

УДК: 338.1:330.5:331.5:517

Горай Оксана Олександрівна
здобувач вищої освіти,
Науковий керівник – Тельнова Ганна Володимирівна
д. е. н., доцент
професор кафедри бізнес-аналітики та цифрової економіки,
Національний авіаційний університет (Київ)

ВИЯВЛЕННЯ ВПЛИВУ НЕОКЛАСИЧНИХ ФАКТОРІВ ЕКОНОМІЧНОГО ЗРОСТАННЯ КРАЇН МЕТОДОМ КЛАСТЕРИЗАЦІЇ K-СЕРЕДНІХ

Анотація. Проаналізовано економічні показники за 2021 рік. Виявлено взаємозв'язки показників.

Ключові слова: економічне зростання, неокласичні фактори, метод k-середніх, кластеризація.

Інтелектуальний аналіз даних (Data Mining) – науковий і технологічний напрямок, пов'язаний з обробкою інформації та виявленням в ній закономірностей і тенденцій, які можуть застосовуватися при підтримці прийняття рішень.

Стадії інтелектуального аналізу даних:

- виявлення закономірностей (вільний пошук);
- використання виявлених закономірностей для прогнозу невідомих значень (прогностичне моделювання);
- аналіз виключень – стадія призначена для виявлення і пояснення аномалій, знайдених у закономірностях.

Кластеризація – це групування об'єктів (спостережень, подій) на основі даних (властивостей), що описують сутність об'єктів. Об'єкти усередині кластера повинні бути "схожими" один на одного і відрізнятися від об'єктів, що увійшли в інші кластери. Чим більше схожі об'єкти всередині кластера і чим більше відмінностей між кластерами, тим точніше кластеризація.

Метод k-середніх (k-means) - найбільш поширений серед неієрархічних методів алгоритм, також названий швидким кластерним аналізом. На відміну від ієрархічних методів, які не вимагають попередніх припущень відносно числа кластерів, для можливості використання цього методу необхідно мати гіпотезу про найбільш вірогідну кількість кластерів.

Перевагами алгоритму k-середніх є: простота та швидкість використання, зрозумілість і прозорість алгоритму.

Недоліки алгоритму k-середніх:

- алгоритм дуже чутливий до викидів, які можуть спотворювати середнє. Можливим вирішенням цієї проблеми є використання модифікації алгоритму - алгоритм k -медіани;

- алгоритм може повільно працювати на великих базах даних. Можливим вирішенням даної проблеми є використання вибірки даних.

У табл. 1 наведено склад досліджуваних показників.

Таблиця 1

Показники для аналізу

Var 1	рівень економічного зростання (GDP growth, %)
Var 2	валове нагромадження капіталу (Gross capital formation, % ВВП)
Var 3	валове заощадження (Gross savings, % ВВП)
Var 4	зайнятість населення старше 15 років (Employment to population ratio, 15+, total, %)
Var 5	приріст населення (Population growth, %)

Значення показників за двадцятьма країнами: Пакистан, Перу, Філіппіни, Польща, Португалія, Парагвай, Румунія, Російська Федерація, Сінгапур, Сальвадор, Сербія, Словацька Республіка, Словенія, Швеція, Таїланд, Туреччина, Україна, Уругвай, Узбекистан, Косово взяті за 2021 рік, збиралися з сайту Світового банку [1].

В результаті обробки даних було отримано характеристики дисперсії показників (рис. 1)

Variable	Between SS	df	Within SS	df	F	signif. p
Var1	47,879	4	115,0105	15	1,56113	0,235556
Var2	605,523	4	224,4772	15	10,11555	0,000356
Var3	852,414	4	308,3863	15	10,36541	0,000313
Var4	1610,079	4	206,0493	15	29,30268	0,000001
Var5	10,703	4	22,5355	15	1,78095	0,185193

Рис. 1. Аналіз дисперсії (Analysis of Variance)

Значення $p < 0,05$ говорить про значну відмінність. В нашому випадку це показники рівень економічного зростання та приріст населення.

На рис. 2 представлено дані щодо відстаней між кластерами, а на рис. 3 – графік середніх значень показників у кластерах.

Так, у перший кластер увійшли країни: Туреччина, Узбекистан; у другий кластер: Пакистан, Філіппіни, Польща, Португалія, Румунія, Сальвадор, Сербія, Словацька Республіка, Словенія, Україна, Уругвай; у третій кластер: Перу, Парагвай; у четвертий кластер: Російська Федерація, Сінгапур, Швеція, Таїланд; у п'ятий кластер: Косово.

Cluster Number	Squared distances above diagonal				
	No. 1	No. 2	No. 3	No. 4	No. 5
No. 1	0,00000	97,68970	135,2985	77,44463	109,9840
No. 2	9,88381	0,00000	46,0451	70,56014	224,7432
No. 3	11,63179	6,78566	0,0000	34,60012	401,0445
No. 4	8,80026	8,40001	5,8822	0,00000	326,9476
No. 5	10,48733	14,99144	20,0261	18,08169	0,0000

Рис. 2. Приведені відстані між класами

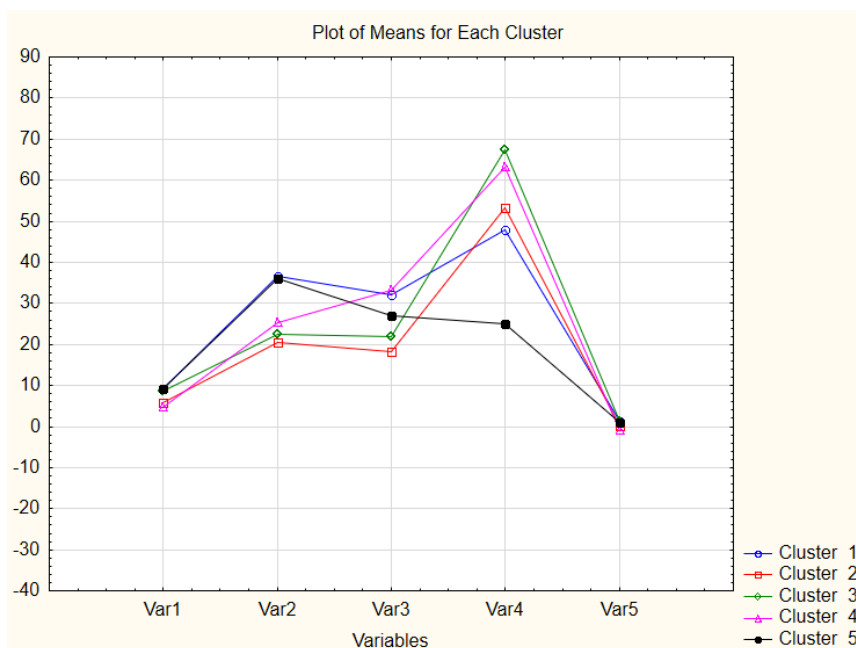


Рис. 3. Графік середніх значень показників у кожному кластері

Рівень економічного зростання показує збільшення обсягів товарів і послуг. Найбільше значення показує 1 кластер – 9,2%, а найменше 4 кластер – 4,7%. В 5 кластері також досить високий показник, тоді як у 2 та 3 кластерах середнє значення.

Валове нагромадження капіталу показує приріст нефінансових активів, які протягом тривалого часу використовуються в процесі виробництва. Найбільше значення показує 1 кластер – 36,5%, а найменше 2 кластер – 20,45%. В 5 кластері також досить високий показник, тоді як у 2 та 3 кластерах середнє значення.

Валове заощадження, тобто заощадження населення, підприємств та держави – головне джерело інвестицій. Найбільше значення показує 4 кластер – 33,25%, а найменше 2 кластер – 18,18%. В 1 кластері також досить високий показник, тоді як у 3 та 5 кластерах середнє значення.

Зайнятість населення старше 15 років показує рівень зайнятості населення. Найбільше значення показує 3 кластер – 67,5%, а найменше 5 кластер – 25,1%. В 4 кластері також досить високий показник, тоді як у 1 та 2 кластерах середнє значення.

Приріст населення показує перевищення народжуваності над смертністю. Найбільше значення показує 1 кластер – 1,4%, а найменше 3 кластер – -0,95%. В 3 кластері також досить високий показник, тоді як у 2 та 5 кластерах середнє значення.

Таким чином, можна зробити висновок, що найбільший рівень економічного зростання спостерігається у країнах, де найбільш високий рівень валового нагромадження капіталу та приріст населення, також в цих країнах низький показник зайнятості населення старше 15 років. Найнижчий рівень валового заощадження спостерігається у країнах з низьким рівнем валового нагромадження капіталу.

Список використаних джерел:

1. Сайт Світового банку. URL: <https://data.worldbank.org/indicator>.