

4. Algaba A., Fernandez-Sanchez F., Freire E., GAmero E., Rodriguez-Luis A.J. Oscillation-sliding in a modified van der Pol-Duffing electronic oscillator. *Journal of Sound and Vibration*. – 2002. Vol 249(9). p. 899-907

5. Murray J.D. *Mathematical biology I. An Introduction* (3rd edn.). Springer. 2002

6. Kaplan V.Z., Gabay I., Saraan G., Biological application of the altered Van der Pol oscillator. *Journal of the Franklin Institute*. 2007, Vol 345(3). P.226-232.

7. Landau I.D., Bouziana F., Bitmead R., Voda A. Analysis of control relevant coupled nonlinear oscillatory systems. *European Journal of Control*. 2008. Vol 10. p. 263-282.

DOI <https://doi.org/10.30525/978-9934-26-506-8-200>

APPLICATION OF MACHINE LEARNING FOR CLUSTERING OF COMPANY CLIENTS

ЗАСТОСУВАННЯ МАШИННОГО НАВЧАННЯ ДЛЯ КЛАСТЕРИЗАЦІЇ КЛІЄНТІВ КОМПАНІЇ

Ivanchenko N.O.,

*PhD (Economics),
Associate Professor, Taras Shevchenko
National University of Kyiv,
Kyiv, Ukraine*

Іванченко Н.О.,

*к.е.н., доцент,
Київський національний університет
імені Тараса Шевченка,
м. Київ, Україна*

Podskrebko O.S.,

*PhD (Economics),
Associate Professor Taras Shevchenko
National University of Kyiv, Kyiv,
Ukraine; LLC "Technical university
"Metinvest polytechnic",
Zaporizhzhia, Ukraine*

Подскребко О.С.,

*к.е.н., доцент,
Київський національний університет
імені Тараса Шевченка;
ТОВ «Технічний університет
«Метінвест політехніка»,
м. Запоріжжя, Україна*

Алгоритми машинного навчання за останні роки стали невід'ємною частиною і потужним інструментом прийняття управлінських рішень. Машинне навчання допомагає ефективно вирішувати широкий спектр бізнес-завдань, а саме:

- **прогнозування, дозволяє компаніям ефективно планувати свою роботу**, оцінювати попит, визначати обсяги продажів та ін;
- **аналіз тенденцій, виявлення прихованих зв'язків, аномалій і повторюваних елементів;**
- **сучасні технології дозволяють достатньо ефективно** розпізнавати фото-, відео– та аудіоконтент, а також виявляти спроби шахрайства, обману, внутрішніх та зовнішніх загроз безпеці.

Машинне навчання в бізнесі активно застосовується для оптимізації різних процесів, зокрема у класифікації та сегментації клієнтів. Кластеризація – метод, який змогу не лише краще розуміти аудиторію, але й ефективніше адаптувати стратегії маркетингу персоналізуючи пропозиції для кожного сегмента. Для проведення кластеризації клієнтів в роботі використано публічний набір даних з Kaggle, якій містить інформацію про клієнтів. Для більш ефективної роботи алгоритму по розподіленню клієнтів на кластери, деякі ознаки були перетворенні на більш релевантні поставлений задачі.

```
df['Age'] = df.Year_Birth.apply(lambda x: end_data.year - x)

df['Month_regist'] = df.Dt_Customer.apply(
    lambda x: get_diff_date(end_data, x, type_date='month'))
```

Рис. 1. Додаткові ознаки, які будуть використанні для кластеризації (поділу на групи) клієнтів

Після проведення всіх необхідних маніпуляцій з даним для проведення кластерного аналізу та побудови моделі було відібрані наступні фактори рис. 2.

Перед безпосередньо побудовою моделі, треба провести певну додаткову підготовку даних, а саме їх перетворення за допомоги інструменту `StandardScaler`, який дозволяє стандартизувати дані, середнім яке дорівнює нулю та одиничною дисперсією. Це один з ефективних інструментів, який застосовується для вирішення задач кластеризації.

```

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

```

Рис. 2. Відібрані ознаки для проведення кластерного аналізу

$$z = \frac{x - \mu}{s}$$

де,

z – стандартизована ознака;

x – значення ознаки певного об'єкту з набору даних;

μ – середнє значення ознаки для всіх об'єктів з набору даних;

s – стандартне відхилення для ознаки x .

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

рис. 3.

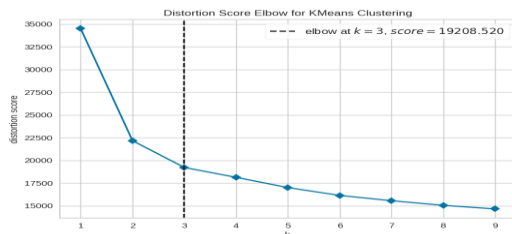


Рис. 3. Встановлення ефективної кількості кластерів за допомогою бібліотеки yellowbrick

Для встановлення ефективної кількості кластерів треба вказати модель, яка буде використовуватись, вказати діапазон кількості кластерів, в даному прикладі це від 1 до 10, а так передати безпосередньо дані. Таким чином, оптимальною кількістю кластерів є 3 зі score 19209.

Результатом побудованою моделі та код отримання номерів кластерів для кожного клієнта наведено на рис.4.

```
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. 4. Результат роботи моделі та код отримання номерів кластерів для кожного клієнта

Таким чином, кластеризація допомагає ідентифікувати групи клієнтів із подібними інтересами або споживацькою поведінкою. Цю інформацію можна використовувати для точного налаштування рекламних кампаній відповідно до характеристик цільової аудиторії.

Використання моделей машинного навчання дає можливість компаніям прогнозувати витрати, автоматизувати обробку даних та ефективно розподіляти ресурси. Вибір алгоритмів визначається залежно від конкретних умов і сфер, які потребують покращення.

Перелік використаних джерел

1. Відкритий набір даних з інформацією по клієнтах однієї з компаній: URL: <https://www.kaggle.com/datasets/imakash3011/customer-personality-analysis>
2. Мікросервісна архітектура як інструмент класифікації клієнтів / Н.О. Іванченко, О.С. Подскребко. *Науковий вісник Полтавського університету економіки і торгівлі*. №2 (108). Полтава: ПУЕТ, 2023. С. 53-59.
3. Machine Learning for Business Analytics: Concepts, Techniques and Applications in RapidMiner 1st Edition / Galit Shmueli, Peter C. Bruce, Amit V. Deokar, Nitin R. Patel. Wiley; 1st edition, 2023. 736 p.
4. Business Intelligence, Analytics, and Data Science: A Managerial Perspective / Ramesh Sharda, Dursun Delen, Efraim Turban. – Pearson; 4th edition, 2017. 512 p.
5. Fundamentals of Machine Learning for Predictive Data Analytics: Algorithms, Worked Examples, and Case Studies / Kelleher John D., Macnamee Brian, D'Arcy Aoife. The MIT Press; 2nd edition, 2020. 856 p.