

УДК 303.22

Веретюк С. М., магістр; Пілінський В. В., к.т.н.; Панченко І. М., бакалавр

ЗАСТОСУВАННЯ МЕТОДУ ГОЛОВНИХ КОМПОНЕНТ ДЛЯ ПОРІВНЯЛЬНОГО АНАЛІЗУ РОЗВИТКУ ІНФОРМАЦІЙНОГО СУСПІЛЬСТВА В УКРАЇНІ

Veretyuk S.M., Pilinsky V.V., Panchenko I.M. Application of principal component analysis for the comparative analysis of the development of information society in Ukraine. The dimensionality of the feature space was reduced by the method of principal components. Cluster analysis of the EU and Ukraine based on the indicators DESI (Digital Economy and Society Index) was conducted. Cluster distances were calculated. The position of Ukraine towards European countries was determined. A qualitative analysis of the technological gap was performed and mechanisms to overcome it was proposed. Cluster analysis in terms of the index of development of the information society (ICT Development Index) was conducted. It is shown Ukraine take a static place when mostly countries go to the more technology advanced cluster. It is shown that Ukraine takes a static place when most of the countries move to the more advanced technological clusters. Gap between Ukraine and technology leaders grows. As result accelerated development of ICT in Ukraine is becoming increasingly important.

Keywords: DESI, the digital economy, information society, PCA, clustering

Веретюк С.М., Пілінський В.В., Панченко І.М. Застосування методу головних компонент для порівняльного аналізу розвитку інформаційного суспільства в Україні. Методом головних компонент зменшено розмірність простору ознак та проведено кластерний аналіз розвитку країн ЄС та України за показниками DESI (Digital Economy and Society Index). Визначено положення України відносно європейських країн. Проведено якісний аналіз технологічного відставання та запропоновано механізми його подолання.

Ключові слова: DESI, цифрова економіка, інформаційне суспільство, метод головних компонент, кластеризація

Веретюк С.М., Пилинский В.В., Панченко И.Н. Применение метода главных компонент для сравнительного анализа развития информационного общества в Украине. Методом главных компонент уменьшено размерность пространства признаков и проведен кластерный анализ развития стран ЕС и Украины по показателям DESI (Digital Economy and Society Index). Определено положение Украины по отношению к европейским странам. Проведен качественный анализ технологического отставания и предложены механизмы его преодоления.

Ключевые слова: DESI, цифровая экономика, информационное сообщество, метод главных компонент, кластеризация

Вступ

В Україні офіційно затверджено концепцію розвитку інформаційного суспільства [1], конкретизація та деталізація якої здійснено в рамках Плану реалізації та Концепції розвитку електронного врядування в Україні [2] без конкретизації у відповідних планових та програмних документах. Одним з базових елементів будь-якої стратегії є розроблення та застосування відповідної системи показників, яка надає вичерпну інформацію щодо ефективності реалізації кроків стратегії та можливість, у разі необхідності, її корегування. У рамках реалізації зазначених ініціатив затверджено Методику формування індикаторів розвитку інформаційного суспільства [3], яка через насиченість абстрактними та неконкретними показниками лише фрагментарно описує поточний стан інформатизації в Україні. Саме тому питання впровадження сучасних методик і засобів оцінювання та моніторингу розвитку цифрової економіки та інформаційного суспільства є надзвичайно актуальними в Україні, адже за відсутності адекватної моделі оцінювання реалізація стратегії є неконтрольованою, а відтак малоефективною. Також варто зауважити, що в умовах глобалізації та з урахуванням пріоритетних напрямів міжнародного співробітництва України, доцільно гармонізувати систему показників розвитку інформаційного суспільства з європейськими стандартами та вимогами.

Європейська ініціатива «Цифровий порядок денний для Європи» [4] ставить за мету глибинну трансформацію соціальних, культурних та економічних відносин в Європейському союзі з застосуванням цифрових та інфотелекомунікаційних технологій. І за суттю є

взірцевим зразком стратегії розвитку інформаційного суспільства. Для оцінювання прогресу кожної країни розроблено відповідну систему показників, які об'єднанні в інтегральний показник – індекс розвитку цифрової економіки та суспільства (Digital Economy and Society Index - DESI) [5]. У роботі [6] за відповідною методикою визначено показник DESI для України. Нагадаємо складові показники індексу DESI та відповідні значення для України, що наведено в табл. 1.

Табл. 1. Значення складових індексу DESI для України [6]

Показник верхнього рівня	Значення
X1. Рівень підключення	0.186
X2. Людський фактор	0.337
X3. Використання Інтернету	0.092
X4. Впровадження цифрових технологій	0.185
X5. Публічні цифрові сервіси	0

Метою статті є порівняльний аналіз стану розвитку цифрової економіки за проведенням кластеризації - об'єднанням країн у кластери за рівнем розвитку цифрової економіки та інформаційного суспільства, а також якісний аналіз технологічного розриву між Україною та країнами ЄС задля визначення рекомендацій для його подолання.

1. Метод головних компонент

Отримані результати надають уявлення про загальне значення розвитку цифрової економіки і суспільства, проте досить часто постає питання щодо об'єднання країн за певними критеріями. Така оцінка надає змогу визначити наскільки стани в країнах подібні один до одного. Відповідь на це питання надає процедура кластеризації – групування країн за певними критеріями.

Нехай X – множина об'єктів, Y – множина кластерів. Задано функцію відстані між об'єктами $\rho(x, x')$ та кінцеву вибірку об'єктів $X_m = \{x_1, \dots, x_m\} \subset X$. Необхідно розбити вибірку на підмножини (кластери), без перетину, таким чином, щоб до складу кластера входили об'єкти близькі за метрикою ρ , а об'єкти різних кластерів суттєво відрізнялись. Кожному об'єкту надають номер кластеру y_i .

Під час розв'язку задач розпізнавання, класифікації та аналізу даних важливе значення має наявність засобів візуалізації багатовимірних даних, що наочно надає уявлення про конфігурації класів, кластерів і розташуванні окремих об'єктів. Задачу візуалізації ускладнено тим, що розмірність простору дорівнює кількості ознак показників верхнього рівня індексу DESI ($x_i \in \mathbb{R}^5$). Для полегшення сприйняття даних доцільно визначити проекцію цього набору точок на площину \mathbb{R}^2 (простір ознак дорівнює 2), так щоб метричні співвідношення між проекціями точок на площині об'єктивно відповідали метричним співвідношенням між ними у вхідному просторі ознак.

Існує ефективний метод лінійної трансформації простору ознак, що із збереженням істотної інформації про структуру даних, надає нові ознаки, зручні та ефективні для розв'язку завдань класифікації та кластеризації. Даний метод називають методом головних компонент, або перетворенням Карунена-Лоева [7]. Метод головних компонент засновано на переході від вхідної безлічі зазвичай корельованих змінних $X_m = \{x_1, \dots, x_m\}$ до нового набору змінних $Z_m = \{z_1, \dots, z_m\}$ з нульовими коефіцієнтами коваріації між ними. Проте таке перетворення переважно призводить до втрати певної кількості інформації. Метрикою для оцінювання об'єму інформації є дисперсія проекцій на напрями нових компонент векторів початкової вибірки.

Сутність методу полягає у переході від базису початкових ознак до базису власних векторів кореляційної матриці даних, які впорядковано за спаданням власних чисел. Ці

вектори називають головними компонентами, а створений ними базис – базисом головних компонент.

2. Реалізація методу головних компонент для показників DESI

Всі обчислення виконано у математичному середовищі **R**. На першому етапі відцентруємо дані за співвідношенням $x_i^c = x_i - \bar{x}$, де \bar{x} – середнє значення. В табл. 2 наведено основні статистичні характеристики – середнє значення та середньоквадратичне відхилення для даних [6].

Табл. 2. Основні статистичні характеристики вхідних даних

Показники	Середнє значення, \bar{x}	Середньоквадратичне відхилення, σ
X1	0.135	0.032
X2	0.132	0.033
X3	0.063	0.015
X4	0.068	0.019
X5	0.069	0.028

Для визначення рівня взаємозв'язку показників обчислюємо кореляційну матрицю для вхідних даних [6] (табл.3).

Табл. 3. Кореляційна матриця

	X1	X2	X3	X4	X5
X1	1.0000	0.7241	0.8130	0.5718	0.5715
X2	0.7241	1.0000	0.7151	0.7035	0.6779
X3	0.8130	0.7151	1.0000	0.5447	0.7089
X4	0.5718	0.7035	0.5447	1.0000	0.6263
X5	0.5715	0.6779	0.7089	0.6263	1.0000

Оскільки всі попарні коефіцієнти кореляції більші за 0,5, то між показниками існує кореляційний зв'язок, а отже використання методу головних компонент є можливим [7].

Для переходу до головних компонент визначаємо власні значення кореляційної матриці (табл. 4).

Табл. 4. Власні значення кореляційної матриці

Компоненти	Власні значення	% загальної дисперсії
K1	3.66835	73.36
K2	0.54130	10.82
K3	0.40451	8.09
K4	0.23968	4.79
K5	0.14615	2.92

З отриманих даних маємо, що перша і друга компонента описують приблизно 73% і 11% загальної дисперсії, разом – 84%, що достатньо для практичного застосування.

Визначаємо власні вектори [7] табл. 5.

Табл. 5. Власні вектори

Показник	K1	K2	K3	K4	K5
X1	-0.45069	-0.49382	-0.37741	0.236667	0.595456
X2	-0.46716	0.110001	-0.23251	-0.84261	-0.07484
X3	-0.46343	-0.46326	0.181483	0.208811	-0.70291
X4	-0.41672	0.692633	-0.36737	0.435896	-0.1471
X5	-0.43613	0.222933	0.797245	0.019627	0.352286

Таким чином рівняння першої головної компоненти:

$$K_1 = -0.45 \cdot X'_1 - 0.467 \cdot X'_2 - 0.463 \cdot X'_3 - 0.416 \cdot X'_4 - 0.436 \cdot X'_5,$$

другої:

$$K_2 = -0.49 \cdot X'_1 + 0.11 \cdot X'_2 - 0.463 \cdot X'_3 + 0.693 \cdot X'_4 + 0.223 \cdot X'_5,$$

де X'_i – нормовані значення i -го показника (перехід до звичних змінних виконують за співвідношенням $X'_i = \frac{X_i}{\sigma_i}$).

Таким чином, зменшено розмірність простору ознак до 2. Первинні данні, а також координати країн у «новому» просторі наведено в табл. 6.

Табл. 6. Нормовані показники DESI та відповідні значення в просторі \mathbb{R}^2

Країни	X_1^C	X_2^C	X_3^C	X_4^C	X_5^C	K1	K2
Austria	0.0050	0.0105	-0.0091	0.0058	0.0061	-0.0016	0.0050
Belgium	0.0590	0.0200	0.0126	0.0250	0.0036	-0.0207	-0.0025
Bulgaria	-0.0258	-0.0524	-0.0034	-0.0190	-0.0326	0.0210	-0.0063
Croatia	-0.0527	-0.0246	-0.0130	0.0079	-0.0124	0.0147	0.0130
Cyprus	-0.0283	-0.0331	0.0002	-0.0060	-0.0070	0.0108	0.0004
Czech Republic	0.0031	0.0045	-0.0046	0.0144	-0.0239	0.0008	0.0045
Denmark	0.0381	0.0512	0.0264	0.0369	0.0576	-0.0372	0.0061
Estonia	-0.0004	0.0169	0.0171	-0.0142	0.0502	-0.0119	-0.0058
Finland	0.0151	0.0633	0.0111	0.0234	0.0446	-0.0262	0.0085
France	-0.0076	0.0138	0.0049	-0.0050	0.0109	-0.0029	-0.0008
Germany	0.0291	0.0192	-0.0060	0.0115	-0.0108	-0.0058	0.0014
Greece	-0.0320	-0.0425	-0.0135	-0.0065	-0.0156	0.0182	0.0039
Hungary	-0.0005	-0.0116	0.0034	-0.0241	-0.0277	0.0102	-0.0123
Ireland	-0.0034	0.0229	-0.0020	0.0263	0.0080	-0.0092	0.0122
Italy	-0.0423	-0.0300	-0.0170	-0.0105	-0.0050	0.0182	0.0062
Latvia	0.0180	-0.0193	0.0114	-0.0298	-0.0141	0.0055	-0.0188
Lithuania	0.0251	-0.0061	0.0120	0.0112	-0.0067	-0.0077	-0.0040
Luxembourg	0.0426	0.0306	0.0131	-0.0023	-0.0196	-0.0106	-0.0117
Malta	0.0096	-0.0082	0.0027	0.0075	0.0003	-0.0027	0.0003
Netherlands	0.0417	0.0349	0.0165	0.0167	0.0498	-0.0268	-0.0001
Poland	-0.0192	-0.0247	-0.0090	-0.0253	-0.0035	0.0149	-0.0048
Portugal	0.0032	-0.0252	-0.0005	0.0065	0.0132	-0.0002	0.0022
Romania	-0.0132	-0.0635	-0.0226	-0.0290	-0.0228	0.0273	-0.0057
Slovakia	-0.0219	0.0003	-0.0034	-0.0098	-0.0281	0.0104	-0.0015
Slovenia	-0.0286	-0.0036	-0.0013	-0.0075	-0.0188	0.0093	0.0004
Spain	-0.0012	-0.0057	-0.0064	0.0042	0.0290	-0.0025	0.0057
Sweden	0.0390	0.0556	0.0265	0.0303	0.0382	-0.0336	0.0021
Ukraine	-0.0882	-0.0473	-0.0496	-0.0307	-0.0687	0.0508	0.0100
United Kingdom	0.0351	0.0496	0.0053	-0.0059	0.0041	-0.0127	-0.0071

3. Візуалізація отриманих результатів

Як зазначено вище, зведення розмірності даних до \mathbb{R}^2 надає можливість відобразити дані на площині (рис. 2).

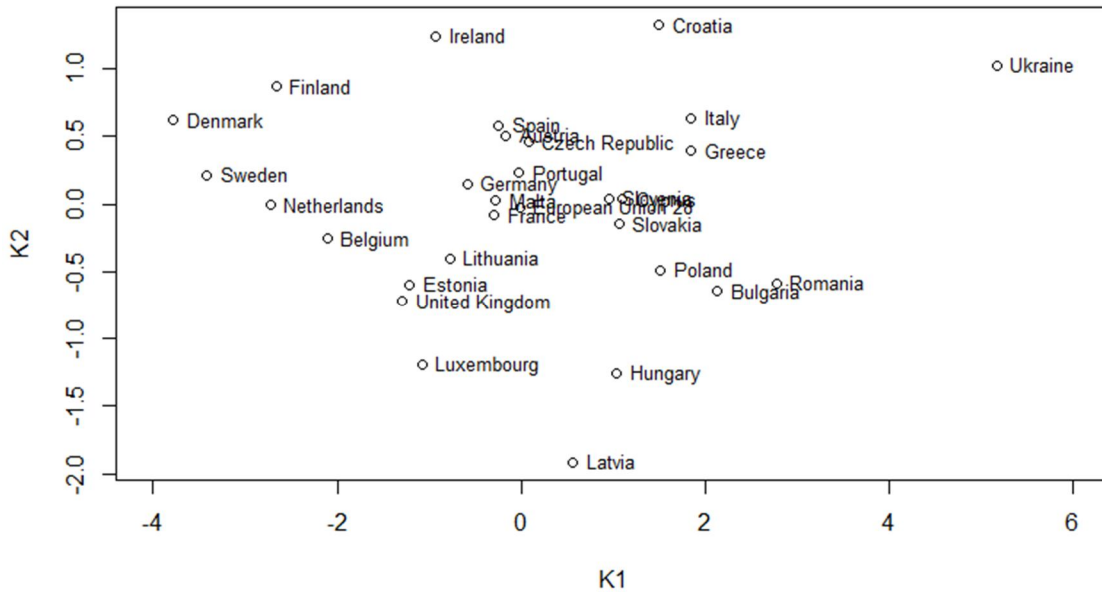


Рис. 2. Розташування країн в ознаковому просторі (K1, K2)

Для кластеризації використаємо метод Варда [8]. Крім вхідної вибірки метод Варда потребує інформацію щодо кількості кластерів. Відповідно до загально прийнятого підходу виконаємо розбиття на 4 кластери, з метою виявити чотири класи країн: високорозвинені, добре розвинені, задовільно розвинені, низько розвинені. За математичним пакетом R виконуємо кластеризацію даних. Дендрограму наведено на рис. 3.

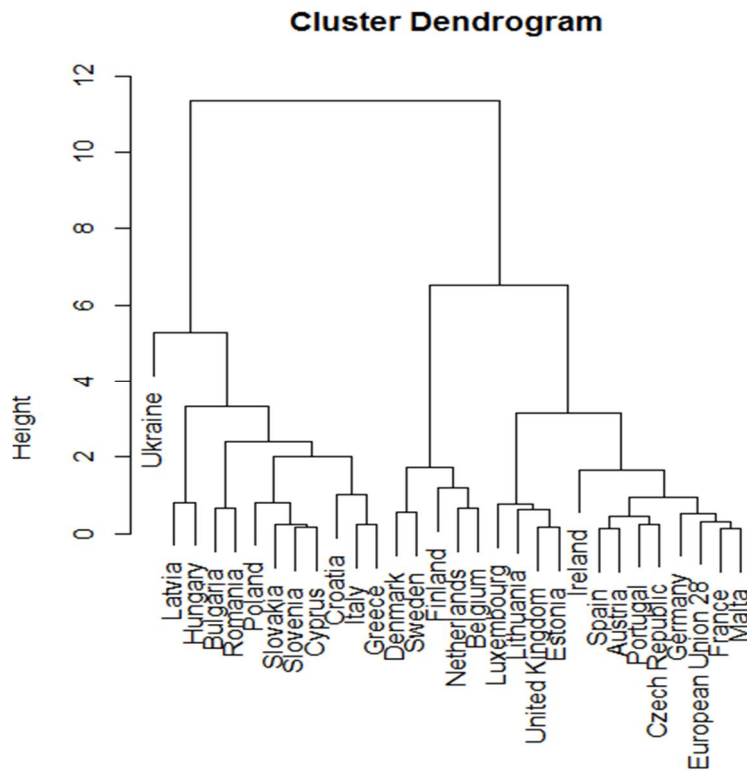


Рис. 3. Дендрограма кластеризації методом Варда

Результати кластеризації у новому ознаковому просторі наведено на рис. 4.

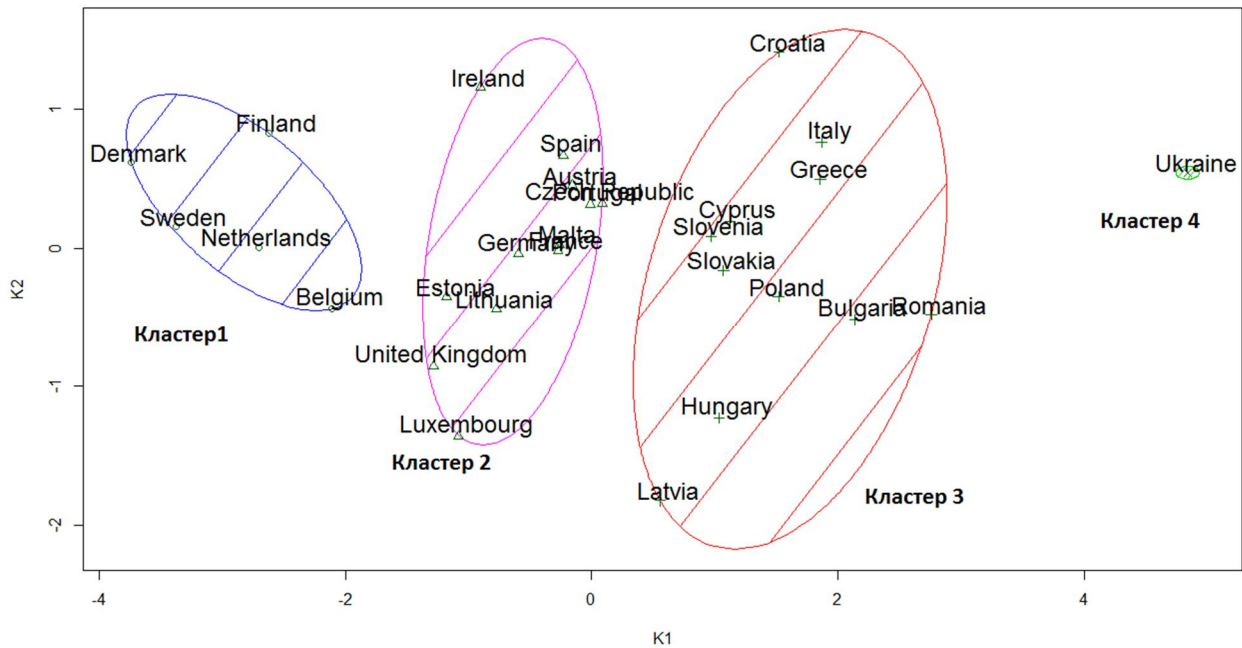


Рис. 4. Кластери країн

В табл. 7 наведено міжкластерні відстані (розрахунки проведені за Евклідовою метрикою):

де (K_{1i}, K_{2i}) та (K_{1j}, K_{2j}) - координати центрів кластерів i та j в просторі

Табл. 7. Міжкластерні відстані

	Кластер 1	Кластер 2	Кластер 3	Кластер 4
Кластер 1	0	2,43	4,46	8,14
Кластер 2	2,43	0	2,02	5,78
Кластер 3	4,46	2,02	0	3,89
Кластер 4	8,14	5,78	3,89	0

Важливим етапом класифікації є інтерпретація отриманих результатів. Отже на підставі кластеризації країн ЄС та України за значеннями показників індексу DESI визначасмо:

- існує група високорозвинених країн (Фінляндія, Данія, Швеція, Голандія, Бельгія), які мають найкращі показники з розвитку цифрової економіки та інформаційного суспільства;
- інші європейські країни формально можна розділити на дві групи – група розвинених та задовільно розвинених країн; зауважимо, що до складу розвинених увійшли Естонія та Латвія;

- між країнами ЄС та Україною існує суттєвий технологічний розрив, що не дозволяє претендувати Україні на перебування в групі навіть задовільно розвинених країн. Проте таке становище має стати **стимулом** для прискореного розвитку цифрової економіки та суспільства;

- за результатами обчислень міжкластерних відстаней, а також з урахуванням того, що всі перетворення в рамках попередніх обчислень є лінійними, можемо стверджувати, що технологічний розрив між Україною і країнами, які входять до третього кластеру (наприклад, Польща) еквівалентний технологічному розриву, який існує між Польщею та країнами-лідерами (Кластер 1).

4. Кластеризація за індексом розвитку ІКТ (ICT Development Index)

Отримані результати надають уявлення щодо технологічного інтервалу між Україною та європейськими країнами. Для аналізу положення України в світі за індексом розвитку ІКТ використовуємо результати, щорічного дослідження розвитку інформаційного суспільства, яке проводить Міжнародний Союз Електрозв'язку [9]. Оскільки, складові індексу розвитку ІКТ здебільшого за своєю суттю співпадають із складовими індексу DESI (мережевий доступ, використання, комп'ютерні навички та освіта), то умовно можна вважати, що індекс розвитку ІКТ є певною мірою індексом розвитку цифрової економіки.

Результати кластеризації методом головних компонент за індексом розвитку ІКТ наведено на рис. 5 та рис. 6 (для досліджень 2008 та 2015 р., що дозволяє оцінити динаміку).

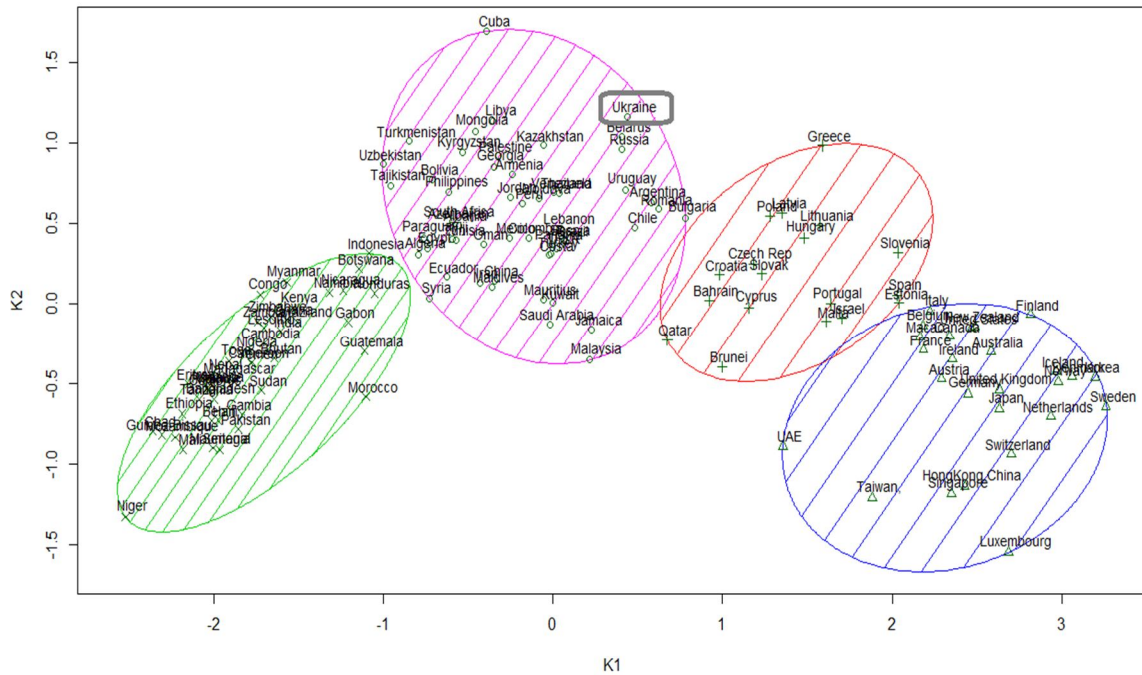


Рис. 5. Кластеризація за індексом розвитку інформаційного суспільства за 2008 р.

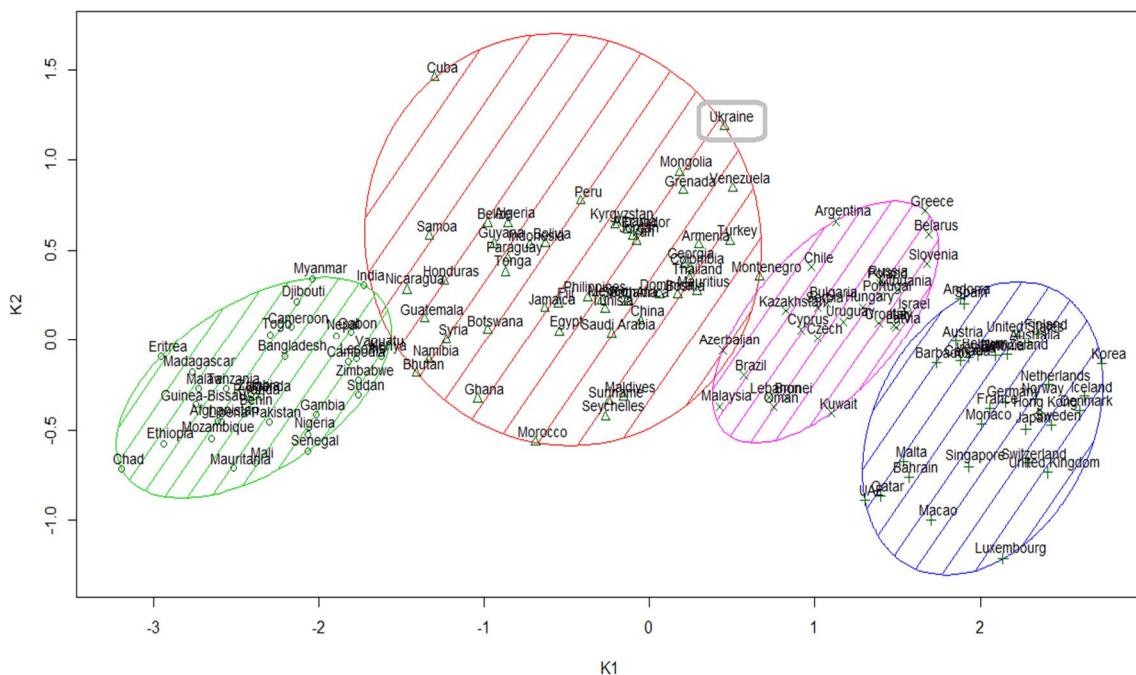


Рис. 6. Кластеризація за індексом розвитку інформаційного суспільства за 2015 р.

З рис. 5 та рис. 6 випливає, що позиція України за 7 років практично не зазнала змін, тоді як інші країни поступово переходять з менш технологічно розвинених кластерів до більш розвинених. На рис. 7 наведено позиції України за інтегральним індексом розвитку ІКТ. Вочевидь, технологічне відставання від розвинених країн поглиблюється. Відставати можна в трьох випадках: всі рухаються – один стоїть на місці; всі рухаються швидко – один повільно; рух в протилежному напрямку.

Статичність положення України на рис. 5 та рис. 6 показує, що держава не встигає за світовими тенденціями, тому є нагальною необхідністю напрацювання та впровадження ефективного та дієвого механізму, який забезпечить розвиток інформаційного суспільства та цифрової економіки.

Висновки

1. Для кластеризації країн за рівнем розвитку цифрової економіки та аналізу показників DESI доцільно використовувати метод головних компонент, оскільки попарні коефіцієнти кореляції більші за 0.5.

2. Використання методу головних компонент зменшує розмірність простору ознак ($X \in \mathbb{R}^5$) \rightarrow ($K \in \mathbb{R}^2$) і дозволяє зберегти 84% загальної первинної інформації, що є цілком достатнім для практичних задач.

3. Визначено аналітичні співвідношення для перетворення $X \rightarrow K$, за якими обчислені координати країн в новому просторі ознак.

4. Методом Варда проведено кластеризацію країн у новому просторі K . За результатами кластеризації країни розділено на чотири умовні класи - високорозвинені, розвинені, задовільно розвинені, недостатньо розвинені.

5. Виконано кластеризацію країн за індексом розвитку ІКТ для 2008 та 2015 рр. Вказано на застій у питанні розвитку інформаційного суспільства в Україні, що призводить до нарощення відставання від розвинених країн.

6. Отримані результати надають можливість визначити порядок та напрям дій, які необхідно виконати для подолання цифрового (технологічного) відставання як між кластерами країн так і безпосередньо між країнами. Важливо врахувати, що отриманий фазовий портрет є статичний, і тому під час аналізу можливих дій важливо враховувати не лише «відрив» сьогодні, а в цілому динаміку, адже, очевидно, що у європейських країн можливостей щодо реалізації стратегій розвитку інформаційного суспільства значно більше.

7. До переліку першочергових дій необхідно включити створення системної державної стратегії розвитку ІКТ, яка може бути сформована за прикладом європейського Цифрового порядку денного. Це дозволить синхронізувати не лише порядок дій, а й методологію визначення ключових показників ефективності, що надасть можливість відстежувати хід реалізації реформи розбудови інформаційного суспільства та цифрової економіки. Важливо не абсолютизувати ІКТ, що може привести до певної атомізації, а навпаки через впровадження сучасних ІКТ рішень «дигіталізувати» галузі промисловості та соціально-економічні відносини в країні. Крім того для ефективного виконання стратегії доцільно створити єдиний центральний орган виконавчої влади відповідальний за розвиток інформаційного суспільства та цифрової економіки (на разі в Україні за розвиток ІКТ відповідають близько 15 відомств (міністерств, служб, агенцій)).

8. Докладний аналіз розвитку інформаційного суспільства в Україні та заходів щодо позитивної динаміки є предметом подальших досліджень.

Література

1. Про затвердження плану заходів з виконання завдань, передбачених Законом України “Про Основні засади розвитку інформаційного суспільства в Україні на 2007-2015 роки” [Електронний ресурс] / Розпорядження Кабінету Міністрів України від 15 серп. 2007 р. № 653-р. - Режим доступу: <http://www.rada.gov.ua>. (22.09.2015)
2. Про схвалення Концепції розвитку електронного урядування в Україні [Електронний ресурс] / Розпорядження Кабінету Міністрів України від 13 груд. 2010 р. № 2250-р. - Режим доступу: <http://www.rada.gov.ua>. (22.09.2015)
3. Про затвердження Методики формування індикаторів розвитку інформаційного суспільства [Електронний ресурс] / Наказ Міністерства освіти і науки 06.09.2013 № 1271. - Режим доступу: <http://www.rada.gov.ua>.
4. Digital agenda for Europe [Електронний ресурс] / Europe 2020 Initiative. - Режим доступу: <http://ec.europa.eu/digital-agenda/en>. (04.04.2015)
5. DESI 2015 Digital Economy and Society Index, Methodological note [Електронний ресурс] / European commission Directorate-General for Communications Networks, Content and Technology Digital Economy and Coordination European Semester and Knowledge Base. - Режим доступу: http://ec.europa.eu/information_society/newsroom/cf/dae/document.cfm?doc_id=8846. (04.04.2015)
6. Пілінський В. В. Аналіз стану інфокомунікаційних технологій в Україні за методологією DESI / В. В. Пілінський, С. М. Веретюк // Наукові записки Українського науково-дослідного інституту зв'язку. – 2015. – №5(39). - с. 5-15.
7. Pearson K. On lines and planes of closest fit to systems of points in space / K. Pearson // Philosophical Magazine. - 1901. - V.2. - P. 559–572.
8. Ward Joe H. Hierarchical Grouping to Optimize an Objective Function / Joe H. Ward, Jr. Source // Journal of the American Statistical Association, Vol. 58, №301 (Mar., 1963), pp. 236-244.
9. ITU Measuring Information Society 2013 [Електронний ресурс] / International Telecommunication Union. - Режим доступу: <http://www.itu.int/en/ITU-D/idi/2015>. (04.04.2015).

Автори статті

Веретюк Сергій Михайлович - магістр, аспірант кафедри звукотехніки та реєстрації інформації, Національний технічний університет України «КПІ», Київ, Україна. Тел. +38 067 405 50 37. E-mail: capserg81@gmail.com

Пілінський Володимир Володимирович - кандидат технічних наук, професор, відмінник освіти України, професор кафедри звукотехніки та реєстрації інформації, Національний технічний університет України «КПІ», Київ, Україна. Тел. +38 095 276 92 70. E-mail: pww@ukr.net

Панченко Ігор Миколайович - бакалавр, студент кафедри звукотехніки та реєстрації інформації, Національний технічний університет України «КПІ», Київ, Україна. Тел. +38 093 509 02 51. E-mail: inpan4enko@gmail.com

Authors of the article

Veretiuk Sergii Myhailovich – master of science, post-graduate student of Department of Audio Engineering and Information Registration, National Technical University of Ukraine “KPI”, Kyiv, Ukraine. Tel. +38 067 405 50 37. E-mail: capserg81@gmail.com

Pilinsky Volodymyr Volodymyrovich – candidate of science (technic), professor, professor of Department Audio Engineering and Information Registration, National Technical University of Ukraine “KPI”, Kyiv, Ukraine. Tel. +38 095 276 92 70. E-mail: pww@ukr.net

Panchenko Ihor Mykolayovych - student, bachelor of science of Department Audio Engineering and Information Registration, National Technical University of Ukraine “KPI”, Kyiv, Ukraine. Tel. +38 093 509 02 51. E-mail: inpan4enko@gmail.com

Дата надходження в редакцію: 17.12.2015 р.

Рецензент: к.т.н., с.н.с. В.О. Гребенніков