

УДК 004.652

DOI: <https://doi.org/10.37734/2409-6873-2024-2-10>

ВИКОРИСТАННЯ МЕТОДІВ МАШИННОГО НАВЧАННЯ ТА МІКРОСЕРВІСНОЇ АРХІТЕКТУРИ В ПРОДУКТОВІЙ АНАЛІТИЦІ

Н. О. ІВАНЧЕНКО

кандидат економічних наук, доцент,
доцент кафедри статистики, інформаційно-аналітичних систем і демографії;

О. С. ПОДСКРЕБКО

кандидат економічних наук, доцент,
доцент кафедри економічної кібернетики,
Київський національний університет імені Тараса Шевченка

Анотація. Метою цієї статті є представлення методології класифікації клієнтів з використанням мікросервісної архітектури, яка може бути використана для прийняття маркетингових рішень і покращення стратегії управління продуктами та взаємодії з клієнтами. **Методика дослідження.** Методологія дослідження передбачає використання наступних методів: логічного узагальнення, кластерного аналізу, аналітичного методу (статистичний аналіз). **Результати.** Розроблений мікросервіс класифікації клієнтів може знаходитися як на локальному хості так і в мережі Інтернет, в модель передаються певні параметри клієнта і вона на цих даних формує прогноз, який по суті є номером ймовірного класу до якого треба віднести цього клієнта. **Практична значущість результатів дослідження.** Результати цього дослідження мають практичне значення для маркетологів та менеджерів з управління продуктами, які шукають інструменти для прогнозування поведінки клієнтів. Запропонована методологія та мікросервіс можуть бути використані для створення системи підтримки прийняття рішень в маркетинговій діяльності.

Ключові слова: класифікація об'єктів, продуктова аналітика, мікросервісна архітектура, кластерний аналіз, машинне навчання.

Постановка проблеми в загальному вигляді та зв'язок із найважливішими науковими чи практичними завданнями. Класифікація є важливим інструментом продуктової аналітики, який дозволяє систематизувати інформацію та групувати об'єкти або явища за спільними ознаками. У контексті продуктової аналітики класифікація використовується для підвищення ефективності взаємодії з продуктами, користувачами, виявлення особливостей використання продукту та інших аспектів, що допомагають в зрозумінні та аналізі даних, які безперервно накопичуються. Основні аспекти класифікації в продуктивній аналітиці представлені в табл. 1.

Класифікація у продуктивній аналітиці допомагає в створенні системи для зручного аналізу, порівняння та прийняття стратегічних рішень. Вона також дозволяє підкреслити ключові аспекти та взаємозв'язки, що сприяє ефективному управлінню продуктом.

Аналіз останніх досліджень і публікацій. Дослідженням у сфері продуктової аналітики, а саме аспектам, що пов'язані з взаємовідносинами з клієнтами займалися багато компаній [1; 2] та науковців: В'ялець О.В. [3], Чміль Г.Л., Джугташвілі Н.М. [4], Ерік Бенджамін Зойферт [5], Джоанна Родрігес [6] та інші. Загалом продуктивні компанії та вчені зробили значний внесок у сферу розвитку продуктової аналітики завдяки своїм дослідженням.

Формування цілей статті (постановка завдання). Мета статті – представлення методології та інструменту класифікації клієнтів з використанням мікросервісної архітектури.

Виклад основного матеріалу дослідження з повним обґрунтуванням отриманих наукових результатів. Методологія класифікації клієнтів – це систематичний підхід до групування клієнтів згідно з різними характеристиками або критеріями для подальшого забезпечення персоналізованого обслуговування та розвитку стратегій взаємодії. Основною метою такої класифікації є пізнання різноманітності клієнтів, їхніх потреб та характеристик для оптимізації комунікації та надання більш ефективних послуг.

Суть методології класифікації клієнтів включає такі ключові елементи представлені на рис. 1.

Важливою задачею є визначення ключових факторів, за якими клієнти будуть класифіковані. Це може включати демографічні характеристики (вік, стать, географічне положення, місто, район проживання), психографію (інтереси, стилі життя), поведінку покупців, рівень використання продукту, та інше. Також, групування клієнтів в окремі сегменти відповідно до визначених критеріїв. Наприклад, можливі сегменти – нові клієнти, постійні клієнти, клієнти з високим оборотом тощо.

Аналіз основних характеристик кожного сегменту або в термінах машинного навчання класу/

Таблиця 1

Аспекти класифікації в продуктивій аналітиці

Класифікація продуктів	Функціональна класифікація	Групування продуктів за їхніми функціональними характеристиками або призначенням
	Сегментація за цільовою аудиторією	Розподіл продуктів відповідно до цільового ринку або клієнтської бази
	Класифікація за життєвим циклом	Розпізнавання етапів життєвого циклу продукту та відповідне позиціонування.
Класифікація користувачів	Демографічна класифікація	Групування користувачів за віком, статтю, освітою та іншими демографічними характеристиками
	Поведінкова сегментація	Розподіл користувачів відповідно до їхньої споживацької поведінки та звичок
	Географічна класифікація	Класифікація користувачів залежно від їхнього місцезнаходження та географічних характеристик
Класифікація аспектів використання	Класифікація за функціональністю:	Розподіл аспектів використання за їхніми функціональними можливостями
	Класифікація за задачами	Групування аспектів використання відповідно до завдань, які вони вирішують
Класифікація за даними та метриками	Класифікація за джерелами даних	Розподіл джерел даних для аналізу та вимірювань.
	Класифікація за ключовими метриками	Групування метрик та показників продуктової ефективності
Класифікація за стратегічними напрямками	Класифікація за стратегією розвитку	Розподіл продуктів відповідно до стратегії розвитку компанії (наприклад, диференціація чи вартісна стратегія).

Джерело: сформовано авторами

кластеру для створення детальних профілів користувачів продукту. Це може включати в себе купівельну історію, поведінку в мережі, надавання переваг певним діям та потреби клієнтів.

Немаловажним є розробка і впровадження стратегій взаємодії для кожного сегменту клієнтів. Це може включати персоналізовану маркетингову комунікацію, спеціальні пропозиції, індивідуальний сервіс тощо.

Постійний моніторинг та адаптація стратегій взаємодії з клієнтами, що спирається на зміни в

споживацькій поведінці, ринкових тенденціях та інших факторах, а також використання аналітичних інструментів та технологій для збору та обробки даних з метою розпізнаванні паттернів, прогнозуванні тенденцій та прийнятті рішень, допомагають забезпечити актуальність та ефективність.

Застосування методології класифікації клієнтів може значно покращити спосіб, яким бізнес взаємодіє з клієнтами, забезпечуючи більш ефективні стратегії маркетингу, обслуговування та утримання клієнтів.

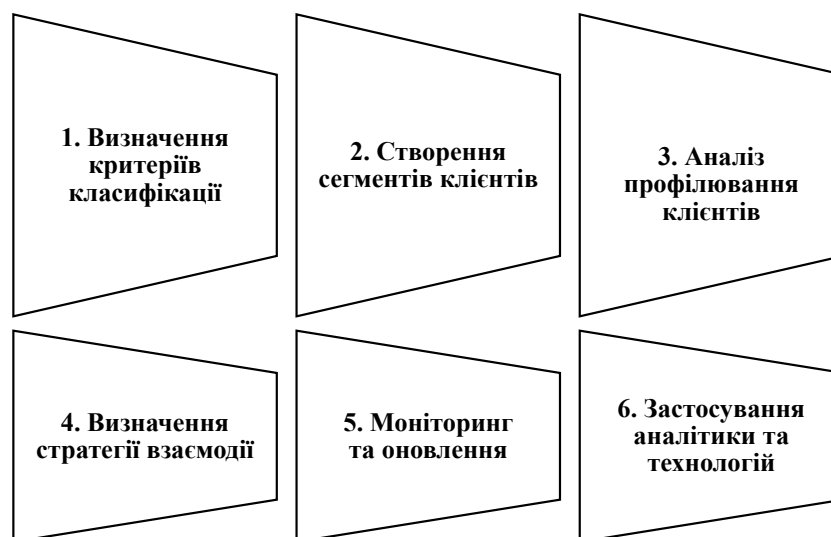


Рис. 1. Методологія класифікації клієнтів

Джерело: сформовано авторами

```
df.head()
```

	ID	Year_Birth	Education	Marital_Status	Income	Kidhome	Teenhome	Dt_Customer	Recency	MntWines	...	NumWebVisitsMonth
0	5524	1957	Graduation	Single	58138.0	0	0	04-09-2012	58	635	...	7
1	2174	1954	Graduation	Single	46344.0	1	1	08-03-2014	38	11	...	5
2	4141	1965	Graduation	Together	71613.0	0	0	21-08-2013	26	426	...	4
3	6182	1984	Graduation	Together	26646.0	1	0	10-02-2014	26	11	...	6
4	5324	1981	PhD	Married	58293.0	1	0	19-01-2014	94	173	...	5

5 rows x 29 columns

Рис. 2. Приклад даних для класифікації клієнтів

Джерело: сформовано [13]

В статті в якості інструменту для класифікації клієнтів запропоновано використати мікросервісну архітектуру.

Концепція розробки мікросервісів – це архітектурний підхід до розробки програмного забезпечення, при якому програма розбивається на невеликі, автономні служби, які називаються мікросервісами. Кожен мікросервіс виконує конкретну функцію або набір пов'язаних функцій і може бути розгорнутий та масштабований незалежно від інших мікросервісів.

У якості даних для проведення класифікації були використані дані з Kaggle [13] про маркетингову кампанію, які мають відкриту ліцензію. Перед початком роботи з даними, треба зробити деякі перетворення даних. Наприклад, перекодувати дату народження клієнта у вік в роках.

Так як дані не мають початкового розбиття на групи (належності конкретного клієнта до певного класу), то перед побудовою моделі класифікації треба

скористатись можливостями кластерного аналізу, це надасть змогу розбити всіх клієнтів на певні кластери.

Після проведення всіх необхідних маніпуляцій з даними для проведення аналізу та побудови моделі було відібрані наступні фактори.

Перед безпосередньо побудовою моделі, треба провести певну додаткову підготовку даних, а саме їх перетворення за допомогою інструменту StandardScaler.

StandardScaler дозволяє стандартизувати дані, з середнім яке дорівнює нулю та одиничною дисперсією. Це один з ефективних інструментів, який дозволяє підвищити якість побудованих моделей за умови, що використання цього виду перетворення даних є доцільним для конкретного класу моделей машинного навчання.

Як зазначалось раніше, перед побудовою моделі треба встановити кількість кластерів. Для цього скористаємось бібліотекою yellowbrick і наступним кодом.

```
num_cols = [
    'Age',
    'Income',
    'Month_regist',
    'Sum_Purchased',
    'Count_Campaings',
    'Recency',
    'NumDealsPurchases',
    'NumCatalogPurchases',
    'NumStorePurchases',
    'NumWebVisitsMonth',
    'NumWebPurchases',
    'MntWines',
    'MntFruits',
    'MntMeatProducts',
    'MntFishProducts',
    'MntSweetProducts',
    'MntGoldProds'
]
```

Рис. 3. Відібрані ознаки для побудови моделі кластеризації клієнтів

Джерело: сформовано авторами

```
import matplotlib.pyplot as plt
from sklearn.preprocessing import StandardScaler

X_scaled = StandardScaler().fit_transform(df_new)
```

Рис. 4. Препроцесінг даних за допомогою бібліотеки sklearn та інструменту StandardScaler

Джерело: сформовано авторами

```
from yellowbrick.cluster import KElbowVisualizer

model = KMeans(random_state = 42, n_init= 'auto')

visual = KElbowVisualizer(model, k = (1, 10), timings = False)

visual.fit(X_scaled) #(X_pca)
visual.show();
```

Рис. 5. Код, який дозволяє встановити кількість класів/кластерів за допомогою методу ліктя

Джерело: сформовано авторами

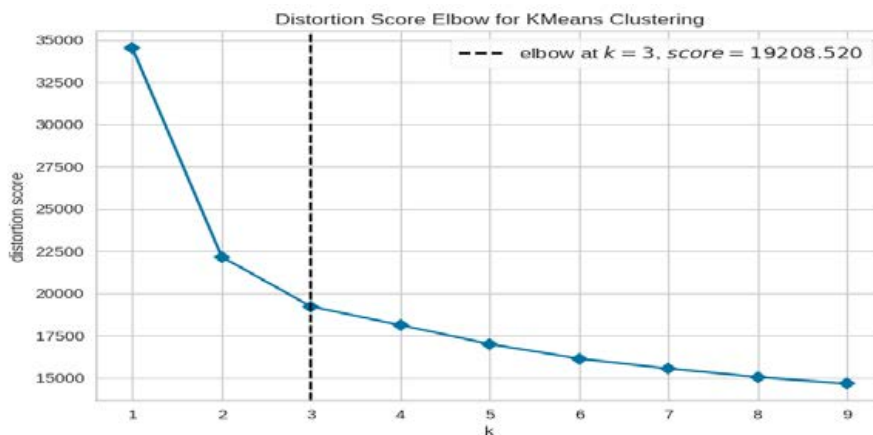


Рис. 6. Встановлення ефективної кількості класів за допомогою бібліотеки yellowbrick

Джерело: сформовано авторами

Для встановлення ефективної кількості класів треба вказати модель, яка буде використовуватись, вказати діапазон кількості кластерів, в даному прикладі це від 1 до 10, а також передати безпосередньо дані. Результат виконання коду наведено на рисунку.

Застосування методів машинного навчання, а саме моделі K-Means дозволило поділити всіх клієнтів на 3 групи, кількість таких груп встановлювалась за допомогою методу ліктя.

Після встановлення ефективної кількості кластерів можна будувати модель. Код побудови моделі наведено на рис. 7.

Результатом побудованою моделі та код отримання номерів класів для кожного клієнта наведено на рисунку.

Отримані labels для кожного клієнта тепер можна використовувати для розробки мікросервісу, який для кожного нового клієнту буде надавати номер класу, тобто вирішувати задачу класифікації.

Для розробки мікросервісу доцільно скористатись зручною та ефективною бібліотекою «streamlit». Ця бібліотека містить багато вбудованих елементів і дозволяє швидко розробляти мікросервіси.

Для роботи з streamlit і побудови мікросервісу зручно скористатись такою IDE як Pycharm, який є зручним середовищем для роботи з Python та розробки різного роду додатків.

```
from sklearn.cluster import KMeans
model = KMeans(n_clusters= 3,
               random_state = 42)

model.fit(x_scaled) # df_new
```

Рис. 7. Побудова моделі K-means з трьома кластерами

Джерело: сформовано авторами

Після завантаження даних та імпорту необхідних бібліотек, необхідно створити модель і зберегти її для подальшого використання і створення мікросервісу. У якості моделі машинного навчання для класифікації клієнтів було обрано логістичну регресію, експериментувати з іншими моделями не було потреби, так як обрана модель показала достатню ефективність та якість.

На рис. 9 наведено приклад роботи розробленого мікросервісу та результат, який повертає модель для клієнта з урахуванням певних параметрів. Тобто, цей конкретний клієнт відноситься до другого класу.

Таким чином, за допомогою розробленого мікросервісу, який може знаходитись як на локальному хості та і в мережі Інтернет, в модель передаються певні параметри клієнта і вона на цих даних формує прогноз, який по суті є номером ймовірного класу (а точніше ймовірністю належності до певного класу) до якого треба віднести цього клієнта.

```
labels = model.predict(x_scaled) # df_new
labels[:20]

array([2, 2, 1, 0, 2, 2, 2, 0, 0, 0, 0, 2, 2, 0, 1, 2, 0, 1, 0, 0],
      dtype=int32)
```

Рис. 8. Результатом роботи моделі та код отримання номерів класів для кожного клієнта

Джерело: сформовано авторами

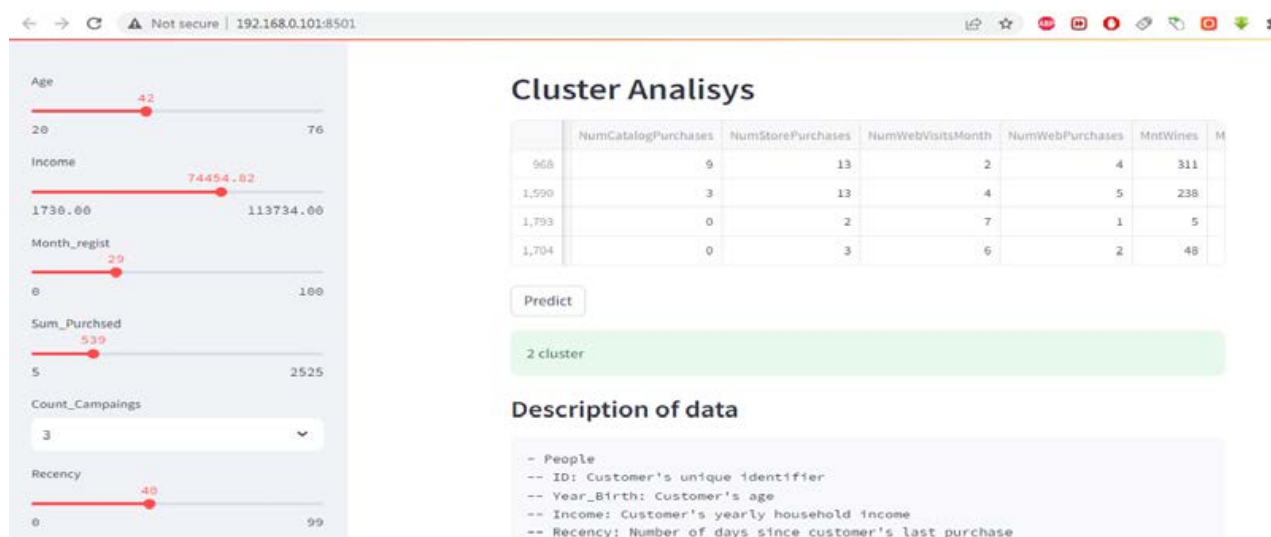


Рис. 9. Приклад роботи мікросервісу з поверненням номеру класу для клієнта за певними параметрами

Джерело: сформовано авторами

Висновки із зазначених проблем і перспективи подальших досліджень у поданому напрямі. Класифікація клієнтів є важливою для продуктової аналітики з багатьох причин, оскільки вона дозволяє власникам продукту та бізнесам розуміти свою аудиторію, налаштовувати стратегії взаємодії та максимізувати ефективність та якість продуктів і послуг.

На основі проведеного аналізу клієнтської бази, побудови моделі класифікації, мови програмування Python та набору спеціалізованих

бібліотек побудовано мікросервіс, який допоможе класифікувати нових клієнтів за певними параметрами. Використання мікросервісної архітектури полегшує розробку, тестування та підтримку складних додатків, забезпечуючи гнучкість та швидкість реагування на зміни у вимогах та ринкових умовах. Однак правильне планування, впровадження та управління мікросервісами є ключовим для уникнення складнощів, пов'язаних із зростанням кількості сервісів та їх взаємодією.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Product Analytics: A Comprehensive Guide. URL: <https://cxl.com/blog/product-analytics/> (дата звернення: 03.02.2024).
2. Mixpanel Guide: The Guide to Product Analytics. A book of questions and answers. URL: <https://mixpanel.com/content/guide-to-product-analytics/report/> (дата звернення: 03.02.2024).
3. В'ялець О.В. Система управління взаємовідносинами з клієнтами: теоретичний аспект. *Наукові праці НУХТ*. 2019. Т. 25. № 3. С. 90–95.
4. Чміль Г.Л., Джугташвілі Н.М. Цифровізація управління клієнтським досвідом у готельно-ресторанній індустрії. *Бізнес Інформ*. 2020. № 8. С. 237–245.
5. Eric Benjamin Seufert. *Freemium Economics: Leveraging Analytics and User Segmentation to Drive Revenue*. Morgan Kaufmann; 1st edition. 2014. 254 p.
6. Rodrigues Joanne. *Product Analytics: Applied Data Science Techniques for Actionable Consumer Insights*. Addison-Wesley Professional, 2020. 360 p.
7. Бортнік А.М. Стратегічне управління взаємовідносинами з клієнтами на основі CRM. *Стратегія економічного розвитку України*. Київ : КНЕУ. 2012. № 30. С. 165–169.
8. Fowler M. *Microservices*. URL: <https://martinfowler.com/articles/microservices.html> (дата звернення: 03.02.2024).
9. Liashenko O., Podskrebko O., Ivanchenko N. (2022) The Impact of Data Analytics on the Nature of Doing Business. 12th International Conference on Advanced Computer Information Technologies (ACIT), Conference. Spišská Kapitula. 2022. P. 331–334.
10. Іванченко Н.О., Подскребко О.С. Інструменти аналізу ефективності взаємодії користувача з продуктом. *Modernization of science and its influence on global processes: collection of scientific papers «SCIENTIA» with Proceedings of the IV International Scientific and Theoretical Conference, November 3, 2023. Bern, Swiss Confederation: International Center of Scientific Research*. P. 21–24.
11. *Microservices are hard*. URL: <https://hackernoon.com/microservices-are-hard-an-invaluable-guide-to-microservices-2d06bd7bcf5d> (дата звернення: 01.02.2024).

12. What is Microservices Architecture? URL: <https://smartbear.com/learn/api-design/what-are-microservices/> (дата звернення: 01.02.2024).
13. Дані з Kaggle про маркетингову компанію. URL: <https://www.kaggle.com/datasets/imakash3011/customer-personality-analysis>. (дата звернення: 01.02.2024).

REFERENCES

1. Product Analytics: A Comprehensive Guide. Available at: <https://cxl.com/blog/product-analytics/> (accessed February 3, 2024).
2. A MIXPANEL GUIDE: The Guide to Product Analytics. A book of questions and answers. Available at: <https://mixpanel.com/content/guide-to-product-analytics/report/> (accessed February 3, 2024).
3. Vialets, O. V. (2019) Systema upravlinnia vzaiemovidnosynamy z kliientamy: teoretychnyi aspekt. *Naukovi pratsi NUKhT*, vol. 25, no. 3, pp. 90–95. [in Ukrainian]
4. Chmil H. L. and Dzhhutashvili N. M. (2020) Digitalization of Customer Experience Management in the hotel and restaurant industry. *Business Inform*, no. 8, pp. 237–245. [in Ukrainian]
5. Eric Benjamin Seufert (2014) Freemium Economics: Leveraging Analytics and User Segmentation to Drive Revenue. Morgan Kaufmann; 1st edition. 254 p.
6. Rodrigues J. (2020) Product Analytics: Applied Data Science Techniques for Actionable Consumer Insights. Addison-Wesley Professional, 360 p.
7. Bortnik A. M. (2012) Stratehichne upravlinnia vzaiemovidnosynamy z kliientamy na osnovi CRM. *Strategy of economic development of Ukraine*. Kyiv: KNEU, no. 30, pp. 165–169. [in Ukrainian]
8. Fowler M. Microservices. Available at: <https://martinfowler.com/articles/microservices.html> (accessed February 3, 2024).
9. Liashenko O., Podskrebko O., Ivanchenko N. (2022) The Impact of Data Analytics on the Nature of Doing Business. 12th International Conference on Advanced Computer Information Technologies (ACIT). Conference. Spišská Kapitula, pp. 331–334. [in Ukrainian]
10. Ivanchenko N. O., Podskrebko O. S. (November 3, 2023) Instrumenty analizu efektyvnosti vzaiemodii korystuvacha z produktom. Modernization of science and its influence on global processes: collection of scientific papers "SCIENTIA" with Proceedings of the IV International Scientific and Theoretical Conference. Bern, Swiss Confederation: International Center of Scientific Research, pp. 21–24. [in Ukrainian]
11. Microservices are hard. Available at: <https://hackernoon.com/microservices-are-hard-an-invaluable-guide-to-microservices-2d06bd7bcf5d> (accessed February 1, 2024).
12. What is Microservices Architecture? Available at: <https://smartbear.com/learn/api-design/what-are-microservices/> (accessed February 1, 2024).
13. Data set from Kaggle about marketing company. Available at: <https://www.kaggle.com/datasets/imakash3011/customer-personality-analysis> (accessed February 1, 2024).

Nadiia Ivanchenko, Oleksandr Podskrebko, Taras Shevchenko National University of Kyiv. Use of machine learning methods and microservice architecture in product analytics.

Abstract. The purpose of this article is to present a customer classification methodology using a microservice architecture, which can be used to make marketing decisions and improve product management strategy and customer interaction. Customer classification is important to product analytics for a number of reasons, as it allows product manufacturers and businesses to understand their audience, adjust engagement strategies, and maximize the effectiveness of products and services. **Methodology of research.** The research methodology involves the use of the following methods: logical generalization, cluster analysis, analytical method (statistical analysis). Customer classification methodology is a systematic approach to grouping customers according to various characteristics or criteria to further provide personalized service and develop interaction strategies. The main purpose of such a classification is to learn about the variety of customers, their needs and characteristics in order to optimize communication and provide more effective services. **Findings.** The developed client classification microservice can be located both on a local host and on the Internet, certain parameters of the client are transferred to the model and it forms a forecast based on these data, which is essentially the number of the probable class to which this client should be assigned. Based on the analysis of the customer base, the construction of a clustering model, the Python programming language, and a set of specialized libraries, a microservice was built that will help classify new customers according to certain parameters. The use of microservices architecture facilitates the development, testing and maintenance of complex applications, providing flexibility and speed of response to changes in requirements and market conditions. However, proper planning, implementation, and management of microservices is key to avoiding the complexities associated with the growing number of services and their interactions. **Practical value.** The results of this study have practical implications for marketers and product management managers looking for tools to predict customer behavior. The proposed methodology and microservice can be used to create a decision support system in marketing activities.

Keywords: object classification, product analytics, microservice architecture, cluster analysis, machine learning.