змінився і має значення 0,614. Це означає стабільний результат, що і потрібно від цієї кластеризації.

Табл. 2. Опис сегментації на п'ять кластерів / Description of segmentation into 5 clusters

	A	B	C	D	E	F
1						
2	Cluster	1	2	3	4	5
3	Shop count	30	24	6	14	5
4	Customer count	174012	174418	10716	122176	1358
5	avg customer count per store	5800.4	7267.4166666667	1786	8726.8571428571	271.6
6	avg age on network	518.87	669.42	169.46	522.17	474.06
7	avg minutes incoming	333.76	398.67	138.81	241.81	354.31
8	avg minutes outgoing	543.38	637.31	215.38	342.37	579.33
9	avg data usage (GB)	7.78	10.74	5.45	5.07	8.90
10	avg SMS count	6.50	8.77	14.12	4.69	3.67
11	ARPU	232.53	281.09	142.63	188.90	219.31
12	churn%	5%	3%	9%	9%	4%
13	churn resident%	3%	2%	3%	3%	3%
14	churn tourist%	2%	1%	6%	6%	1%
15	new user%	9%	7%	23%	12%	23%
16	new user Resident%	6%	5%	20%	4%	22%
17	new user tourist%	4%	2%	3%	9%	1%
18	QoA inactive%	2%	2%	2%	3%	4%
19	QoA 30 days active%	92%	89%	69%	78%	74%
20	QoA 60 days active%	78%	87%	73%	70%	74%
21	QoA 90 days active%	74%	82%	62%	55%	60%
22	profitable users%	74%	83%	42%	72%	49%
23	users with paid plan 6-12 Months%	50%	60%	29%	29%	50%
24	users with paid plan Monthly%	12%	11%	21%	13%	20%
25	users with paid plan PAYG%	36%	27%	47%	52%	29%
26	users with paid plan VIP%	2%	2%	2%	6%	1%
27	total resident users%	83%	90%	80%	64%	97%
28	total tourist users%	17%	10%	20%	36%	3%
29						

Табл. 3. Розподіл пунктів продажу та користувачів нових даних по кластерах / Distribution of stores and new user data across clusters

Cluster	1	2	3	4	5	Internet orders
Shop count	34	22	7	19	8	1
Customer count	148,640	205,250	10,796	169,728	8,340	145676
avg customer count per store	4,372	9,330	1,542	8,933	1,043	145676

Індекс Данна – це метрика, яку використовують для оцінювання якості кластеризації.

Формула індексу Данна виглядає так:

$$D = \frac{\min_{1 \leq i < j \leq k} \delta(C_i, C_j)}{\max_{1 \leq l \leq k} \Delta(C_l)}, \quad (1)$$

де $\min_{1 \leq i < j \leq k} \delta(C_i, C_j)$ – мінімальна відстань між кластерами; $\max_{1 \leq l \leq k} \Delta(C_l)$ – максимальний діаметр кластера.

Це найбільша відстань між двома елементами всередині одного кластера. Це міра компактності кластера.

Обидва експерименти підтвердили, що вибрана модель кластеризації ефективна та може успішно застосовуватися для аналізу пунктів продажу мобільного оператора. Ці результати підкреслюють значення правильного підбору кількості кластерів та роботи з актуальними даними для досягнення найкращих результатів у кластеризаційному аналізі.

Після реалізації програмного забезпечення, розробленого під час цього дослідження, отримано такі результати.

Згідно з результатами, поданими в табл. 4, алгоритм кластеризації розділив дані на п'ять кластерів. Один з цих кластерів великий, два менші, ще два – малі.

Це підтверджує успішне виконання кластеризації, оскільки дані поділено на групи різних розмірів. Відсутність великої кількості аномалій або відхилень

свідчить про ефективність алгоритму та правильність вибору характеристик для кластеризації пунктів продажу мобільного оператора.

Також успіх кластеризації підтверджує значення метрики індексу Данна, яка показує, що велика частина даних є гомогенними та добре розподіленими.

Обговорення результатів дослідження. Отримані результати уможливають подальший аналіз кожного кластеру окремо, виявлення закономірностей та залежності між пунктами продажу в межах кожного кластера. Такий підхід може допомогти в прийнятті рішень щодо оптимізації роботи пунктів продажу та впровадженні стратегій, спрямованих на поліпшення якості обслуговування клієнтів та збільшення прибутковості мобільного оператора.

У табл. 4 кластери розмежовано за кожною з характеристик.

За допомогою табл. 4 можна чітко визначити різні сегменти, які відповідають різним рівням прибутковості. Розмежування відбулося успішно, оскільки значення характеристик кожного кластера відрізняються, вказуючи на його зарахування до певної категорії. Це підтверджує ефективність здійсненої кластеризації та її здатність виділяти прибуткові та менш прибуткові сегменти на основі аналізу відповідних характеристик.

Табл. 4. Опис характеристик сегментів пунктів продажу / Description of Store Segment Characteristics

Cluster	1	2	3	4	5	Internet orders
Shop count	34	22	7	19	8	1
Customer count	148,640	205,250	10,796	169,728	8,340	145676
avg customer count per store	4,372	9,330	1,542	8,933	1,043	145676
total resident users%	92.4%	93.0%	87.2%	60.7%	99.9%	99.8%
total tourist users%	7.6%	7.0%	12.8%	39.3%	0.1%	0.2%
avg age on network, days	785.30	878.64	211.80	544.11	173.57	1163.72
2023 : GA's (average / month)	8884	10380	1494	22713	1202	5550
Mean of the cluster (average / store / month)	261	472	213	1195	150	5550
2023 : Resident GA's (average / month)	5417	5539	262	1541	1194	5545
Mean of the cluster : Resident (average / store / month)	159	252	37	81	149	5545
2023 : Tourist GA's (average / month)	3467	4841	1232	21172	9	6
Mean of the cluster : Resident (average / store / month)	102	220	176	1114	1	6
Average Monthly Revenue (2022)	34,488,717.33	56,285,461.27	1,561,277.18	31,582,876.45	1,772,380.37	34,197,209.72
New Payments (Monthly avg 2022)	4,485,100.91	4,845,633.33	429,852.78	5,312,806.67	837,308.57	4,436,700.00
avg minutes incoming	351.84	341.72	164.26	184.65	209.35	293.28
avg minutes outgoing	584.72	525.92	251.19	280.72	318.91	402.15
avg data usage (GB)	9.69	14.17	7.40	6.47	16.78	11.05
profitable users%	80.7%	85.7%	44.5%	74.2%	51.5%	81.8%
profitable residents%	76.3%	80.9%	36.6%	45.5%	51.4%	81.7%
ARPU	232.03	274.23	144.62	186.08	212.52	234.75
churn%	3.4%	2.9%	6.4%	8.7%	2.2%	2.3%
churn resident%	2.2%	2.0%	4.1%	2.2%	2.2%	2.3%
churn tourist%	1.2%	1.0%	2.3%	6.5%	0.0%	0.04%

Опишемо словесно визначені кластери.

Середній кластер (кластер 1) – середнє використання хвилин та інтернету та помірний середній прибуток на користувача, рівень відтоку громадян та туристів майже однаковий. Низький рівень нових користувачів, велика кількість прибуткових користувачів. Більшість клієнтів використовують PAYG або платний план на 6–12 місяців, висока якість нових користувачів.

Кластер з високим прибутковим потенціалом (кластер 2) – найбільше використання інтернету та хвилин серед усіх кластерів та найвищий середній прибуток на користувача, найменший відтік та найменша кількість нових користувачів. Більшість користувачів мають платний план на 6–12 місяців. Для нього характерні найбільша кількість прибуткових користувачів серед усіх кластерів та найвища якість нових користувачів.

Кластер з низьким прибутковим потенціалом (кластер 3) – найменше використання хвилин, інтернету та найнижчий середній прибуток на користувача. Велика кількість нових користувачів (громадян), найменше прибуткових користувачів. Якість нових користувачів нижча від середньої.

Кластер туристів (кластер 4) – найвищий рівень відтоку та найбільша кількість нових користувачів-туристів. Половина користувачів використовує PAYG. Загальні кількості громадян та туристів майже однакові. Найвищий відсоток користувачів з VIP планом серед усіх кластерів. Найнижча якість нових користувачів.

Кластер громадян (кластер 5) – найвища кількість нових користувачів-громадян та найменше нових користувачів-туристів. Мала кількість прибуткових ко-

ристувачів. Середнє використання інтернету та хвилин, невеликий рівень відтоку. Середня якість нових користувачів.

Такий опис кластерів дає змогу легко та швидко зрозуміти, які пункти продажу містяться в певному кластері. Такі описи дають змогу, не заглиблюючись в опис кластерів, швидко зрозуміти, яка суть кластеру та що він собою являє. Щоб детальніше проаналізувати опис кластерів, розроблено спеціальні таблиці опису кластерів.

У нашому дослідженні ми зосередили увагу на комплексному аналізі роботи пунктів продажу мобільного оператора, інтегрувавши широкий спектр оперативних показників для управління роздрібними точками. Це контрастує з підходом, застосованим у дослідженні з використанням нейронної мережі SOM, яке було зосереджене на класифікації поведінки користувачів ADSL послуг, надаючи важливу, але вузьку перспективу, що фокусується на маркетинговій стратегії та візуалізації ринкових сегментів [4]. Візуалізація допомагає ідентифікувати поведінкові патерни клієнтів, тоді як наше дослідження розширює межі аналізу, охоплюючи такі ключові аспекти, як внутрішня логістика та управління, що відіграють критичну роль у повсякденній експлуатації пунктів продажу.

Акцентуючи на практичній застосовності, ми не лише класифікуємо пункти продажу за різними рівнями прибутковості, а й пропонуємо аналіз основних показників для оптимізації кожної точки продажу. Це надає інструменти для формування стратегій на рівні пункту продажу, що необхідно для досягнення вищої ефективності та прибутковості. У центрі уваги SOM-

орієнтованого аналізу – розуміння клієнтських відносин та їх категоризація [4].

Однак важливо зазначити, що, хоча наш підхід пропонує ширше розуміння діяльності пунктів продажу, SOM здатний відіграти важливу роль у виявленні відмінностей у поведінці користувачів, які можуть бути використані для налаштування цільового маркетингу. Об'єднання обох підходів може привести до ще глибших знань і кращих результатів у керуванні як клієнтським досвідом, так і оперативною діяльністю пунктів продажу.

Отже, за результатами досліджень можна сформулювати наукову новизну та практичну цінність роботи.

Наукова новизна отриманих результатів дослідження полягає у подальшому розвитку методу класифікації K-середніх, який ґрунтується на розподілі об'єктів за сегментами відповідно до встановлених атрибутів та індексів подібності, що дає змогу підвищити точність і ефективність аналізу ринкової поведінки й оптимізації бізнес-процесів у сфері мобільного зв'язку.

Практична значущість результатів дослідження – розроблено систему автоматизованої сегментації пунктів продажу мобільного оператора на основі класифікації, що дасть змогу операторам ефективно групувати пункти продажу, визначати їхню продуктивність, відстежувати зміни з часом і адаптувати стратегії розвитку для задоволення потреб клієнтів та підвищення конкурентоспроможності на ринку.

Висновки / Conclusions

Розроблено інформаційну технологію, яка спрямована на моніторинг і аналіз сегментів пунктів продажу мобільного оператора. Результати дослідження дають змогу мобільному оператору визначити найприбутковіше та найперспективніші пункти продажу, що може сприяти розробленню ефективних маркетингових стратегій та підвищенню конкурентоспроможності на ринку мобільного зв'язку.

Дослідження має важливу практичну значущість, оскільки сприяє підвищенню ефективності функціонування та прибутковості бізнесу в сфері мобільного зв'язку, а також дає змогу адаптувати стратегії розвитку до змінних ринкових умов та потреб клієнтів.

Дослідження підтвердило, що модель здатна ефективно адаптуватися до нових даних та виявляти різноманітні сегменти пунктів продажу для подальшого вдосконалення бізнес-стратегій. Виявлено різні типи кластерів, що дає змогу глибше аналізувати та опти-

мізувати бізнес-процеси, зорієнтовані на різні сегменти користувачів.

Результати цього дослідження можуть бути використані не тільки в сфері мобільного зв'язку, але й у інших галузях, де актуальне питання сегментації, аналізу та оптимізації мережі точок продажу або сервісу.

References

- [1] Chiang, M. T., & Mirkin, B. (2010). Intelligent Choice of the Number of Clusters in K-Means Clustering: An Experimental Study with Different Cluster Spreads. *J. Classif.*, 27, 3–40. <https://doi.org/10.1007/s00357-010-9049-5>
- [2] Collica, R. S. (2021). Segmentation Analytics with SAS Viya: An Approach to Clustering and Visualization. SAS Institute.
- [3] Doroshenko, A. (2020). “Analysis of the Distribution of COVID-19 in Italy Using Clustering Algorithms”, 2020 IEEE Third International Conference on Data Stream Mining & Processing (DSMP), Lviv, Ukraine, 2020, 325–328, <https://doi.org/10.1109/DSMP47368.2020.9204202>
- [4] Hanafizadeh, P., & Mirzazadeh, M. (2011). Visualizing market segmentation using self-organizing maps and Fuzzy Delphi method – ADSL market of a telecommunication company. *Expert Systems with Applications*, 38(1), 198–205. <https://doi.org/10.1016/j.eswa.2010.06.045>
- [5] Maddumala, V. R., Chaikam, H., Velanati, J. S., Ponnaganti, R., & Enuguri, B. (2022). Customer Segmentation using Machine Learning in Python. 2022 7th International Conference on Communication and Electronics Systems (ICCES), Coimbatore, India, 1268–1273. <https://doi.org/10.1109/ICCES54183.2022.9836018>
- [6] Mim, S. S., & Logofatu, D. (2022). A Cluster-based Analysis for Targeting Potential Customers in a Real-world Marketing System. 2022 IEEE 18th International Conference on Intelligent Computer Communication and Processing (ICCP), Cluj-Napoca, Romania, 159–166. <https://doi.org/10.1109/ICCP56966.2022.10053985>
- [7] Nandapala, E. Y. L., & Jayasena, K. P. N. (2020). The practical approach in Customers segmentation by using the K-Means Algorithm. 2020 IEEE 15th International Conference on Industrial and Information Systems (ICIIS), RUPNAGAR, India, 344–349. <https://doi.org/10.1109/ICIIS51140.2020.9342639>
- [8] Rosário, A. T., Dias, J. C., & Ferreira, H. (2023). Bibliometric Analysis on the Application of Fuzzy Logic into Marketing Strategy. *Businesses*, 3(3), 402–423. <https://doi.org/10.3390/businesses3030025>
- [9] Stiad, M. (2022). Market segmentation analysis in food selection. *Jurnal Ekonomi*, 11(03), 169–173. <https://doi.org/10.1016/j.eswa.2010.06.045>
- [10] Wu, J., Xiong, H., & Chen, J. (2009). Adapting the right measures for K-means clustering. In Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD '09). Association for Computing Machinery, New York, NY, USA, 877–886. <https://doi.org/10.1145/1557019.1557115>

O. O. Narushynska, V. I. Motrunych, M. V. Arzubov, V. M. Teslyuk

Lviv Polytechnic National University, Lviv, Ukraine

INFORMATION TECHNOLOGY FOR THE ANALYSIS OF MOBILE OPERATOR SALES OUTLETS BASED ON CLUSTERING METHODS

This research presents the development and implementation of information technology for monitoring and analyzing segments of a mobile operator's stores using clustering methods. The study addresses a pertinent issue in marketing and business optimization, namely the enhancement of strategies for the network of mobile communication stores.

The research paper presents a novel approach to segmenting mobile operator stores using clustering algorithms. A software product was developed that includes machine learning algorithms for clustering stores according to critical parameters. A comprehensive analysis of the mobile operator's database was conducted to identify critical characteristics

of the stores, such as profitability, patterns of mobile operator service usage, the number of new and lost customers, geographical location, and other vital indicators.

Particular attention was paid to developing tools for preparing and processing input data, ensuring the accuracy of subsequent clustering. With the created product, the mobile operator can identify the most profitable stores, uncover growth opportunities, and develop targeted strategies for each segment.

By applying the developed technology, the mobile operator gains the ability not only to identify crucial and profitable sales points but also to develop focused strategies for different groups of stores, taking into account their unique characteristics. This approach strengthens the company's market position, increasing customer satisfaction and profitability.

Additionally, when examining the possibilities of analyzing store dynamics over time, it is necessary to consider the ever-evolving business environment. Such a tool can assist the operator in swiftly adapting strategies and responding to new trends and challenges while preserving stability and profitability.

Similar innovative approaches not only facilitate the management of a mobile operator's store network but also enable the establishment of more open and flexible customer relationships. By providing personalized services and responding to their needs, businesses can enhance customer loyalty and increase their profits.

In conclusion, this research endeavour carries significant implications for the realms of marketing and mobile operator development. Its findings can be harnessed to enhance the efficiency of operations and profitability within this industry.

Keywords: machine learning; unsupervised learning; data processing; K-means; consumer behavior patterns.

Інформація про авторів:

Нарушинська Ольга Олександрівна, асистент, кафедра автоматизованих систем управління.

Email: narushynska@gmail.com

Мотрунич Владислав Іванович, студент, кафедра інформаційних систем та мереж.

Email: onistarvlaad@gmail.com

Арзузов Максим Валерійович, асистент, кафедра автоматизованих систем управління.

Email: marzubov@gmail.com

Теслюк Василь Миколайович, канд. техн. наук, доцент, кафедра автоматизованих систем управління.

Email: vasyt.teslyuk@gmail.com; <https://orcid.org/0000-0002-5974-9310>

Цитування за ДСТУ: Нарушинська О. О., Мотрунич В. І., Арзузов М. В., Теслюк В. М. Інформаційна технологія для аналізу пунктів продажу мобільного оператора на основі методів кластеризації. *Український журнал інформаційних технологій*. 2023. Т. 5, № 2. С. 105–113.

Citation APA: Narushynska, O. O., Motrunych, V. I., Arzubov, M. V., & Teslyuk, V. M. (2023). Information technology for the analysis of mobile operator sales outlets based on clustering methods. *Ukrainian Journal of Information Technology*, 5(2), 105–113. <https://doi.org/10.23939/ujit2023.02.105>