

Л.О. Чаговець¹, В.В. Чаговець²

¹Харківський національний економічний університет ім. С. Кузнеця, Україна

²Державний біотехнологічний університет, Україна

МОДЕЛІ ІДЕНТИФІКАЦІЇ ТА ПРОГНОЗУВАННЯ СТАНУ ЦИФРОВІЗАЦІЇ КРАЇН У СВІТОВОМУ ПРОСТОРІ

У статті виконано розробку моделей ідентифікації та прогнозування стану цифровізації країн світу на базі просторово-динамічної вибірки та проаналізовано їх розподіл за кластерами. Методами багатовимірною аналізу та Data Science знайдено позиціонування України відносно країн світу. Для прогнозування стану цифровізації країн проведено експерименти з побудови різних видів моделей дерев класифікації й отримано модель за CART-аналізом. Доведено її адекватність.

Ключові слова: модель, дерева класифікації, прогнозування, цифровізація, індикатори цифровізації, телекомунікаційний розвиток.

Постановка проблеми

Проблеми та перспективи цифровізації країн набувають актуального значення на сучасному етапі розвитку технологічної сфери людства. Цифровізація стає найважливішим фактором економічного зростання будь-якої країни і взагалі є сучасним трендом її розвитку. Несвоєчасне усунення проблем цифровізації здатне спричинити негативні наслідки в масштабах усієї економіки країни. Тому щорічно розраховуються різні рейтинги країн щодо цифровізації.

Так, нова редакція звіту Digital Riser від Європейського центру цифрової конкурентоспроможності, оприлюдненого у 2021 році, містить глобальний рейтинг країн, який охоплює дані зі 140 країн світу [1]. У звіті наводяться дані про цифрову конкурентоспроможність держав, порівнюючи їх динаміку у конкретному регіоні впродовж останніх трьох років. За цими даними серед країн G7 Канада досягла найбільших успіхів у цифровій конкурентоспроможності між 2018 та 2020 роками і посіла перше місце у своїй групі країн, Японія та Німеччина навпаки найбільше опустилися в рейтингу в межах G7, а Італія піднялась з останнього місця в попередньому році до другого у 2021 році. Першими трьома переможцями серед G20 стали Китай, Саудівська Аравія та Бразилія. Внизу рейтингу – Німеччина, Японія та Індія. В Євразії за останні три роки найбільше покращила свій сектор цифрових технологій Грузія. Україна в цьому рейтингу піднялась на одну сходинку вище за Росію (рис. 1). Успіхам країн у цифровому секторі, на думку Digital Riser, сприяє «слідування комплексним планам з пріоритетними цілями та

зосередження на підприємстві» [1].

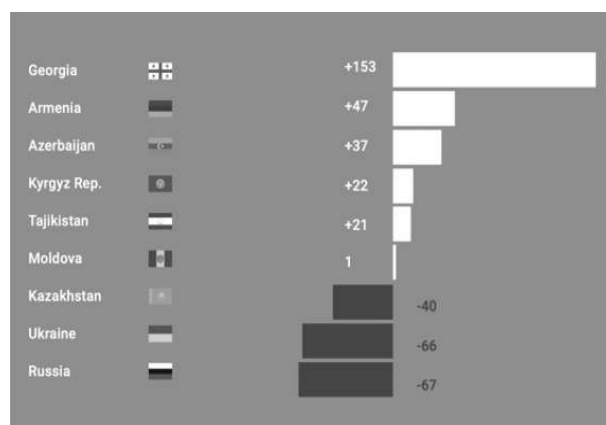


Рис. 1. Зміна рейтингу країн у цифровій конкурентоспроможності

Формування України як цифрової держави відбувається швидкими темпами, незважаючи на тяжкі умови війни. Так, у 2022 році в рейтингу за технологічними компетенціями, оприлюдненим у щорічному дослідженні Global Skill Report 2022 від Courser, Україна посіла восьме місце [2]. Проте наявність значної кількості як зовнішніх, так і внутрішніх факторів впливу на розвиток цифровізації держави свідчить про необхідність системного підходу до дослідження цього питання.

Перспективи цифрового розвитку країн та відповідні моделі ідентифікації й прогнозування стану їх цифровізації потребують особливого дослідження, тому що це дозволяє виявити чинники цифрового зростання. Слід зазначити також, що обґрунтування з точки зору системного підходу та оцінки сталого цифрового розвитку країн із

застосуванням сучасних методів системного аналізу та багатовимірному аналізу Data Science, як одних з найефективніших у виконанні математичного моделювання станів досліджуваних процесів, не отримало достатньо розвитку в наукових роботах і потребує подальшої розробки.

Аналіз останніх досліджень і публікацій

Глобальні проблеми цифровізації країн світу знайшли відображення у роботах багатьох закордонних вчених, серед яких слід особливо відзначити праці Aytekin Ç., Appiah-Otoo I., Değerli A., Demissie D., Francisco J., Ftiti Z., Haddaji A., Kallal R., Karagiannis G., Na Song, Osman I. H., Pardo T., Ravanos P., Rorissa A., Zablith F., Whitmore A. A. Питання оцінки стану цифрового та телекомунікаційного розвитку України досліджувалися у наукових роботах вітчизняних вчених: Апалькова В. В., Білик О. І., Варламова М., Гавкалова Н. Л., Дем'янова Ю., Левицька Я. В., Ночвай В., Подольчак Н. Ю., Пищуліна О., Сенкевич О. Ф., Семенов А. Ю., Ткалич Т. А., Тутова О. В. та інших.

Проведений аналіз існуючих останніх досліджень і публікацій в оцінюванні рівня телекомунікаційного прогресу в контексті сучасних процесів цифровізації країн [3 – 10] дозволив виявити та охарактеризувати різні підходи до дослідження питань розвитку цифровізації країн, побудови їх ІКТ-рейтингу (The Information and Communication Technology Development Ranking). Так, по-перше, встановлено, що існуючі моделі оцінки рівня цифровізації країн, які базуються на традиційних індикаторах економічного добробуту на основі внутрішнього валового продукту, не враховують інші важливі соціальні чинники, і це зумовило створення низки спеціальних узагальнюючих індикаторів цифрового розвитку: глобального індексу розвитку електронного уряду, індексу мережної готовності та індексу розвитку інформаційних і комунікаційних технологій [7, 8, 9]. Проте у системі індикаторів цифровізації країн, як зазначає багато дослідників, існують все ж таки суттєві протиріччя і найбільш повним з точки зору оцінки темпів розвитку інформаційних технологій, моніторингу та порівняння розвитку інформаційно-комунікаційних технологій (ІКТ) між країнами у часі вважається індекс розвитку ІКТ (The ICT Development Index). Він є інтегральним індексом, який об'єднує в собі одинадцять підіндексів та використовується для оцінювання:

- рівня розвитку ІКТ країн відносно інших;
- прогресу розвитку ІКТ як у розвинутих країнах, так і в країнах, що розвиваються;
- цифрового розриву, тобто відмінностей між країнами відносно рівня розвитку ІКТ;

- потенціалу розвитку ІКТ та ступеню використання їх для посилення розвитку в контексті наявних можливостей та навичок.

Досліджуючи світові рейтинги та аналізуючи місце України в світлі цих глобальних індексів [10], слід зазначити, що всі країни у цих рейтингах займають різне місце. Це ускладнює сприйняття загальної ситуації і визначення позиціонування України за рівнем телекомунікаційного розвитку на макрорівні. Тому вважається доцільним побудова такої класифікації країн за цифровізацією, яка була б корисною для прогнозування стану цифровізації нашої країни під час формування та ухвалення управлінських рішень з імплементації стратегій цифрового розвитку.

По-друге, у результаті аналізу останніх публікацій встановлено, що для виконання оцінки стану цифровізації країн застосовується різний інструментарій. Звичайно використання інструментальних засобів пакетів прикладних програм та систем дозволяє виконати велику кількість розрахункових операцій, що є особливо важливим під час побудови математичних моделей оцінки цифрового розвитку країн. Найвагомішою перевагою застосування сучасних інформаційних систем і технологій є здобуття кількісно обґрунтованих та якісних рішень з формування та ухвалення напрямків розвитку цифровізації країн. Активне використання сучасного програмного інструментарію забезпечує реалізацію наукового пошуку серед моделей, достатній рівень інформаційного менеджменту, швидкий доступ до баз даних та багатовимірних кубів, виконання візуалізації результатів досліджень, що виконуються та підвищення їх статичної значущості й адекватності побудованих моделей.

Метою статті є удосконалити комплекс математичних моделей оцінки й аналізу стану цифровізації країн світу, який, на відміну від інших, буде побудований на базі методів інтелектуального аналізу багатовимірних об'єктів. Для цього маємо вирішити такі завдання: виконати і проаналізувати розподіл країн світу за кластерами на базі просторово-динамічної вибірки даних та виявити позиціонування України у світовому просторі методами багатовимірному аналізу та Data Science; отримати модель дерева класифікації для прогнозування стану цифровізації країн у майбутньому, а також довести адекватність цієї моделі.

Викладення основного матеріалу

Для досягнення поставленої мети авторами використано такий програмний засіб, як система Statistica 12 та її програмні модулі: Multiple Regression, Multivariate Exploratory Technics/Cluster

Analysis, Factor Analysis, Classification Trees. Також для порівняння низки альтернативних алгоритмів побудови комплексу моделей оцінки цифрового розвитку країн краще використовувати вбудовані модулі, які реалізують такі методи машинного навчання та штучного інтелекту, як методи нейронних мереж та ансамблі дерев рішень.

Аналіз даних у системі проводиться в режимі діалогових вікон аналізу, які розроблено так, що на першій вкладці розташовано необхідні опції, а на наступних вкладках – поглиблені методи та спеціальні опції. У модулі виконання кластерного аналізу реалізовано методи *k*-середніх, ієрархічної кластеризації та двохходова кластеризація. Дані можуть надходити як у вихідному вигляді, і у вигляді матриці відстаней між об'єктами. Спостереження або змінні можна кластеризувати, використовуючи різні метрики – евклідову відстань, квадрат евклідової відстані, манхеттенську відстань, відстань Чебишева, відсоток незгоди та 1-коефіцієнт кореляції Пірсона, а також різні правила об'єднання (ближнього сусіда, повного зв'язку, незважене та зважене попарне середнє за групами, незважене та зважене відстань між центрами, метод Уорда. Під час проведення кластерного аналізу методом *k*-середніх існує можливість задавати початкове розташування центрів кластерів. Алгоритм кластеризації дозволяє виконувати в ієрархічному групуванні обчислення матриці до 90 тис. відстаней. Крім стандартних результатів кластерного аналізу, у модулі є доступним різноманітний набір описових статистик та розширених діагностичних методів (схема поєднання з пороговими рівнями ієрархічної кластеризації, таблиця дисперсійного аналізу методом *k*-середніх). Інформація про належність об'єктів до кластерів може бути додана до файлу даних та використовуватись у подальшому аналізі. Графічні можливості модуля містять дендрограми, двохходові діаграми об'єднань, графічне подання схеми об'єднання, діаграму середніх за методом *k*-середніх.

Модуль факторного аналізу містить набір статистик ієрархічного факторного аналізу з розширеною діагностикою. Тут виконують аналіз головних компонент (загальний та ієрархічний косокутний) та аналіз головних факторів для наборів даних, що містять до 300 змінних). Система також містить програму для аналізу головних компонент. Вихідними датасетами є власні значення (звичайні, кумулятивні та відносні), навантаження факторів та коефіцієнти факторних балів (які можна додати до файлу вхідних даних, переглянути на піктографіку та в інтерактивному режимі перекодувати), а також спеціальні методи обертання факторів: варімакс, біквартімакс, квартімакс та еквімакс (за нормалізованими або початковими

навантаженнями), а також косокутні обертання. Простір факторів можна візуально подавати у дво- або тривимірних діаграмах розсіювання; графіки «каменистого осипу», діаграми розсіювання, гістограми, лінійні графіки залишкової дисперсійної матриці.

У модулі реалізації методу канонічних кореляцій передбачена робота як із файлами вихідних даних, так і з кореляційними матрицями; обчислюються всі стандартні статистики канонічної кореляції (власні вектори та власні значення, коефіцієнти надмірності, канонічні вагові коефіцієнти, навантаження, дисперсії, критерії значущості для кожного з коренів), а також розширені статистики. Для кожного спостереження обчислюються значення канонічних змінних. Набір графічних засобів містить діаграми власних значень та канонічних кореляцій, діаграми значень канонічних змінних [12].

Модуль дерев класифікації містить реалізацію розроблених методів побудови конфігурації дерев та їх тестування (метод дерев класифікації є ітераційним способом прогнозування класу об'єкту за значеннями передикторних змінних. Древа класифікації можна будувати за категоріальними або порядковими показниками, а також за набором показників обох типів на підставі алгоритмів розгалужень за окремими змінними або їх лінійними комбінаціями.

З метою виконання поставлених завдань дослідження, частина яких є чітко алгоритмізована, постає необхідність залучення програмних модулів та систем обробки великих обсягів даних. Формування первісної сукупності індикаторів оцінки цифрового розвитку країн виконаємо на підставі монографічного аналізу та результатів обробки наборів відкритих статистичних даних та рейтингів. Аналіз літературних джерел дозволив виокремити таку групу головних індикаторів оцінки інформаційно-комунікаційного та цифрового розвитку країн:

$x_1(NRI)$ – індекс мереженої готовності (The Networked Readiness Index),

$x_2(EGI)$ – індекс розвитку електронного уряду (E-Government Development Index),

$x_3(ICT)$ – індекс розвитку інформаційно-комунікаційних технологій (The ICT Development Index),

$x_4(FH)$ – індекс мережної свободи країни.

Так як оцінка стану цифрового розвитку здійснюється для країн світу, які мають різний рівень соціально-економічного розвитку, та відповідно до класифікації Світового банку їх відносять до категорій з високим, середнім, вище середнього, нижче середнього та низьким рівнем доходів у перерахунку на душу населення, тому

розширимо загальний пул індикаторів такими показниками групи індексів соціально-економічного та інноваційного розвитку країн:

x_5 (*GII*) – міжнародний індекс інновацій (Global Innovation Index),

x_6 (*GCI*) – індекс глобальної конкурентоспроможності (Global Competitiveness Index)

x_7 (*SPI*) – індекс соціально-економічного прогресу (Social Progress Index).

Саме ці індикатори характеризують стан цифровізації за співвідношенням інституційних

індикаторів, які фактично відображають загрози втрати інформаційної безпеки, індексів цифрового розвитку та індексів оцінки стану соціально-економічного розвитку будь-якої країни.

Для висунення гіпотези по можливої кількості виокремлених кластерів – груп країн за станом цифровізації використаємо різні методи ієрархічної агломеративної кластеризації. По-перше, методом Уорда побудуємо розподіл країн за кластерами на базі просторово-динамічної вибірки даних. Аналіз графіку дендрограми свідчить про можливість виокремлення чотирьох кластерів (рис. 2).

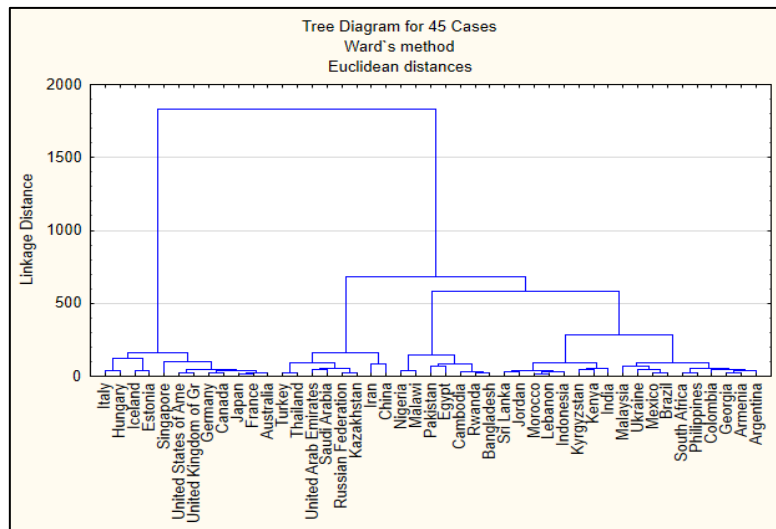


Рис. 2. Дендрограма кластеризації

Аналіз центроїдів кластерів та результати дисперсійного аналізу підтвердили гіпотезу про доцільність формування чотирьох груп об'єктів. Візуальне зображення центрів кластерів (центроїдів) подано у вигляді графіку *k*-середніх на рис. 3. Як видно із графіку, існує чітке розмежування кластерів як за набором індикаторів цифровізації та телекомунікаційного розвитку, так і за соціально-економічними індикаторами, які однозначно ідентифікують стан цифрового розвитку країн.

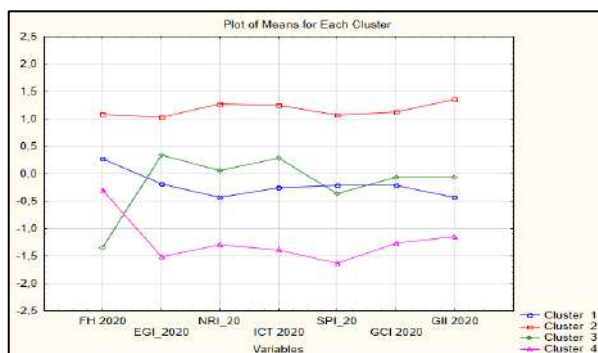


Рис. 3. Графік *k*-середніх (центроїдів) кластерів

Аналіз внутрішньогрупових та міжгрупових дисперсій також свідчить про чітке розмежування та розділення кластерів. Значення *F*-критерію Фішера та *p*-рівень підтверджують статистичну значущість оцінених індексів на рівні 99 % та адекватність проведеної кластеризації в цілому (рис. 4).

Variable	Analysis of Variance (даные страны все стандарт)					
	Between SS	df	Within SS	df	F	signif. p
FH 2020	40,79081	3	14,58501	52	48,4772	0,000000
EGI_2020	42,22464	3	10,69347	52	68,4430	0,000000
NRI_20	44,25632	3	6,54215	52	117,2566	0,000000
ICT 2020	45,07359	3	10,47830	52	74,5613	0,000000
SPI_20	45,48290	3	7,16040	52	110,1014	0,000000
GCI 2020	36,27549	3	13,37599	52	47,0078	0,000000
GII 2020	43,65920	3	9,82539	52	77,0208	0,000000

Рис. 4. Дисперсійний аналіз групування у 2020 році

Аналіз виконаного групування вимагає та дозволяє надати характеристику кожному кластеру:

1. Кластер, який об'єднує країни з високим (В) станом цифровізації, телекомунікаційним та цифровим розвитком. На графіку середніх це кластер № 2. Він характеризується одночасним

високим рівнем мережної свободи та мережної готовності, високим рівнем розвитку інформаційних технологій та високим рівнем соціально-економічного розвитку. До цього кластеру належать переважно високо конкурентні країни світу, Європейського Союзу (Австралія, Канада, Естонія, Франція, Німеччина, Угорщина, Ісландія, Італія, Японія, Республіка Корея, Сінгапур, Сполучене Королівство Великої Британії та Північної Ірландії), також Сполучені Штати Америки. Тобто економіка цих країн характеризується як високим рівнем цифровізації, так і потужним темпом промислового розвитку в цілому, підтримкою та впровадженням новітніх цифрових інструментів в реальному секторі економіки, розвинутим ринком інформаційних продуктів та послуг усіх сегментів B2B, B2C, B2G тощо, високими темпами розвитку електронної комерції. Високі соціальні стандарти надають можливості залучення, навчання і утримання висококваліфікованих ІТ-спеціалістів. Усі ці аспекти синергетично діють та збільшують окремі ефекти цифрової економіки.

2. Кластер країн із середнім (С) станом цифровізації та одночасно із середнім станом економічного розвитку, який описується відповідно до класифікації Світового банку як стан країн із середнім та нижче середнього доходів на душу населення (на графіку середніх – кластер № 3). Особливістю країн, які потрапляють до цього кластеру, є те, що, не зважаючи на достатні умови для економічного та соціального розвитку, вони мають цифрову інфраструктуру, яка є обмеженою. Це переважно країни із низьким ступенем мережної та інтернет-свободи. До кластеру увійшли такі країни, як Китай, Іран, Казахстан, Малайзія, Російська Федерація, Саудівська Аравія, Таїланд, Туреччина, Об'єднані Арабські Емірати. Ці обмеження є гальмівними для інтенсивного та природного цифрового розвитку, зменшують темпи розвитку ринку інновацій та соціальних трансфертів.

3. Кластер країн із середнім станом цифровізації та одночасними темпами соціального розвитку нижче середнього (НС). На графіку середніх це кластер № 1. До кластеру увійшли Аргентина, Вірменія, Бразилія, Колумбія, Грузія, Угорщина, Індія, Індонезія, Ліван, Мексика,

Марокко, Філіппіни, Південна Африка, Шрі Ланка, Україна. За класифікацією Світового банку це країни із доходом на душу населення нижче середнього, але вони мають достатні потужності та промисловий потенціал цифрового розвитку. До такого кластеру належить і Україна. Якщо виділити головні загрози для нашої країни з точки зору цифровізації, то до них слід віднести недостатнє залучення венчурних інвестицій в підприємства та стартапи, низьке вітчизняне фінансування НДДКР, недостатність мобільного та широкополосного інтернету у малих територіальних громадах, недостатнє заохочення інвестицій в цифрові підприємства, високі темпи експорту ІТ-кадрів.

Перевагою цього кластеру над попереднім є більш високий рівень свободи в інтернеті та доступності мереж, демократичність цифрових взаємодій та трансформацій. Країнам цього кластеру слід зробити пріоритетом ефективне управління в напрямку міграційних процесів, розвиток соціально-економічних умов для стримування іноземної релокації фахівців з діджиталізації та ІТ, нарощування темпів цифровізації через успішну адаптацію найкращих досвідів розвинутих країн першого кластеру.

4. Кластер країн низького (Н) стану цифровізації (на графіку середніх – кластер № 4). До них належать Бангладеш, Камбоджа, Єгипет, Кенія, Малаві, Нігерія, Пакистан, Руанда. Він поєднує країни, для яких є характерним як низькі темпи цифрового розвитку, так і соціально-економічного. Необхідно приділяти суттєву увагу початковим передумовам телекомунікаційного розвитку – нарощуванню економічних темпів промислового сектору, галузей фінансового та реального секторів економіки. Кризові явища у таких економіках не дозволяють здійснити швидкі темпи зростання внаслідок критичного відставання в промисловому, якісному соціально-демографічному розвитку, існують високі ризики реалізації загроз економічній безпеці країни.

Кластеризація країн за індексами цифровізації у розрізі 2013 – 2020 рр. дозволяє встановити стан цифрового розвитку для кожної окремої країни. Країни, для яких не було достатньої статистичної бази у вибраному часовому періоді, виключимо із моделювання (табл. 1).

Таблиця 1

Результати групування країн за станом ЦРК

Країна	ПД	Рік							
		2020	2019	2018	2017	2016	2015	2014	2013
1	2	3	4	5	6	7	8	9	10
Аргентина	НС	НС	НС	НС	НС	НС	НС	НС	НС
Арменія	НС	НС	НС	НС	НС	НС	НС	НС	НС

Продовження табл. 1

1	2	3	4	5	6	7	8	9	10
Австралія	В	В	В	В	В	В	В	В	В
Бангладеш	Н	Н	Н	Н	Н	Н	Н	Н	Н
Бразилія	НС	НС	НС	НС	НС	НС	НС	НС	НС
Камбоджа	Н	Н	Н	Н	Н	Н	Н	Н	Н
Канада	В	В	В	В	В	В	В	В	В
Китай	С	С	С	С	С	С	С	С	С
Колумбія	НС	НС	НС	НС	НС	НС	НС	НС	НС
Еквадор	-	НС	НС	НС	НС	НС	-	-	НС
Єгипет	Н	С	Н	Н	Н	Н	Н	Н	С
Естонія	В	В	В	В	В	В	В	В	В
Ефіопія	-	Н	Н		Н	Н	Н	Н	Н
Франція	В	В	В	В	В	В	В	В	В
Грузія	НС	НС	НС	НС	НС	НС	НС	НС	НС
Німеччина	В	В	В	В	В	В	В	В	В
Угорщина	НС	В	НС	НС	НС	НС	НС	НС	НС
Ісландія	В	В	В	В	В	В	В	В	В
Індія	НС	НС	НС	НС	НС	Н	Н	НС	С
Індонезія	НС	НС	НС	НС	НС	НС	НС	НС	НС
Іран	С	С	С	С	С	С	Н	Н	С
Італія	В	В	В	В	В	В	В	Не	Не
Японія	В	В	В	В	В	В	В	В	В
Йорданія	НС	НС	НС	НС	НС	НС	НС	НС	НС
Казахстан	С	С	С	С	С	С	С	С	НС
Кенія	Н	Н	Н	Н	Н	Н	Н	НС	Н
Киргизстан	НС	НС	НС	НС	НС	НС	НС	Н	Н
Ліван	НС	НС	НС	НС	НС	НС	НС	НС	НС
Малаві	Н	Н	Н	Н	Н	Н	Н	Н	Н
Малайзія	С	В	С	С	С	С	С	С	НС
Мексика	НС	НС	НС	НС	НС	НС	НС	НС	НС
Марокко	НС	НС	НС	НС	НС	НС	НС	НС	НС
Нігерія	Н	Н	Н	Н	Н	Н	Н	Н	Н
Пакистан	Н	Н	Н	Н	Н	Н	Н	Н	Н
Філіппіни	НС	НС	НС	НС	НС	НС	НС	НС	НС
Республіка Корея	-	В	В	В	-	-	-	-	В
Російська Федерація	С	С	С	С	С	С	С	С	НС
Руанда	Н	Н	Н	Н	Н	Н	Н	Н	Н
Саудівська Аравія	С	С	С	С	С	С	С	С	НС
Сінгапур	В	В	В	В	В	В	В	В	В
Південна Африка	НС	НС	НС	НС	НС	НС	НС	НС	НС
Шрі Ланка	НС	НС	НС	НС	НС	НС	НС	НС	С
Таїланд	С	С	С	С	С	С	С	С	С
Туніс	-	НС	НС	НС	-	-	-	-	Н
Туреччина	С	С	С	С	С	С	С	С	НС
Уганда	-	Н	-	Н	Н	Н	Н	Н	Н

Продовження табл. 1

1	2	3	4	5	6	7	8	9	10
Україна	НС	НС	НС	НС	НС	НС	НС	НС	НС
ОАЕ	С	С	С	С	С	С	С	С	НС
Англія	В	В	В	В	В	В	В	В	В
США	В	В	В	В	В	В	В	В	В
Уругвай	-	С	С	С	-	-	-	-	НС
В'єтнам	-	С	-	-	-	-	-	-	С
Замбія	-	Н	-	-	-	-	-	-	Н
Зімбабве	-	Н	Н	Н	Н	-	Н	Н	Н

Як видно із таблиці протягом років відбувається стабілізація розподілу країн між кластерами. Детальний аналіз групування країн за просторово-динамічною вибіркою підтверджує гіпотезу про сталість розбиття, тому можна дійти висновку про високу інерційність у процесах цифрового та соціально-економічного розвитку країн.

Відповідно до концептуальної схеми дослідження наступним кроком є розробка моделі прогнозування стану ЦРК. Для виконання такої задачі, як було зазначено, існує низка методів класифікації з навчанням або методів супервайзорного навчання, до яких належать методи дискримінантного аналізу та дерев класифікації, методи глибинного навчання. У попередніх роботах [9, 10] було показано, що методи дерев класифікації є одними з найкращих для ідентифікації нових об'єктів оцінки стану цифровізації. Розглянемо розробку такої моделі.

Побудову моделі прогнозування стану ЦРК виконаємо такими етапами:

- завдання критерію точності розгалуження та прогнозу;
- вибір варіанту розгалуження;
- визначення моменту, коли подальше розгалуження необхідно завершити;
- визначення оптимального розміру дерева.

Розглянемо їх детальніше.

На першому етапі у якості критерію точності побудови дерева вибрано G-індекс Джині, який є найбільш уживаним та дає найменшу помилку навчання дерева. Для навчання та тестування резервування об'єктів вибірки здійсимо у співвідношенні 60 % та 30 %. Також зарезервуємо 10 % для глобальної крос-перевірки дерева.

Для вибору найкращої конфігурації проведемо експерименти із побудови різних моделей дерев з використанням таких альтернативних методів, як одновимірне розгалуження, CART-аналіз, розгалуження за лінійною комбінацією змінних. Також проведемо моделювання методом Random Forest. Аналіз результатів застосованих методів

побудови дерева класифікації дозволив дійти висновку, що найкращою за критерієм точності побудови дерева є модель за CART-аналізом.

Зазвичай на першому етапі вибір найкращої конфігурації дерева здійснюють за найменшою ціною навчання та найменшим співвідношенням між ціною навчання та ціною крос-перевірки. У нашому випадку графік послідовності цін навчання та крос-перевірки дозволив встановити найкращу конфігурацію – дерево № 1 з чотирьох можливих конфігурацій (рис. 5).

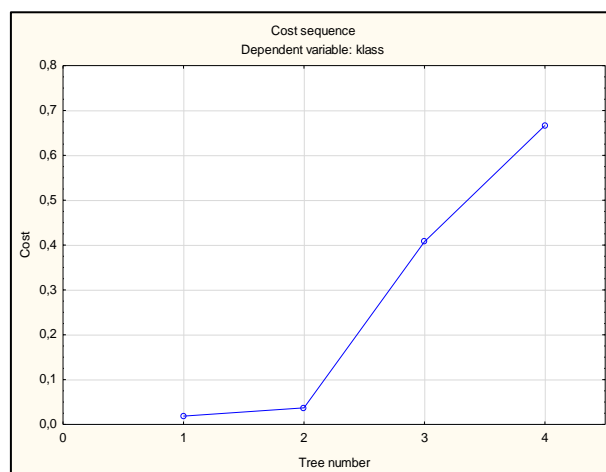


Рис. 5. Графік вартості помилок сформованого набору дерев класифікації

Для проведення другого етапу побудови моделі дерева – аналізу важливості змінних для класифікації, розрахуємо їх ранги. Змінна, яка є найважливішою для побудови моделі класифікації та за якою відбуватиметься подальше розгалуження спостережень, набуває найбільшого значення. Такими змінними знайдено індекс цифрової свободи та мережної готовності. Найнижчий ранг отримав індекс глобальної конкурентоспроможності, тому його було виключено під час побудови конфігурації дерева. Графічне зображення розподілу рангів змінних подано на рис. 6.

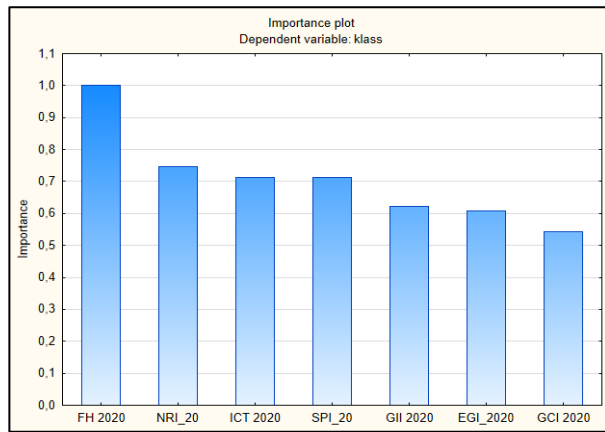


Рис. 6. Ранги класифікаційних атрибутів

На третьому етапі побудови дерева – вибору методу зупинки алгоритму розгалуження – необхідно зробити відсікання за помилкою класифікації або відсікання за варіацією, тоді необхідним стає проведення аналізу можливих конфігурацій дерев. Для цього проаналізуємо отриману таблицю послідовності дерев за помилками навчання та складності побудови конфігурації (рис. 7).

	Tree sequence (данніе страны без пробелов)		
	Dependent variable: class		
	Terminal node	Resubstitution cost	Node complexity
Tree 1	5	0,018519	0,000000
Tree 2	4	0,037037	0,018519
Tree 3	2	0,407407	0,185185
Tree 4	1	0,666667	0,259259

Рис. 7. Значення помилки навчання та складності побудови конфігурації

Як видно зі значення ціни навчання Resubstitution cost = 0,018 та складності конфігурації дерева Node complexity = 0,000, найкращою слід вважати підсумкову конфігурацію дерева під номером 1 (рис. 7). Отриманий граф побудованого дерева класифікації, за методом розгалуження CART наведено на рис. 8.

В інформаційній частині вікна бачимо граф, який має чотири вершини з розгалуженням і п'ять термінальних вершин, подальше розгалуження яких є неможливим.

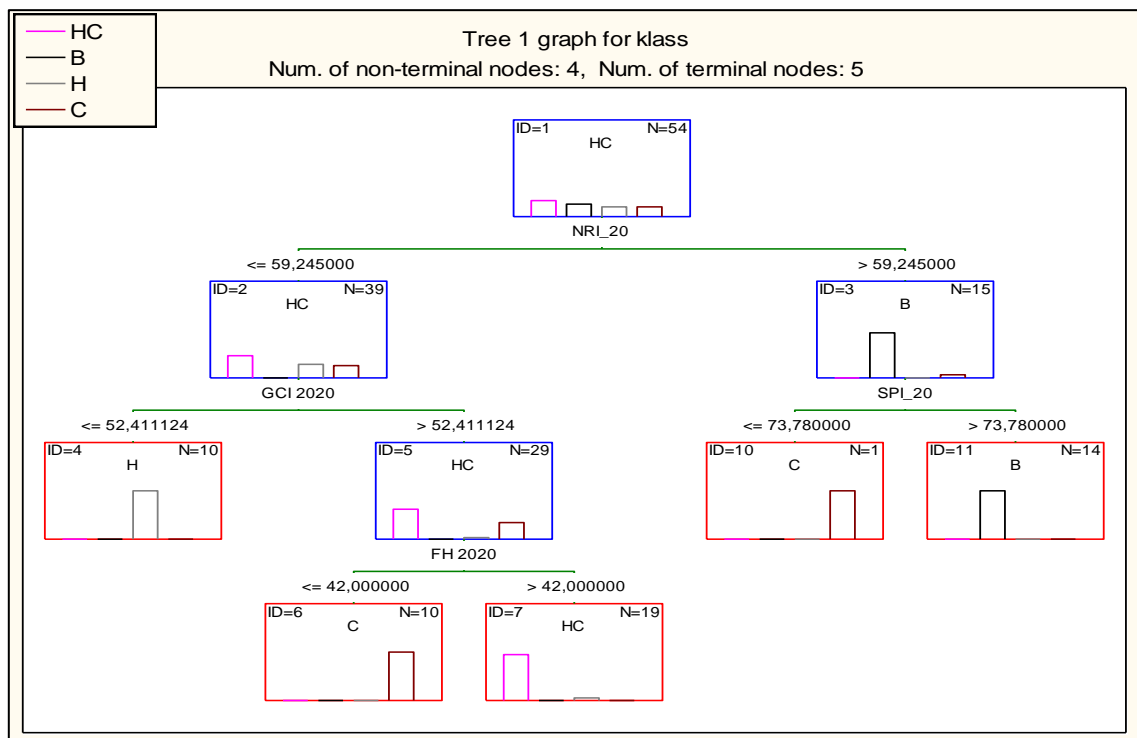


Рис. 8. Граф побудованого дерева

У кожній вершині наведено гістограму розподілу спостережень за класами, а також номер вершини та прогнозний клас, який формується за найбільшою кількістю класифікованих

спостережень у вузлі. Аналіз побудованого дерева класифікації можна виконати за допомогою таблиці структури дерева, (табл. 2).

Таблиця 2

Структурний розподіл дерева

Номер вузла	Відгук дерева								Спліт-константа	Значення спліт-константи
	Ліва гілка	Права гілка	Розмір вузла	Об'єкти у класі 1	Об'єкти у класі 2	Об'єкти у класі 3	Об'єкти у класі 4	Прогнозний клас		
1	2	3	54	18	14	11	11	НС	NRI_20	59,2
2	4	5	39	18	0	11	10	НС	GCI 2020	52,4
4			10	0	0	10	0	Н		
5	6	7	29	18	0	1	10	НС	FH 2020	42,0
6			10	0	0	0	10	С		
7			19	18	0	1	0	НС		
3	10	11	15	0	14	0	1	В	SPI_20	73,8
10			1	0	0	0	1	С		
11			14	0	14	0	0	В		

У таблиці наведено інформацію про номери вузлів, номери дочірніх вершин на лівій та правій гілках, кількість об'єктів з навчальної вибірки, приналежних кожному класу, прогнозні класи вершин, константа розгалуження та змінна розгалуження. За отриманою константою та змінною розгалуження формуємо логічне правило віднесення спостережень до певного кластеру країн або стану цифровізації. Вирішальні правила розпізнавання мають такий вигляд:

якщо $x_1(NRI) \leq 59,245 \cap x_6(GCI) \leq 52,41 \Rightarrow 4$ кластер – низький стан ЦРК;

якщо $\{x_1(NRI) \leq 59,245 \cap x_6(GCI) > 52,41 \cap x_4(FH) \leq 42\} \cup \{x_1(NRI) > 59,245 \cap x_7(SPI) > 73,78\} \Rightarrow 2$ кластер – середній стан ЦРК;

якщо $x_1(NRI) > 59,245 \cap x_6(GCI) > 52,41 \cap x_4(FH) > 42 \Rightarrow 3$ кластер – стан ЦРК нижче середнього;

якщо $x_1(NRI) > 59,245 \cap x_7(SPI) > 73,78 \Rightarrow 1$ кластер – високий стан ЦРК.

Визначені апостеріорні ймовірності розподілу розпізнаних об'єктів дозволили встановити, що розпізнавання відбулось на рівні 95 % значущості, тому таку класифікацію слід вважати статистично адекватною (рис. 9).

Для того, щоб отримати інформацію про кількість спостережень, які для кожного з прогнозних класів було помилково віднесено до іншого класу, обчислимо відповідну матрицю розподілень спостережень за прогнозними та

реальними класами.

Dependent variable: class							Options: Categorical response, Tree number 1, Analysis sample						
Observed value	Predicted value	Probability for HC	Probability for B	Probability for H	Probability for C	Terminal node							
Argentina	HC	HC	0,947368	0,000000	0,052632	0,000000	7						
Armenia	HC	HC	0,947368	0,000000	0,052632	0,000000	7						
Australia	B	B	0,000000	1,000000	0,000000	0,000000	11						
Bangladesh	H	H	0,000000	0,000000	1,000000	0,000000	4						
Brazil	HC	HC	0,947368	0,000000	0,052632	0,000000	7						
Cambodia	H	H	0,000000	0,000000	1,000000	0,000000	4						
Canada	B	B	0,000000	1,000000	0,000000	0,000000	11						
China	C	C	0,000000	0,000000	0,000000	1,000000	6						
Colombia	HC	HC	0,947368	0,000000	0,052632	0,000000	7						
Ecuador	HC	HC	0,947368	0,000000	0,052632	0,000000	7						
Egypt	C	C	0,000000	0,000000	0,000000	1,000000	6						
Estonia	B	B	0,000000	1,000000	0,000000	0,000000	11						
Ethiopia	H	H	0,000000	0,000000	1,000000	0,000000	4						
France	B	B	0,000000	1,000000	0,000000	0,000000	11						
Georgia	HC	HC	0,947368	0,000000	0,052632	0,000000	7						
Germany	B	B	0,000000	1,000000	0,000000	0,000000	11						
Hungary	B	B	0,000000	1,000000	0,000000	0,000000	11						
Iceland	B	B	0,000000	1,000000	0,000000	0,000000	11						
India	HC	HC	0,947368	0,000000	0,052632	0,000000	7						

Рис. 9. Апостеріорні ймовірності розподілу розпізнаних об'єктів

З рисунку 10 видно, що значення міжкласифікаційних помилок розпізнавання дорівнює 2 (1 об'єкт у кластері з низьким станом ЦРК і 1 об'єкт у кластері з середнім станом ЦРК).

Це ті об'єкти, які було помилково віднесено до інших кластерів. Такий розподіл свідчить про високу якість розпізнавання за моделлю.

Dependent variable: class						Options: Categorical response, Tree number 1					
Node #	Class HC	Class B	Class H	Class C	Gain						
4	0	0	10	0	10,00000						
6	0	0	0	10	10,00000						
7	18	0	1	0	19,00000						
10	0	0	0	1	1,00000						
11	0	14	0	0	14,00000						

Рис. 10. Розподіл міжкласифікаційних помилок

Додатковий аналіз якості дерева проведемо за критерієм приросту інформації, розрахованого для кожного кластеру об'єктів. Він є подібним до ROC-кривої, загальноприйнятої для оцінки підгонки моделі дерева. Приклад графіку кумулятивних значень для четвертого кластеру наведено на рис. 11.

Як видно з рисунку, величина показника приросту інформації – накопичений відсоток розпізнавання елементів кластеру з найнижчим станом цифрового розвитку – є швидкозростаючим на рівні 18 перцентилів за моделлю дерева рішень, що свідчить про якісне виокремлення об'єктів кластеру. Аналогічного висновку можна дійти за іншими кластерами. Це підтверджує загальну адекватність моделі розпізнавання стану ЦРК.

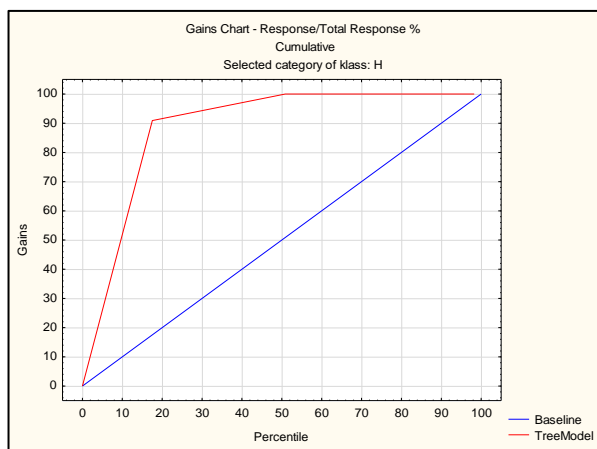


Рис. 11. Графік показника приросту інформації за кластером 4

Висновки

Проведений у роботі аналіз багатовимірної структури індикаторів цифровізації країни на підставі розроблених моделей дозволив виокремити однорідні групи країн за станом цифрового розвитку та виконати позиювання України відносно країн світу. Встановлено, що Україна відноситься до країн із середнім станом цифровізації та одночасно нижче середнього темпами соціально-економічного розвитку. Детальний аналіз зв'язків множин індексів цифровізації та соціально-економічного розвитку країн дозволяє дійти висновку про високий рівень їх взаємного впливу та важливість формування за цими напрямками системи заходів із забезпечення належного рівня цифрового розвитку країни.

Література

1. *Digital Riser Report: які країни найкраще розвивають цифрові технології у 2021 році.* [Електронний ресурс]. – Режим доступу: <https://tyzhden.ua/digital-riser-report-iaki-krainy-najkrashche-rozvyvaiut-tyfrovi-tekhnologii-u-2021-rotsi/>.
2. *Україна увійшла у топ-10 країн світу в рейтингу*

технологічних навичок [Електронний ресурс]. – Режим доступу: <https://ck-oda.gov.ua/novyny-cherkaskoyi-oblasti/ukrayina-uvijshla-u-top-10-krain-svitu-v-rejtingu-technologichnix-navichok/>.

3. *Проблеми та перспективи гармонізації цифрового ринку України з ринками ЄС та країн СхП. Інформаційно-аналітичний звіт [Електронний ресурс] / В. Ночвай, М. Корявець, І. Кульчицький, Л. Олексюк, О. Приходько, К. Гороховський, В. Грига. – Режим доступу: <https://www.civic-synergy.org.ua/analytics/problemy-ta-perspektyvy-garmonizatsiyi-tyfrovrogo-rynku-ukrayiny-z-rynkaty-yes-ta-krayin-shp>.*
4. *Семенов А. Ю. Аналіз світових рейтингів оцінки формування та розвитку цифрової економіки та місце України в них [Електронний ресурс] / А. Ю. Семенов // Науковий вісник Міжнародного гуманітарного університету С. 38–43. – Режим доступу: <http://www.vestnik-econom.mgu.od.ua/journal/2020/43-2020/8.pdf>.*
5. *Україна 2030E – країна з розвинутою цифровою економікою [Електронний ресурс]. – Режим доступу: <https://strategy.uifuture.org/kraina-z-rozvinutoyu-cifrovoyu-ekonomikoyu.html>.*
6. *Цифрова адженда України – 2020 («Цифровий порядок денний» – 2020). Концептуальні засади [Електронний ресурс]. – Режим доступу: <https://ucci.org.ua/uploads/files/58e78ee3c3922.pdf>.*
7. *Network Readiness Index 2020. [Електронний ресурс]. – Режим доступу: <https://networkreadinessindex.org/nri-2020-countries/#navigate>.*
8. *The ICT Development Index (IDI): conceptual framework and methodology. [Електронний ресурс]. – Режим доступу: <https://www.itu.int/en/ITU-D/Statistics/Pages/publications/mis/methodology.aspx>.*
9. *2020 Social Progress Index. Global Index: Methodology. [Електронний ресурс]. – Режим доступу: <https://www.socialprogress.org/index/global/methodology>.*
10. *Чаговець Л. О. Концептуальний базис моделювання телекомунікаційного розвитку регіонів методами системного аналізу [Текст] / Л. О. Чаговець, С. В. Прокопович, С. М. Вознюк, В. В. Чаговець // Комунальне господарство міст. – 2021, т. 1. – вип. 161. – С. 230-240.*
11. *Chagovets L. Machine Learning Methods Applications for Estimating Unevenness Level of Regional Development [Текст] / L. Chagovets, V. Chahovets, N. Chernova // Data-Centric Business and Applications. Evolutions in Business Information Processing and Management (Volume 3): Springer; Cham, 2020. – Pp. 115–139.*
12. *10 краєвих програм та інструментів для статистики у 2022 році. [Електронний ресурс]. – Режим доступу: <https://ua.softlist.com.ua/articles/10-luchshikh-program-i-instrumentov-dlia-statisiki-v-2022-godu/>.*

References

1. *Digital Riser Report: yaki krainy najkrashche rozvyvaiut tyfrovi tekhnologii u 2021 rotsi.* Retrieved from <https://tyzhden.ua/digital-riser-report-iaki-krainy-najkrashche-rozvyvaiut-tyfrovi-tekhnologii-u-2021-rotsi/>.
2. *Ukraina uviishla u top-10 krain svitu v reitynhu tekhnolohichnykh navychok.* Retrieved from <https://ck-oda.gov.ua/novyny-cherkaskoyi-oblasti/ukrayina-uvijshla-u-top-10-krayin-svitu-v-rejtingu-technologichnix-navichok/>.
3. *Nochvai, V., Koriavets, M., Kulchytskyi, I., Oleksiuk, L., Prykhodko, O., Horokhovskiy, K. & Hryha, V. Harmonization problems and prospects of the Ukraine digital market according to the EU markets and Eastern countries.* Retrieved from <https://www.civic-synergy.org.ua/analytics/>

problemy-ta-perspektyvy-garmonizatsiyi-tsyfrovogo-rynku-ukrayiny-z-rynkamy-yes-ta-krayin-shp/ [in Ukrainian].

4. Semenoh, A. Yu.(2020). Analysis of the world rankings assess the formation and development of the digital economy and place them in Ukraine. *Scientific Bulletin of the International Humanities University*, 38–43. Retrieved from <http://www.vestnik-econom.mgu.od.ua/journal/2020/43-2020/8.pdf> [in Ukrainian].

5. Ukraine 2030E – the country with a developed digital economy. Retrieved from <https://strategy.uifuture.org/kraina-z-rozvinutoyu-cifrovoyu-ekonomikoyu.html> [in Ukrainian].

6. Digital Agenda of Ukraine – 2020 Conceptual principles. Retrieved from <https://ucci.org.ua/uploads/files/58e78ee3c3922.pdf> [in Ukrainian].

7. Network Readiness Index 2020. Retrieved from <https://networkreadinessindex.org/nri-2020-countries/#navigate>.

8. The ICT Development Index (IDI): conceptual framework and methodology. Retrieved from <https://www.itu.int/en/ITU-D/Statistics/Pages/publications/mis/methodology.aspx>.

9. 2020 Social Progress Index. Global Index: Methodology. Retrieved from <https://www.socialprogress.org/index/global/methodology>.

10. Chagovets, L., Prokopovych, S., Voznyk, S., & Chahovets, V. (2021). Modelling conceptual basis of regional telecommunications development by system analysis methods. *Municipal Economy of Cities*, 1(161), 230–240.

11. Chagovets, L., Chahovets, V. & Chernova, N. (2020). Machine Learning Methods Applications for Estimating

Unevenness Level of Regional Development. In *Data-Centric Business and Applications. Evolvments in Business Information Processing and Management*, Vol. 3 (pp. 115–139). Cham: Springer.

12. 10 krashchykh prohram ta instrumentiv dlia statystyky u 2022 rotsi. Retrieved from <https://ua.softlist.com.ua/articles/10-luchshikh-programm-i-instrumentov-dlia-statisiki-v-2022-godu/>.

Рецензент: д-р. екон. наук, проф. Т.В. Шабельник, Харківський національний економічний університет ім. С. Кузнеця, Україна.

Автор: ЧАГОВЕЦЬ Любова Олексіївна
кандидат економічних наук, доцент, доцент
кафедри економічної кібернетики і системного
аналізу
Харківський національний економічний університет
імені С. Кузнеця
E-mail – liubov.chahovets@hneu.net
ID ORCID: <http://orcid.org/0000-0003-4064-9712>

Автор: ЧАГОВЕЦЬ Віта Віталіївна
кандидат економічних наук, доцент, доцент
кафедри кібернетики та інформаційних технологій
Державний біотехнологічний університет
E-mail – chagovec.v@ukr.net
ID ORCID: <http://orcid.org/0000-0003-0066-2760>

IDENTIFICATION AND FORECASTING MODELS OF COUNTRIES DIGITALIZATION STATUS IN WORLD SPACE

L. Chagovets¹, V. Chahovets²

¹Simon Kuznets Kharkov National University of Economics, Ukraine

²State Biotechnological University, Ukraine

The article is devoted to development issues of identification and forecasting models of countries digitalization status in world space. The problems and prospects of countries digitization are becoming relevant at the current stage of the technological sphere development of mankind was proved. Digitization is becoming the most important factor in the economic growth of any country and is generally a modern trend in its development. Untimely elimination of digitalization problems can cause negative consequences in the scale of the entire country's economy. Therefore, the prospects of the digital development of countries and the corresponding models of identification and forecasting of their digitalization status require special research. It will reveal the factors of digital growth. In addition, the work shows that the justification based on the system approach and assessment of the sustainable digital development of countries using modern methods of system analysis and multidimensional analysis Data Science, as one of the most effective in performing mathematical modeling of the researched processes status, has not received sufficient development in scientific works and needs further development. Therefore, the purpose of the article is to improve the complex of mathematical models for estimation and analyzing countries digitalization status in world, which, unlike others, will be built on the basis of intellectual analysis methods of multidimensional objects. Based on the set purpose, the distribution of the world countries by clusters on the basis of spatial and dynamic data sampling is carried out and analyzed in the article. Using the methods of multidimensional analysis and Data Science, the positioning of Ukraine in relation to the countries of the world was found. It has been established that Ukraine belongs to the countries with an average status of digitization and at the same time below average rates of socio-economic development. In order to predict the status of countries digitization, experiments were conducted on the construction of various types of classification tree models and the best model was obtained. A detailed analysis of the links between the set of digitalization and countries socio-economic development indices allowed us to come to a conclusion about the high level of their mutual influence and the importance of forming a system of measures in these directions to ensure the proper level of country digital development.

Keywords: model, classification trees, forecasting, digitalization, indicators of digitalization, telecommunication development.