

Використання моделей експертного аналізу для оцінювання стану ІТ-галузі України

Анотація. *Сформовано тривимірний набір показників, що відображає стан ІТ-індустрії України та включає інвестиційну привабливість регіонів, внутрішній стан ІТ-ринку та зарплату. Використовуючи принципи побудови моделі багатокритеріального прийняття рішень за детермінованими вихідними даними, реалізується етап синтезу методології прийняття рішень в умовах багатокритеріальної та інтервальної невизначеності. Побудовано математичну модель стану ІТ-індустрії України, яка представлена як комплексний показник. Розрахунки для визначення оцінки показників стану ІТ-індустрії виконуються на основі побудованої моделі. За допомогою кореляційно-регресійного аналізу визначено силу взаємозв'язку між обраними факторами та комплексним показником, а також значимість впливу факторів на комплексний показник. За допомогою критерію Фішера було виявлено, що запропонована математична модель адекватна описаним процесам. Проведено класифікацію регіонів з використанням кластерного аналізу.*

Ключові слова: *ІТ-індустрія, багатокритеріальність, оптимізація, нормалізація, інвестиційна привабливість, тривимірний кортеж, узагальнена скалярна оцінка якості можливих рішень, комплексний показник, коефіцієнт кореляції, коефіцієнт детермінації, кореляційний і регресійний аналіз, множинне регресійне рівняння, кластерний аналіз.*

Abstract. *The article formed a three-dimensional tuple of indicators that reflects the state of the IT industry in Ukraine, and includes the investment attractiveness of the regions, the internal state of the IT market, and salary. Using the principles of constructing a model of multi-criteria decision making under deterministic output data, the stage of synthesis of a decision-making methodology under conditions of multi-criteria and interval uncertainty is implemented. A mathematical model of the state of the IT industry in Ukraine has been built, which is presented as a complex indicator. The calculations for determination the assessment of indicators of the state of the IT industry are performed on the basis of the constructed model. Using correlation and regression analysis, the strength of the relationship between the selected factors and the complex indicator were determined, as well as the significance of the influence of factors on the complex indicator. Using the Fisher*

criterion, it was found that the proposed mathematical model is adequate to the processes that it describes. The classification of regions with using cluster analysis.

Keywords: *IT-industry, multicriteriality, optimization, normalization, investment attractiveness, three-dimensional tuple, generalized scalar assessment of the quality of feasible solutions, a complex indicator, coefficient of correlation, coefficient of determination, correlation and regression analysis, multiple regression equation, cluster analysis.*

Вступ та постановка задачі

Український ІТ-ринок має величезний потенціал до зростання, проте нестабільність податкового регулювання, нестача спеціалістів і низка регуляторних бар'єрів є стримуючими факторами для розвитку галузі. У зв'язку з обранням Україною курсу на євроінтеграцію, сфера розробки програмного забезпечення є чи не єдиною експортною галуззю України, що стабільно зростає. Лише за 2018 р. приріст збільшився на 26 %. Також галузь виступає потужним економічним мультиплікатором: 1 робоче місце в ІТ створює ще 3 у суміжних професіях. Заробітна плата ІТ-спеціаліста у 5 разів вища за середню в країні. При цьому високий рівень оплати зберігається не лише в столиці, але й в регіонах.

Але й у цьому бізнесі є свої проблеми. Зокрема, спостерігається серйозна нестача молодих спеціалістів, які спроможні вирішувати нові завдання, що постійно з'являються та змінюються. Ця нестача пов'язана з тим, що одночасно існують проблеми як з підготовки конкурентоспроможних спеціалістів профільними навчальними закладами, так й «вимивання» кваліфікованих кадрів не тільки з менш благополучних регіонів в крупні вітчизняні ІТ-центри, а й за межі країни. Отже, проведення аналізу такої ситуації за допомогою сучасних математичних підходів та оцінка впливу виявлених факторів на результат, є актуальними питаннями.

Процеси, що відбуваються на ІТ-ринку праці, важливі і привертають увагу не тільки вчених, але й представників бізнесу. Багато питань щодо прогнозів розвитку, перегріву ІТ-ринку, створення ІТ-кластерів, визначення переваг і недоліків української ІТ-галузі, і багато іншого піднімаються в роботах Р. А. Винничук і Т. В. Склярчук [1], Е. Рубіна [2], в інтерв'ю з представниками українського бізнесу, такими як Дарина Стремetsька [3] та Анна Грабовська [4], ІТ-співтовариством dou і іншими. В своїй роботі Н. О. Бринза та А. А. Гаврилова [5] провели дослідження стану ІТ-галузі на основі виділення показників, але не за всіма регіонами України. Це ставить під сумнів зроблені висновки стосовно ступеня впливу кожного показника на результат та

адекватної оцінки запропонованої цими авторами моделі.

Метою дослідження в даній роботі є проведення аналізу стану ІТ-галузі за більшою кількістю регіонів, ніж це було зроблено в роботі Н. О. Бринзи та А. А. Гаврилової. В основу дослідження покласти сформовану модель експертної оцінки показників [5], які характеризують дану галузь, виявити значність впливу факторів на результат, перевірити адекватність наведеної моделі, а також провести упорядкування регіонів в порівняно однорідні групи із застосуванням математичних методів.

Основна частина

Позитивні зміни в ІТ-галузі України пов'язані з появою якісних навчальних ІТ-програм у ЗВО, привабливі умови праці ІТ-компаній в Україні стрімко збільшує інтерес до галузі серед фахівців різних сфер [5].

Аналітиками спільноти програмістів на сайті dou.ua були розміщені результати опитування експертів (топ-менеджери відомих ІТ-компаній, які працюють в Україні), що працюють в ІТ-бізнесі, який був спрямований на виявлення позитивних та негативних факторів ведення ІТ-бізнесу в регіонах України. На підставі результатів даних опитувань були виділені показники, що характеризують стан ІТ-галузі (2015-2018 рр.): ІТ-кластер, ЗВО та ІТ, перегрів ІТ-ринку, вимивання кадрів, авіасполучення, співпраця ІТ-індустрії й місцевої влади, коворкінг, рівень кваліфікації ІТ-кадрів, комфортність проживання та ведення бізнесу, превалювання в напрямку діяльності, близькість до Євросони, рівень володіння англійською мовою, актуальність знань, придбаних в ЗВО по ІТ-профілю, конкуренція на ІТ-ринку, близькість до вітчизняних ІТ-центрам, наявність інвесторів, попит на ІТ-кадри, кількість профільних ЗВО, кількість фахівців ІТ-сфери, рівень зарплати [5].

В роботі [5] для кожного показника стану ІТ-галузі були визначені 23 критерії оцінювання, які в свою чергу були розподілені на три категорії: «Інвестиційна привабливість регіонів України», «Внутрішній стан ІТ-ринку України», «Заробітна плата». Вхідні дані за всіма критеріями, відповідно до категорії представлено у табл. 1 – 2.

Багатофакторна оцінка показників, за якими можна оцінити стан розвитку ІТ-галузі, оцінюється кортежем, що враховує інвестиційну привабливість регіонів України (F_j^{IN}), внутрішній стан ІТ-ринку України (F_j^{VN}) та заробітну плату (F_j^{ZP}) [5]:

$$F_j = \langle F_j^{IN}, F_j^{VN}, F_j^{ZP} \rangle, \quad (1)$$

Таблиця 1

Дані групи «Внутрішній стан ІТ-ринку України» за Рівненським, Полтавським та Херсонським регіонами

Критерій \ Регіон	ІТ-кластер	ЗВО та ІТ	перегрів ІТ-ринку	вимивання кадрів	співпраця ІТ-індустрії та влади	коворкінг	рівень кваліфікації ІТ-кадрів	превалювання в напрямку діяльності	актуальність знань	кількість профільних ЗВО/факультетів	конкуренція на ІТ-ринку	близькість до ІТ-центрів	кількість фахівців ІТ-сфери, люд.	наявність інвесторів	попит на ІТ-кадри
Рівне	+	+	+	+	-	-	середній	А і П	не відповідають вимогам	4	-	+	1000	+	+
Полтава	-	+	-	+	-	-	низький	А	не відповідають вимогам	5	-	+	1000	-	+
Херсон	+	+	-	+	-	-	середній	А і П	не відповідають вимогам	3	-	-	1000	+	+

Таблиця 2

Дані груп «Інвестиційна привабливість регіонів України» та «Заробітна плата, USD» за Рівненським, Полтавським та Херсонським регіонами

Регіон Критерій	Інвестиційна привабливість регіонів України								Заробітна плата, USD		
	авіаспо- лучення	еколо- гія	транс- портна розв'язка	рівень дозвілля (відпо- чинок)	рівень безпе- ки прожи- вання	рівень життя	близь- кість до Євро- зони	рівень воло- діння англій- ською	зарпла- та Junior	зарпла- та Middle	зарпла- та Senior
Рівне	-	сприят- лива	гарна	середній	висо- кий	доступ- ний	-	низький	450	1150	2250
Полтава	-	сприят- лива	гарна	середній	висо- кий	доступ- ний	-	низький	300	1200	-
Херсон	+	сприят- лива	погана	середній	висо- кий	доступ- ний	-	середній	400	700	1900

З урахуванням цього завдання оцінка показників перетворюється в задачу багатокритеріальної оптимізації виду [5]:

$$O_j^o = \arg G[\langle F_j(O) \rangle], \quad (2)$$

де G – оператор, який визначає структуру моделі; $\langle F_j(O_j) \rangle$ – тривимірний кортеж, компонентами якого є різномірні по розмірності, за інтервалами зміни і вимірювальним шкалами величини.

У загальному випадку завдання багатокритеріальної оптимізації є некоректними за Адамаром, тому що не мають однозначної відповіді [6] і вимагають регуляризації [7].

Загальний підхід до регуляризації задач багатокритеріальної оптимізації полягає в їх скаляризації. В даний час існує безліч різних способів регуляризації: принцип головного критерію, послідовна оптимізація, функціонально-вартісний аналіз і таке інше, але найбільш аргументованим є підхід, заснований на теорії корисності [8], який полягає у формуванні узагальненої скалярної оцінки (функції корисності) допустимих рішень. В даному випадку функція корисності має вигляд [5]:

$$P(O_j^o) = Q[\langle \lambda_j \rangle, \langle F_j(O_j) \rangle], \quad j = \overline{1, n}, \quad (3)$$

де $\langle \lambda_j \rangle$ – кортеж коефіцієнтів ізоморфізму; $\langle F_j(O_j) \rangle$ – кортеж різномірних оцінок показників розвитку ІТ-галузі.

Для конструктивного використання моделі (3) необхідно вирішити задачу її структурно-параметричної ідентифікації, тобто визначити структуру (вид оператора Q) і параметри (значення коефіцієнтів ізоморфізму λ_j). Рішення зазначених завдань є евристичними процедурами, заснованими на реалізації методології експертного оцінювання.

В якості узагальненої оцінки ефективності локальної категорії приймають адитивну модель виду [5]:

$$P(O_j) = \sum_{j=1}^3 \lambda_j F_{ji}(O_i), \quad (4)$$

де $F_{ji}(O_i)$, $j = \overline{1, 3}$ – інвестиційна привабливість регіонів України ($j=1$), внутрішній стан ІТ-ринку України ($j=2$) і заробітна плата ($j=3$) оцінки, виміряні в натуральних показниках; λ_j – коефіцієнти ізоморфізму.

Коефіцієнти ізоморфізму λ_j в моделі (4) виконують дві функції: призводять різномірні величини до однієї розмірності; враховують різну відносну важливість.

Така подвійність ускладнює експертну параметричну ідентифікацію моделі (4).

Для подолання цих труднощів модель перетворено шляхом приведення її до безрозмірною форми і представлено у вигляді [5]:

$$P(O_j) = \sum_{j=1}^n a_j F_{ji}(O_i), \quad (5)$$

де $F_{ji}(O_i)$ – значення j -ої категорії з врахуванням і показників; a_j – безрозмірні коефіцієнти відносної важливості, що відповідають наступним обмеженням:

$$0 \leq a_j \leq 1, \forall j = \overline{1,3}; \sum_{j=1}^3 a_j = 1. \quad (6)$$

Модель оцінювання $F_{ji}(O_i)$ має вигляд [9]:

$$F_{ji}(O_i) = \sum_{i=1}^n b_{ji} O_{ji}^H, \quad (7)$$

де O_{ji}^H – нормалізовані критерії j -ої категорії; b_{ji} – безрозмірні коефіцієнти відносної важливості нормалізованих приватних критеріїв. Їх нормалізація проводиться за формулою [5]:

$$O_{ji}^H(x) = \left[\frac{O_{ji} - O_{ji}^G}{O_{ji}^K - O_{ji}^G} \right], \quad (8)$$

де O_{ji} – значення критерію; O_{ji}^K, O_{ji}^G – відповідно найкраще і найгірше значення критерію, яке він приймає на області допустимих рішень.

З урахуванням проведеного раніше перетворення функцій, модель оцінки категорій, цільова функція має вигляд [5]:

$$O_j^o = \arg \sum_{j=1}^3 a_j \left\{ \sum_{i=1}^n b_{ji} \left[\frac{O_{ji} - O_{ji}^G}{O_{ji}^K - O_{ji}^G} \right] \right\}. \quad (9)$$

Таким чином, за допомогою використання принципів побудови моделі прийняття багатокритеріальних рішень в умовах детермінованих вихідних даних, реалізовано етап синтезу методології прийняття рішень в умовах багатокритеріальності і інтервальної невизначеності.

На основі моделі (7) були виконані розрахунки по визначенню оцінки показників стану ІТ-галузі на основі узагальнених даних за опитуванням експертів.

Кожний з критеріїв має свою шкалу оцінювання [5]. Взагалі використовувалися двохрівневі та трьохрівневі шкали.

Результати розрахунку за Рівненським, Полтавським та Херсонським регіонами подано у табл. 3.

Таблиця 3

Значення показників стану ІТ-галузі за кожною категорією за Рівненським, Полтавським та Херсонським регіонами

Регіон	Внутрішній стан ІТ-ринку України (F_j^{VN})	Інвестиційна привабливість регіонів України (F_j^{IN})	Заробітна плата (F_j^{ZP})
Вінниця	0,371889401	0,675371224	0,465240642
Миколаїв	0,180832907	0,455709165	0,490813108
Запоріжжя	0,299982932	0,339477727	0,406965583
Дніпро	0,519952210	0,206093190	0,804538599
Одеса	0,413807817	0,473630312	0,941176471
Луцьк	0,274074074	0,444444444	0,431372549
Івано-Франківськ	0,550605905	0,615207373	0,587686823
Житомир	0,344427377	0,283922171	0,506101741
Хмельницький	0,268817204	0,559139785	0,116550117
Харків	0,866666667	0,333333333	0,703962704
Тернопіль	0,474534904	0,611879160	0,304675716
Ужгород	0,368509985	0,780849974	0,094405594
Львів	0,764225977	0,878648233	0,776402029
Черкаси	0,378067930	0,228878648	0,279720280
Чернівці	0,285936167	0,729646697	0,536404772
Суми	0,443659328	0,282642089	0,290004114
Чернігів	0,584400068	0,338197645	0,505635541
Рівне	0,591039427	0,448028674	0,385986562
Полтава	0,298446834	0,448028674	0,177156177
Херсон	0,450298686	0,503584229	0,155354449

За кожною групою показників максимальні та мінімальні значення, що впливає на комплексний показник, не змінилися [5]. Так, за групою «Внутрішній стан ІТ-ринку» мінімальне значення спостерігається у Миколаївському регіоні (0,180833), а максимальне – у Харківському (0,866667). За групою «Інвестиційна привабливість регіонів» максимум належить

Львівському регіону (0,878648), а мінімум – Дніпропетровському регіону (0,206093). За групою «Заробітна плата» максимальне значення за Одеським регіоном (0,9411765), а мінімальне – за Ужгородським (0,0944056).

Визначення комплексного показника стану ІТ-галузі регіону за моделлю (5) має вигляд [5]:

$$P(O_i) = \frac{F_i^{VN} + F_i^{IN} + F_i^{ZP}}{3}, i = \overline{1, 20}. \quad (10)$$

Результати розрахунку комплексного показника стану ІТ-галузі за кожним регіоном України представлені в табл. 4.

Таблиця 4

Значення комплексного показника стану ІТ-галузі з врахуванням Рівненського, Полтавського та Херсонського регіонів

Регіон	Значення показника
Львів	0,8064254
Харків	0,634654
Одеса	0,609538
Івано-Франківськ	0,584500
Чернівці	0,517000
Дніпро	0,510190
Вінниця	0,504167
Чернігів	0,476078
Рівне	0,475018
Тернопіль	0,463697
Ужгород	0,414589
Луцьк	0,383297
Житомир	0,378150
Миколаїв	0,375785
Херсон	0,369746
Запоріжжя	0,348809
Суми	0,338769
Хмельницький	0,314836
Полтава	0,307877
Черкаси	0,295556

Із введенням в дослідження додатково трьох регіонів тенденції стосовно найменшого та найбільшого значення комплексного показника стану ІТ-галузі, не змінились: попереду Львів (0,806425), Харків (0,634654) та Одеса (0,609538), слабкіший розвиток – Хмельницький (0,314836), Полтава (0,307877) та Черкаси (0,295556) (рис. 1).

За результатами сформованої моделі та проведеними за нею розрахунками, необхідно з'ясувати силу зв'язку між обраними факторами та комплексним показником, а також значність впливу факторів на комплексний показник. Таке дослідження було проведено за допомогою кореляційно-регресійного аналізу.

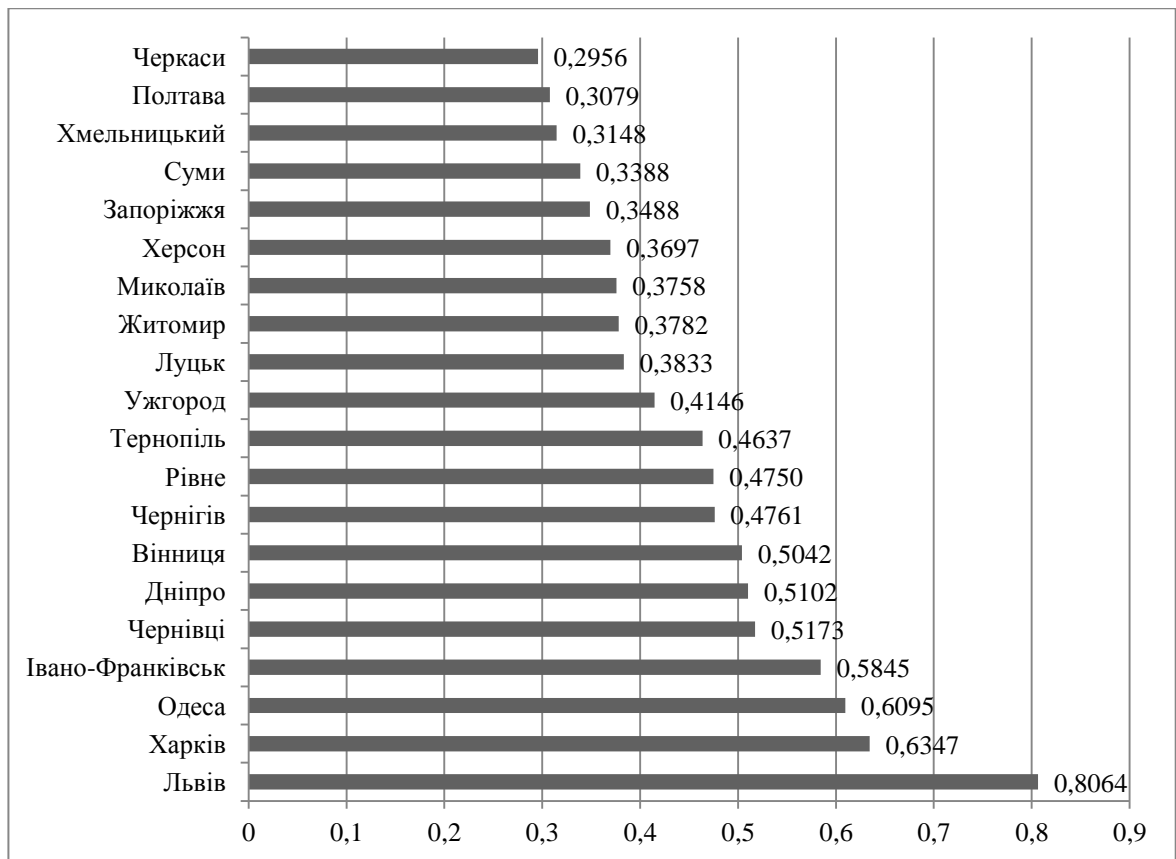


Рис. 1. Ранжування регіонів України за комплексним показником стану ІТ-галузі

Розрахуємо показник тісноти зв'язку. Таким показником є вибірковий лінійний коефіцієнт кореляції, який розраховується за формулою [10]:

$$r_{xy} = \frac{\overline{x * y} - \bar{x} * \bar{y}}{s_x * s_y} \quad (11)$$

де $\overline{x * y}$, \bar{x} , \bar{y} – середні значення ($\bar{x} = \frac{\sum x_i}{n}$, $\bar{y} = \frac{\sum y_i}{n}$, $\overline{x * y} = \frac{\sum (x_i * y_i)}{n}$);

s_x , s_y – середньоквадратичні відхилення ($s_x = \sqrt{\frac{\sum x_i^2}{n} - \bar{x}^2}$,

$$s_y = \sqrt{\frac{\sum y_i^2}{n} - \bar{y}^2}).$$

Коефіцієнти парної кореляції між факторами «Внутрішній стан ІТ-ринку України» (F_j^{VN}), «Інвестиційна привабливість регіонів України» (F_j^{IN}), «Заробітна плата» (F_j^{ZP}) та комплексним показником стану ІТ-галузі регіону та їх інтерпретація наведені в табл. 6.

Таблиця 6

Результати розрахунку коефіцієнтів парної кореляції за двадцятьма регіонами України

Назва показника	Значення коефіцієнта кореляції	Оцінка значення коефіцієнта кореляції за шкалою Чеддока [10]
Внутрішній стан ІТ-ринку України	0,726	сильна та пряма
Інвестиційна привабливість регіонів України	0,459	помірна та пряма
Заробітна плата	0,769	сильна та пряма

Тісноту спільного впливу чинників на результат оцінює індекс множинної кореляції. На відміну від парного коефіцієнта кореляції, який може приймати негативні значення, він приймає значення від 0 до 1 [10]. Чим щільніше фактичні значення y_i розташовуються щодо лінії регресії, тим менше залишкова дисперсія i , отже, більше величина $R_y(x_1, \dots, x_m)$. Таким чином, при значенні R близькому до 1, рівняння регресії краще описує фактичні дані і фактори сильніше впливають на результат. При значенні R близькому до 0 рівняння регресії погано описує фактичні дані і фактори чинять слабкий вплив на результат [10]:

$$R = \sqrt{1 - \frac{S_e^2}{\sum (y_i - \bar{y})^2}} \quad (12)$$

Коефіцієнт множинної кореляції $R = \sqrt{1 - \frac{0}{0,792}} = 1$, отже зв'язок між ознакою Y та факторами X_i сильний.

Рівняння множинної регресії наступне:
 $Y = 0,000021 + 0,3335X_1 + 0,3332X_2 + 0,3332X_3$.
 Можлива економічна інтерпретація параметрів моделі така: збільшення X_1 на 1 од. вим. призводить до збільшення Y в середньому на 0,3335 од. вим.; збільшення за X_2 та X_3 на 1 од. вим. призводить до збільшення Y в середньому на 0,3332 од. вим. за кожним фактором.

Перейдемо до статистичного аналізу отриманого рівняння регресії: перевірімо значимості рівняння і його коефіцієнтів. Об'єктивною оцінкою є скоригований коефіцієнт детермінації [10]:

$$\bar{R}^2 = 1 - \frac{1 - R^2}{n - m - 1} \quad (13)$$

Чим ближче цей коефіцієнт до одиниці, тим більше рівняння регресії пояснює поведінку Y . В нашому випадку він дорівнює 1. За максимальним коефіцієнтом $\beta_3=0,601$ можна зробити висновок, що найбільший вплив на результат Y оказує фактор X_3 .

Також необхідно перевірити побудовану модель на адекватність реальному процесу шляхом порівняння даних, розрахованих за моделлю і визначених експериментально.

Для перевірки адекватності моделей обраний критерій Фішера [10]:

$$F_e = \frac{S_{ad}^2}{S_e^2}, \quad (14)$$

що є відношенням двох дисперсій – відтворюваності та адекватності.

Дисперсія відтворюваності [10]:

$$S_e^2 = \frac{\sum_{i=1}^n S_{ni}^2}{n}, \quad (15)$$

де S_{ni}^2 – середньоквадратичне відхилення при вимірюванні i -го результату.

Розсіювання експериментальних точок щодо розрахункових характеризується залишковою дисперсією або дисперсією адекватності [11]

$$S_{ad}^2 = \frac{\sum_{i=1}^m y_i - y_{pi}^2}{m - 1 - 1}, \quad (16)$$

де y_i, y_{pi} – розрахункові та експериментальні значення; m – число порівнюваних значень; 1 – число коефіцієнтів, визначених за дослідними даними.

Теоретична залежність покладається адекватною дослідній, якщо отримане експериментальне значення критерію Фішера F_e менше табличного $F_{табл}$, обраного за кількістю ступенів свободи чисельника і знаменника, а також довірчої ймовірності α :

$$F_e < F_{табл}. \quad (17)$$

В даному випадку табличне значення при ступенях свободи $k_1 = 3$ та $k_2 = n - m - 1 = 20 - 3 - 1 = 16$, $F_{табл}(3;16) = 3,24$, а експериментальне – $F_e = 1,25$.

Експериментальне значення критерію Фішера не перевищує табличного значення. Тому розроблену модель можна вважати адекватною і покласти в основу подальших досліджень.

Для того, щоб з'ясувати які з регіонів за своїми характеристиками (факторами) близькі один до одного, а які ні, використано кластерний аналіз, в якому використовується деяка міра подібності або відстань між об'єктами. Кластерний аналіз проводили в пакеті STATISTICA [12].

Типовим результатом такої кластеризації є ієрархічне дерево. Процес вибору кількості кластерів найбільш ефективно відображає процедура ієрархічної класифікації об'єктів (рис. 2).

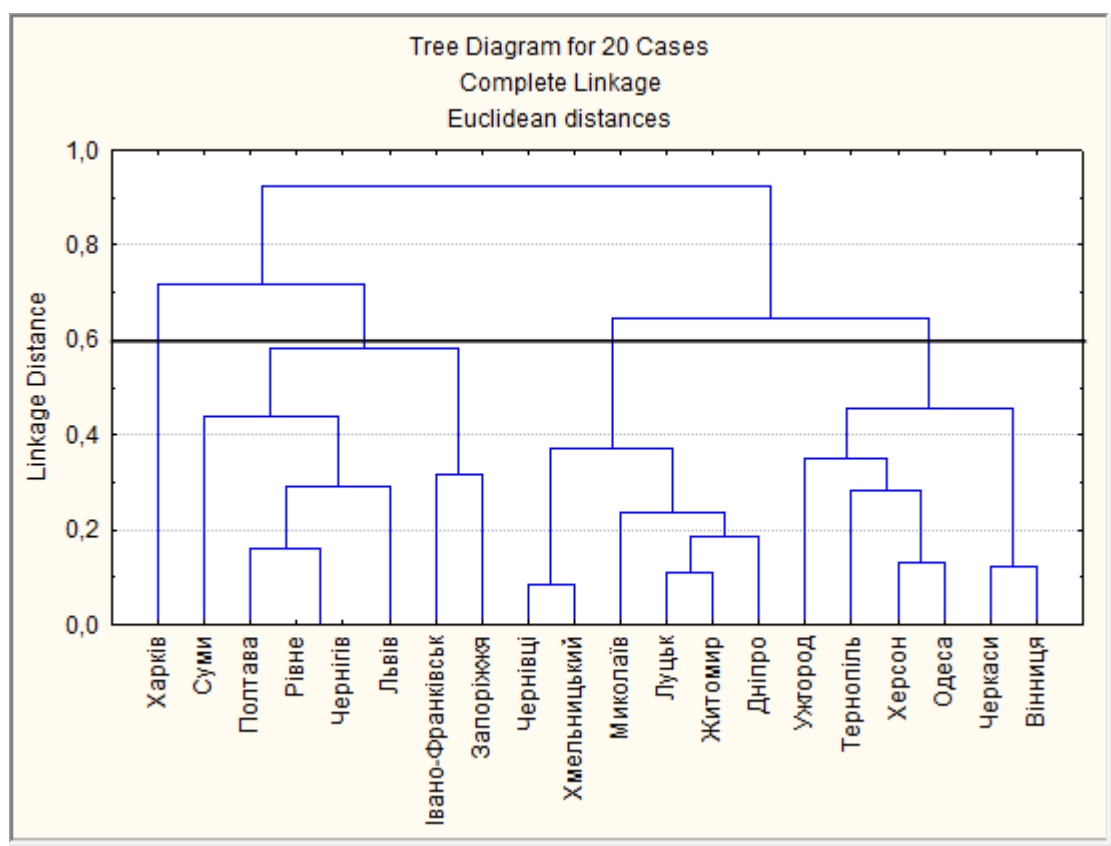


Рис. 2. Дендрограма ієрархічної агломеративної кластеризації регіонів за комплексним показником стану ІТ-галузі

На даній діаграмі горизонтальна ось являє собою спостереження, вертикальна ось представляє відстань об'єднання. Для кожного вузла в графі (там, де формується новий кластер) представлена величина відстані, для якого відповідні фактори зв'язуються в новий єдиний кластер [12]. Коли дані мають ясну "структуру" в термінах кластерів об'єктів, схожих між собою, тоді ця структура, швидше за все, повинна бути відображена в ієрархічному дереві

різними гілками. В результаті успішного аналізу методом об'єднання з'являється можливість виявити кластери (гілки) і інтерпретувати їх.

Метод деревовидної кластеризації було використано при формуванні кластерів несходства або відстані між об'єктами. Найбільш прямий шлях обчислення відстаней між об'єктами в багатовимірному просторі полягає в обчисленні евклідових відстаней [12]:

$$d_{Eij} = \left(\sum_{l=1}^m x_i^l - x_j^l \right)^2 \Bigg)^{\frac{1}{2}}. \quad (18)$$

Правило, за яким об'єднувались кластери – одиночний зв'язок (відстань «Найближчого сусіда») [12]:

$$\rho_{\min} K_i, K_j = \min_{x_i \in K_i, x_j \in K_j} \rho_{x_i, x_j} \quad (19)$$

З рис. 2 видно, що кількість кластерів при 60% розрізі дорівнює трьом: перший кластер включає такі об'єкти, як Харків, Суми, Полтава, Рівне, Чернігів, Львів, Івано-Франківськ, Запоріжжя; до другого кластеру входять Ужгород, Тернопіль, Херсон, Одеса, Черкаси, Вінниця; до третього кластеру – Чернівці, Хмельницький, Миколаїв, Луцьк, Житомир, Дніпро.

Для уточнення щодо отриманого розподілу регіонів за кластерами при проведенні дослідження також було використано метод k-середніх (k-means clustering), який відрізняється від ієрархічних агломеративних методів тим, що вже до проведення розрахунків відомо кількість кластерів (3 в нашому випадку) та ці кластери розташовані на великій відстані один від одного (рис. 3).

Members of Cluster Number 1 and Distances from Respective Cluster contains 7 variables		Members of Cluster Number 2 and Distances from Respective Cluster contains 5 variables		Members of Cluster Number 3 and Distances from Respective Cluster contains 8 variables	
Variable	Distance	Variable	Distance	Variable	Distance
Запоріжжя	0,166924	Вінниця	0,094008	Дніпро	0,130886
Івано-Франківськ	0,185370	Одеса	0,135125	Житомир	0,039378
Львів	0,115235	Тернопіль	0,077643	Луцьк	0,072346
Суми	0,163105	Ужгород	0,136502	Миколаїв	0,091464
Харків	0,264853	Черкаси	0,144089	Хмельницький	0,096540
Чернігів	0,126594			Чернівці	0,085837
Рівне	0,126594			Полтава	0,146465
				Херсон	0,125221

Рис. 3. Розподілення регіонів за групами методом k-середніх за комплексним показником стану ІТ-галузі

За результатами проведеного аналізу даним методом розподіл регіонів за кластерами наступний: перший кластер – Харків, Суми, Рівне, Чернігів, Львів, Івано-Франківськ, Запоріжжя; другий кластер – Ужгород, Тернопіль, Одеса,

Черкаси, Вінниця; третій кластер – Чернівці, Хмельницький, Миколаїв, Луцьк, Житомир, Дніпро, Полтава, Херсон.

З представленою аналізу видно, що до складу сукупності, що аналізували, входили чотири великі ІТ-центри України (Харків, Львів, Дніпро, Одеса) з п'яти (за Києвом інформація відсутня). Не всі кластери за двома методами кластерного аналізу включають однаковий склад регіонів, але розподіл флагманів за кластерами збігається у обох випадках, що дає привід стверджувати про наявність трьох шляхів розвитку ІТ-індустрії в Україні та обґрунтовує подальше вивчення цього питання.

Висновок

В ході дослідження було проведено опитування експертів за додатковими трьома регіонами, оцінено за цими регіонами кожний з показників стану ІТ-галузі та проведено розрахунки за моделлю оцінки регіонів згідно з критеріями оцінювання. За допомогою кореляційно-регресійного аналізу з'ясовано силу зв'язку між обраними факторами та комплексним показником і виявлено значність впливу факторів (показників) на комплексний показник з урахуванням внесених змін в набір даних. Отримані результати дозволили стверджувати, що фактори, які входять до моделі, мають значущий вплив на комплексний показник та однозначно визначають його стан навіть з врахуванням внесених змін. Також було визначено, що й при доданих показниках дана модель за критерієм Фішера є адекватною. Для розподілу регіонів за розрахованими комплексними показниками стану ІТ-галузі було використано кластерного аналізу у межах математичного пакету STATISTICA. Отримані результати підтверджують гіпотезу про наявність трьох шляхів розвитку ІТ-індустрії в Україні та надають підґрунтя для проведення додаткових досліджень.

Список літератури

1. Vynnychuk, R. O. & Skljарuk, T. V. (2015) Особливості розвитку ІТ-ринку в Україні: стан та тенденції [*Osoblyvosti rozvytku IT-rynku v Ukrajinі: stan ta tendenciji*]. Visnyk Nacionaljnogho universytetu "Ljvivsjka politekhnika". Loghistryk, .833, 3-8.

2. Rubin, Je. (2018) Перемога под елочку: ІТ-кластер координує програму облсовета «ІТ-Харьковщина» [*Peremoga pod elochku: IT-klaster koordiniruet programmu oblsoveta «IT- Kharjkovshhyna»*]. Retrieved from <https://dou.ua/lenta/columns/it-kharkiv/>.

3. Stremetsjka, Daryna. (2017) Регулювання і перспективи ІТ-ринку України [*Reghuljuvannja i perspektyvu IT-rynku Ukrainy*]. Retrieved from <https://dou.ua/forums/topic/22075/>.

4. Grabovskaja, Anna. (2019) Токсичный HR: действия, которые отравляют команду. [*Toksichnyj HR: dejstvija, kotorye otravljajut komandu*]. Retrieved from <https://dou.ua/lenta/articles/toxic-hr/>.

5. Brynza, N. A. & Gavrilova, A. A. (2018) Многофакторная оценка показателей развития ИТ-отрасли в регионах Украины [*Mnogofaktornaja ocenka pokazatelej razvitija IT-otrasli v regionah Ukrainy*]. Zbirnyk naukovykh pracj Kharkivskogo nacionaljnogo universytetu Povitrjanykh Syl, 2(56), 159-169.

6. Podinovskiy, V.V.& Nogin, V.D. (Ed.) (1982). Парето-оптимальные решения многокритериальных задач [*Pareto-optimalnye reshenija mnogokriterial'nyh zadach*]. (254 p.). Moscow: Nauka.

7. Tikhonov, A.N. & Arsenin, V.Ya. (Ed.) (1986). Методы решения некорректных задач [*Metody reshenija nekorrektnyh zadach*] (286 p.). Moscow: Nauka.

8. Fishburn, P. (Ed.) (1978). Теория полезности для принятия решений [*Teorija poleznosti dlja prinjatija reshenij*] (352 p.). Moscow: Nauka.

9. Kuzmin, I.V. & Petrov, E.G. & Alferov, I.A. & Evseev, V.V. & Migunova, L.V. (Ed.) (1978). Автоматизированные системы управления городским хозяйством [*Avtomatizirovannye sistemy upravlenija gorodskim hozjajstvom*] (144 p.). Kyiv, Budiveljnuk.

10. Aliev, T. A. (1991) Экспериментальный анализ [*Jeksperimental'nyj analiz*] (272 p.). Moscow: Mashinostroenie.

11. Tavanjuk, T.Ja. & Nikolaenko, A.P. & Romanchenko, A.V. & Shumakova T.A. (2017) Оценка адекватности математических моделей характеристик электрогидравлических следящих приводов [*Ocenka adekvatnosti matematicheskikh modelej harakteristik jelektrogidravlicheskih sledjashhih privodov*]/. Visnyk skhidnoukrajins'kogho nacionaljnogo universytetu imeni Volodymyra Dalja, 7 (237), 89-94.

12. Буреева Н. Н. Многомерный статистический анализ с использованием ППП “STATISTICA”. Учебно-методический материал по программе повышения квалификации «Применение программных средств в научных исследованиях и преподавании математики и механики». Нижний Новгород, 2007, 112 с.