

УДК 519.2

**М.Тененський- студент СНм-61, С. Гарматій, аспірант**

Тернопільський національний технічний університет імені Івана Пулюя, Україна

Науковий керівник: Н.Гарматій, канд. економ. наук, доц.

## МОДЕЛЮВАННЯ ДІЯЛЬНОСТІ АГРОФІРМ УКРАЇНИ МЕТОДОМ КЛАСТЕРИЗАЦІЇ

**M.Tenenskyi, S.Harmatii**

Ternopil Ivan Puluj National Technical University, Ukraine

Supervisor: N.Harmatii, Ph.D., Assoc. Prof.

### MODELING OF AGRICULTURAL FIRMS BY CLUSTERING METHOD IN UKRAINE

Класична задача в проведенні кластерний аналізу полягає в розбитті заданої вибірки об'єктів на певні підмножини – кластери. , так, щоб кожен кластер складався з схожих об'єктів, а об'єкти різних кластерів істотно відрізнялися. Завдання кластеризації можна віднести до статистичної обробки, а також до широкого класу завдань некерованого навчання. Відповідно до цього справедливим є твердження, що кластерний аналіз – це не один конкретний алгоритм, а загальна задача, яка може бути вирішена багатьма підходами. Так, суттєво можуть відрізнятися способи побудови кластерів та їх ефективного пошуку [1].

Однією з найпопулярніших концепцій кластеризації є утворення груп елементів на основі їх відстаней між собою, щільності у просторі даних чи інтервалах тощо. Тому кластеризацію доцільно буде сформулювати як задачу багатокритеріальної оптимізації. Інакше кажучи, основна мета кластерного аналізу – знаходження груп схожих об'єктів серед заданої вибірки даних. Застосовують кластерний аналіз в багатьох сферах: медицина, археологія, хімія, соціологія, державне управління, маркетинг тощо [1].

Розглянемо вісім з топ десяти агрофірм України відповідно до рейтингу Latifundist [3] та проведемо кластерний аналіз на основі даних, поданих в таблиці 1 (числові значення доходу та прибутку подано в млрд. грн).

Таблиця 1

Дані про дохід та прибуток агрофірм України за 2021 та 2022 роки

Назва	Дохід 2021	Прибуток 2021	Дохід 2022	Прибуток 2022
Кренел Трейд	85.8800	6.8800	68.08	5.1800
МХП	46.6838	-2.3440	46.78	-0.4440
АДМ Юкрейн	38.8800	-1.8950	25.18	-1.5200
Катеринопільський елеватор	21.1800	-0.4376	23.58	-0.0976
Ерідон	26.5800	0.3930	20.68	0.2640
НІБУЛОН	40.6800	-11.7200	15.28	-10.4200
Вайтерра Україна	18.8800	0.5280	13.78	1.0800
Сантрейд	33.6800	-2.6970	13.68	-1.7200

Першим кроком на шляху вирішення задачі кластерного аналізу є нормалізація вхідних даних, поданих в таблиці 1. Для цього варто скористатись наступними формулами:

$$z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j}, \quad (1)$$

$$\bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ij}, \quad (2)$$

$$s_j = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2} \quad (3)$$

Таким чином здійснюємо перехід від початкової матриці даних (див. табл. 1) до матриці нормованих значень з Z елементами (див. табл 2).

Таблиця 2

Нормовані дані, отримані шляхом використання формул 1-3

Назва	Дохід 2021	Прибуток 2021	Дохід 2022	Прибуток 2022
Кренел Трейд	2.354317	1.720813	2.198746	1.495975
МХП	0.383549	-0.193513	1.019066	0.125653
АДМ Юкрейн	-0.008823	-0.100329	-0.177229	-0.136520
Катеринопільський елеватор	-0.898771	0.202136	-0.265843	0.210056
Ерідон	-0.627262	0.374517	-0.426457	0.298162
НБУЛОН	0.081680	-2.139385	-0.725531	-2.305059
Вайтера Україна	-1.014414	0.402535	-0.808607	0.496985
Сантрейд	-0.270277	-0.266744	-0.814145	-0.185252

Наступним кроком потрібно побудувати матрицю зважених відстаней. Існує багато методів її побудови, серед яких варто виділити так звану «зважену» евклідову відстань. Вона слугує мірою відмінностей між кластерами та сприяє збільшенню контрастності між ними:

$$\rho(z_i, z_v) = \sqrt{\sum_l^4 w_l (z_{il} - z_{vl})^2}, \quad (4)$$

де  $w_l$  – це «вага показника», при чому має виконуватись умова  $0 \leq w_l \leq 1$  [2].

Для поданих даних варто використати набір із чотирьох ваг для кожного із наведених показників в таблиці 1. Так, дані за 2021 рік матимуть значення ваги рівне 0.15, а дані за 2022 рік – 0.35, адже вони є більш новими, відповідно більш важливими. Скориставшись формулою 4, отримаємо симетричну матрицю відстаней (див. рис. 1).

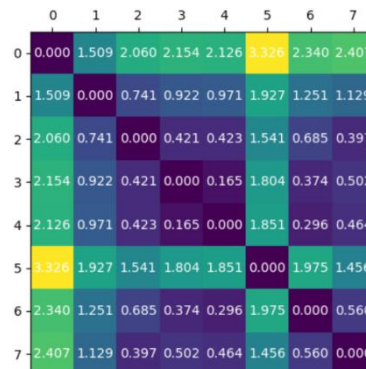


Рис. 1 Матриця «зважених» евклідових відстаней між кластерами

Відповідно до рисунку 1 розуміємо, що, використовуючи матрицю відстаней, можна реалізувати агломеративну ієрархічну процедуру кластерного аналізу. Відстань між двома сусідніми кластерами визначають за багатьма різними принципами, серед яких найчастіше виділяють принцип найменшої відстані (в даному випадку за відстань між кластерами приймається відстань між найближчими елементами цих кластерів) [2].

Його ідея полягає в тому, що на першому кроці кожне спостереження  $z_i$  розглядається як окремий кластер. В наступних кроках проводиться об'єднання двох найближчих кластерів та будується нова матриця відстаней, розмір якої зменшується на одиницю. Алгоритм завершується тоді, коли всі спостереження об'єднуються в один кластер.

Для прикладу, на рисунку 1, агрофірма номер 3 та 4 (Катеринопільський елеватор та Ерідон) перший кластер, адже відстань між ними є мінімальною серед усіх та становить 0.165. Нову матрицю «зважених» відстаней наведено на рисунку 2.

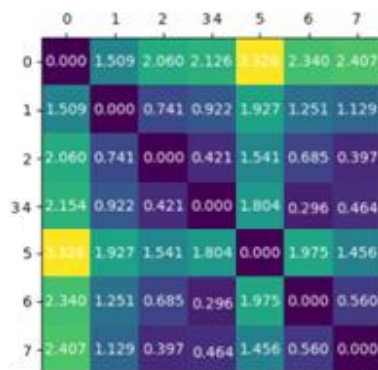


Рис. 2 Матриця «зважених» евклідових відстаней між кластерами після першого об'єднання кластерів з найменшою відстанню

З рисунку 2 бачимо, що далі слід об'єднати новоутворений перший кластер з агрофірмою 6 (Вайтерра Україна), адже відстань між ними є мінімальною та становить 0.296. Таке об'єднання, як вже було сказано раніше, проводиться доти, доки всі спостереження не буде об'єднано в один кластер. Для цього можна використати метод *linkage* з бібліотеки *scipy.cluster.hierarchy* для мови програмування *Python* [4]. Він не тільки завершить об'єднання за принципом найменшої відстані, але й побудує відповідну дендрограму (див. рис. 3).

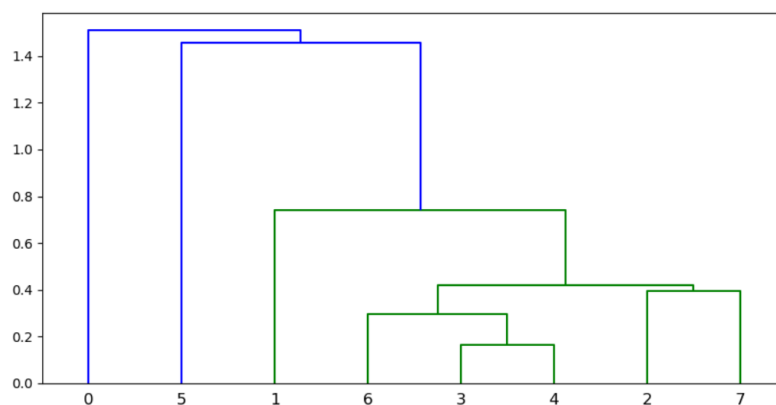


Рис. 3 Побудована дендрограма кластерів

Відповідно до рисунку 3, робимо наступні висновки щодо проведеного моделювання агрофірм України методом кластеризації:

- Відстань між «Катеринопільський елеватор» та «Ерідон» становить 0.1651 та разом вони формують перший кластер.
- Відстань між «Вайтерра Україна» та першим кластером становить 0.2958 та разом вони формують другий кластер.
- Відстань між «АДМ Юкрейн» та «Сантрейд» становить 0.3965 та разом вони формують третій кластер.
- Відстань між другим кластером та третім кластером становить 0.421 та разом вони формують четвертий кластер.
- Відстань між «МХП» та четвертим кластером становить 0.7411 та разом вони формують п'ятий кластер.
- Відстань між «НІБУЛОН» та п'ятим кластером становить 1.456 та разом вони формують шостий кластер.
- Відстань між «Кернел Трейд» та шостим кластером становить 1.5088 та разом вони формують сьомий кластер.

Варто розуміти також і те, що рішення задачі кластеризації є принципово неоднозначним. Причиною цьому є те, що результат кластеризації істотно залежить від ваг, вибір яких є

суб'єктивним та може змінюватись різними експертами. Також важливим фактором є те, що будь-які алгоритми кластеризації можуть давати як схожі, так і абсолютно різні кінцеві результати. Причиною цьому є, як не дивно, універсальність застосування кластерного аналізу, яка привела до появи надзвичайно великої кількості різних методів та підходів до його виконання та способів інтерпретації кінцевих результатів.

### **Література**

1. Методичні вказівки для виконання лабораторних робіт з дисципліни "Методи оптимального управління в економіці" / уклад. Н. М. Гарматий. Тернопіль: ТНТУ ім. І. Пулюя, 2017. 54 с.
2. Названо ТОП-10 найприбутковіших аграрних компаній України 2022 року. LATIFUNDIST.com. Головний сайт про агробізнес. 12 квіт., 2023. URL: <https://latifundist.com/novosti/61326-nazvano-top-10-najpributkovishih-agrarnih-kompanij-ukrayini-2022-roku>
3. Scipy cluster hierarchy linkage. URL: <https://docs.scipy.org/doc/scipy/reference/generated/scipy.cluster.hierarchy.linkage.html>