УДК 519.862.6:(338.1+004.08) JEL Classification C44, O40

АНАЛІЗ ТА МОДЕЛЮВАННЯ СОЦІАЛЬНО-ЕКОНОМІЧНОГО РОЗВИТКУ КРАЇН З УРАХУВАННЯМ РІВНЯ ЇХ КІБЕРБЕЗПЕКИ[1]

Яровенко Г.М.,
д.е.н., доцентка, доцентка кафедри економічної кібернетики
Сумського державного університету
h.yarovenko@biem.sumdu.edu.ua
Кочережченко Р.Д.,
студент кафедри економічної кібернетики
Сумського державного університету
r.kocherezhchenko@student.sumdu.edu.ua

Статтю присвячено актуальній темі аналізу й моделювання соціально-економічного розвитку країн з урахуванням рівня їх кібербезпеки. Дана проблематика обумовлена зростанням рівня кіберзлочинів, які набувають глобальних масштабів та їх наслідки призводять до дестабілізації економічних, соціальних та політичних процесів у суспільстві. Дослідження проводилося на основі статистичних даних 141 країни світу за 2019 рік за допомогою мови програмування Python. Національний індекс кібербезпеки було обрано як індикатор, що характеризує рівень країн протидіяти різного роду кіберзагрозам. У якості показників соціально-економічного розвитку було обрано 11 макроекономічних індексів, які характеризують ВВП на душу населення, рівень інфляції, легкість ведення бізнесу, рівень безробіття, тощо. Методика дослідження проводилася у шість етапів. За результатами першого кроку було виявлено, що масив даних не містить пропущених значень, але за рядом показників, таких як рівень інфляції, рівень безробіття, витрати уряду на освіту, дохід за винятком грантів, експорт високих технологій, витрати на кінцеве споживання державного бюджету, ВВП, спостерігаються викиди. Головною причиною цього факту є існування значного розриву між рівнями соціально-економічного розвитку найменш розвинених та розвинених країн. Проведений кореляційний аналіз виявив існування помітної та високої кореляції між факторами: національний індекс кібербезпеки, загальна очікувана тривалість життя при народженні, легкість ведення бізнесу, ВВП на душу населення, наймані працівники та вразлива зайнятість. Їх було обрано для подальших розрахунків, оскільки інші соціально-економічні показники не мають кореляції із національним індексом кібербезпеки. На третьому етапі було використано метод головних компонент для усунення мультиколінеарності, що дозволило сформувати три статистично значущі компоненти. На четвертому етапі було проведено кластеризацію країн за методом k-means, в результаті чого було отримано 5 кластерів країн в залежності від рівня їх кібербезпеки та соціально-економічного розвитку. В результаті сешменти було сформовано країнами, які мають близькі значення, як національного індексу кібербезпеки, так й показників соціально-економічного розвитку. Проведення передискретизації даних на п'ятому етапі дозволило збалансувати спостереження в залежності від обраних класифікаційних груп-кластерів. На шостому етапі було побудовано класифікаційну модель дерева рішень, яка має високі показники загальної точності та для кожної класифікаційної групи, а також яку можна застосовувати для прогнозування ймовірних сегментів соціально-економічного розвитку країн з урахуванням рівня їх кібербезпеки.

Ключові слова: *кібербезпека, економічний розвиток, соціальний розвиток, дерево рішень, кластерний аналіз, метод головних компонент.*

DOI: 10.21272/ 1817-9215.2022.1-5

ПОСТАНОВКА ПРОБЛЕМИ

Швидкі темпи цифровізації багатьох процесів у суспільстві призвели до того, що з'явилися нові загрози, пов'язані із кіберзлочинністю, що проявляється у вигляді здійснення кібервійн, кібертероризму, кібершпигунства, масштабних кіберзлочинів, хактивізму, атак на електронний уряд, націлених на порушення функціонування інфраструктури держави та суб'єктів економіки, заволодіння їх фінансовими ресурсами, технологіями, даними. Це призводить до політичної, економічної, соціальної дестабілізації у суспільстві, військових конфліктів, зниження довіри до уряду, іміджу країни на міжнародній арені. Наприклад, у 2016 році відбулася масштабна хакерська атака шляхом відправлення електронних листів із фішингом до Федерального резервного банку Нью-Йорка з метою викрадення 15,25 млн. дол.

[1] *Робота виконана в рамках держбюджетної науково-дослідної роботи «Національна безпека через конвергенцію систем фінансового моніторингу та кібербезпеки: інтелектуальне моделювання механізмів регулювання фінансового ринку» №0121U109559*

золотовалютних резервів [1]. У 2019 році було здійснено викрадення записів платіжних карток з муніципальної платіжної системи Click2Gov США [2]. У 2018 році більше ніж 100 компаній, серед яких були такі компанії-гіганти, як Ford, Tesla, Toyota, General Motors, Fiat Chrysler, Volkswagen, ThyssenKrupp та інші, отримали величезні збитки за рахунок масштабного витоку конфіденційних даних [3]. Ці й інші приклади свідчать, що кіберзлочинність є масштабним явищем не залежно від рівня суб'єктів економіки, видів діяльності, територіального розташування тощо. Всесвітній економічний форум у 2021 році визначив кіберзлочинність як п'ятий за величиною ризик у світі після економічних протистоянь, внутрішньополітичної поляризації, екстремальної спеки, руйнування природних екосистем [4]. Саме тому для більшості країн світу одним із пріоритетних завдань є підвищення рівня їх кіберзахисту не тільки на рівні окремого суб'єкта господарювання, але й на національному рівні.

Оскільки рівень кіберзахисту країни значно впливає на її соціальний та економічний розвиток, який також з іншого боку обумовлює формування відповідного рівня кібербезпеки, то цей факт є важливим при розробці стратегії розвитку країни. Тому є потреба у визначенні відповідних факторів соціально-економічного розвитку, які мають найбільш тісний зв'язок з процесами цифровізації та створення ефективної системи кіберзахисту. Відповідно, це дозволить розробити більш якісні та точні моделі, які дозволять прогнозувати можливі варіанти соціально-економічного розвитку країн з урахуванням рівня їх кібербезпеки. Тому є потреба у вивченні даної проблематики, розробці відповідного науково-методичного підходу та проведені розрахунків.

АНАЛІЗ ОСТАННІХ ДОСЛІДЖЕНЬ І ПУБЛІКАЦІЙ

Дослідженнями в сфері кібербезпеки в останнє десятиліття займається досить багато науковців, але її поєднанню з проблемами соціально-економічного розвитку країн присвячено не досить багато публікацій. Це пов'язано із тим, що це питання виносять у розряд суто технічних та програмних досліджень. Але останнім часом науковці вивчають також її вплив на різні сфери суспільно-економічного життя. Так, Цао Ю., Пін Ю., Тао С., Чен Ю., Чжу Ю. розглядають аспекти та специфіку політики контролю доступу для кібер-фізично-соціальних просторів [5]. Прейс Б. та Сасскінд Л. досліджують практику кібербезпеки на муніципальному рівні та аналізують її фінансово-економічні ризики [6]. Такі вчені, як Кошутич Д. та Піні Ф., запропонували концептуальну модель конкурентних переваг кібербезпеки, яка дозволяє оцінити можливості потенційного нарощення обсягів кібербезпеки для досягнення довгострокової конкурентної переваги суб'єктів економіки [7]. Ван Кемп К. та Пітерс В. розглядають потенційні контрзаходи для виявлення та попередження кіберзагроз, такі як програми швидкого реагування та політичні заходи [8]. Бабу К.-Е.-К. та Бакар Сіддік М.А. досліджують найбільш поширені способи кібершахрайств у соціальних мережах, аналізують наслідки та можливі заходи протидії [9]. Автори Сальві А., Спаньолетті П., Нурі Н.С. запропонували модель кіберстійкості критичних кіберінфраструктур на основі впровадження цифрового близнюка в електроенергетиці, яка дозволить мінімізувати час реагування та зменшити вплив кібератак на організації та суспільство в цілому [10]. Мазух К., Греве М., Транг С., Кольбе Л.М. вивчають ступінь впливу заходів реагування щодо порушення даних компаній на вартість їх акцій, що може впливати на коливання, які відбуваються на ринку цінних паперів [11]. Цименідіс С., Лагкас Т., Рантос К. досліджують Інтернет речей та ті виклики, які стоять перед кіберзахистом даної сфери діяльності, а також пропонують використання моделей глибокого навчання для виявлення вторгнень [12]. Кузьменко О.В., Кубалек Я., Боженко В.В., Кушнерьов О.С., Віда І. здійснили оцінку факторів кіберзлочинності для фінансового сектору на основі моделі навчання з наглядом із асоційованим навчанням [13]. Кіанпур М., Ковальський С.Д., Овербі Х. дослідили економічні моделі та виявили, що більшість з них є неперевіреними, спрощеними та нереалістичними, що ускладнює можливості їх подальшого

використання для вирішення потреб практики [14]. Не дивлячись на широкий спектр проблем, які аналізуються вченими, але питання побудови моделей прогнозування соціально-економічного розвитку країн з урахуванням рівня їх кібербезпеки є недостатньо вивченим і потребує подальших досліджень.

ПОСТАНОВКА ЗАВДАННЯ

Мета статті полягає у проведенні аналізу індикаторів соціально-економічного розвитку країн з урахуванням рівня їх кібербезпеки на основі розрахунку базових статистик та проведення кластерного аналізу, а також побудові прогнозної моделі у вигляді дерева рішень.

ВИКЛАД ОСНОВНОГО МАТЕРІАЛУ ДОСЛІДЖЕННЯ

Дослідження передбачає вибір набору індикаторів, які будуть характеризувати рівень кібербезпеки країн та рівень їх соціально-економічного розвитку. У якості першої ознаки було обрано національний індекс кібербезпеки (National Cyber Security Index), який розраховується компанією «e-Governance Academy Foundation» та характеризує можливості країни протидіяти різного роду кіберзагрозам, передбачувати потенційні кіберінциденти шляхом формування відповідних умов в країні – фінансових, інфраструктурних, програмних, підготовки відповідних спеціалістів та інших. Його високе значення свідчить про забезпечення потужних заходів кібербезпеки на різних рівнях державного управління. Найнижче значення показує слабкі темпи розвитку країни у напрямку організації системи кібербезпеки.

Для визначення рівня соціально-економічного розвитку обрано ряд показників, які було відібрано на основі проведеного аналізу наукових публікацій. Так, сюди увійшли: ВВП на душу населення (в поточних доларах США) (GDP per capita); загальна очікувана тривалість життя при народженні (в роках) (Life expectancy at birth, total); легкість ведення бізнесу (Ease of doing business score); наймані працівники (Wage and salaried workers, total); вразлива зайнятість (Vulnerable employment, total); рівень безробіття (Unemployment, total); рівень інфляції, у споживчих цінах (Inflation, consumer prices); витрати уряду на освіту (Government expenditure on education, total); дохід, за винятком грантів (Revenue, excluding grants); експорт високих технологій (High-technology exports); витрати на кінцеве споживання державного бюджету (General government final consumption expenditure).

Базу даних дослідження було сформовано на основі фактичних даних обраних показників для 141 країни світу за 2019 рік, які було узято з офіційного сайту Світового банку (World Bank) та Академії електронного врядування (e-Governance Academy Foundation). Всі розрахунки проводилися із використанням мови програмування Python.

На першому етапі дослідження було проведено аналіз базових статистик вхідних даних з метою попередньої ідентифікації вибірки. Його результати представлені на рисунку 1.

	National Cyber Security Index	GDP per capita	Inflation, consumer prices	Unemployment, total	Vulnerable employment, total	Wage and salaried workers, total	Government expenditure on education, total	Revenue, excluding grants	High-technology exports	General government final consumption expenditure	Life expectancy at birth, total
count	141.000000	141.000000	141.000000	141.000000	141.000000	141.000000	141.000000	1.410000e+02	1.410000e+02	1.410000e+02	141.000000
mean	44.957447	15524.531072	4.953204	7.588511	34.927660	61.849716	13.357412	1.799748e+13	1.740076e+10	1.071350e+11	73.662259
std	25.765212	20675.890125	14.015177	4.895578	26.419543	25.859923	5.652098	1.658264e+14	7.005420e+10	3.661371e+11	6.955050
min	1.300000	238.990726	-2.540315	0.310000	0.140000	7.400000	0.000000	0.000000e+00	0.000000e+00	0.000000e+00	54.239000
25%	22.080000	2250.601164	0.506690	4.370000	10.840000	43.830002	10.502530	2.336316e+01	1.069741e+07	1.667513e+09	69.870000
50%	45.450000	5917.262575	2.524621	6.240000	28.590001	68.050003	13.092890	7.983206e+10	2.120509e+08	6.819717e+09	74.914634
75%	63.640000	20233.641348	4.201793	9.390000	52.289999	84.950000	16.214991	1.023396e+12	4.172700e+09	4.378328e+10	78.497561
max	96.100000	116014.602497	150.322724	28.740000	92.250000	99.589996	34.327251	1.955102e+15	7.576827e+11	3.077990e+12	84.356341

Рисунок 1 – Базові статистики вхідних даних (розраховано авторами)

На рисунку 1 представлені мінімальне, максимальне, середнє значення, стандартне відхилення та квартилі для вхідних даних. Загальна кількість спостережень по кожному показнику дорівнює 141, що свідчить про відсутність пропусків для даних. Аналіз мінімальних значень показав, що по ряду факторів існують спостереження, які дорівнюють нулю (витрати уряду на освіту, дохід, за винятком грантів, експорт високих технологій, витрати на кінцеве споживання державного бюджету). Це обумовлено тим, що для ряду країн, які є найменш розвиненими, їх значення фактично є нульовим. Результати отриманих значень квартилів свідчать, що по ряду показників є викиди, тобто або значення для відповідної країни є вкрай низьким, або вкрай високим у порівнянні із середнім по вибірці. Це обумовлено тим, що дані є просторовими та країни мають значні розриви у соціально-економічному розвитку. Наприклад, валовий внутрішній продукт на душу населення для Бурунді дорівнює 238,99, а для Люксембургу – 116014,60, при чому межа 75% квартилю дорівнює 20233,64, а середнє значення – 15524,53. Тобто має місце значна нерівність даного показника для аналізованих країн, що викликана значною різницею у ступені соціально-економічного розвитку цих країн. Аналогічна картина спостерігається й для наступних факторів: рівень інфляції, у споживчих цінах, рівень безробіття, витрати уряду на освіту, дохід, за винятком грантів, експорт високих технологій, витрати на кінцеве споживання державного бюджету. Загальна очікувана тривалість життя при народженні та легкість ведення бізнесу для більшості спостережень мають значення, що відповідають встановленим межам квартилів, тільки для декількох країн спостерігається аномальне значення.

На другому етапі проведемо кореляційний аналіз шляхом побудови кореляційної матриці за допомогою інструменту «Heatmap» (рисунок 2). З отриманих даних аналізу можна побачити, що між такими показниками, як національний індекс кібербезпеки, загальна очікувана тривалість життя при народженні, легкість ведення бізнесу, ВВП на душу населення, наймані працівники та вразлива зайнятість існує помітний кореляційний зв'язок від 0,5 до 0,7, а для деяких показників високий – від 0,7 – 0,9. Між іншими показниками кореляція або є слабкою, або помірною. Тому для подальшого дослідження приймає рішення залишити тільки ті фактори, які є корельованими, особливо з національним індексом кібербезпеки.

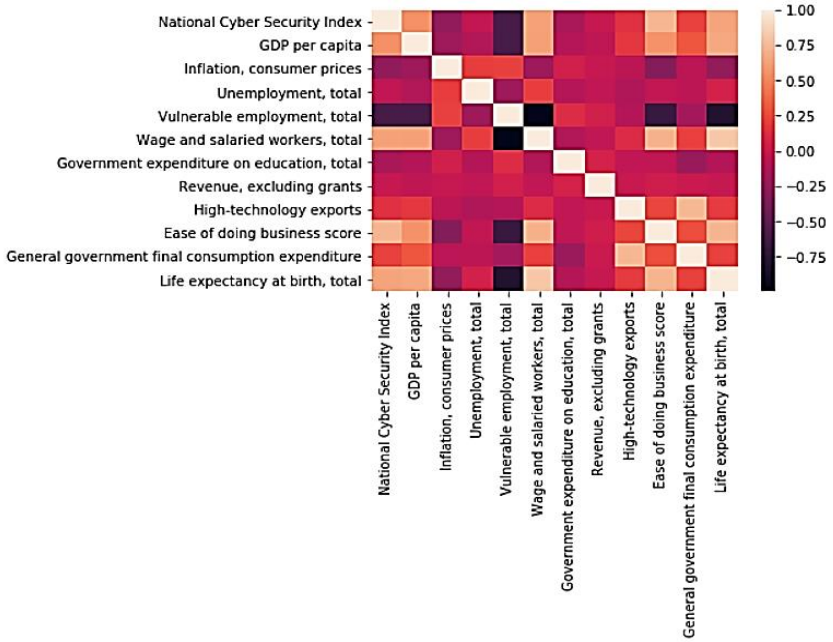
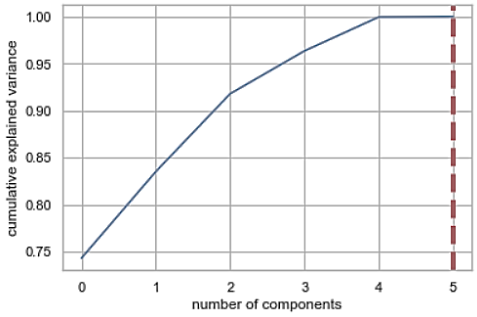


Рисунок 2 – Кореляційна матриця (розраховано авторами)

На третьому етапі застосуємо метод головних компонент, який дозволить позбавитися мультиколінеарності між факторами. Ця умова є необхідною умовою для проведення кластерного аналізу. Також дана процедура дозволить знизити розмірність вхідних даних. Результати застосування методу головних компонент представлені на рисунку 3, де можна побачити, що статистично значущими є три компоненти, для кожної з яких рівень пояснюючої варіації перевищує 0,05, що забезпечує 0,9178 рівня накопиченої варіації. Тому отримані результати для трьох перших компонент буде використано для проведення кластерного аналізу, яке було виконано із використанням методу k-means.



	Cumulative Variance Ratio	Explained Variance Ratio
0	0.742990	0.742990
1	0.835041	0.092052
2	0.917824	0.082783
3	0.963390	0.045565
4	0.999476	0.036086
5	1.000000	0.000524

Рисунок 3 – Результати застосування методу головних компонент (розраховано авторами самостійно)

На четвертому етапі було визначено оптимальну кількість кластерів, для чого було розраховано їх значення шляхом застосування «Ліктьового методу». В результаті встановлено, що можлива кластеризація на 5 або 6 секторів. Але подальші розрахунки показали, що при застосуванні шестикластерної моделі неможливо отримати адекватне дерево рішення, тому краще використати п’ять кластерів. Отримані результати було візуалізовано у вигляді карти країн із зазначенням відповідного кластеру (рисунок 4).

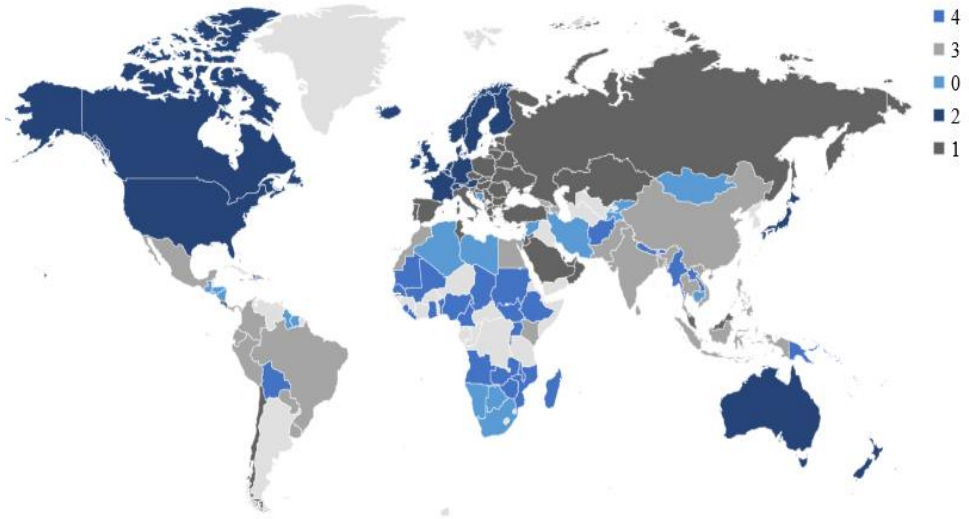


Рисунок 4 – Результати кластерного аналізу (розраховано авторами самостійно)

На рисунку 4 можна побачити, що нульовий кластер було сформовано 23-ма країнами, до яких відносяться: Алжир, Барбадос, Камбоджа, Гватемала, Іран, Монголія, Нікарагуа, Суринам, Таджикистан, Тонга та інші. Країни даного кластеру

відносяться до тих, що розвиваються, при чому переважна їх більшість за рівнем ВВП відноситься до другого квартилю, а за рівнем кібербезпеки їх значення значно нижче середнього за вибіркою, що свідчить про те, що темпи розвитку системи кіберзахисту в даних країнах значно відстають від темпів розвитку їх економіки. До першого кластеру увійшли 36 країн, переважна більшість яких є країнами з економікою, що розвивається, а саме: Чилі, Хорватія, Малайзія, Польща, Російська Федерація, Україна, Чорногорія, Оман, Білорусь, Туреччина, та інші. Середнє значення ВВП для країн першого кластеру знаходиться на рівні середнього по вибірки. Щодо рівня кібербезпеки, то дані країни підпадають до 3 та 4 квартилів. Тобто даний сектор сформували країни, які мають рівень економічного розвитку на середньому рівні, але рівень кібербезпеки є вище середнього. Наприклад, Мальта, для якої національний індекс кібербезпеки дорівнює 50,65, але рівень ВВП – 27884,64, та Україна, яка має рівень кібербезпеки рівний 75,32 та ВВП – 3726,93. До другого кластеру увійшли 22 країни найбільш економічно розвинені та із високим рівнем кібербезпеки: Австралія, Канада, Німеччина, Франція, США, Швейцарія, Велика Британія, Нідерланди, Японія, Норвегія та інші. До третього кластеру увійшли 28 країн, куди увійшли також й нові індустріальні країни – Китай, Індія, Бангладеш, Пакистан, Індонезія, Мексика, Філіппіни, Таїланд, В'єтнам. За рівнем економічного розвитку та кібербезпеки ці країни належать до 2-го та 3-го квартилів, що свідчить про близькість тенденцій їх розвитку, як економічного, так і в сфері забезпечення належного рівня кіберзахисту. До 4-го кластеру увійшли країни, які є найменш розвиненими, або економіка яких знаходиться на стадії розвитку, який відбувається досить повільними темпами. Для них також є характерним низький рівень кібербезпеки. Отримані результати кластерного аналізу використаємо в якості класифікаційної ознаки для побудови дерева рішення – майбутньої прогнозної моделі.

На п'ятому етапі проведемо передискретизацію даних. Це необхідно для покращення результатів класифікації, а також це пов'язано із нерівномірністю розподілених значень за кластерами (рисунок 5). Різниця між найменшим кластером та найбільшим сягає 10%, тому процедура передискретизації дозволить збалансувати вибірку. Її результат представлено на рисунку 5, де можна побачити, що кластери мають однаковий розмір.

Результати класифікації до передискретизації	*Результати класифікації після передискретизації*
Class=4, n=32 (22.695%)	Class=4, n=36 (20.000%)
Class=3, n=28 (19.858%)	Class=3, n=36 (20.000%)
Class=0, n=23 (16.312%)	Class=0, n=36 (20.000%)
Class=2, n=22 (15.603%)	Class=2, n=36 (20.000%)
Class=1, n=36 (25.532%)	Class=1, n=36 (20.000%)

Рисунок 5 – Результати класифікації до та після передискретизації (розраховано авторами самостійно)

На шостому етапі побудуємо дерево рішень, результати якості для якого представлені на рисунку 6. Загальна якість моделі по всім класам є досить високою і дорівнює 0,91, тобто побудована модель з ймовірністю 90,74% зробить вірний прогноз. Точність для позитивних класів коливається від 0,8 до 1,0, що також свідчить про високу ймовірність моделі робити багато коректних позитивних прогнозів та меншу кількість невірних позитивних класифікацій. Параметр чутливості для п'яти класів знаходиться від 0,78 до 1,00, що підтверджує високу спроможність моделі правильно ідентифікувати позитивні класи.

Побудована модель дерева рішень представлена на рисунку 7. Модель починає класифікацію із показника легкості ведення бізнесу. Якщо його значення менше або дорівнює 68,122, то перехід відбувається до лівого вузла, якому відповідає індикатор вразливої зайнятості, в протилежному випадку до правого – ВВП.

```
Confusion Matrix:
[[13  1  0  0  0]
 [ 1  7  0  1  0]
 [ 0  0 12  0  0]
 [ 0  0  0  9  2]
 [ 0  0  0  0  8]]
Classification Report:
              precision    recall  f1-score   support

           0       0.93      0.93      0.93        14
           1       0.88      0.78      0.82         9
           2       1.00      1.00      1.00        12
           3       0.90      0.82      0.86        11
           4       0.80      1.00      0.89         8

    accuracy                           0.91        54
   macro avg       0.90      0.90      0.90        54
weighted avg       0.91      0.91      0.91        54

Accuracy: 0.9074074074074074
```

Рисунок 6 – Результати якості побудови дерева рішень (розраховано авторами)

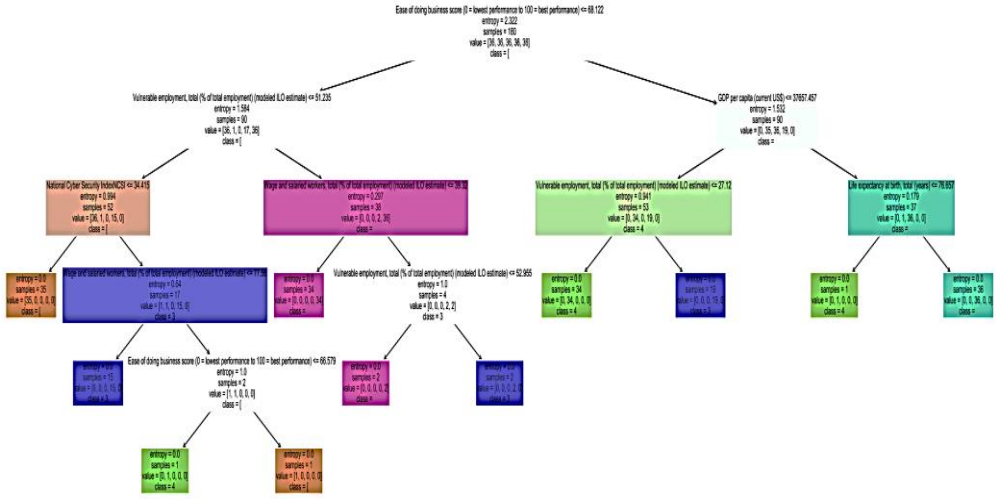


Рисунок 7 – Результати побудови дерева рішень (розраховано авторами самостійно)

Якщо його результат менше або дорівнює 37657,46, то спостереження класифікуються за рівнем вразливої зайнятості. Для значень, які є менше або дорівнюють 27,12, дані відносяться до четвертого кластеру, в іншому випадку – до третього. Якщо результат ВВП більше ніж 37657,46, то це можуть бути кластери 0, 1, 2, 3 або 5. Це визначатиметься за рівнем очікуваної тривалості життя. Для його значень менше ніж 76,657, спостереження класифікуються як четвертий сектор, в протилежному випадку – це ймовірно можуть бути 0, 1, 2, 3 або 5. Якщо результат індикатора легкості ведення бізнесу менше або дорівнює 68,122, то на наступному рівні по лівому вузлу класифікація відбувається за індикатором вразливої зайнятості. Для його значень, які є менше або дорівнюють 51,235, класифікація здійснюється за національним індексом кібербезпеки. Якщо його результат менше або дорівнює 34,415, спостереження ідентифікується за одним із наступних кластерів – 0, 1, 2, 4, 5, в протилежному випадку сектор є третім і класифікація відбувається за показником найманих працівників. Його значення нижче 77,56 відповідає третьому сегменту, в

іншому випадку класифікація здійснюється за індикатором легкості ведення бізнесу. Його результат менше 66,579 – це 4 кластер, в протилежному випадку – це всі інші сегменти. Якщо індикатор вразливої зайнятості більше 51,235, класифікуємо на третьому рівні за показником найманих працівників. Його результат менше або дорівнює 39,32, на четвертому рівні може класифікувати спостереження за кластерами 0, 1, 2, 3, 5, в іншому випадку ідентифікується 3-й сектор за індикатором вразливої зайнятості. Його значення менше або дорівнює 52,955 ймовірно класифікує спостереження за 0, 1, 2, 4, 5 кластерами, в протилежному випадку – за третім.

ВИСНОВКИ

Дане дослідження присвячене проблемі прогнозування соціально-економічного розвитку країн з урахуванням їх національного рівня кібербезпеки, що дозволяє враховувати їх можливості протистояти масовим кіберзагрозам, виникнення яких призводить до негативних економічних та суспільних наслідків. Для здійснення розрахунків було обрано національний індекс кібербезпеки та ряд факторів, які характеризують рівень соціально-економічного розвитку країн. Дані було обрано для 141 країни світу за 2019 рік. Дослідження проводилося із застосування мови програмування Python. В результаті здійснення первинного аналізу базових статистик було визначено відсутність пропущених значень та існування викидів для таких показників, як: рівень інфляції, рівень безробіття, витрати уряду на освіту, дохід за винятком грантів, експорт високих технологій, витрати на кінцеве споживання державного бюджету, ВВП. Це обумовлено існуванням значного розриву між рівнями соціально-економічного розвитку найменш розвинених та розвинених країн. За результатами кореляційного аналізу виявлено, що найбільш корельованими факторами є національний індекс кібербезпеки, загальна очікувана тривалість життя при народженні, легкість ведення бізнесу, ВВП на душу населення, наймані працівники та вразлива зайнятість, які було залишено для подальшого дослідження. Для усунення мультиколінеарності було застосовано метод головних компонент, за результатами якого було визначено три найбільш статистично значущі компоненти. Дана процедура дозволила здійснити кластеризацію країн в залежності від рівня їх кібербезпеки та соціально-економічного розвитку. За методом k-means було сформовано 5 кластерів. До відповідних секторів увійшли країни, близькі як за значеннями національного індексу кібербезпеки, так й факторами соціально-економічного розвитку. Проведено передискретизацію даних для збалансування класифікаційних груп, обраних за результатами кластерного аналізу. На останньому кроці побудовано класифікаційну модель дерева рішень, яку можна застосовувати в процесі прогнозування відповідних кластерів країн.

SUMMARY

Yarovenko H., Kocherezhchenko R. **Analysis and modeling of the countries socio-economic development with considering the level of their cyber security.**

The article is devoted to the topical issue of analysis and modelling of countries socio-economic development with considering the level of their cybersecurity. This issue is due to the growing level of cybercrime, which is gaining global scale, and its consequences lead to destabilization of economic, social and political processes in society. The study was conducted based on statistics from 141 countries in 2019 using Python as the programming language. The National Cyber Security Index was chosen as an indicator of the countries level to respond to various types of cyber threats. Eleven macroeconomic indices were selected as indicators of socio-economic development, which characterize GDP per capita, inflation rate, ease of doing business, unemployment rate, etc. The research methodology was conducted in six stages. The first step revealed that the data set did not contain missing values. Still, some indicators, such as inflation, unemployment, government spending on education, income excluding grants, exports of high technology, final consumption expenditures, GDP, are observed like anomalies. The main reason for this fact is the existence of a significant gap between the levels of socio-economic development of the least developed and developed countries. The correlation analysis revealed a substantial and high correlation between factors: the national cybersecurity index, overall life expectancy at birth, ease of doing business, GDP per capita, employees and vulnerable employment. They were selected for further calculations because other socio-economic indicators do not correlate with the national cybersecurity index. In the third stage, the authors used the principal components method to eliminate multicollinearity, forming three statistically significant components. In the fourth stage, countries were clustered using the k-means method, resulting in 5 sectors depending on the level

of countries cybersecurity and socio-economic development. As a result, the segments were formed by countries with similar values, both the national cybersecurity index and indicators of socio-economic development. The data were resampled in the fifth stage to balance the observations depending on the selected classification groups-clusters. In the sixth stage, a decision tree classification model was built, which has high indicators of overall accuracy for each classification group. The model can be used to predict probable segments of socio-economic development of countries based on their cybersecurity.

Keywords: cybersecurity, economic development, social development, decision tree, cluster analysis, principal components method.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Bangladesh may soon get $15.25 million of stolen reserve money. *The Economic Times* : website. URL: https://economictimes.indiatimes.com/news/international/world-news/bangladesh-may-soon-get-15-25-million-of-stolen-reserve-money/articleshow/ 55295915.cms (дата звернення: 03.01.2022).

2. Riley D. Payment card records stolen in latest attack targeting municipal payments system. *SiliconANGLE* : website. URL: https://siliconangle.com/ 2019/09/22/payment-card-records-stolen-latest-attack-targeting-municipal-payments-system/ (дата звернення: 03.01.2022).

3. Short Circuit: How a Robotics Vendor Exposed Confidential Data for Major Manufacturing Companies. *UpGuard* : website. URL: https://www.upguard.com/breaches/short-circuit-how-a-robotics-vendor-exposed-confidential-data-for-major-manufacturing-companies (дата звернення: 03.01.2022).

4. Wild Wide Web. Consequences of Digital Fragmentation. *World Economic Forum* : website. URL: https://reports.weforum.org/global-risks-report-2020/wild-wide-web/ (дата звернення: 03.01.2022).

5. Cao Y., Ping Y., Tao S., Chen Y., Zhu Y. Specification and adaptive verification of access control policy for cyber-physical-social spaces. *Computers and Security*. 2022, №114. Article number 102579. DOI: https://doi.org/10.1016/j.cose.2021.102579.

6. Preis B., Susskind L. Municipal Cybersecurity: More Work Needs to be Done. *Urban Affairs Review*. 2022, №58(2). P. 614–629. DOI: https://doi.org/10.1177/1078087420973760.

7. Kosutic D., Pigni F. Cybersecurity: investing for competitive outcomes. *Journal of Business Strategy*. 2022, № 43(1). P. 28–36. DOI: https://doi.org/10.1108/JBS-06-2020-0116.

8. Van Camp C., Peeters W. A World without Satellite Data as a Result of a Global Cyber-Attack. *Space Policy*. 2022. Article number 101458. DOI: https://doi.org/10.1016/j.spacepol.2021.101458.

9. Babu K.-E.-K., Bakar Siddik M.A. Cybercrime in the social media of Bangladesh: An analysis of existing legal frameworks. *International Journal of Electronic Security and Digital Forensics*. 2022, №14(1). P. 1–18. DOI: https://doi.org/10.1504/IJESDF.2022.119998.

10. Salvi A., Spagnoletti P., Noori N.S. Cyber-resilience of Critical Cyber Infrastructures: Integrating digital twins in the electric power ecosystem. *Computers and Security*. 2022, №112. Article number 102507. DOI: https://doi.org/10.1016/j.cose.2021.102507.

11. Masuch K., Greve M., Trang S., Kolbe L.M. Apologize or justify? Examining the impact of data breach response actions on stock value of affected companies? *Computers and Security*. 2022, №112. Article number 102502. DOI: https://doi.org/10.1016/j.cose.2021.102502.

12. Tsimenidis S., Lagkas T., Rantos K. Deep Learning in IoT Intrusion Detection. *Journal of Network and Systems Management*. 2022, №30(1). Article number 8. DOI: https://doi.org/10.1007/s10922-021-09621-9.

13. Kuzmenko O.V., Kubálek J., Bozhenko V.V., Kushneryov O.S., Vida I. An approach to managing innovation to protect financial sector against cybercrime | [Podejście do zarządzania innowacjami w celu ochrony sektora finansowego przed cyberprzestępczością]. *Polish Journal of Management Studies*. 2021, №24(2). P. 276–291. DOI: https://doi.org/10.17512/pjms.2021.24.2.17.

14. Kianpour M., Kowalski S.J., Øverby H. Systematically understanding cybersecurity economics: A survey. *Sustainability (Switzerland)*. 2021, №13(24). Article number 13677. DOI: https://doi.org/10.3390/su132413677.

REFERENCES

1. Bangladesh may soon get $15.25 million of stolen reserve money. *The Economic Times*. Available at: https://economictimes.indiatimes.com/news/international/world-news/bangladesh-may-soon-get-15-25-million-of-stolen-reserve-money/articleshow/ 55295915.cms (accessed 03 January 2022).

2. Riley D. Payment card records stolen in latest attack targeting municipal payments system. *SiliconANGLE*. Available at: https://siliconangle.com/ 2019/09/22/payment-card-records-stolen-latest-attack-targeting-municipal-payments-system/ (accessed 03 January 2022).

3. Short Circuit: How a Robotics Vendor Exposed Confidential Data for Major Manufacturing Companies. *UpGuard.* Available at: https://www.upguard.com/breaches/short-circuit-how-a-robotics-vendor-exposed-confidential-data-for-major-manufacturing-companies (accessed 03 January 2022).

4. Wild Wide Web. Consequences of Digital Fragmentation. *World Economic Forum.* Available at: https://reports.weforum.org/global-risks-report-2020/wild-wide-web/ (accessed 03 January 2022).

5. Cao Y., Ping Y., Tao S., Chen Y., Zhu Y. (2022). Specification and adaptive verification of access control policy for cyber-physical-social spaces. *Computers and Security*, vol. 114, no. 102579. DOI: 10.1016/j.cose.2021.102579.

6. Preis B., Susskind L. (2022). Municipal Cybersecurity: More Work Needs to be Done. *Urban Affairs Review*, vol. 58, no. 2, pp. 614–629. DOI: 10.1177/1078087420973760.

7. Kosutic D., Pigni F. (2022). Cybersecurity: investing for competitive outcomes. *Journal of Business Strategy*, vol. 43, no. 1, pp. 28–36. DOI: 10.1108/JBS-06-2020-0116.

8. Van Camp C., Peeters W. (2022). A World without Satellite Data as a Result of a Global Cyber-Attack. *Space Policy*, no. 101458. DOI: 10.1016/j.spacepol.2021.101458.

9. Babu K.-E.-K., Bakar Siddik M.A. (2022). Cybercrime in the social media of Bangladesh: An analysis of existing legal frameworks. *International Journal of Electronic Security and Digital Forensics*, vol. 14, no. 1, pp. 1–18. DOI: 10.1504/IJESDF.2022.119998.

10. Salvi A., Spagnoletti P., Noori N.S. (2022). Cyber-resilience of Critical Cyber Infrastructures: Integrating digital twins in the electric power ecosystem. *Computers and Security*, vol. 112, no. 102507. DOI: 10.1016/j.cose.2021.102507.

11. Masuch K., Greve M., Trang S., Kolbe L.M. (2022). Apologize or justify? Examining the impact of data breach response actions on stock value of affected companies? *Computers and Security*, vol. 112, no. 102502. DOI: 10.1016/j.cose.2021.102502.

12. Tsimenidis S., Lagkas T., Rantos K. (2022). Deep Learning in IoT Intrusion Detection. *Journal of Network and Systems Management*, vol. 30, no. 1, article number 8. DOI: 10.1007/s10922-021-09621-9.

13. Kuzmenko O.V., Kubálek J., Bozhenko V.V., Kushneryov O.S., Vida I. (2021). An approach to managing innovation to protect financial sector against cybercrime | [Podejście do zarządzania innowacjami w celu ochrony sektora finansowego przed cyberprzestępczością]. *Polish Journal of Management Studies*, vol. 24, no. 2, pp. 276–291. DOI: 10.17512/pjms.2021.24.2.17.

14. Kianpour M., Kowalski S.J., Øverby H. (2021). Systematically understanding cybersecurity economics: A survey. *Sustainability (Switzerland)*, vol. 13, no. 24, article number 13677. DOI: 10.3390/su132413677.