

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ
імені ІГОРЯ СІКОРСЬКОГО»

Сталий інноваційний розвиток

Аналіз, моделювання і прогнозування розвитку суспільства: Візуалізація показників сталого розвитку

Навчальний посібник

Рекомендовано Методичною радою КПІ ім. Ігоря Сікорського
як навчальний посібник для здобувачів ступеня магістра

Укладач: І. М. Джигирей

Електронне мережне навчальне видання

Київ
КПІ ім. Ігоря Сікорського
2022

Рецензент: *Пишинограєв, І. О., канд. фіз-мат. наук, доцент,*
кафедра математичного моделювання економічних систем
КПІ ім. Ігоря Сікорського

Відповідальний
редактор *Бендюг, В. І., канд. техн. наук, доцент*

*Гриф надано Методичною радою КПІ ім. Ігоря Сікорського
(протокол № 5 від 26.05.2022 р.)
за поданням Вченої ради Навчально-наукового інституту прикладного системного аналізу
(протокол № 4 від 25.04.2022 р.)*

У навчальному виданні викладено методичні рекомендації до самостійної роботи з опанування веб-сервісів СЦД-Україна візуалізації та аналізу даних сталого розвитку. Видання забезпечує вивчення студентами теоретичного матеріалу, підготовку до семінарського заняття та буде корисним для самостійної роботи.

Реєстр. № НП 21/22-410. Обсяг 1,1 авт. арк.

Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»
проспект Перемоги, 37, м. Київ, 03056
<https://kpi.ua>

Свідоцтво про внесення до Державного реєстру видавців, виготовлювачів
і розповсюджувачів видавничої продукції ДК № 5354 від 25.05.2017 р.

© КПІ ім. Ігоря Сікорського, 2022

Зміст

Передмова	4
План семінару	5
Методичні рекомендації до проведення семінарського заняття	5
Індекс сталого розвитку	6
Веб-сервіс СЦД-Україна «Sustainable Development Index»	16
Кластерний аналіз	17
Завдання для самостійної роботи студента	19
Рекомендована література	22
Додаток А. <i>Вимоги до оформлення звіту з виконання завдань СРС</i>	23
Додаток Б. <i>Зразок титульного листа звіту</i>	24
Додаток В. <i>Приклад кластеризування методом k-середніх в ППП STATISTICA</i>	25

Передмова

Цей навчальний посібник є частиною інформаційно-методичного забезпечення навчальної дисципліни «Сталий інноваційний розвиток», а саме семінарського заняття «Аналіз, моделювання і прогнозування розвитку суспільства», і розроблений відповідно до програм підготовки магістрів спеціальностей 054 Соціологія, 081 Право, 111 Математика, 113 Прикладна математика, 121 Інженерія програмного забезпечення, 122 Комп'ютерні науки, 123 Комп'ютерна інженерія, 124 Системний аналіз, 126 Інформаційні системи та технології, 132 Матеріалознавство, 136 Металургія, 142 Енергетичне машинобудування, 143 Атомна енергетика, 144 Теплоенергетика, 151 Автоматизація та комп'ютерно-інтегровані технології, 163 Біомедична інженерія, 172 Телекомунікації та радіотехніка, 186 Видавництво та поліграфія, 231 Соціальна робота, 281 Публічне управління та адміністрування.

Кредитний модуль «Сталий інноваційний розвиток» відноситься до циклу загальної підготовки і є нормативним у підготовці магістрів вказаних спеціальностей. Зазначене семінарське заняття кредитного модуля є основою формування знань і умінь студентів щодо загальних питань оцінювання, аналізу, моделювання і прогнозування у сфері сталого розвитку. Робота студента над учбовим матеріалом у межах семінарського заняття «Аналіз, моделювання і прогнозування розвитку суспільства» містить такі види роботи: вивчення матеріалу за навчальними посібниками і веб-ресурсами, підготування до семінарського заняття, виконання завдань винесених на самостійне опрацювання, індивідуальні консультації.

Представлені матеріали мають на меті закріплення знань та набуття вміння застосовувати навички, отримані в процесі вивчення питань, які розглядають на семінарському занятті «Аналіз, моделювання і прогнозування розвитку суспільства» дисципліни «Сталий інноваційний розвиток». У виданні подано перелік питань, які студент повинен розглянути та з обговоренням яких проводиться семінарське заняття, надано методичні

вказівки до виконання завдань для самостійної роботи, наведено основні теоретичні положення. Навчальний посібник містить вимоги до оформлення звіту з виконання завдань для самостійної роботи і контрольні питання для самопідготовки студентів.

Метою семінарського заняття «*Аналіз, моделювання і прогнозування розвитку суспільства*» є ознайомлення з оцінюванням стану і розвитку країн та регіонів світу за допомогою агрегованих систем оцінювання (індексів), зокрема, індексу сталого розвитку, індексу екологічної керованості, індексу процвітання тощо, розгляд успішних закордонних та національних форсайт-досліджень різного рівня, а також ознайомлення із веб-сервісами візуалізації та аналізу даних Світового центру даних з геоінформатики та сталого розвитку (СЦД-Україна).

План семінару

1. Агреговані системи оцінювання розвитку суспільства у прикладах.
2. Глобальний і національний аналіз якості та безпеки життя людей.
3. Форсайт: системні світові конфлікти та глобальний прогноз на XXI століття.
4. Сценарії репрезентативних траєкторій концентрацій (RCPs) та спільних соціоекономічних шляхів (SSPs) МГЕЗК.

Методичні рекомендації до проведення семінарського заняття

1. Проводиться дискусія на тему «*Індекси сталого розвитку*». Метою обговорення є критичний аналіз особливостей, переваг і недоліків різних агрегованих систем оцінювання сталості або вимірів сталого розвитку, зокрема, індексу сталого розвитку, індексу екологічної керованості, індексу живої планети, індексу процвітання, екологічного сліду, індексу щасливої планети тощо.

2. Проводиться представлення студентами та обговорення прикладів актуальних агрегованих систем оцінювання розвитку Природи і Людини. Результат обговорення – розширення розуміння важливості, переваг і наслідків використання індексів для оцінюванням стану і розвитку країн та регіонів світу.

3. Проводиться ознайомлення студентів з веб-сервісами візуалізації та аналізу даних СЦД-Україна.

Індекс сталого розвитку¹

Методологія оцінювання та аналізу сталого розвитку містить модель сталого розвитку, яка є міждисциплінарним узагальненням моделей, відомих в природничих, економічних та соціальних галузях науки, та методіку застосування формальних статистичних методів та методів експертного оцінювання для аналізу процесів сталого розвитку.

Згідно з цією методологією, процес сталого розвитку будемо характеризують двома основними складниками: безпекою (C_{sl}) і якістю (C_{ql}) життя людей, а узагальнену міру сталого розвитку визнають за допомогою кватерніону:

$$\{Q\} = j \cdot w_{sl} \cdot C_{sl} + w_{ql} \cdot \|\vec{C}_{ql}\|. \quad (1)$$

Кватерніон $\{Q\}$ містить уявну зважену скалярну частину $j \cdot w_{sl} \cdot C_{sl}$, яка описує безпеку життя людей і зважену дійсну векторну частину, яка описує якість життя людей у просторі трьох вимірів: економічного (I_e), екологічного (I_{ec}) і соціально-інституційного (I_s). При цьому j набуває значення дійсної одиниці для нормального, регулярного стану розвитку суспільства за $C_{sl} > 0$, і значення уявної одиниці, коли суспільство переходить у стан конфлікту ($C_{sl} = 0$).

Згідно з результатами аналізу стану країн світу Гейдельберзького інституту досліджень міжнародних конфліктів використовують такі рівні конфліктів:

¹ методологія оцінювання сталого розвитку як кватерного функціоналу якості та безпеки життя людей [1]

- 0 – у країні відсутні конфлікти;
- 1 – політичний конфлікт без людських втрат;
- 2 – політичний конфлікт, в якому одна з сторін погрозувала розправою;
- 3 – слабо інтенсивний конфлікт з невеликими людськими втратами, в якому використовується тільки легка зброя;
- 4 – інтенсивний конфлікт з середніми людськими втратами, в якому або використовується тяжка зброя або задіяно значну кількість учасників;
- 5 – інтенсивний конфлікт з великими людськими втратами, в якому використовується тяжка зброя та задіяно значну кількість учасників.

Тому j можна визначити таким чином: $j = 1$ для рівнів конфлікту 1-4, $j = \sqrt{-1}$ для рівню конфлікту 5. Вагові коефіцієнти w_{sl} і w_{ql} у формулі (1) використано з метою вирівнювання масштабів компонент безпеки і якості життя (у випадку оцінювання сталого розвитку країн світу – $w_{sl} = 1/\sqrt[3]{11}, w_{ql} = 1/\sqrt{3}$).

Таким чином, під *індексом сталого розвитку* розуміють кількісну міру сталого розвитку, яка враховує безпеку та якість життя людей і обчислюється за формулою (1).

Для кожної країни евклідову норму радіус вектора якості життя людей (\vec{C}_{ql}) подають у такому вигляді:

$$\|\vec{C}_{ql}\| = \sqrt{I_{ec}^2 + I_e^2 + I_s^2}. \quad (2)$$

Кут відхилення α радіус вектора (\vec{C}_{ql}) від ідеального вектора (1, 1, 1) визначають через значення вимірів I_{ec}, I_e, I_s у такий спосіб:

$$\alpha = \arccos \frac{I_{ec} + I_e + I_s}{\sqrt{3} \cdot \sqrt{I_{ec}^2 + I_e^2 + I_s^2}}, \quad (3)$$

$$0 \leq \alpha \leq \arccos \frac{1}{\sqrt{3}},$$

Отже, норма радіус вектора характеризує якість життя людей, а просторове положення вектора у системі координат (I_{ec}, I_e, I_s) характеризує міру «гармонійності» сталого розвитку. Рівновіддаленість вектора від кожної із зазначених координат буде відповідати найбільшій гармонійності сталого

розвитку, а наближення цього вектора до однієї з координат буде вказувати на пріоритетний розвиток за відповідним виміром і нехтування двома іншими.

Назвемо величину $G = 1 - \alpha$ *ступенем гармонізації* сталого розвитку. Вона зростатиме у разі наближення G до 1 і зменшуватиметься у разі наближення G до 0.

Таким чином, *компонента якості життя* – інтегральна оцінка, яка враховує сумісно усі три виміри сталого розвитку, і, тим самим, відображає взаємозв'язок між трьома нероздільними сферами розвитку суспільства: економічною, екологічною та соціальною (табл. 1, 2). Ступінь гармонізації сталого розвитку відображає баланс між його економічним, екологічним та соціально-інституційним вимірами. Оцінки вимірів сталого розвитку, які використано для визначення компоненти якості життя повинні ґрунтуватися на даних про широкий спектр явищ різної природи. Також такі оцінки повинні бути інтегральними, тобто характеризувати певну сторону життєдіяльності людини, як цілісну систему. З метою кількісного оцінювання вимірів сталого розвитку використано принципи побудови ієрархічної системи показників та індексів, які визначаються як L_1 -норми:

$$I_i = \sum_{j=1}^n w_j x_{i,j}, i = \overline{1, m}, \sum_{j=1}^n w_j = 1 \quad (4)$$

в просторі показників $X^1 \times X^2 \times \dots \times X^m$, що характеризують економічний, екологічний та соціально-інституціональний розвиток кожної i -ї країни. Вагові коефіцієнти w_j в формулі (4) визначають експертним шляхом.

Використання формули (4) потребує узгодження різноманітних даних, як за одиницями вимірювання, так і за діапазоном значень. Тому, якщо більші значення показника X^i відповідають кращому стану сталого розвитку, то застосовують логістичне нормування значень показників за формулою:

$$C_{norm}(x_{i,j}) = \left(1 + e^{\frac{a-x_{i,j}}{b}} \right)^{-1}, \quad (5)$$

в якій параметри a і b обчислюються як середнє значення та стандартне відхилення за вибіркою країн, що аналізуються.

Таблиця 1 – Структура компоненти якості життя для оцінювання країн світу

	<i>Індекс</i>	<i>Об'єкт</i>	<i>Категорія політики</i>	<i>Індикатор</i>
Економічний вимір (33 %)	Глобальна конкурентоспроможність (50 %)	Ефективність стимуляції (50 %)	Розмір ринку (20 %)	Внутрішнє кредитування приватного сектору (30 %)
				Індекс обсягів імпорту (30 %)
				Витрати на кінцеве споживання домогосподарств (40 %)
			Рівень розвитку фінансового ринку (20 %)	Зовнішня заборгованість (50 %)
				Процентна ставка за депозитом (50 %)
			Технологічна готовність (20 %)	Гранти технічного співробітництва (50 %)
				Високотехнологічний експорт (50 %)
			Ефективність ринку праці (20 %)	ВВП на одного працюючого персоналу (50 %)
				Рівень зайнятості населення (50 %)
			Ефективність товарів та послуг (20 %)	Експорт товарів і послуг (40 %)
		Імпорт товарів і послуг (40 %)		
		Ринкова капіталізація лістингових компаній (20 %)		
		Основні вимоги (40 %)	Макроекономічне середовище (50 %)	Прямі іноземні інвестиції (20 %)
				Портфельні інвестиції (40 %)
			Інфраструктура (50 %)	Загальні резерви (40 %)
				Сільське господарство (20 %)
		Інновації (10 %)		Промисловість (30 %)
				ВВП на душу населення (50 %)
				Патентні заявки (30 %)
				Технічні спеціалісти, зайняті у сфері досліджень і розробок (40 %)
	Плата за використання інтелектуальної власності, платежі (30 %)			
	Економічна свобода (50 %)	Верховенство права (25 %)		Права власності (8 %)
				Цілісність уряду (8 %)
				Судова ефективність (8 %)
			Розмір уряду (25 %)	Державні витрати (8 %)
				Податковий тягар (8 %)
				Фінансовий ризик (8 %)
		Ефективність регулювання (25 %)	Свобода бізнесу (8 %)	
			Свобода праці (8 %)	
			Грошова свобода (8 %)	
		Відкриті ринки (25 %)	Свобода торгівлі (8 %)	
			Свобода інвестицій (8 %)	
Фінансова свобода (8 %)				
Соціальний вимір (33 %)	Здоров'я, оздоровлення та основні потреби (20 %)		Кількість пацієнтів у розрахунку на одного лікаря (11,1 %)	
			Лікарняні ліжка (11,1 %)	
			Рівень дитячої смертності (11,1 %)	
			Тривалість життя (11,1 %)	
			Витрати на охорону здоров'я (11,1 %)	
			Поширеність недоїдання (11,1 %)	
			Масштаби дефіциту продовольства (11,1 %)	
			Коефіцієнт материнської смертності (11,1 %)	
	Смертність (11,1 %)			
	Освіта (20 %)		Рівень грамотності (20 %)	
			Охоплення дітей шкільною освітою (20 %)	
			Охоплення дітей середньою шкільною освітою (20 %)	
			Охоплення населення вищою освітою (20 %)	
			Державні витрати на освіту (20 %)	

Закінчення табл. 1

	<i>Індекс</i>	<i>Об'єкт</i>	<i>Категорія політики</i>	<i>Індикатор</i>	
Соціально-інституційний вимір (33 %)	Особисті права та свобода (15 %)			Свобода слова	
	Особиста безпека (15 %)			Умисні вбивства (50 %)	
				Політичний терор (50 %)	
	Сприйняття корупції (20 %)				
	Соціально-інституційний вимір (10 %)			Абоненти мобільного зв'язку (12,5 %)	
				Користувачі Інтернету (12,5 %)	
				Телефони (12,5 %)	
				Транспортні засоби (12,5 %)	
				Кількість аеропортів (12,5 %)	
				Судноплавні водойми (12,5 %)	
			Асфальтовані магістралі (12,5 %)		
			Залізничні лінії (12,5 %)		
Екологічний вимір (33 %)	Екологічне здоров'я людини (40%)	Якість повітря (50 %)		Вплив твердих часток розміром менше 2,5 мкм (55 %)	
				Використання твердих видів палива у домогосподарствах (40 %)	
				Вплив озону (5 %)	
			Санітарія та питна вода (40 %)		Безпечна санітарія (40 %)
				Безпечна питна вода (60 %)	
		Важкі метали (5 %)		Вплив свинцю (100 %)	
		Керування відходами (5 %)		Контрольовані тверді відходи (100 %)	
		Сільське господарство (5 %)		Індекс сталого керування азотом (100 %)	
	Життє-здатність екосистем (60 %)	Викиди забруднювальних речовин (5 %)		Скореговані темпи зростання викидів оксидів азоту (50 %)	
				Скореговані темпи зростання викидів двоокису сірки (50 %)	
		Зміна клімату (40 %)		Скореговані темпи зростання викидів фторвмісних газів (10 %)	
				Скореговані темпи зростання інтенсивності викидів парникових газів (5 %)	
				Скореговані темпи зростання викидів двоокису вуглецю (55 %)	
				Скореговані темпи зростання викидів закису азоту (5 %)	
				Скореговані темпи зростання викидів метану (15 %)	
				Скореговані темпи зростання викидів сажистого вуглецю (5 %)	
				Темпи зростання викидів двоокису вуглецю, пов'язані зі зміною ґрунтово-рослинного покриву (2,5 %)	
				Викиди парникових газів на одну особу (2,5 %)	
		Водні ресурси (5 %)		Очищення стічних вод (100 %)	
		Рибний промисел (10 %)		Водні живі ресурси, виловлені траленням (30 %)	
				Морський трофічний індекс (35 %)	
				Стан запасів водних живих ресурсів (35 %)	
		Екосистемні послуги (10%)		Втрата луків (5 %)	
			Втрата лісового покриву (90 %)		
	Втрата водно-болотних угідь (5 %)				
Біорізноманіття і хабітат (25 %)		Морські природоохоронні території (20 %)			
		Захист наземного біому в національному контексті (20 %)			
		Захист наземного біому в глобальному контексті (20 %)			
		Індекс захисту видів (10 %)			
		Індекс репрезентативності природоохоронних територій (10 %)			
		Індекс захисту середовища існування видів (10 %)			
		Індекс захисту середовищ існування з високим рівнем біорізноманіття (10 %)			

Таблиця 2 – Структура компоненти якості життя для оцінювання регіонів України

	<i>Індекс</i>	<i>Категорія</i>	<i>Індикатор</i>
Індекс економічного виміру (33 %)	Індекс Інноваційно-кадрового потенціалу (50 %)	Інвестиційно-інноваційні можливості (70 %)	Капітальні інвестиції (33 %)
			Реалізація інноваційного потенціалу (33 %)
			Інноваційна діяльність (33 %)
		Ринок праці (30 %)	Потреба ринку праці (33 %)
			Економічна активність населення (33 %)
			Заробітна плата (33 %)
	Індекс конкурентоспроможності (50 %)	Базові потреби (60 %)	Валовий національний продукт (33 %)
			Виробнича сфера (33 %)
			Сфера послуг (33 %)
		Підприємницька діяльність (40 %)	Підприємницька активність (33 %)
			Споживчий ринок (33 %)
			Зовнішньоекономічна діяльність (33 %)
Індекс екологічного виміру (33%)	Стан довкілля (33%)	Атмосферне повітря (20%)	Комплексний індекс забруднення атмосфери (100%)
		Земельні ресурси (20%)	Екологічна стабільність території (100%)
		Природно-заповідний фонд (20%)	Частка земель заповідників та національних природних парків (100%)
		Стан природної та техногенної безпеки (20%)	Частка населення у зоні можливого хімічного забруднення (30%)
			Пожежі у лісах (40%)
			Пожежі на відкритих територіях (20%)
			Пожежі на торфовищах (10%)
	Відходи (20%)	Накопичено відходів I-III класів небезпеки (50%)	
		Накопичено відходів IV класів небезпеки (50%)	
	Екологічне навантаження і небезпеки (33%)	Викиди в атмосферне повітря (33%)	Викиди забруднювальних речовин від стаціонарних та пересувних джерел
		Водне навантаження (33%)	Інтенсивність забирання прісної води (50%)
		Утворення відходів (33%)	Тиск скидання забруднених зворотних вод (50%)
			Відходи I-III класів небезпеки (50%)
	Екологічне керування (33%)	Змінювання клімату (33%)	Відходи IV класу небезпеки (50%)
			Викиди діоксиду вуглецю на млн грн ВРП*
		Водоефективність (33%)	Водоемність ВРП* (50%)
			Економія водоспоживання у промисловості (50%)
			Керування відходами (33%)
	Утилізування і перероблення відходів IV класу небезпеки (50%)		
	Індекс соціально-інституційного виміру (33 %)	Стан соціальної інфраструктури (35 %)	Оздоровчі заклади (17 %)
			Забезпеченість населення лікарями (17 %)
Забезпеченість житлом (17 %)			
Абоненти мобільного зв'язку (17 %)			
Забезпеченість об'єктами спорту та ФК (17 %)			
Дошкільні навчальні заклади (17 %)			
Якість життя (33 %)			Медико-біологічні НС (33 %)
		Потерпілі від злочинів (33 %)	
		Кількість злочинних організацій (33 %)	
Економічне благополуччя (30 %)		Валовий регіональний продукт (20 %)	
		Витрати на харчування (20 %)	
		Нерівність розподілу матеріальних благ (20 %)	
		Низький прожитковий мінімум (20 %)	

* у цінах останнього звітного року

Закінчення табл. 2

	<i>Індекс</i>	<i>Категорія</i>	<i>Індикатор</i>
Індекс соціально-інституційного виміру (33 %)	Інституційний розвиток (33 %)	Ефективність державної влади (35 %)	Рівень довіри до Президента (14 %)
			Рівень довіри до Верховної Ради (14 %)
			Рівень довіри до місцевої влади (14 %)
			Рівень довіри до Кабміну (14 %)
			Державне управління (14 %)
			Індекс сприйняття корупції (14 %)
			Готовність до протестів (14 %)
		Розвиток інститутів громадянського суспільства (25 %)	Кількість членів громадських організацій (33 %)
			Рівень довіри до ГО (33 %)
			Активність на останніх виборах (33 %)
	Розвиток інформаційного суспільства (30 %)	Рівень довіри до ЗМІ (25 %)	
		Абоненти мережі Інтернет (25 %)	
		Розподіл газет (25 %)	
	Вплив релігійних інституцій (10 %)	Наявність ПК (25 %)	Кількість релігійних організацій (50 %)
			Довіра до церкви (50 %)
			Здоров'я людей (33 %)
	Розвиток людського потенціалу (33 %)	Розвиток здоров'я (30 %)	ВІЛ-інфіковані (33 %)
			Захворюваність на туберкульоз (33 %)
			Розвиток освіти (30 %)
		Відтворення населення (20 %)	
Гідна праця (20 %)			

У протилежному випадку, коли найбільші значення показника X^i відповідають гіршому стану сталого розвитку, використовується обернена до обрахованої за формулою (5) величина:

$$C_{norm}(x_{i,j}) = 1 - \left(1 + e^{\frac{a-x_{i,j}}{b}}\right)^{-1} \quad (6)$$

Сумарний вплив сукупності глобальних загроз на різні країни й групи країн світу оцінюють за допомогою компоненти безпеки життя людей C_{sl} як складової індексу сталого розвитку в формулі (1).

Кожній країні j ставлять у відповідність вектор:

$$Tr_j = (t_i^j), j = \overline{1, n} \quad (7)$$

з координатами $t_i^j \in [0,1], i = \overline{1, n}$, які характеризують ступінь прояву кожної з загроз.

В якості глобальних загроз виділено ті, що визначено головними в XXI сторіччі такими авторитетними міжнародними організаціями, як ООН, Всесвітня організація охорони здоров'я (ВООЗ), Всесвітній економічний форум (ВЕФ), Трансперенсі Інтернешнл, Глобальна мережа екологічного

слідую, Міжнародне енергетичне агентство, Інститут світових ресурсів, компанія BP p.l.c. та іншими:

ES – глобальне зниження енергетичної безпеки (вимірюють за допомогою індексу енергетичної безпеки, який враховує складник, що характеризує динаміку вичерпаності ресурсів, та складник, що характеризує обсяги використання відновлюваних джерел в національних енергосистемах);

BB – порушення балансу між біологічною можливістю Землі й потребами людства в біосфері у контексті змінювання демографічної структури світу (розраховують як відношення біоємності території та споживання біоресурсів країною);

GINI – наростаюча нерівність між людьми й країнами на Землі, вимірюють Gini-індексом, який змінюється в діапазоні від 0 до 100, де 0 – мінімальна нерівність, 100 – максимальна нерівність;

GD – поширення глобальних інфекційних хвороб, вимірюють сумарною кількістю померлих людей (мільйони на рік) від діарейних хвороб, СНІДу, туберкульозу й малярії;

IG – інформаційний виклик формується двома визначальними процесами сучасного інформаційного суспільства:

1. Людство постійно створює гігантські обсяги нових даних та інформації. У той же час людство спроможне осмислити, систематизувати, обробити і задокументувати значно менші обсяги нових даних та інформації. Таким чином, накопичується неосмислена і необроблена інформація, свого роду «інформаційна чорна діра». Цією інформаційною невизначеністю будуть пояснюватися непередбачувані і наперед неосмислені явища в природі і суспільстві. Цей складник інформаційного виклику оцінено з використанням ICT Development Index (IDI).

2. Вразливість тієї, чи іншої країни, чи території світу до дії кібератак. Цей складник інформаційного виклику оцінено з використанням Global Cybersecurity Index.

CP – наростання корупції (оцінено за допомогою індексу сприйняття корупції, який змінюється в діапазоні від 0 до 100, де 0 – максимальний рівень корупції, 100 – мінімальний);

WA – обмеженість доступу до питної води (відсоток населення країни, що не має доступу до питної води), яка визначається оберненою величиною до індикатора доступу до чистої питної води;

GW – глобальне потепління (вимірюється обсягом викидів вуглекислого газу в метричних тонах);

SF – нестабільність держави, вимірюється індексом вразливості держави (Failed States Index);

ND – індекс вразливості до природних катастроф, який розраховується на основі даних Міжнародної бази даних з катастроф Центру досліджень епідеміології катастроф;

CI – рівень конфлікту в країні, який визначає Гейдельберзьким інститутом досліджень міжнародних конфліктів (від 0 до 5);

PI – ступінь розповсюдження зброї, вимірюють комплексним, розробленим СЦД-Україна, індексом нерозповсюдження зброї, що складається з показників рівня озброєності та розвитку країни, розвитку наукового потенціалу та ступеню мілітаризації.

Значення показників, що характеризують прояв зазначених загроз, нормують за допомогою формул (5)–(6) таким чином, щоб більшому прояву загрози відповідали значення, ближчі до 1.

Значення компоненти безпеки життя C_{sl} для j -ї країни визначається як норма Мінковського вектора $\vec{S}_j = (s_j^i)$, $s_j^i = 1 - t_j^i$, $i = \overline{1, n}$:

$$C_{sl} = \|\vec{S}_j\| = \left(\sum_{i=1}^n (s_j^i)^p \right)^{\frac{1}{p}} \quad (8)$$

з параметром $p = 3$.

Введено також значення вразливості країни до сукупності глобальних загроз, яке є оберненою величиною до компоненти безпеки життя C_{sl} :

$$I_{vul} = \sqrt[3]{n} - C_{sl} \quad (9)$$

Таким чином, компонента безпеки життя – інтегральна оцінка, яка враховує сумарний вплив сукупності глобальних загроз на сталий розвиток країн світу, а індекс вразливості країни до впливу сукупності глобальних загроз відображає ступінь наближення цієї країни одночасно до усіх загроз у просторі, який визначається нормою Мінковського.

Кількісне оцінювання та аналіз загроз сталому розвитку для регіонів України здійснюється з комплексним використанням індикаторного та програмно-цільового підходу. Згідно із застосованою методикою експертним шляхом визначено чинники загроз з урахуванням особливостей регіонів України, а також вибрано кількісні показники, які відображають ступінь негативного впливу факторів загроз на сталий розвиток цих адміністративно-територіальних утворень (табл. 3).

Таблиця 3 – Структура компоненти безпеки життя для оцінювання регіонів України

<i>№ пп</i>	<i>Індекс</i>	<i>Загроза</i>	<i>Показник</i>	<i>Одиниця вимірювання</i>
1	Соціальна безпека	Зниження тривалості життя	Середня очікувана тривалість життя при народженні за регіонами	роки
2		Злочинність	Коефіцієнт злочинності за регіонами	кількість зареєстрованих злочинів на 100 тис. наявного населення
3		Погіршення здоров'я	Індекс смертності населення від туберкульозу, від новоутворень, від хвороб систем кровообігу, від СНІДу	умов. од
4		Корупція	Індекс сприйняття корупції	%
5		Соціальна нерівність	Індекс Джині (коефіцієнт нерівності в отриманні доходів)	умов. од
6	Економічна безпека	Рівень безробіття	Рівень зареєстрованого безробіття	% на кінець року
7		Технологічна відсталість	Ступінь зносу основних засобів	%
8		Зниження добробуту населення	Наявний дохід населення у розрахунку на одну особу	грн
9	Екологічна безпека	Забруднення довкілля	Індекс забруднення атмосферного повітря та поверхневих вод	умов. од
10		Дефіцит водних ресурсів	Сукупний обсяг місцевого річкового стоку і запасів підземних вод	тис. м ³ /км ²
11		Вплив на змінювання клімату	Викиди діоксиду вуглецю	т СО ₂ екв./км ²
12	Інші загрози	Техногенна небезпека	Середня сума збитків від ТНС за рік	тис. грн

Веб-сервіс СІД-Україна «Sustainable Development Index»

Сервіс *sdi.wdc.org.ua* надає зручний та відкритий доступ до даних зі сталого розвитку. Користувачу надано дані за окремими країнами² або регіонами України³ та можливість їх порівняння із результатами інших країн (регіонами) в табличній та графічній формах.

Меню «*Methodology*» (<http://sdi.wdc.org.ua/global/#tab-5470-2>,
<http://sdi.wdc.org.ua/ukraine/#tab-2235-2>)

Цей веб-ресурс містить опис методології оцінювання сталого розвитку як кватерного функціоналу якості та безпеки життя людей у глобальному та національному контекстах.

Меню «*Countries Ranking*» (<http://sdi.wdc.org.ua/global/#tab-8222-1>,
<http://sdi.wdc.org.ua/global/#tab-8222-2>) та «*Region Ranking*»
(<http://sdi.wdc.org.ua/ukraine/#tab-8948-1>, <http://sdi.wdc.org.ua/ukraine/#tab-8948-2>)

Ці веб-сервіси дають змогу переглянути ранжування країн світу та регіонів України як за головними агрегованими показниками метрики, так і за будь-яким складником представленої вище системи оцінювання.

Меню «*Countries Profiles*» (<http://sdi.wdc.org.ua/global/#tab-5470-4>) та «*Region Profiles*» (<http://sdi.wdc.org.ua/ukraine/#tab-2235-4>)

Ці веб-сервіси дають змогу відобразити профілі країн світу та регіонів України, побудовані на основі головних складників індексу сталого розвитку.

Меню «*Compare Countries*» (<http://sdi.wdc.org.ua/global/#tab-5470-5>) та
«*Compare Regions*» (<http://sdi.wdc.org.ua/ukraine/#tab-2235-5>)

Ці веб-сервіси відображають у вигляді пелюсткових діаграм показники сталого розвитку верхнього рівня країн світу або регіонів України, а саме індекси вимірів компоненти якості життя людей та показники загроз компоненти безпеки життя людей. Поля введення «Choose

² <http://sdi.wdc.org.ua/global/>

³ <http://sdi.wdc.org.ua/ukraine/>

countries» / «Choose regions» дають змогу обрати один і більше об'єктів для відображення.

Меню «*SDI Trends*» (<http://sdi.wdc.org.ua/global/#tab-4858-6>,
<http://sdi.wdc.org.ua/ukraine/#tab-4551-6>)

Ці веб-сервіси візуалізують значення індексу сталого розвитку та його компонент, а також індексів вимірів компоненти якості життя людей та показників загроз компоненти безпеки життя людей, за наявними даними з початку 2000-х років.

Меню «*Maps*» (<http://sdi.wdc.org.ua/global/#tab-4858-7>,
<http://sdi.wdc.org.ua/ukraine/#tab-4551-8>)

Ці веб-сервіси дають змогу переглянути значення усіх складників системи оцінювання у вигляді тематичних карт світу або України (див. табл. 1-3).

Меню «*Custom charts*»

Це сукупність веб-сервісів візуалізації, яка дає змогу побудувати вибірккові часові серії та бульбашкові діаграми для будь-якого складника системи оцінювання сталого розвитку.

Меню «*Countries Growth*» (<http://sdi.wdc.org.ua/global/#tab-4858-10>)

Цей веб-сервіс дає змогу переглянути зміни заданого користувачем показника протягом визначеного періоду часу.

Меню «*Clusters*» (<http://sdi.wdc.org.ua/global/#tab-4858-9>,
<http://sdi.wdc.org.ua/ukraine/#tab-4551-9>)

Цей веб-сервіс призначено для кластерного аналізу країн світу та регіонів України на основі заданих користувачем показників методом *k*-середніх.

Кластерний аналіз

Прикладну статистику можна охарактеризувати як наукову дисципліну про прийоми, математичні методи й моделі оброблення статистичних даних з метою аналізування й ухвалювання рішень. Прикладна статистика охоплює статистику випадкових величин; багатомірний статистичний аналіз;

статистику часових рядів та випадкових величин; статистику об'єктів нечислової природи.

Багатомірний статистичний аналіз вивчає оброблення багатомірних статистичних даних з метою виявлення взаємозв'язків, їх характеру та структури. До методів багатомірного статистичного аналізу можна віднести множинний кореляційний та регресійний аналіз, методи *зниження розмірності* багатомірного простору (наприклад, метод факторного аналізу), методи багатомірного *класифікування* (наприклад, методи кластерного й дискримінантного аналізу) та інші.

Метод класифікування полягає у розбиванні великої кількості об'єктів на однорідні групи. *Класифікація* – це певна закономірність, яка дає змогу зробити висновок щодо визначення характеристик цих груп. Розрізняють одномірне⁴ і багатомірне класифікування. Для класифікування використовують різні методи: дерева розв'язків; баєсове класифікування; генетичні алгоритми; статистичні методи, зокрема, лінійну регресію; штучні нейронні мережі та багато інших.

Кластерний аналіз (автоматичне класифікування «без вчителя», класифікаційний аналіз, числова таксономія) об'єднує методи класифікування багатомірних об'єктів або подій у відносно однорідні групи (кластери). Об'єкти всередині кластеру повинні бути схожі один на одного у більшій мірі ніж на об'єкти інших кластерів і відрізнятися від об'єктів інших кластерів більше ніж від об'єктів власного кластеру. Характеристиками кластеру є внутрішня однорідність та зовнішня ізольованість.

Існує багато методів кластеризування: ієрархічні методи (агломеративні та дивизимні), алгоритми розбивання, зокрема ітераційні методи (метод k-середніх; метод «сходження на пагорб»), факторні методи, нейромережні методи, методи ущільнення об'єктів, грид-методи, методи моделювання та інші. Також вирізняють двохідне об'єднання, коли кластеризування здійснюють одночасно і за об'єктами, і за змінними, які ці об'єкти

⁴ за однією ознакою

характеризують.

Метод k -середніх (швидкий кластерний аналіз). Алгоритм k -середніх будує k кластерів, розташованих на максимальній відстані один від одного. Середні у кластері для усіх змінних максимально відрізняються.

Крок 1. Початковий розподіл об'єктів за кластерами. Вибір числа k та початкових об'єктів-центроїдів: а) вибір k -спостережень для максимізування початкової відстані; б) сортування відстаней і задавання спостережень на постійних інтервалах; в) вибір перших k -спостережень. Результатом цього кроку є призначення кожного об'єкту певному кластеру.

Крок 2. Ітераційний процес. Обчислення центрів кластерів та перерозподілення об'єктів. Процедуру повторюють доки стабілізуються кластерні центри або досягнуто максимальну кількість ітерацій.

Крок 3. Перевірення якості кластеризування та ступеню впливу ознак. У якості критерію якості кластеризування можна взяти відношення середньої внутрішньокластерної відстані до середньої міжкластерної відстані. Чим ближче значення такого критерію до нуля, тим краще проведено кластеризування.

Ступінь впливу ознак визначають як відношення функції Фішера, розрахованої емпірично на основі внутрішньо- і міжкластерних відстаней у деякому наборі об'єктів, до функції Фішера, обчисленої теоретично за відповідних ступенів свободи вибірки для внутрішньо- і міжкластерних відстаней за умови рівності їхніх дисперсій. Чим більше значення такого критерію для певної змінної, тим вищою є ступінь впливу цієї ознаки на хід та результати кластеризування.

Завдання для самостійної роботи студента

(самостійне ознайомлення з веб-сервісом <http://sdi.wdc.org.ua/>)

Послідовність виконання та заповнення звіту.

1. Перейдіть до веб-сервісу World SDI (<http://sdi.wdc.org.ua/global>) якщо Ваш номер у списку академічної групи є непарним (далі див. завдання з позначкою «а»), і до веб-сервісу Ukraine SDI (<http://sdi.wdc.org.ua/ukraine>),

якщо – парним (далі див. завдання з позначкою «б»).

1.а. За допомогою меню «*Countries Ranking / Custom indeces*» отримайте таблицю країн світу для будь-якого показника та року та відсортуйте за спаданням або зростанням (Rank " " ▲).

1.б. За допомогою меню «*Regions Ranking / Custom indeces*» отримайте таблицю регіонів України для будь-якого показника та року та відсортуйте за спаданням або зростанням (Rank " " ▲).

[1] Додайте у звіт скріншот перших п'яти-десяти позицій таблиці.

2.а. За допомогою меню «*Countries Profile*» отримайте профіль сталого розвитку будь-якої країни світу для останнього наявного звітного року.

2.б. За допомогою меню «*Regions Profile*» отримайте профіль сталого розвитку будь-якого регіону України для останнього наявного звітного року.

[2] Додайте у звіт скріншот пелюсткової діаграми індексів вимірів компоненти якості життя людей та пелюсткової діаграми показників загроз компоненти безпеки життя людей.

3.а. За допомогою меню «*Compare Countries*» отримайте профіль сталого розвитку для України та будь-якої країни світу для останнього наявного звітного року.

3.б. За допомогою меню «*Compare Regions*» отримайте профіль сталого розвитку для м. Київ та будь-якого регіону України для останнього наявного звітного року.

[3] Додайте у звіт скріншот порівняльного профілю (пелюсткова діаграма індексів вимірів компоненти якості життя людей, пелюсткова діаграма показників загроз компоненти безпеки життя людей та таблиця основних показників).

4.а. За допомогою меню «*SDI Trends*» отримайте діаграми трендів показників системи оцінювання сталого розвитку для будь-яких трьох країн

світу.

4.6. За допомогою меню «*SDI Trends*» отримайте діаграми трендів показників системи оцінювання сталого розвитку для будь-яких трьох регіонів України.

- [4] Додайте у звіт скріншот будь-яких двох діаграм трендів. Зробіть висновки.

5. За допомогою меню «*Maps*» отримайте тематичну карту за деяким показником для останнього наявного звітного року.

- [5] Додайте у звіт скріншот будь-яких двох тематичних карт.

6.a. За допомогою меню «*Custom charts / Scatter Plots*» отримайте бульбашкову діаграму для країн світу, де, наприклад, вісь абсцис представляє компоненту якості життя (англ. Quality of Life Component), вісь ординат – компоненту безпеки життя (англ. Security of Life Component), а розмір бульбашки ВВП у розрахунку на одну особу (англ. GDP per capita), або інший показник.

6.б. За допомогою меню «*Custom charts / Scatter Plots*» отримайте бульбашкову діаграму для регіонів України, де, наприклад, вісь абсцис представляє компоненту якості життя, вісь ординат – компоненту безпеки життя, а розмір бульбашки валовий регіональний продукт, або інший показник.

- [6] Додайте у звіт скріншот діаграми. Зробіть висновки.

7.a. За допомогою меню «*Clusters*» отримайте класифікацію країн світу на три, чотири і п'ять кластерів для певного показника або показників, наприклад, індексу сталого розвитку або сукупно індексів вимірів сталого розвитку. Кількість отримуваних кластерів можна вказати у полі «*Number of clusters*» ліворуч.

7.б. За допомогою меню «*Clusters*» отримайте класифікацію регіонів України на три, чотири і п'ять кластерів для певного показника або показників, наприклад, індексу сталого розвитку або сукупно індексів

вимірів сталого розвитку. Кількість отримуваних кластерів можна вказати у полі «Number of clusters» ліворуч.

- [7] Додайте у звіт скріншоти трьох карт світу з розбиттям на три, чотири і п'ять кластерів. Зробіть висновки щодо віднесення України (м. Київ або іншого регіону для варіанту «б») до певного кластеру для кожної класифікації.

8. У висновку необхідно представити основні складники веб-сервісу, що розглянуто, критично проаналізувати ефективність візуальної підтримки аналізу даних, зручність і повноту веб-сервісу.

Використані та рекомендовані джерела

1. Аналіз сталого розвитку — глобальний і регіональний контексти / Міжнародна рада з науки (ISC) та ін.; наук. кер. проекту М. З. Згуровський. — К. : КПІ ім. Ігоря Сікорського, 2019. — Ч. 1. Глобальний аналіз якості і безпеки життя (2019). — 216 с. URL:
<http://wdc.org.ua/sites/default/files/SD2019-P1-FULL-UA.pdf>
2. Аналіз сталого розвитку — глобальний і регіональний контексти / Міжнародна рада з науки (ISC) та ін.; наук. кер. проекту М. З. Згуровський. — К. : КПІ ім. Ігоря Сікорського, 2019. — Ч. 2. Україна в індикаторах сталого розвитку (2019). — 112 с. URL:
<http://wdc.org.ua/sites/default/files/SD2019-P2-FULL-UA.pdf>
3. Веб-сервіс «Sustainable Development Index» / СІД-Україна. URL:
<http://sdi.wdc.org.ua/>
4. Соціально-економічний потенціал управління станом довкілля. Комп'ютерний практикум [Електронний ресурс] : навч. посіб. для здобувачів ступеня магістра за освітньою програмою «Комп'ютерний моніторинг та геометричне моделювання процесів і систем» спеціальності 122 «Комп'ютерні науки» / КПІ ім. Ігоря Сікорського ; уклад.: Караєва Н.В., Варава І.А. – Електронні текстові дані (1 файл: 6,39 Мбайт). – Київ : КПІ ім. Ігоря Сікорського, 2020. – 67 с. URL:
<https://ela.kpi.ua/handle/123456789/32716>

Додаток А. Вимоги до оформлення звіту з виконання завдань СРС

Звіт повинен містити титульний аркуш, хід виконання завдань та висновки.

Звіт оформлюється на аркушах формату А4 шрифтом Times New Roman 14 розміру (інтервал абзаців – одинарний (1,0) або полуторний (1,5), вирівнювання – за шириною, поля сторінок: ліве, верхнє, нижнє – 20 мм, праве – 10 мм).

Першою сторінкою звіту є титульний аркуш (Додаток Б), який включають до загальної нумерації сторінок, але номер сторінки не ставлять. На наступних сторінках номер проставляють у верхньому правому куті.

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ ІМЕНІ ІГОРЯ СІКОРСЬКОГО»
ХІМІКО-ТЕХНОЛОГІЧНИЙ ФАКУЛЬТЕТ

Кафедра кібернетики хіміко-технологічних процесів

Звіт

з виконання завдань самостійної роботи студента

«Візуалізація показників сталого розвитку»

з кредитного модуля «Сталий інноваційний розвиток»

Виконав(ла)
студент(ка) гр. **ХА-11**
Гончаренко М.І.

Прийняв(ла): _____

Кількість балів: _____

Дата: _____

Підпис: _____

Київ 2022

Кластерний аналіз у пакеті STATISTICA проводять за допомогою модуля «Cluster Analysis», який можна викликати за допомогою підменю «Multivariate Exploratory Techniques» меню «Statistics». Якщо ознаки об'єктів представлено у різних одиницях вимірювання варто виконати попереднє нормування даних за допомогою підменю «Standardize Block» меню «Edit».

У діалоговому вікні «Clustering Method» представлено такі методи

- «Joining (tree clustering)» – агломеративні методи;
- «K-means clustering» – метод k -середніх;
- «Two-way joining» – двовхідне об'єднання.

У разі вибору агломеративних методів з'являється діалогове вікно «Cluster Analysis: Joining (tree clustering)», у якому можна вибрати алгоритм об'єднання кластерів («Single Linkage», «Complete Linkage», «Unweighted pair-group average», «Weighted pair-group average», «Unweighted pair-group centroid», «Weighted pair-group centroid» або «Ward's method») та міру відстаней («Squared Euclidean distances», «Euclidean distances», «City-block (Manhattan) distances», «Chebychev distance metric», «Power», «Percent disagreement⁶», «1-Person r »). Опцію «Distance matrix» у рядку «Input file» передбачено для випадку, якщо початкові дані представлено у вигляді мір подібності. «Блок MD deletion» призначено для вибору способу оброблення неповних даних («Casewise» – неповні дані за рядками видаляють; «Mean substitution» – пропущені дані замінюють середніми значеннями).

У вікні «Joining Results» можна переглянути хід об'єднання у вигляді горизонтальної та вертикальної дендрограми, схеми об'єднання і графіка схеми об'єднання, отримати вихідну матрицю відстаней і описові

⁵ факультативно

⁶ для категоріальних даних

характеристики, зокрема середніми значеннями та стандартними відхиленнями для кожного об'єкта.

У разі вибору методу k -середніх з'являється діалогове вікно «Cluster Analysis: K-Means Clustering», яке дає змогу налаштувати параметри кластеризування цим методом: показники, за якими проходитиме кластеризування, кількість кластерів, максимальну кількість ітерацій, спосіб задавання початкових центрів кластерів. За варіантом «Choose observations to maximize initial between-cluster distances» обирають перші k об'єктів, які стають центрами кластерів. Центри замінюють наступними об'єктами, якщо найменша відстань до будь-якого з них більша ніж найменша відстань між кластерами. Таким чином, початкові відстані між кластерами максимізовано. За варіантом «Sort distances and take observations at constant intervals» відстані між об'єктами відсортовують та обирають в якості початкових центрів кластерів об'єкти на постійних інтервалах. За варіантом «Choose the first N (Number of cluster)» у якості початкових центрів кластерів обирають перші k спостережень.

У вікні «K-Means Clustering Results» можна переглянути результати класифікування. У верхній частині діалогового вікна представлено вихідні дані, параметри кластеризування та кількість ітерацій, за які досягнуто стабілізування центрів кластерів. У нижній частині розташовано кнопки виведення результатів процедури: результати дисперсійного аналізу, середні значення кластерів, отримані відстані, графік середніх, статистичні характеристики кожного кластеру, об'єкти кожного кластеру та відстані від кожного з членів кластеру до центру кластеру.

Вихідні дані представлено у табл. В.1 та на рис. В.1. Відобразимо об'єкти у двомірному просторі xu (див. рис. В.1).

Прийmemo $k=2$ та скористаємось інструментами пакету STATISTICA для отримання проміжних результатів. Відкриваємо підменю «Multivariate Exploratory Techniques» меню «Statistics» і натискаємо команду «Cluster Analysis». У діалоговому вікні «Clustering Method» обираємо «K-Means

clustering», а у діалоговому вікні «Cluster Analysis» вказуємо, що класифікуватимемо об'єкти («Cluster: Cases (rows)») з отриманням двох кластерів («Number of clusters: 2»). У вікні «K-Means Clustering Results» обираємо команду «Members of each cluster & distances» на вкладці «Advanced».

Таблиця В.1 – Вихідні дані для розрахунку

Об'єкт	Ознаки	
	x	y
1	1	1
2	2	1
3	2	2
4	3	2
5	5	4
6	4	5
7	5	5
8	6	5
9	5	6
10	6	6

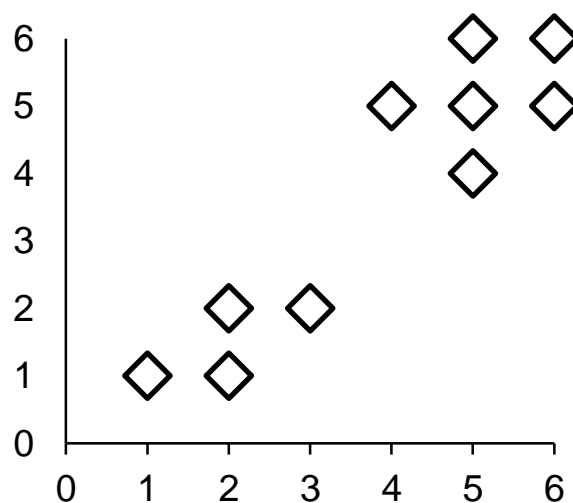


Рисунок В.1 – Об'єкти для класифікування

Бачимо, що один кластер утворили об'єкти 1, 2, 3 та 4, а другий – решта об'єктів. Розрахуємо суму внутрішньокластерних відстаней

$$d_w = (0,79 + 0,35 + 0,35 + 0,79) + (0,83 + 0,83 + 0,17 + 0,60 + 0,60 + 0,83) = 6,1$$

для $n_w = 10$. У вікні «K-Means Clustering Results» обираємо команду «Summary: Cluster means & Euclidean distances» та під головною діагоналлю отриманої матриці відстаней знаходимо відстань між двома кластерами $d_b = 3,4$ для $n_b = 1$. Розраховуємо значення критерію якості кластеризування за формулою

$$Q = \frac{d_w/n_w}{d_b/n_b},$$

де d_w та d_b – суми внутрішніх та міжкластерних відстаней, n_w та n_b – кількість внутрішніх та міжкластерних відстаней.

$$Q = \frac{6,1/10}{3,4/1} = 0,18,$$

Прийmemo $k=3$ та за допомогою пакету STATISTICA отримаємо, що перший кластер утворили об'єкти 1, 2, 3 та 4, другий – об'єкти 5, 6 та 7, а третій – об'єкти 8, 9 та 10. Сума внутрішньокластерних відстаней становить

$$d_w = (0,79 + 0,35 + 0,35 + 0,79) + (0,53 + 0,53 + 0,33) + (0,53 + 0,53 + 0,33) = 5,1$$

для $n_w = 10$, сума відстаней між кластерами $d_b = 2,9 + 1,0 + 3,9 = 7,8$ для $n_b = 3$. Розраховуємо значення критерію якості кластеризування

$$Q = \frac{5,1/10}{7,8/3} = 0,19.$$

Таким чином, розбиття на два кластери є якіснішим, а частка залишкової дисперсії є меншою ніж для розбиття на три кластери.

Визначимо, яка з ознак більше вплинула на результати кластеризування.

$$K = \frac{F}{F_{кр}}, F_{кр} = F(\alpha, \max(f_b, f_w), \min(f_b, f_w)),$$

де f_b та f_w – число ступенів свободи міжкластерних та внутрішньокластерних відстаней. Для $k=2$ у вікні «K-Means Clustering Results» обираємо команду «Analysis of variance» та отримуємо таблицю результатів дисперсійного аналізу, за якою приймаємо $f_b = 1$, $f_w = 8$, $F_x = 39,8$ та $F_y = 67,3$. Отже, для рівня значимості $\alpha = 0,01$ отримуємо⁷

$$F_{кр} = F(0,01; 8; 1) = 5981, K_x = \frac{F_x}{F_{кр}} = \frac{39,8}{5981} = 0,007, K_y = \frac{F_y}{F_{кр}} = \frac{67,3}{5981} = 0,011.$$

Таким чином, змінна y має вищий ступінь впливу на результати кластеризування ніж змінна x .

⁷ для визначення значення можна скористатись функцією MS Excel FРАСПОБР()