

DOI <https://doi.org/10.30525/978-9934-26-542-6-2>

HIERARCHICAL CLUSTERING AS A TECHNOLOGY FOR INTELLIGENT DATA ANALYSIS OF ENTERPRISE PRODUCTION INDICATORS

ІЄРАРХІЧНА КЛАСТЕРИЗАЦІЯ ЯК ТЕХНОЛОГІЯ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ДАНИХ ВИРОБНИЧИХ ПОКАЗНИКІВ ПІДПРИЄМСТВ

Holovnia S. A.

*Postgraduate Student at the Software
Engineering Department
National Technical University
of Ukraine
“Lutsk National Technical University
Lutsk, Ukraine*

Головня С. А.

*аспірант кафедри інженерії
програмного забезпечення
Національний технічний університет
України
“Луцький національний технічний
університет”
м. Луцьк, Україна*

Koval I. M.

*Postgraduate Student at the Software
Engineering Department
National Technical University
of Ukraine
“Lutsk National Technical University
Lutsk, Ukraine*

Коваль І. М.

*аспірант кафедри інженерії
програмного забезпечення
Національний технічний університет
України
“Луцький національний технічний
університет”
м. Луцьк, Україна*

У сучасних умовах економічного розвитку інтелектуальний аналіз даних (Data Mining) набуває все більшого значення. Зростаючі обсяги інформації вимагають від підприємств ефективних методів обробки та аналізу даних для прийняття обґрунтованих рішень. Застосування інтелектуального аналізу даних дозволяє виявляти приховані закономірності, прогнозувати тенденції та оптимізувати бізнес-процеси, що сприяє підвищенню конкурентоспроможності та ефективності підприємств.

Кластеризація є одним із ключових методів інтелектуального аналізу даних, який полягає у групуванні об'єктів на основі їхньої схожості [1, с. 11–12]. Основна мета кластеризації – виявлення структур у даних без попереднього знання про їхню природу. Існують різні види кластеризації [2, с. 25]. *Ієрархічна кластеризація*: формує вкладені кластери у вигляді дерева (дендрограми), де кожен вузол відповідає об'єднанню більш дрібних кластерів. *Кластеризація*

на основі розбиття (центроїдів): алгоритми, такі як k-середніх, розподіляють дані на k кластерів. **Класифікація на основі щільності:** алгоритми, як-от DBSCAN, виявляють кластери на основі щільності розташування точок даних, що дозволяє знаходити кластери довільної форми та ігнорувати шуми.

Ієрархічна кластеризація: застосування, переваги та недоліки.

Ієрархічна кластеризація є методом, який будує ієрархію вкладених кластерів у вигляді дендрограми [3]. Цей підхід широко використовується в біоінформатиці для класифікації генів, у маркетингу для сегментації клієнтів, у мануфактурі для групування виробництва та в інших галузях, де важливо розуміти багаторівневу структуру даних [4].

Переваги ієрархічної кластеризації [5]:

- Не вимагає попереднього визначення кількості кластерів.
- Дозволяє виявляти вкладені структури в даних.
- Забезпечує візуалізацію у вигляді дендрограми, що спрощує інтерпретацію результатів.

Недоліки:

- Висока обчислювальна складність, особливо для великих наборів даних.
- Чутливість до шумів та викидів у даних.
- Неможливість перегрупування об'єктів після об'єднання або розбиття кластерів.

Порівняння ієрархічної кластеризації з іншими методами (табл. 1).

Таблиця 1

Аналіз методів кластеризації даних

Характеристика	Ієрархічна кластеризація	К-середніх	DBSCAN
Потреба у визначенні кількості кластерів	Ні	Так	Ні
Виявлення кластерів довільної форми	Так	Ні	Так
Стійкість до шумів та викидів	Низька	Низька	Висока
Обчислювальна складність	Висока	Низька	Середня
Можливість перегрупування об'єктів	Ні	Так	Так

Висновок: ієрархічна кластеризація є більш гнучкою у виявленні вкладених структур та не вимагає попереднього визначення кількості кластерів, що робить її корисною для дослідження даних з невідомою структурою [6]. Однак, для великих наборів даних або даних з шумами

можуть бути більш ефективними інші методи, такі як DBSCAN або k-середніх.

Ієрархічна кластеризація широко застосовується в інтелектуальному аналізі даних для: сегментації ринку – визначення груп споживачів з подібними характеристиками для розробки цільових маркетингових стратегій [7]; аналізу тексту – групування документів за тематикою для покращення пошуку та організації інформації.

Основним засобом візуалізації результатів ієрархічної кластеризації є *дендрограма* – деревоподібна структура, яка показує послідовність об'єднання або розбиття кластерів, її широко використовують у маркетингу та інших галузях, де необхідно проаналізувати взаємозв'язки між об'єктами.

Існують два основних підходи до побудови дендрограм ієрархічної кластеризації:

– Агломеративний метод (знизу вгору) – кожен об'єкт спочатку розглядається як окремий кластер, після чого поступово об'єднується з найближчими кластерами, поки не утвориться єдина структура.

– Дивізивний метод (зверху вниз) – спочатку всі об'єкти розглядаються як один кластер, який потім поступово розділяється на підкласи до досягнення бажаної деталізації.

Основними метриками для обчислення відстані між кластерами є методи: одиночного зв'язку, метод повного зв'язку, метод середнього зв'язку, центроїдний метод, метод Уорда [8, с. 77].

Для демонстрації можливостей ієрархічної кластеризації було проаналізовано базу даних виробничих показників фармацевтичного підприємства, що включає інформацію про 151 найменування виробничих товарів (рис. 1), а також: назву продукту (Name), категорію (Category), ціну (Price), обсяги продажів в різні квартали з 2014 до 2015 року через аптеки та інтернет.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
	Name	Category	Price	2014_1quarter	2014_2quarter	2014_3quarter	2014_4quarter	2015_1quarter	2015_2quarter	2015_3quarter	2015_4quarter	2016_1quarter	2016_2quarter	2016_3quarter	2016_4quarter	2017_1quarter	2017_2quarter	2017_3quarter	2017_4quarter
1	Алтей корінь	1	158,55	15955	8161	13442	4537	6122	3341	12995	5184	15359	8779	13615	7602	7805	5213	11329	7604
2	Березові бруньки	1	200	19566	8181	10010	7737	8041	2029	11812	3240	19304	9082	11542	7797	9527	6022	16895	5868
3	Бузина чорної квітки	1	248	11434	10696	11340	7370	8832	4785	8880	3295	18260	12568	13938	7629	8181	4533	12339	7648
4	Валериана кореневиця з	1	185	12493	5535	10146	9033	9506	2888	9062	4999	22523	10568	9525	7173	7619	5854	9747	6854
5	Валериана кореневиця з	1	220,8	11029	11612	10258	7514	5812	2868	7087	4882	15239	8264	10760	6087	11199	4372	14059	8503
6	Вішня сушених	1	182	20243	7825	11527	6952	6615	4838	10750	6675	16523	8895	10312	8344	10104	4315	10902	8768
7	Галена	1	209,5	19911	8620	11710	6983	5759	1934	7141	5424	15978	10816	13623	6340	8247	5838	9987	5580
8	Голод листів кайлий	1	193	18568	8139	8415	9148	4512	1911	10808	3612	22747	10436	11207	7961	9002	3506	14688	7319
9	Горобини чорноплодні	1	246	16620	6491	13818	9879	4190	3777	12470	3412	13633	8447	11825	8871	10354	4261	14543	8888
10	Грицивія трава	1	157	15206	7902	9527	9671	7470	1879	9247	6190	16246	11013	14488	7679	10774	5960	15616	8540
11	Деревиця трава	1	388,5	19556	8874	9123	9424	5196	3294	15080	3889	13383	8102	15115	6454	6418	3552	14770	7750
12	Дуб кора	1	216	13264	8493	7850	5615	7875	4500	11518	4491	20214	9002	15075	6952	9648	5352	14483	5332
13	Евалантія прутковидного	1	205	16682	6739	12623	7719	4002	1673	7664	3098	17043	10452	11904	8440	6579	6728	9957	8429
14	Зіробою трава	1	192,65	13220	9388	14721	7436	4596	4151	13457	4385	13636	11699	11801	6595	10060	6197	15785	7929

Рис. 1. База даних фармацевтичного підприємства

На основі проведеного аналізу продукції методом ієрархічної кластеризації отримана дендрограма показала наявність трьох основних кластерів (рис. 2).

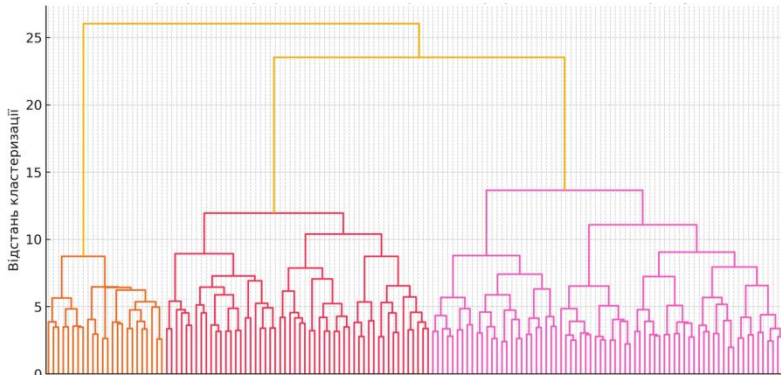


Рис. 2. Дендрограма ієрархічної кластеризації

Основні виявлені кластери:

- Кластер 1 (Популярні та високооб'ємні продукти). Включає лікарські засоби з високими обсягами продажів як через аптеки, так і через інтернет. Ці продукти мають стабільний попит упродовж усіх кварталів. Приклад: "Алтеї корені", "Березові бруньки".

- Кластер 2 (Продукти середнього рівня популярності). Препарати, які мають помірні обсяги продажів та демонструють сезонні коливання. Часто купуються в аптеках більше, ніж онлайн. Приклад: "Вільхи супліддя", "Валеріани кореневища з коренями".

- Кластер 3 (Нішеві або малопопулярні товари). Продукти з низьким попитом, які продаються в незначних обсягах. Ймовірно, їхня популярність залежить від маркетингових кампаній або специфічного використання. Приклад: деякі трав'яні збори або рідкісні препарати.

Взаємозв'язки між кластерами:

- Кластер 1 і Кластер 2 демонструють тісніший взаємозв'язок, оскільки їхні обсяги продажів мають спільні риси (стабільність або сезонність).

- Кластер 3 значно віддалений на дендрограмі, що свідчить про його відмінність за характеристиками продажів.

Висновки. Ієрархічна кластеризація дозволяє глибше аналізувати виробничі показники та визначати групи товарів із подібними характеристиками. Це дає змогу ухвалювати стратегічні рішення щодо оптимізації виробництва, коригування маркетингових стратегій та

управління асортиментом продукції. Використання дендрограм дозволяє наочно представляти структуру даних та виявляти приховані закономірності, що робить цей метод потужним інструментом інтелектуального аналізу даних.

Література:

1. Ланде Д., Субач І., Бояринова Ю. Основи теорії і практики інтелектуального аналізу даних у сфері кібербезпеки / Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського». Київ, 2018. 300 с.
2. Коверзнев Д. Ю. Інтелектуальна система кластеризації у комп'ютерних іграх для персоналізованого контенту. Чорноморський національний університет імені Петра Могили, 2024. 80 с.
3. Єсипенко Р. В. Ієрархічна кластеризація. *Scribd*. 20.11.2020. URL: https://www.scribd.com/document/609477892/%D0%86%D1%94%D1%80%D0%B0%D1%80%D1%85%D1%96%D1%87%D0%BD%D0%B0-%D0%BA%D0%BB%D0%B0%D1%81%D1%82%D0%B5%D1%80%D0%B8%D0%B7%D0%B0%D1%86%D1%96%D1%8F?language_settings_changed=English
4. Методологічні основи кластерного аналізу регіональної економіки. *Ucluster*. 21.08.2020. URL: <https://ucluster.org/universitet/klastery-ukraina/2005-study/competitiveness-theory/>
5. Тихонова К. В., Тихонов Є. С. Аналіз існуючих алгоритмів кластеризації даних. переваги та недоліки, 2020. *Слово науковця*.
6. Гуменюк К., Хмелівський Ю. Порівняння ієрархічної кластеризації та методу k-середніх. *Прикладні інформаційні технології*, Вінниця, Україна, 19 трав. 2023.
7. Роскладка Н., Дзигман О. Кластерний аналіз клієнтської бази даних підприємств сфери послуг. *Центральноукраїнський науковий вісник*. 2019. Економічні науки. С. 151–159.
8. Болюбаш Н. М. Інтелектуальний аналіз даних. Миколаїв : Чорноморський національний університет імені Петра Могили, 2023. 316 с.