

ВИКОРИСТАННЯ КЛАСТЕРНОГО АНАЛІЗУ ДЛЯ ГРУПУВАННЯ МІЖНАРОДНИХ КОМПАНІЙ

CLUSTER ANALYSIS USE FOR INTERNATIONAL COMPANIES GROUPING

Пономаренко І.В.

кандидат економічних наук, доцент,
доцент кафедри економічної кібернетики та маркетингу,
Київський національний університет технологій та дизайну

Унтура А.В.

магістр,
Київський національний університет технологій та дизайну

Стаття присвячена дослідженню застосування кластерного аналізу для виокремлення груп під час проведення аналізу різноманітних соціально-економічних явищ та процесів. Обґрунтовано доцільність використання зазначеного статистичного підходу як ефективного методу для виділення груп з подібними одиницями сукупності. На основі рейтингу "Forbes" за 2017 р. проведено кластеризацію міжнародних компаній, що дасть змогу на підставі якісного аналізу виокремлених груп розробити ефективні стратегії розвитку глобального економічного середовища.

Ключові слова: "Forbes", дендрограма, кластерний аналіз, міжнародне економічне середовище, транснаціональна корпорація.

Статья посвящена исследованию применения кластерного анализа для выделения групп при проведении анализа различных социально-экономических явлений и процессов. Обоснована целесообразность использования указанного статистического подхода как эффективного метода для выделения групп с подобными единицами совокупности. На основе рейтинга "Forbes" за 2017 г. проведена кластеризация международных компаний, что позволит на основании анализа выделенных групп разработать эффективные стратегии развития глобальной экономической среды.

Ключевые слова: "Forbes", дендрограмма, кластерный анализ, международная экономическая среда, транснациональная корпорация.

The article is devoted to the study of the cluster analysis application to distinguish groups in the analysis of various socio-economic phenomena and processes. The expediency of using this statistical approach as an effective method for allocating groups with similar aggregate units is substantiated. Based on the "Forbes" 2017 ranking, a clustering of international companies has been conducted, which will allow, based on qualitative analysis of isolated teams, to develop effective strategies for the development of the global economic environment.

Key words: "Forbes", dendrogram, cluster analysis, international economic environment, transnational corporation.

Постановка проблеми. Глобальне економічне середовище генерує значні обсяги статистичної інформації, що виражаються в різній формі, мають різноманітну структуру та ступінь деталізації. В процесі прийняття ефективних управлінських рішень наявні дані слугують цінним ресурсом, оскільки дають змогу комплексно оцінити досліджувані явища та процеси. Статистична наука володіє значною кількістю методів для дослідження соціально-економічних явищ та процесів, вибір яких базується на специфіці наявних даних, рівні їх деталізації, обсязі сукупності тощо. Слід зазначити, що існує цілий клас методів, орієнтований на ідентифікацію певних

груп у досліджуваній сукупності. Класифікація дає змогу на основі наукової методології, яка містить систему ідентифікуючих критеріїв, розподілити сукупність та групи. Серед методів групування слід виділити кластерний аналіз, який володіє низкою переваг. Для проведення кластеризації розроблено значну кількість методів, що дає змогу в процесі дослідження розглядати різноманітні методи групування та вибирати найкращий варіант з огляду на специфіку аналізу.

Аналіз останніх досліджень і публікацій. Дослідженню питань використання методів кластерного аналізу присвячено праці таких іноземних вчених, як М. Алдендерфер, Д. Гар-

сон, Г. Джеймс, Б. Еверіт, А. Кассамбара, Р. Кінг, С. Ландау, М. Лісі, М. Мейла, Ч. Ромесбург, Д. Стахл, К. Хеннінг. Розробленість представленого методу статистичного аналізу та його універсальність передбачають його застосування на постійній основі для виділення груп з типовими одиницями сукупності під час проведення комплексних наукових досліджень будь-яких соціально-економічних явищ та процесів.

Формулювання цілей статті (**постановка завдання**). Метою статті є дослідження застосування кластерного аналізу для виокремлення груп під час проведення аналізу різноманітних соціально-економічних явищ та процесів; проведення кластеризації міжнародних компаній на основі рейтингу "Forbes" за 2017 р. для розроблення ефективних стратегій розвитку глобального економічного середовища.

Виклад основного матеріалу дослідження. Інтенсифікація запровадження інноваційних технологій приводить до генерування значних масивів статистичної інформації в глобальному середовищі. Одним з важливих джерел даних виступають компанії, які функціонують у міжнародному економічному середовищі. З використанням методів статистичного аналізу можна провести дослідження діяльності цих компаній за різноманітними напрямками. З огляду на цілі дослідження та наявну інформацію передбачається застосування конкретних методів. В практиці міжнародних досліджень важливе місце посідають рейтинги компаній ("Fortune Global 500", "Forbes Global 2000" тощо), які дають змогу оцінити ситуацію на глобальному та національних ринках, визначити лідерів та аутсайдерів. Представлені рейтинги публікуються щорічно та дають можливість оцінити зміни в позиціях різноманітних транснаціональних корпорацій. Окрім побудови рейтингу компаній, слід згрупувати їх на основі наявної системи показників. Одним з ефективних методів, який може використовуватись для виокремлення груп та характеризується відсутністю суворих обмежень щодо якості даних і простотою інтерпретації отриманих результатів, є кластерний аналіз [1; 2].

Американська редакція видання "Forbes" щорічно публікує рейтинг "Тор 2000", в якому представляє найбільші міжнародні компанії. Методологія ранжування передбачає застосування системи показників, які зважуються на відповідні коефіцієнти з метою об'єктивного оцінювання. Для проведення кластерного аналізу використано чотири показники:

1) продажі, тобто вартість реалізованих компанією товарів або послуг у поточному році (млрд. дол. США);

2) ринкова вартість, тобто загальна вартість компанії на глобальному ринку в поточному році (млрд. дол. США);

3) прибуток, тобто отриманий компанією в поточному році прибуток (млрд. дол. США);

4) активи, тобто загальна вартість усіх ресурсів, які належать компанії (млрд. дол. США) [3].

Для якісного оцінювання отриманих груп також можна використати такі показники:

- назва компанії;
- рейтинг компанії;
- країна базування компанії;
- вид економічної діяльності, до якого належить компанія.

Для проведення кластеризації для 2 000 компаній з рейтингу "Forbes 2000" використано дані за 2017 р. Представлені дані є у вільному доступі на сайті платформи "Kaggle", яка розміщує масиви інформації про різноманітні соціально-економічні явища та процеси з метою проведення змагань з аналітики та передбачуваного моделювання [4].

Перший етап дослідження передбачає визначення кількості кластерів, на яку необхідно розділити досліджувану сукупність. В практиці кластерного аналізу використовуються різноманітні методи визначення кількості груп (метод «ліктя», або «кам'яного осипу», метод GAP-статистики, метод силуетів тощо). У цьому разі використано метод «ліктя» (рис. 1). Отримані результати показують, що доцільно розподілити підприємства, які увійшли в рейтинг "Forbes" у 2017 р., на чотири групи, оскільки за зазначеного методу кількість кластерів визначається в місці інтенсивного згасання суми квадратів обчислюваного показника [5–7].

В процесі роботи з первинними даними серед 2 000 компаній вибрано 1 314. На основі стандартизованих значень досліджуваних показників для вибраних міжнародних компаній у 2017 р. здійснено їх віднесення до певного кластеру. В процесі дослідження виявлено, що найкращі результати зі створення кластерів дає ієрархічний кластерний аналіз. Цей метод передбачав вибір двох основних параметрів.

1) Відстань. Вибрана Евклідова відстань, що є найбільш загальним типом відстані. Це геометрична відстань в багатовимірному просторі, яка обчислюється за формулою:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}.$$

2) Метод кластеризації. Найкращий результат показав метод мінімуму дисперсії Варда. Представлений підхід використовує методи дисперсійного аналізу для оцінювання відстаней між кластерами та мінімізує суму квадратів відстаней до центрів кластерів. На першому етапі для всіх виявлених кластерів обчислюються середні значення показників. Розраховуються квадрати евклідових відстаней від окремих представників кожного кластеру до середнього значення досліджуваного кластеру, після чого отримані відстані підсумовуються. Новий кластер утворюється завдяки об'єднанню наявних кластерів, в яких виявлено найменший приріст загальної суми відстаней [8; 9].

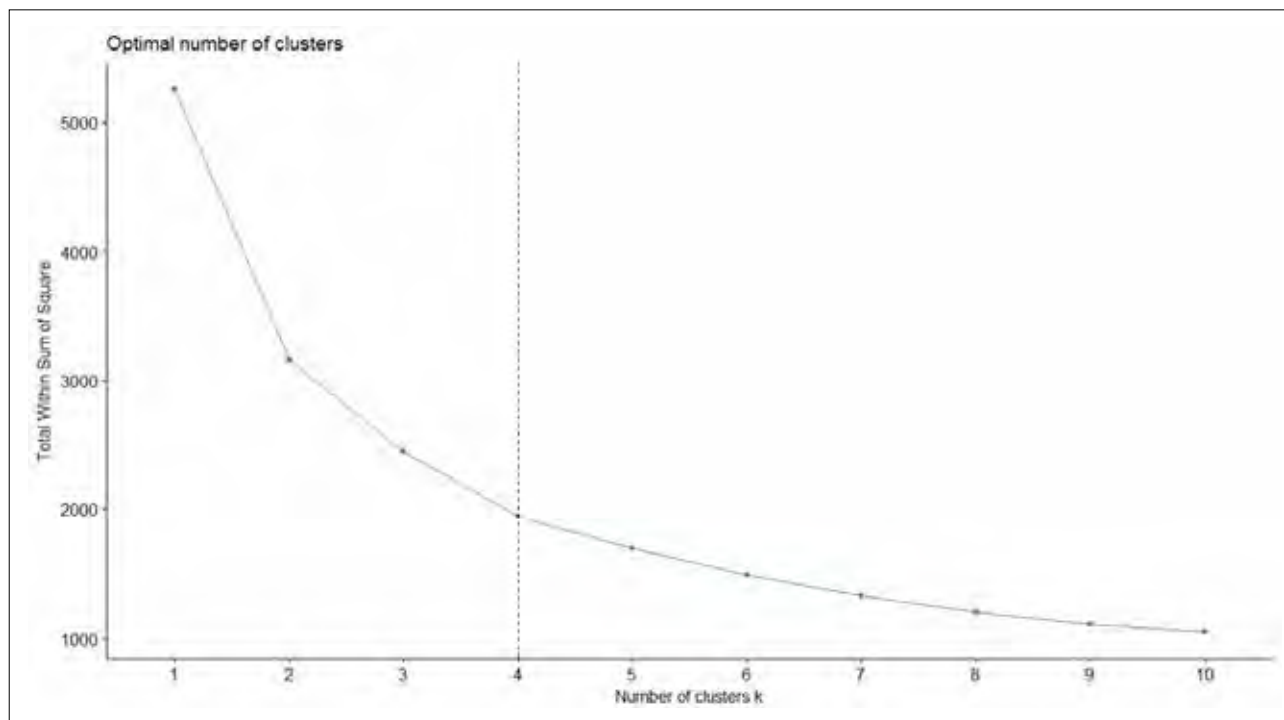


Рис. 1. Виділення кількості кластерів за допомогою методу «ліктя» для ТОП-2000 компаній з рейтингу “Forbes” у 2017 р.

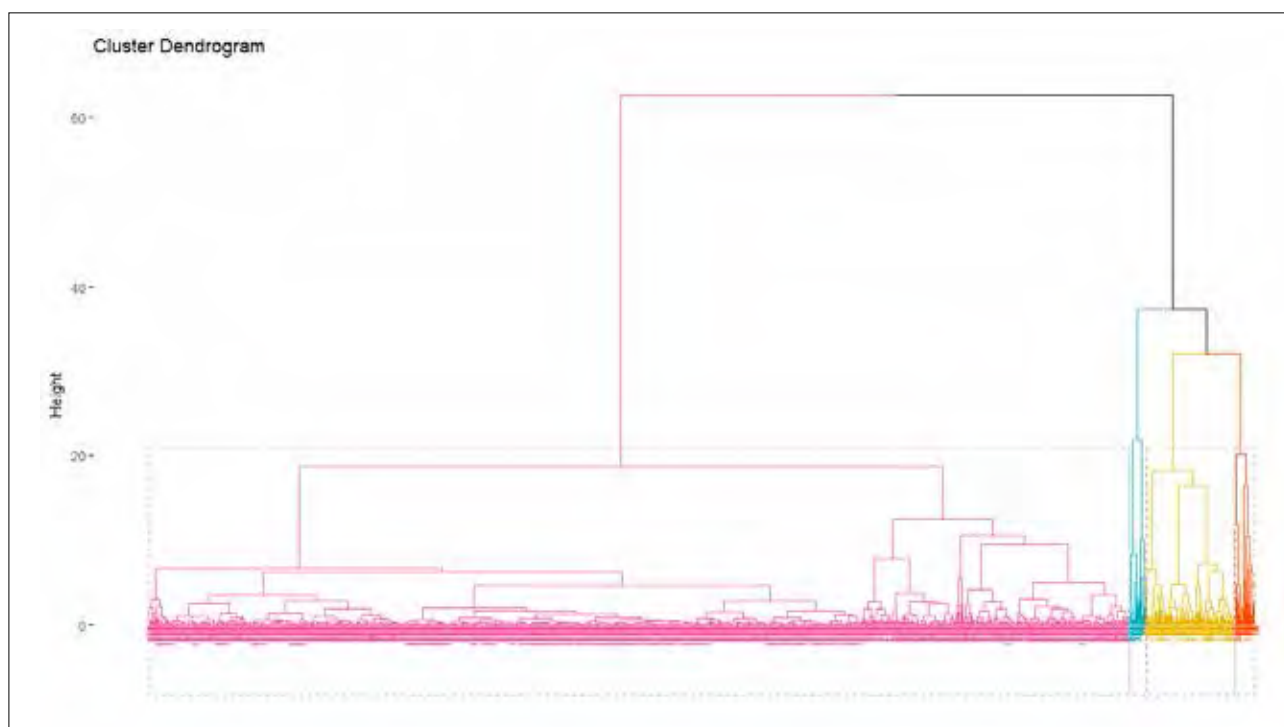


Рис. 2. Дендрограма компаній, які увійшли до рейтингу “Forbes” у 2017 р.

На рис. 2 представлено дендрограму для чотирьох кластерів. До першого кластеру увійшли 20 компаній, до другого – 24, до третього – 105, до четвертого – 1 165.

В табл. 1 представлено мінімальні та максимальні значення, середню, моду, а також перший та третій квартилі за кожним з досліджуваних показників загалом для 1 314 компаній та для

Таблиця 1

Основні показники для виокремлених кластерів у 2017 р., млрд дол. США

Показники	Продажі	Ринкова вартість	Прибуток	Активи
<i>Загалом</i>				
Мінімум	0,031	0,363	-13,000	0,289
Q ₁	6,000	9,200	0,478	15,100
Медіана	11,600	16,950	0,865	29,650
Середнє	22,796	32,363	1,707	115,920
Q ₃	24,375	33,300	1,875	77,550
Максимум	485,300	752,000	45,200	3 473,200
<i>Кластер 1</i>				
Мінімум	18,600	1,500	-1,900	1 458,000
Q ₁	40,850	43,270	3,875	1 736,000
Медіана	79,350	82,200	8,300	2 194,000
Середнє	77,910	116,640	13,430	2 250,000
Q ₃	109,500	173,350	22,475	2 595,000
Максимум	151,400	306,600	42,000	3 473,000
<i>Кластер 2</i>				
Мінімум	27,600	50,800	0,115	65,000
Q ₁	124,400	148,800	6,675	191,000
Медіана	175,800	227,100	9,750	234,400
Середнє	179,200	268,300	12,259	267,500
Q ₃	218,800	359,200	16,500	350,000
Максимум	485,300	752,000	45,200	620,900
<i>Кластер 3</i>				
Мінімум	11,700	11,900	-0,497	33,900
Q ₁	38,000	43,700	2,900	87,300
Медіана	65,200	62,200	4,900	193,000
Середнє	68,660	87,790	5,397	367,100
Q ₃	93,200	109,800	7,400	649,600
Максимум	153,700	277,100	15,000	1 412,400
<i>Кластер 4</i>				
Мінімум	0,031	0,363	-13,000	0,289
Q ₁	5,500	8,600	0,429	13,500
Медіана	10,400	14,700	0,761	25,500
Середнє	14,494	21,060	0,955	53,518
Q ₃	18,300	26,000	1,400	55,000
Максимум	81,100	123,800	7,800	986,900

кожного з виокремлених кластерів. Отримані результати дають можливість охарактеризувати кожен з виокремлених кластерів, порівняти їх між собою та визначити відмінності порівняно з усією сукупністю компаній. Аналіз показників загалом свідчить про значний розмах їх значень у грошовій формі, що пояснюється великою кількістю досліджуваних міжнародних компаній.

Для проведення порівняльного аналізу між виокремленими кластерами доцільно зіставити середні значення чотирьох показників (рис. 3). Найбільші відмінності між кластерами спостерігаються за розміром активів. Перший клас-

тер за цим показником істотно переважає інші кластери, що пояснюється входженням в цю групу найбільших банківських інституцій світу ("Agricultural Bank of China", "Bank of America", "Bank of China", "Barclays", "BNP Paribas", "China Construction Bank").

За середнім значенням ринкової вартості та продажами лідируючі позиції займає другий кластер, до складу якого увійшли компанії у сфері інноваційних технологій ("Alphabet", "Apple", "AT&T"), очолювана Вореном Бафітом американська холдингова компанія "Berkshire Hathaway", лідер у сфері інтернет-торгівлі "Amazon.com",

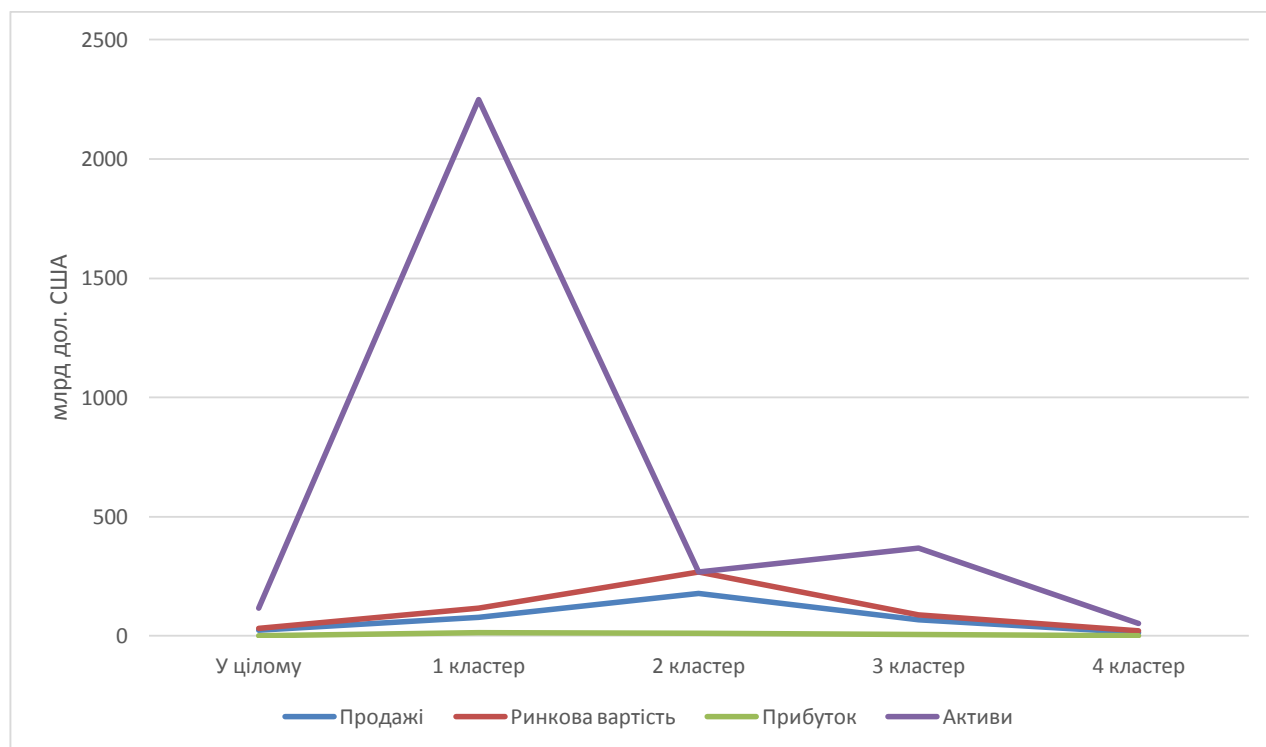


Рис. 3. Дендрограма компаній, які увійшли до рейтингу “Forbes” у 2017 р.

а також нафтогазова, нафтохімічна та вугільна транснаціональна корпорація “BP”.

Необхідно зазначити, що до складу третього кластеру увійшли такі компанії, як “Alibaba”, “Allergan”, “Allianz”, “Altria Group”, “Amerisource Bergen” та “Anheuser-Busch InBev”, а до четвертого – “Zi Group”, “77 Bank”, “AAC Technologies Holdings”, “ABB”, “Abbott Laboratories” та “AbbVie”.

Висновки. Кластерний аналіз є ефективним статистичним методом, який доцільно використовувати для розподілу досліджуваної сукупності на групи. Представлений підхід завдяки існуванню більше 100 алгоритмів кластеризації дає можливість розглянути різноманітні варіанти розбивки сукупності на кластери.

В процесі дослідження можна використовувати різноманітні комбінації наявних статистичних показників для вивчення будь-якого соціально-економічного процесу, у медичній сфері, для розгляду фізичних явищ тощо. Отримані в результаті кластеризації групи міжнародних компаній з рейтингу “Forbes” характеризуються схожими рисами для кожного виокремленого кластеру. В процесі дослідження глобального середовища сформовані групи компаній можуть детально аналізуватись для виявлення специфіки розвитку міжнародної економічної системи та прийняття відповідних управлінських рішень на корпоративному, національному та наднаціональному рівнях.

БІБЛІОГРАФІЧНИЙ СПИСОК:

1. Чубукова О.Ю., Ралле Н.В. Структурні інноваційної економіки – освіта, технологічні уклади, когнітивні технології. Науковий вісник Полісся. 2016. № 3 (7). С. 130–133.
2. Rzepka A., Ślusarczyk B. Correlation and dependence between: Business-Globalization-Information Society and Global Society. International Journal of Management Invention. 2016. Vol. 5. Issue 1. January. P. 39–45.
3. GLOBAL 2000: The World's Largest Public Companies. URL: <https://www.forbes.com/global2000/#2bf429a4335d>.
4. Forbes Top 2000 Companies. URL: <https://www.kaggle.com/ash316/forbes-top-2000-companies>.
5. Шитиков В.К., Мастицкий С.Э. Классификация, регрессия, алгоритмы Data Mining с использованием R. 2017. URL: <https://ranalytics.github.io/data-mining>.
6. Поиск структуры в данных. URL: <https://ru.coursera.org/learn/unsupervised-learning>.
7. Everitt B.S., Landau S., Leese M., Stahl D. Cluster Analysis. 5th Edition. John Wiley & Sons, 2011. 346 p. URL: <https://pdfs.semanticscholar.org/b05a/ca28ced3751d6302d8ce7396f80c90d612fd.pdf>.
8. Кластерный анализ. URL: <http://statsoft.ru/home/textbook/modules/stcluan.html>.
9. Kassambara A. Practical Guide to Cluster Analysis in R. New York: STHDA, 2017. 187 p.