

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Сумський державний університет
Навчально-науковий інститут бізнесу, економіки та менеджменту
Кафедра економічної кібернетики

«До захисту допущено»
Завідувач кафедри

_____ Койбічук В.В.

«___» грудня 2024 р.

КВАЛІФІКАЦІЙНА РОБОТА
на здобуття освітнього ступеня магістр
(бакалавр / магістр)
зі спеціальності 051 «Економіка»,
(код та назва)
освітньо-професійної програми «Економічна кібернетика»
(освітньо-професійної / освітньо-наукової) (назва програми)
на тему: «Математичне моделювання впливу цифрової трансформації на
економічне зростання країни»
Здобувача групи ЕК.м-31 Оголя Дмитра Олексійовича
(шифр групи) (прізвище, ім'я, по батькові)

Кваліфікаційна робота містить результати власних досліджень. Використання ідей, результатів і текстів інших авторів мають посилання на відповідне джерело.

(підпис)

Дмитро ОГОЛЬ
(Ім'я та ПРІЗВИЩЕ здобувача)

Керівник

доцентка, д.е.н., доцентка, **Ганна ЯРОВЕНКО** _____
(посада, науковий ступінь, вчене звання, Ім'я та ПРІЗВИЩЕ) (підпис)

Суми – 2024

Міністерство освіти і науки України
Сумський державний університет
Навчально-науковий інститут бізнесу, економіки та менеджменту
Кафедра економічної кібернетики

ЗАТВЕРДЖУЮ
Завідувач кафедри
к.е.н., доцент
_____ В.В. Койбічук
“ ___ ” _____ 2024 р.

ЗАВДАННЯ НА КВАЛІФІКАЦІЙНУ РОБОТУ МАГІСТРА
спеціальність 051 «Економіка (Економічна кібернетика)»
студента 6 курсу, групи ЕК.м-31
Оголя Дмитра Олексійовича

1. Тема роботи «МАТЕМАТИЧНЕ МОДЕЛЮВАННЯ ВПЛИВУ ЦИФРОВОЇ ТРАНСФОРМАЦІЇ НА ЕКОНОМІЧНЕ ЗРОСТАННЯ КРАЇНИ».
затверджена наказом по університету від «22» жовтня 2024 року № 1082-VI.
2. Термін подання студентом закінченої роботи «09» грудня 2024 року.
3. Мета кваліфікаційної роботи – розробка та застосування математичних моделей для оцінки впливу цифрової трансформації на економічне зростання, зокрема через такі показники, як ВВП на душу населення та очікувана тривалість життя в країнах Європи.
4. Об'єкт дослідження – процеси цифрової трансформації та їх вплив на економічний розвиток країн.
5. Предмет дослідження – математичні моделі та методології для оцінки залежності між цифровими показниками (DESI) та ключовими економічними індикаторами.
6. Кваліфікаційна робота виконується на матеріалах наукових статей, монографій, статистичних та аналітичних даних міжнародних організацій, а також на інформаційно-аналітичних джерелах, пов'язаних із цифровізацією економіки та індикаторами економічного розвитку країн.
7. Орієнтовний план кваліфікаційної роботи, терміни подання розділів керівникові та зміст завдань для виконання поставленої мети:

Розділ 1. ТЕОРЕТИЧНІ ОСНОВИ ЦИФРОВИХ ТРАНСФОРМАЦІЙ В КРАЇНІ – 23 жовтня 2024 р.

У розділі 1 провести аналіз сучасних підходів до цифровізації економіки, визначити її ключові аспекти та бар'єри на шляху впровадження цифрових технологій у різних країнах світу.

Розділ 2. РОЗРОБКА МЕТОДОЛОГІЇ ДОСЛІДЖЕННЯ – 15 листопада 2024 р.

У розділі 2 розробити та обґрунтувати математичну методологію аналізу впливу цифрових трансформацій на економічне зростання, включаючи аналіз вхідних даних, перевірку мультиколінеарності, факторний аналіз і вибір моделей для моделювання.

Розділ 3. АНАЛІЗ І ВЕРИФІКАЦІЯ МОДЕЛЕЙ ВПЛИВУ ЦИФРОВИХ ТРАНСФОРМАЦІЙ НА ЕКОНОМІКУ КРАЇНИ – 05 грудня 2024 р.

У розділі 3 провести оцінку якості побудованих моделей, здійснити їх верифікацію через порівняння результатів їх роботи на тренувальних та тестових даних, а також розробити рекомендації на основі отриманих результатів для підвищення ефективності цифрової трансформації.

8. Консультації з роботи:

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		Завдання видав	Завдання прийняв
1			
2			
3			

9. Дата видачі завдання: «18» жовтня 2024 року

Керівник кваліфікаційної роботи _____ Яровенко Г.М.
(підпис) (ініціали, прізвище)

Завдання до виконання одержав _____ Оголь Д.О.
(підпис) (ініціали, прізвище)

АНОТАЦІЯ

кваліфікаційної магістерської роботи на тему:
«МАТЕМАТИЧНЕ МОДЕЛЮВАННЯ ВПЛИВУ ЦИФРОВОЇ
ТРАНСФОРМАЦІЇ НА ЕКОНОМІЧНЕ ЗРОСТАННЯ КРАЇНИ»

студента Оголя Дмитра Олексійовича

Актуальність теми зумовлена тим, що цифрова трансформація є ключовим чинником економічного розвитку, підвищуючи ефективність, продуктивність і якість життя. Відсутність комплексного підходу до оцінки її впливу ускладнює ефективне впровадження, тому дослідження з використанням математичних методів дозволяє визначити чинники, що сприяють сталому зростанню.

Метою даної роботи є розробка та застосування математичних моделей для оцінки впливу цифрової трансформації на економічне зростання європейських країн, зокрема на ВВП на душу населення та очікувану тривалість життя.

Об'єктом дослідження є процеси цифрової трансформації та їх вплив на економічний розвиток країн.

Предметом дослідження є математичні методи та моделі для оцінки залежності між цифровими індикаторами (індекс DESI) та економічними показниками.

Задачами дослідження є розроблення рекомендацій для підвищення ефективності цифрової трансформації з метою забезпечення економічного зростання.

Методи дослідження: для аналізу панельних даних у роботі застосовано факторний аналіз, регресійний аналіз (OLS, Ridge, Lasso), методи машинного навчання (Random Forest, XGBoost, Support Vector Regression).

Інформаційною базою кваліфікаційної магістерської роботи є статистичні дані міжнародних організацій, зокрема показники економічного розвитку (World

Development Indicators) та показники індексу цифрової економіки і суспільства (DESI).

Основний науковий результат кваліфікаційної магістерської роботи полягає у такому: при моделюванні впливів цифрових трансформацій на ВВП на душу населення найточніші результати показали моделі Lasso-регресії та Support Vector Regression; на очікувану тривалість життя – моделі Ridge-регресії та Support Vector Regression. Ці моделі дозволяють прогнозувати вплив цифрових трансформацій з мінімальними похибками

Одержані результати можуть бути використані для вдосконалення цифрової політики, зокрема в частині розвитку інфраструктури, підвищення цифрової грамотності та підтримки інновацій.

Результати кваліфікаційної роботи апробовані на VII Міжнародній науково-практичній конференції «GRUNDLAGEN DER MODERNEN WISSENSCHAFTLICHEN FORSCHUNG» (Цюріх, Швейцарія) і є частиною держбюджетної науково-дослідної роботи № 0124U000544 «Кібербезпекові та цифрові трансформації економіки країни воєнного часу: боротьба із кіберзлочинами, корупцією та тіньовим сектором».

Ключові слова: цифрова трансформація, індикатори цифрового розвитку, ВВП на душу населення, очікувана тривалість життя, регресійний аналіз, факторний аналіз, методи машинного навчання.

Основний зміст кваліфікаційної роботи викладено на 52 сторінках. Список використаних джерел із 35 найменувань, який розміщено на 4 сторінках.

Робота містить 2 таблиці, 30 рисунків, а також 9 додатків.

Рік виконання кваліфікаційної роботи – 2024 рік.

Рік захисту роботи – 2024 рік.

SUMMARY

The study delves into the theoretical and methodological aspects of digital transformation within the economy. Digital transformation is described as a comprehensive integration of digital technologies into all spheres of economic and social life, fundamentally altering traditional business models, improving productivity, and enhancing societal well-being.

The use of modern digital technologies contributes to the optimization of production processes and increase business efficiency. Tools such as artificial intelligence, the internet of things, blockchain, and big data help automate routine operations, reduce costs, and improve productivity. As a result, companies can enhance product quality, reduce time to market for new products, and maintain competitiveness in the dynamic conditions of the modern market.

The thesis emphasizes the key drivers of digital transformation, such as the rapid development of information and communication technologies, globalization, and the growing need for innovation and efficiency in economic processes. It also examines the transformative impact of digitalization on labor markets, economic structures, and public services.

Particular attention is paid to the socio-economic consequences of digital transformation, including both opportunities and challenges. On the one hand, digital transformation fosters economic growth, facilitates innovation, and improves access to essential services. On the other hand, it raises concerns about digital inequality, cybersecurity threats, and the need for upskilling the workforce.

Additionally, the study reviews the methodological tools used to analyze and measure digital transformation, emphasizing the importance of advanced econometric techniques for understanding its effects on economic indicators. This sets the foundation for further analysis of the role of digital technologies in shaping economic development across countries.

The research focuses on examining the effects of digital transformation on key economic indicators, specifically GDP per capita (reflects economic productivity and wealth distribution) and average life expectancy (countries with high life expectancy usually have a developed economy, an efficient healthcare system, an educated population, and stable social policies), using advanced mathematical and econometric approaches.

The chosen research topic, dedicated to mathematical modeling of the impact of digital transformation on economic growth, is highly relevant. It allows not only to assess the benefits of implementing digital tools but also to identify potential factors that have the most significant impact on economic growth.

The study seeks to understand how digital transformation influences economic development of countries by quantifying its impact on economic and social metrics.

Subject of research is processes of digital transformation and their impact on the economic indicators of countries.

Object of study economic systems of 27 European countries during the period from 2017 to 2022 which initially implies diverse levels of technological advancement, variations in governmental policies on digitalization, different socioeconomic conditions influencing the outcomes of digital adoption.

Therefore, the research tasks include:

Conducting a primary analysis of input data and preparing it for modeling. It starts with collecting and cleaning the data. This involves identifying missing or inconsistent data points and addresses them through imputation or removal. After we normalize or standardize variables to ensure comparability, especially when working with machine learning models. Next is visualizing the data using histograms, scatter plots, and correlation matrices to understand distributions and relationships.

Investigating multicollinearity and addressing it via hierarchical clustering. For this reason, we calculate the correlation matrix for input variables to detect multicollinearity (when independent variables are highly correlated). Thereafter we use

hierarchical clustering techniques to group highly correlated variables into clusters. Select representative variables from each cluster to reduce redundancy.

The next step is performing factor analysis to identify latent variables that explain the observed correlations among the original variables. For that purpose we extract factors that account for most of the variance in the data, reducing the dimensionality while retaining meaningful information. In the end, we use these factors to eliminate the residual effects of multicollinearity and simplify subsequent modeling steps.

Build mathematical models and make forecasts using Ridge and Lasso regressions, Random Forest, XGBoost, and Support Vector Regression.

After that we evaluate models accuracy: we assess the performance of each model using metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), Mean Absolute Percentage Error (MAPE), Mean Absolute Deviation (MAD), and R-squared.

In conclusion, we are identifying the most effective models, to do so we compare the predictive accuracies of all models to determine which performs best in explaining the impact of digital transformation on GDP per capita and life expectancy.

Need to say, that methodologically, the study employs a combination of traditional regression techniques and modern machine-learning approaches. Models such as Ordinary Least Squares (OLS), Ridge and Lasso regression, Random Forest, XGBoost, and Support Vector Regression are utilized to analyze panel data from 27 European countries over a period from 2017 to 2022. These models evaluate the relationship between the Digital Economy and Society Index (DESI) and economic indicators like GDP per capita and life expectancy. The study also integrates processes like data preprocessing, analysis of multicollinearity, and factor analysis to ensure robustness.

The results reveal significant insights – digital transformation variables, particularly those related to digital competencies, infrastructure, and ICT specialization,

demonstrate strong positive correlations with GDP per capita. Similarly, aspects such as broadband penetration and connectivity have notable impacts on life expectancy.

Among the models tested, Support Vector Regression demonstrated the best performance for both GDP per capita and life expectancy. It consistently provided highly accurate predictions, with minimal deviations from the actual values, even for extreme observations. Support Vector Regression was particularly effective in modeling the relationships between digital transformations and economic indicators, accounting for both intergroup differences and temporal variability.

XGBoost showed strong performance for predicting GDP per capita, providing more accurate forecasts than Random Forest. However, it struggled with extreme values and some variations in the data. Nevertheless, XGBoost's ability to adapt to dynamic changes and data variability highlights its suitability for panel data modeling.

Both Ridge and Lasso regressions showed high data approximation, with Lasso providing slightly more accurate predictions than Ridge. The use of these regularized models helped mitigate issues related to inflated parameter estimates and multicollinearity, improving the reliability of the predictions.

Although Random Forest generally captured the trends in GDP per capita, it showed significant deviations for certain observations, particularly in countries with extreme GDP values. This suggests that Random Forest may not fully account for the complexities inherent in panel data, such as time-series relationships and intergroup differences.

The models that performed best, including Support Vector Regression and XGBoost, demonstrated the importance of handling the specific characteristics of panel data, such as temporal changes and intergroup variations. These models were able to adapt more effectively to the variability in the data, providing more reliable forecasts.

The results highlight the significant role of digital transformations in influencing economic growth and life expectancy. Models that successfully incorporated digitalization metrics were better able to explain and predict changes in GDP per capita

and life expectancy, emphasizing the importance of digitalization in shaping the future of economies.

Among the models used, Lasso regression and Support Vector Regression provide the most accurate predictions for GDP per capita, while Ridge regression and Support Vector Regression excel in forecasting life expectancy. These findings underscore the importance of tailored digital policies that address specific national needs.

Practical implications of the study include recommendations for improving digital policy, particularly in the areas of infrastructure development, digital literacy enhancement, and innovation support. Specifically, to accelerate economic growth, it is advised to focus on developing digital skills among the population, expanding access to high-speed internet, and fostering business innovation.

Developing countries are encouraged to actively attract investments in digital infrastructure, reduce digital inequality, and support IT sector startups.

Thus, the results of this work confirm that digital transformation is one of the key drivers of economic growth in the modern economy. The proposed mathematical models can serve as tools for evaluating the impact of digital initiatives and making informed managerial decisions in the field of digital policy.

The results of the qualification work were presented at the VII International Scientific and Practical Conference «GRUNDLAGEN DER MODERNEN WISSENSCHAFTLICHEN FORSCHUNG» (Zurich, Switzerland) and are part of the state-funded research project No. 0124U000544, «Cybersecurity and Digital Transformations of the Economy During Wartime: Combating Cybercrime, Corruption, and the Shadow Economy».

The thesis not only advances the understanding of how digital transformation drives economic growth but also provides actionable insights for policymakers aiming to leverage digital technologies for sustainable development. Its comprehensive approach and rigorous analysis make it a valuable resource for academics and practitioners alike.

Keywords: digital transformation, digital development indicators, GDP per capita, life expectancy, regression analysis, factor analysis, and machine learning methods.

ЗМІСТ

ВСТУП.....	14
1. ТЕОРЕТИЧНІ ОСНОВИ ЦИФРОВИХ ТРАНСФОРМАЦІЙ В КРАЇНІ.....	16
1.1. Сутність та значення цифрових трансформацій для економіки країн....	16
1.2. Сучасний стан цифрової економіки країн (світу).....	18
1.3. Проблеми та бар'єри на шляху цифрової трансформації.....	22
2. РОЗРОБКА МЕТОДОЛОГІЇ ДОСЛІДЖЕННЯ.....	25
2.1. Проведення первинного аналізу вхідних даних.....	25
2.2. Дослідження мультиколінеарності вхідних змінних.....	28
2.3. Здійснення факторного аналізу.....	32
2.4. Характеристика основних методів моделювання.....	36
2.4.1. Побудова OLS регресії.....	36
2.4.2. Побудова Ridge та Lasso регресій.....	37
2.4.3. Побудова Random Forest.....	38
2.4.4. Побудова XGBoost.....	40
2.4.5. Побудова Support Vector Regression.....	41
3. АНАЛІЗ І ВЕРИФІКАЦІЯ МОДЕЛЕЙ ВПЛИВУ ЦИФРОВИХ ТРАНСФОРМАЦІЙ НА ЕКОНОМІКУ КРАЇНИ.....	43
3.1. Аналіз моделей та їх якості для ВВП на душу населення.....	43
3.1.1. OLS, Ridge та Lasso регресії.....	43
3.1.2. Random Forest.....	47
3.1.3. XGBoost.....	48

3.1.4. Support Vector Regression.....	49
3.1.5 Оцінка якості прогнозів.....	50
3.2. Аналіз моделей та їх якості для очікуваної тривалості життя.....	51
3.2.1. OLS, Ridge та Lasso регресії.....	51
3.2.2. Random Forest.....	56
3.2.3. XGBoost.....	57
3.2.4. Support Vector Regression.....	58
3.2.5 Оцінка якості прогнозів.....	59
ВИСНОВКИ.....	61
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ.....	63
ДОДАТКИ.....	67

ВСТУП

Актуальність даної теми обумовлена стрімким зростанням впливу цифрових технологій на різні аспекти соціально-економічного розвитку країн. В умовах сьогодення діджиталізація стає основним чинником економічного зростання, інновацій та конкурентоспроможності держав у глобальному контексті.

Впровадження цифрових технологій сприяє оптимізації виробничих процесів, підвищенню продуктивності праці, покращенню якості послуг та зниженню витрат. Однак зростаюча складність цифрової економіки породжує нові виклики, такі як кіберзагрози, дисбаланс доступу до технологій та необхідність формування нових підходів до регулювання.

Обрана тема дослідження, присвячена математичному моделюванню впливу цифрової трансформації на економічне зростання, є надзвичайно актуальною. Вона дозволяє не лише оцінити переваги впровадження цифрових інструментів, але й виявити потенційні чинники, які мають найбільш суттєвий вплив на економічне зростання.

Мета роботи полягає у розробці та застосуванні математичних моделей для оцінки впливу цифрової трансформації на економічне зростання.

Об'єкт дослідження – процеси цифрової трансформації та їх вплив на економічний розвиток країн.

Предмет дослідження – математичні моделі та методології для оцінки залежності між цифровими показниками (DESI) та економічними індикаторами що відображають рівень економічного розвитку країни (ВВП на душу населення та очікувана тривалість життя).

Завдання дослідження:

- вивчити теоретичні аспекти цифрової трансформації та її значення для економіки;

- розробити методологію дослідження, включаючи підготовку вхідних даних, аналіз мультиколінеарності змінних, факторний аналіз для зменшення розмірності даних, застосування методів моделювання (OLS, Ridge, Lasso регресії, Random Forest, XGBoost, SVR).
- провести моделювання впливу цифрової трансформації на економічне зростання та якість життя;
- верифікувати побудовані моделі та оцінити їх прогнозну точність.

Ця робота охоплює як теоретичні, так і практичні аспекти, що дозволяє забезпечити всебічний аналіз та розробити рекомендації для ефективної інтеграції цифрових технологій в економіку країни.

Результати кваліфікаційної роботи апробовані на VII Міжнародній науково-практичній конференції «GRUNDLAGEN DER MODERNEN WISSENSCHAFTLICHEN FORSCHUNG» (Цюріх, Швейцарія) [35] і є частиною держбюджетної науково-дослідної роботи № 0124U000544 «Кібербезпекові та цифрові трансформації економіки країни воєнного часу: боротьба із кіберзлочинами, корупцією та тіншовим сектором».

1. ТЕОРЕТИЧНІ ОСНОВИ ЦИФРОВИХ ТРАНСФОРМАЦІЙ В КРАЇНІ

1.1. Сутність та значення цифрових трансформацій для економіки країн

Цифрова трансформація – це складний і багатогранний процес, що включає глибоке впровадження сучасних цифрових технологій у всі сфери економіки, державного управління та суспільного життя. Вона передбачає не лише використання новітніх ІТ-рішень, але й переосмислення підходів до ведення бізнесу, надання державних послуг і взаємодії громадян із державою. Завдяки цьому процесу можна значно підвищити ефективність, продуктивність і конкурентоспроможність у різних секторах, що сприятиме сталому економічному розвитку країни.

Основою цифрового розвитку економіки та суспільного життя стало швидке поширення інтернету та стільникового (рис.1.1). Завдяки мобільним мережам і широкосмуговому доступу до інтернету зросло охоплення цифрових технологій навіть у віддалених регіонах. Це сприяє розвитку електронного урядування, онлайн-освіти, телемедицини та електронної комерції. Стільниковий зв'язок, особливо з появою технологій 4G і 5G, забезпечує швидкість і надійність з'єднань, необхідних для впровадження інновацій, таких як інтернет речей та штучний інтелект. Усе це створює нові можливості для економічного та соціального розвитку.

Використання сучасних цифрових технологій сприяє оптимізації виробничих процесів та підвищенню ефективності бізнесу. Інструменти, такі як штучний інтелект, інтернет речей, блокчейн і великі дані, допомагають автоматизувати рутинні операції, зменшувати витрати та підвищувати продуктивність. Завдяки цьому компанії можуть покращувати якість продукції,

скорочувати час на впровадження нових продуктів на ринок і забезпечувати конкурентоспроможність у динамічних умовах сучасного ринку.

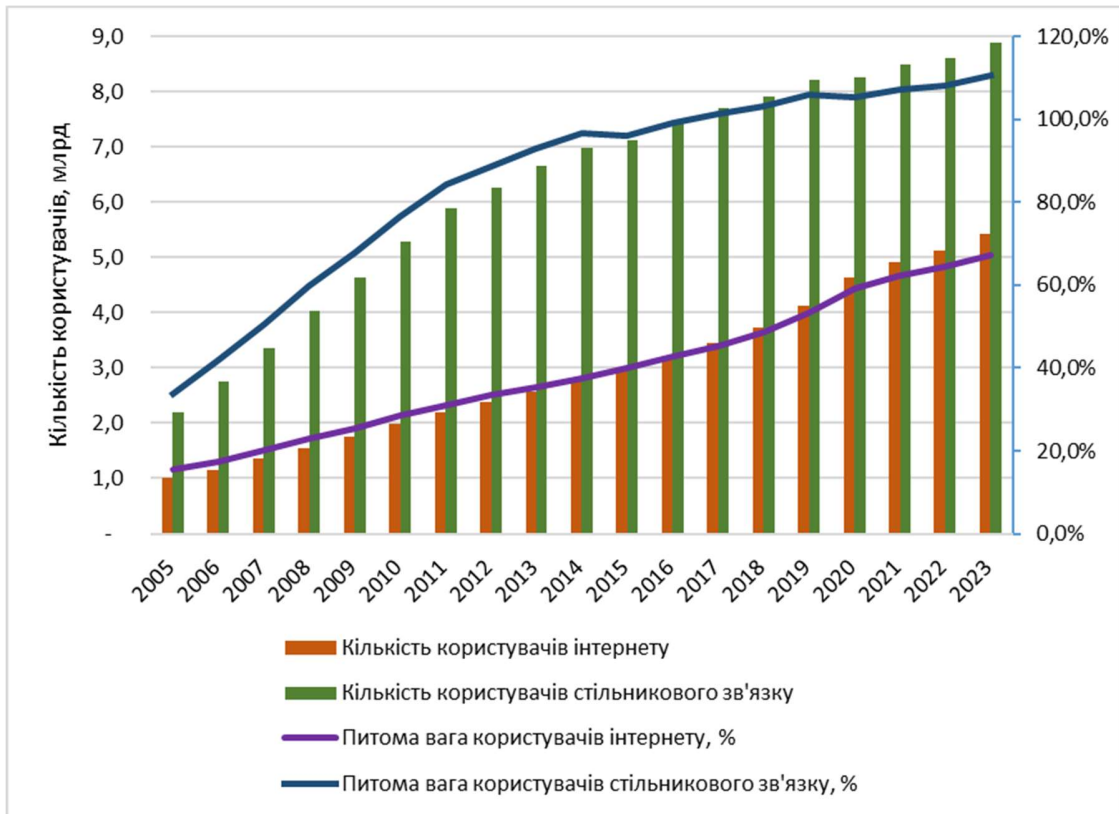


Рисунок 1.1 – Кількість користувачів інтернету та стільникового зв'язку у світі

Сучасні цифрові інструменти допомагають компаніям глибше розуміти потреби клієнтів, аналізуючи дані про їхні вподобання, поведінку та запити. Це дозволяє створювати персоналізовані пропозиції, підвищувати якість обслуговування та покращувати клієнтський досвід. Наприклад, чат-боти та системи підтримки, що працюють на основі штучного інтелекту, забезпечують швидке реагування на запити клієнтів і надають цілодобову підтримку, сприяючи підвищенню їхньої задоволеності.

Цифрова трансформація відкриває широкі можливості для інновацій, стимулюючи створення нових бізнес-моделей та розвиток стартап-екосистеми.

Вона сприяє економічному зростанню завдяки підвищенню ефективності використання ресурсів і залученню інвестицій у високотехнологічні сектори. Окрім того, цифровізація сприяє збільшенню кількості робочих місць у галузях інформаційних технологій та наукоємних сферах.

Інтеграція цифрових технологій у державне управління допомагає зменшити бюрократичні бар'єри, спростити процедури та скоротити час отримання адміністративних послуг. Електронне урядування й онлайн-платформи підвищують прозорість управлінських процесів, що сприяє зниженню рівня корупції. Наприклад, система електронних торгів забезпечує прозорість у державних закупівлях, підвищуючи довіру громадян до державних інституцій.

Цифрові технології значно покращують доступ до освітніх, медичних і соціальних послуг. Наприклад, телемедицина дозволяє пацієнтам отримувати консультації лікарів дистанційно, що є особливо важливим для жителів віддалених регіонів. У сфері освіти цифровізація сприяє розвитку онлайн-курсів та платформ для дистанційного навчання, відкриваючи нові можливості для саморозвитку та підвищення кваліфікації.

Цифрова трансформація є ключовим елементом сучасного розвитку, створюючи конкурентні переваги як для приватного сектору, так і для держави. Це не просто технологічний тренд, а стратегічний напрям, що охоплює всі сфери життя, стимулюючи економічне зростання та покращення якості життя суспільства.

1.2. Сучасний стан цифрової економіки країн (світу)

Провідні країни світу активно інтегрують цифрові технології в економічні процеси, адже в сучасному світі цифрова економіка є ключовим чинником конкурентоспроможності. Держави, які інвестують у розвиток таких технологій,

як штучний інтелект, аналіз великих даних, інтернет речей і блокчейн, досягають значного прогресу в оптимізації виробничих процесів, покращенні якості послуг і підвищенні ефективності державного управління.

Такі підходи не лише сприяють економічному зростанню, але й зміцнюють позиції цих країн на глобальній арені, дозволяючи їм адаптуватися до викликів сучасної економіки та впроваджувати інноваційні рішення в усіх сферах життя.

Світовим лідером у розвитку цифрової економіки за станом на 2022 рік є Данія – для неї характерні високі показники цифровізації як національної економіки так і загалом усього суспільства. Крім Данії високі місця у рейтингу цифрової конкурентоспроможності посідають найбільш інноваційно розвинені країни, адже вони насамперед безпосередньо продукують сучасні інформаційно-комунікаційні технології. [3] За даними міжнародних організацій, США є лідером у впровадженні штучного інтелекту та аналізу великих даних. Це дозволяє американським компаніям, таким як Google, Microsoft та Amazon, випереджати конкурентів і створювати нові бізнес-моделі. Китай, зі свого боку, робить ставку на розвиток технологій інтернету речей та інновацій у фінансових технологіях, активно використовуючи блокчейн для створення цифрової національної валюти (цифрового юаня). Німеччина фокусується на впровадженні Індустрії 4.0, що передбачає глибоку інтеграцію інтернету речей у виробничі процеси, зокрема в автомобільній та машинобудівній галузях. Південна Корея, одна з найбільш цифровізованих країн світу, активно розвиває мережі 5G, що дозволяють впроваджувати інноваційні рішення у сферах охорони здоров'я, освіти та міського управління.

Останнім часом швидко розвиваються та впроваджуються в економічні та інші сфери технології штучного інтелекту. У цифровій економіці вони сприяють персоналізації послуг, аналізу великих даних, прогнозуванню ринкових трендів, а також відкривають нові можливості для створення інноваційних продуктів і сервісів, стимулюючи економічне зростання і конкурентоспроможність. За

дослідженнями McKinsey, проведеними у 2024 році до 34% компаній в залежності від напрямку діяльності вже сьогодні застосовують технології штучного інтелекту в своїй операційній діяльності (рис. 1.2) [25].

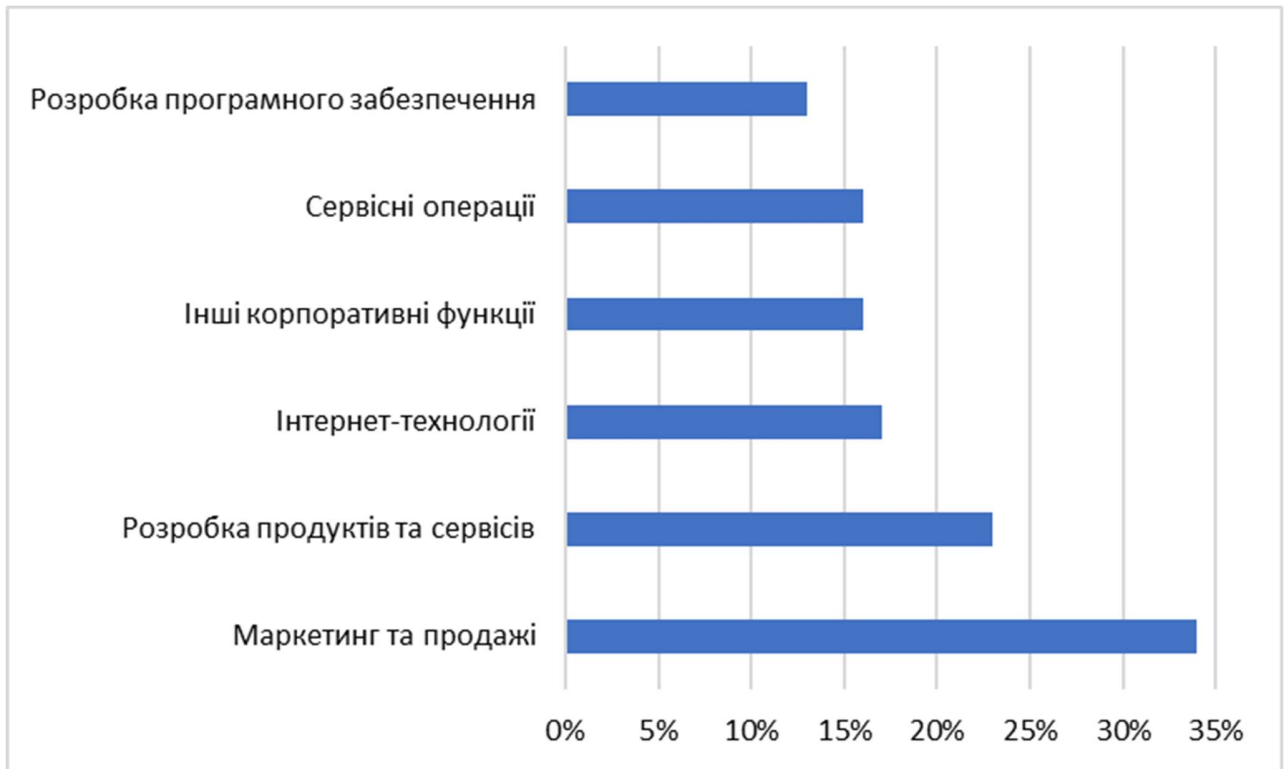


Рисунок 1.2 – Використання компаніями технологій штучного інтелекту за напрямками діяльності у 2024 році

Впровадження цифрових технологій дозволяє країнам прискорити економічний розвиток завдяки підвищенню продуктивності праці, оптимізації ресурсів та зниженню витрат на виробництво. Це позитивно впливає на зростання ВВП країн та створює нові можливості для залучення інвестицій. Цифрові платформи та онлайн-сервіси сприяють розвитку малого та середнього бізнесу, дозволяючи їм швидше адаптуватися до змін на ринку та виходити на міжнародні ринки. За оцінками McKinsey, до 2030 до 30% робочих місць можна буде автоматизувати або суттєво підвищити їх продуктивність використовуючи потужності генеративного штучного інтелекту.

Цифровізація відкриває широкі можливості для міжнародної торгівлі та кооперації між країнами. Завдяки інтернет-платформам і хмарним технологіям компанії можуть швидко масштабувати свою діяльність і залучати клієнтів з усього світу. Це сприяє інтеграції глобальної економіки, розвитку міжнародного ринку праці та створенню умов для співпраці з висококваліфікованими фахівцями незалежно від їхнього географічного розташування.

Хоча автоматизація викликає побоювання щодо скорочення робочих місць, цифрова економіка одночасно створює нові можливості. Зростає попит на спеціалістів у сферах інформаційних технологій, кібербезпеки, аналітики даних, управління проектами, а також розробки й підтримки цифрових платформ, штучного інтелекту та машинного навчання. Це відкриває перспективи для професійного розвитку, перекваліфікації та освоєння нових компетенцій, що відповідають вимогам сучасного ринку праці.

Попри стрімкий розвиток цифрових технологій, значний розрив у доступі до них між розвиненими країнами та тими, що розвиваються, залишається серйозною проблемою. Брак інфраструктури, низький рівень цифрової грамотності та обмежений доступ до інтернету стримують розвиток цифрової економіки в багатьох регіонах світу. Це сприяє поглибленню глобальної економічної нерівності, адже країни, які не встигають за темпами цифровізації, втрачають можливості для економічного зростання та залучення інвестицій.

Для подолання цифрового розриву необхідні масштабні інвестиції в інфраструктуру, програми підвищення цифрової грамотності населення та активізація міжнародного співробітництва у сфері технологій. Ключову роль можуть відіграти міжнародні організації, впроваджуючи ініціативи для підтримки цифровізації в країнах, що розвиваються. Це включає надання технічної допомоги, фінансування проектів у галузях освіти, охорони здоров'я та електронного урядування, що сприятиме інтеграції цих країн у глобальну цифрову економіку.

Цифрова економіка стає основним рушієм сучасного економічного розвитку, і країни, які успішно впроваджують цифрові технології, отримують суттєві конкурентні переваги на глобальному ринку. Водночас важливо, щоб переваги цифровізації були доступними для всіх, щоб уникнути поглиблення нерівності та забезпечити стале економічне зростання у всьому світі.

1.3. Проблеми та бар'єри на шляху цифрової трансформації

Цифровізація, попри численні переваги, супроводжується низкою ризиків і викликів, таких як забезпечення кібербезпеки, захист персональних даних і подолання цифрової нерівності. Для успішної цифрової трансформації необхідно створити надійний правовий і технологічний фундамент, що запобігатиме зловживанням, гарантуватиме конфіденційність інформації та захищатиме права користувачів.

Однією з головних перешкод на шляху цифрової трансформації є недостатній рівень цифрових навичок, особливо серед старших поколінь та мешканців сільських регіонів. Брак обізнаності у використанні цифрових технологій обмежує здатність людей адаптуватися до змін на ринку праці, що може призводити до соціальної ізоляції та зниження конкурентоспроможності.

Розвиток цифрової грамотності має стати одним із ключових пріоритетів. Важливу роль у цьому відіграють програми навчання для школярів і перекваліфікації для дорослих. Найбільш технологічно розвинені країни вже сьогодні активно інвестують у цифрову освіту, забезпечуючи високу компетентність населення у використанні сучасних технологій. Їхній досвід демонструє, як стратегічний підхід до навчання може сприяти загальному прогресу суспільства в умовах цифрової епохи.

У багатьох країнах, що розвиваються, обмежений доступ до інтернету, зокрема до широкопasmового зв'язку, є значною перешкодою для впровадження цифрових рішень. Недостатній розвиток інфраструктури, зокрема відсутність високошвидкісних мереж та якісного мобільного покриття, призводить до цифрової ізоляції певних регіонів, що обмежує можливості для розвитку електронної комерції, дистанційної освіти та охорони здоров'я. Брак інвестицій у цифрову інфраструктуру також гальмує впровадження інноваційних технологій, таких як інтернет речей, які могли б значно підвищити ефективність економічних процесів.

Окрім технічних проблем, багато країн стикаються з дефіцитом кваліфікованих фахівців у сфері інформаційних технологій, що ускладнює впровадження та підтримку цифрових систем. Попит на спеціалістів у таких галузях, як кібербезпека, штучний інтелект, аналіз даних та програмування, значно перевищує пропозицію, що призводить до зростання заробітних плат і «витоку мізків» із країн з менш конкурентоспроможними економіками.

Рішення проблеми дефіциту кваліфікованих кадрів вимагає активної співпраці між освітніми установами та бізнесом для розробки програм підготовки, що відповідають потребам сучасного ринку праці. Важливим є створення освітніх ініціатив, які не тільки допоможуть заповнити вакансії в технологічних сферах, але й сприятимуть сталому розвитку цифрової економіки, надаючи людям навички для адаптації до нових вимог.

З одночасним розвитком цифрових технологій зростає кількість кіберзагроз, серед яких атаки на критичну інфраструктуру, витоки конфіденційних даних та фінансові шахрайства. Чим більше компанії та державні установи використовують хмарні рішення та інтернет речей, тим вищий ризик кібератак, що можуть призвести до значних фінансових втрат та зниження довіри з боку споживачів.

Для захисту від кіберзагроз необхідно активно інвестувати в системи кібербезпеки, розробляти надійні стратегії захисту та постійно навчати персонал, щоб мінімізувати ризики і забезпечити безпеку цифрових операцій на всіх рівнях.

Стрімкий розвиток таких технологій, як штучний інтелект, блокчейн і великі дані, ставить перед регулюючими органами нові виклики, оскільки вони часто не встигають адаптувати нормативно-правову базу до нових реалій. Це призводить до виникнення правових та етичних питань, пов'язаних із захистом персональних даних, конфіденційністю, правами людини та використанням автоматизованих систем для ухвалення рішень. Наприклад, алгоритми штучного інтелекту, що використовуються у правосудді чи рекрутингу, можуть стати джерелом упередженості і дискримінації, що вимагає розробки чітких етичних стандартів та належного контролю з боку держави.

Багато компаній не готові до цифрової трансформації через необхідність значних інвестицій у модернізацію процесів та зміну корпоративної культури. Традиційні галузі, такі як важка промисловість, енергетика та сільське господарство, часто зустрічають спротив до змін через побоювання автоматизації та скорочення робочих місць. Наприклад, підприємства, які десятиліттями поклалися на ручну працю, можуть відчувати труднощі при впровадженні автоматизованих систем, що знижує їх конкурентоспроможність. Для подолання цього бар'єра необхідно проводити роз'яснювальні кампанії, демонструючи, що цифровізація не лише оптимізує процеси, але й відкриває нові можливості для бізнесу та створення нових робочих місць.

Для успішного впровадження цифрових технологій важливий комплексний підхід, який включає розвиток цифрової інфраструктури, підвищення рівня цифрової грамотності населення, зміцнення кібербезпеки та адаптацію законодавства до нових реалій. Крім того, необхідно створювати стимули для бізнесу, щоб компанії могли інвестувати у цифрові рішення, підвищувати ефективність і адаптуватися до змін у глобальній економіці.

2. РОЗРОБКА МЕТОДОЛОГІЇ ДОСЛІДЖЕННЯ

2.1. Проведення первинного аналізу вхідних даних

Для дослідження були використані дані за 2017-2022рр. для 27 європейських країн, а саме:

- дані показників економічного розвитку World Development Indicators [32], а саме ВВП на душу населення та середня очікувана тривалість життя для різних країн світу, оскільки вони є ключовими індикаторами для оцінки рівня економічного розвитку, добробуту населення та якості життя у країні;

- показники індексу Європейської Комісії оцінки рівня цифрової економіки та цифрового суспільства в країнах Європейського Союзу (DESI, digital economy and society Index) [9], оскільки вони характеризують готовність країни до автоматизації, її підготовленість до впровадження цифрових технологій, а також характеризують легкість та доступність взаємодії з державними органами.

Отримані дані будуть використані для оцінки взаємозв'язків між рівнем економічного розвитку та цифровою трансформацією країн, що дозволить виявити потенційні фактори, які впливають на їхню конкурентоспроможність у сучасній цифровій економіці, а також дозволять побудувати прогнозні моделі. Концептуальна модель проведеного дослідження представлена в додатку А.

Перед реалізацією математичного моделювання, вхідні дані було попередньо оброблено та перевірено на наявність пропусків та аномальних значень. Показники DESI були закодовані у змінні Var1-Var82 для спрощення роботи з ними в моделі, при цьому кожна змінна відповідає конкретному аспекту цифрової трансформації (наприклад, доступ до високошвидкісного інтернету, рівень цифрових навичок, інтеграція цифрових технологій у бізнес тощо) (див. Додаток Б).

Було виявлено, що по деяким змінним була пропущена інформація (Рисунок 2.1).

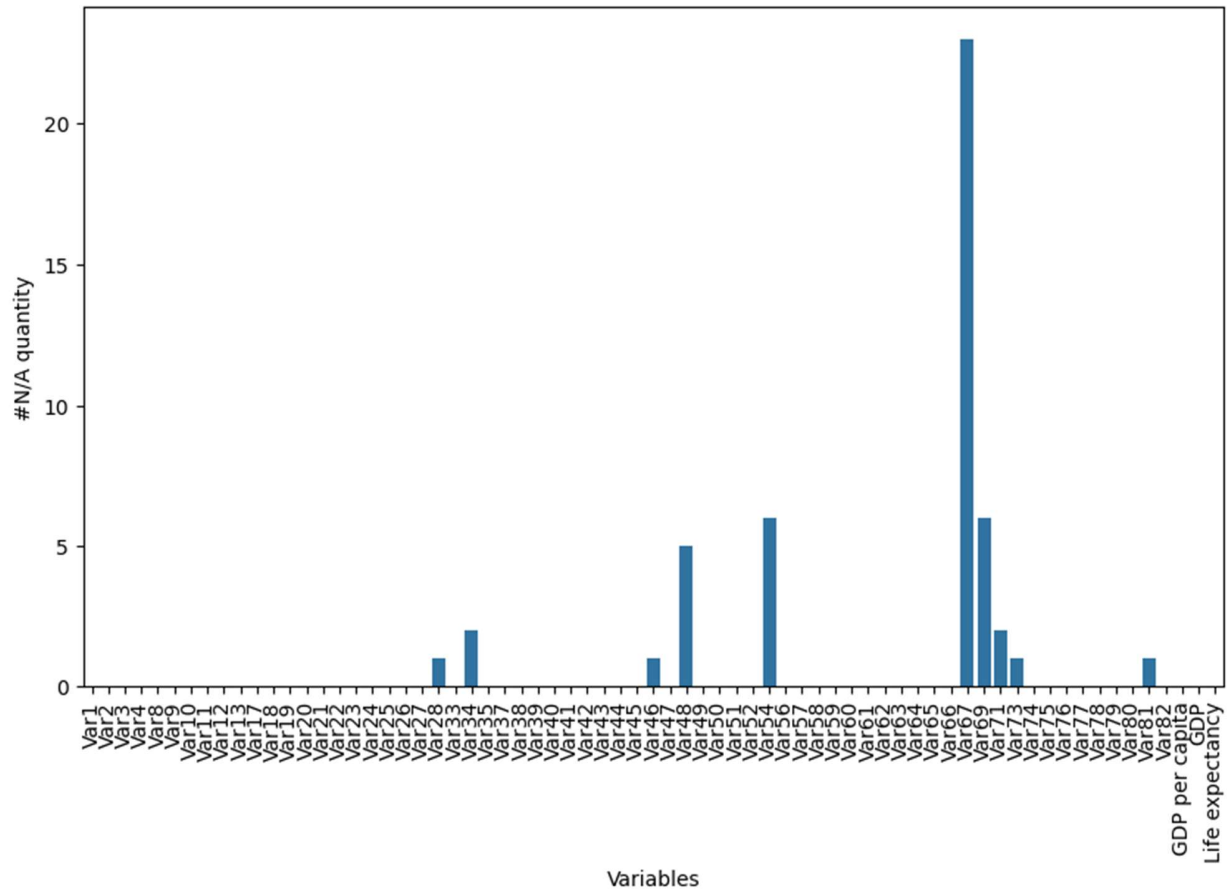


Рисунок 2.1 – Гістограма змінних з пропущеними значеннями

В результаті було прийнято рішення щодо заміни цих значень на їх усереднене значення, а в деяких випадках на нульове, щоб зменшити помилку в процесі моделювання. Результат даної процедури представлений на рисунку 2.2.

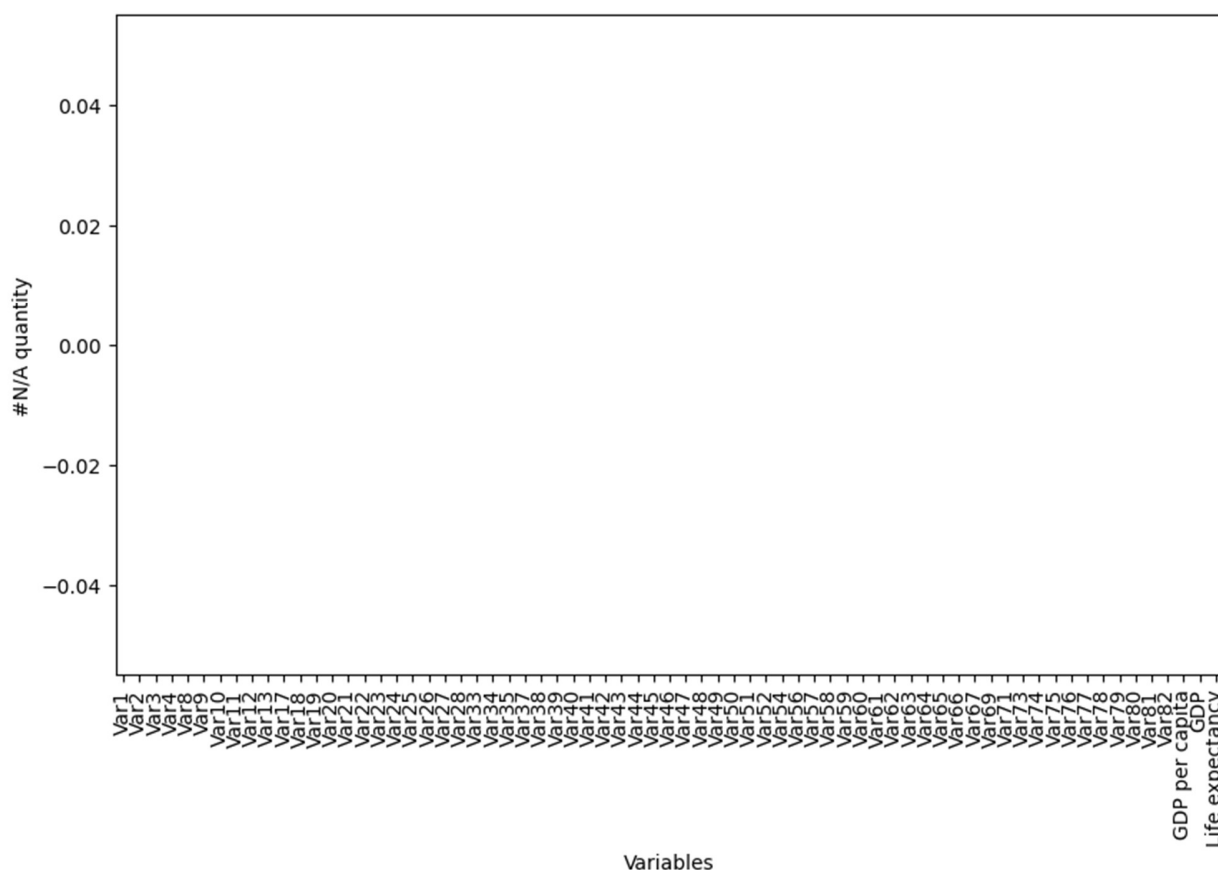


Рисунок 2.2 – Гістограма змінних з відновленими даними

Побудова гістограм розподілу змінних (Додаток В) дозволила побачити, що змінні мають різний розподіл і більшість з них наближається до нормального. Щодо інших, то це обумовлюється тим, що змінні брались з урахуванням часу та для різних країн. Тому в подальших дослідженнях буде здійснено кластерний аналіз, який повинен врахувати розкид даних та сформувати групи з близьким рівнем цифрового розвитку.

Отриманий набір даних є підготовленим для перевірки на наявність мультиколінеарності між змінними за допомогою показника фактору інфляції варіансу (VIF, Variance Inflation Factor), і, за необхідності, певні змінні будуть усунуті або трансформовані. Також буде проведено факторний аналіз для зменшення кількості змінних, зберігаючи при цьому основні фактори, які

пояснюють варіативність даних, що може покращити ефективність моделювання. Далі, застосовуючи інструментарій Python, будуть побудовані математичні моделі основними методами моделювання: OLS регресії, Ridge та Lasso регресій, Random Forest, XGBoost та Support Vector Regression.

2.2. Дослідження мультиколінеарності вхідних змінних

Дослідження мультиколінеарності вхідних змінних є важливим етапом під час побудови регресійних і не тільки моделей, оскільки вона може значно вплинути на точність і стабільність оцінок параметрів моделі. Мультиколінеарність виникає, коли дві або більше вхідних змінних (незалежних змінних) у моделі мають високий рівень кореляції між собою. Це може призвести до проблем з інтерпретацією результатів, нестабільності коефіцієнтів регресії та зниження точності прогнозів.

Перед перевіркою мультиколінеарності дані необхідно стандартизувати, для чого буде використано бібліотека Python StandardScaler (Рисунок 2.3).

```
#Стандартизація даних
from sklearn.preprocessing import StandardScaler

scaler = StandardScaler()
df = pd.DataFrame(scaler.fit_transform(df), columns=df.columns)
```

Рисунок 2.3 – Стандартизація вхідних даних

Для перевірки даних на мультиколінеарність застосуємо бібліотеку аналізу і обробки даних Python – `variance_inflation_factor` (Рисунок 2.4).

```
#Перевірка мультиколінеарності за допомогою VIF
from statsmodels.stats.outliers_influence import variance_inflation_factor

vif_data = pd.DataFrame()
vif_data["feature"] = df.columns
vif_data["VIF"] = [variance_inflation_factor(df.values, i) for i in range(df.shape[1])]

pd.set_option('display.max_rows', None)
pd.set_option('display.max_columns', None)
```

Рисунок 2.4 – Перевірка мультиколінеарності вхідних даних

Отримані результати обчислень було візуалізовано у вигляді гістограми 2.5. Вони вказують на значну мультиколінеарність досліджуваних змінних, оскільки значення показника VIF зависокі. Лише три з 82 досліджуваних показників свідчать про відсутність кореляції, тобто їх взаємозв'язок не є настільки сильним, щоб суттєво вплинути на точність оцінки інших змінних в моделі. Але дуже високі значення VIF інших 79 змінних є серйозною проблемою, і потрібно вжити заходів для її виправлення.

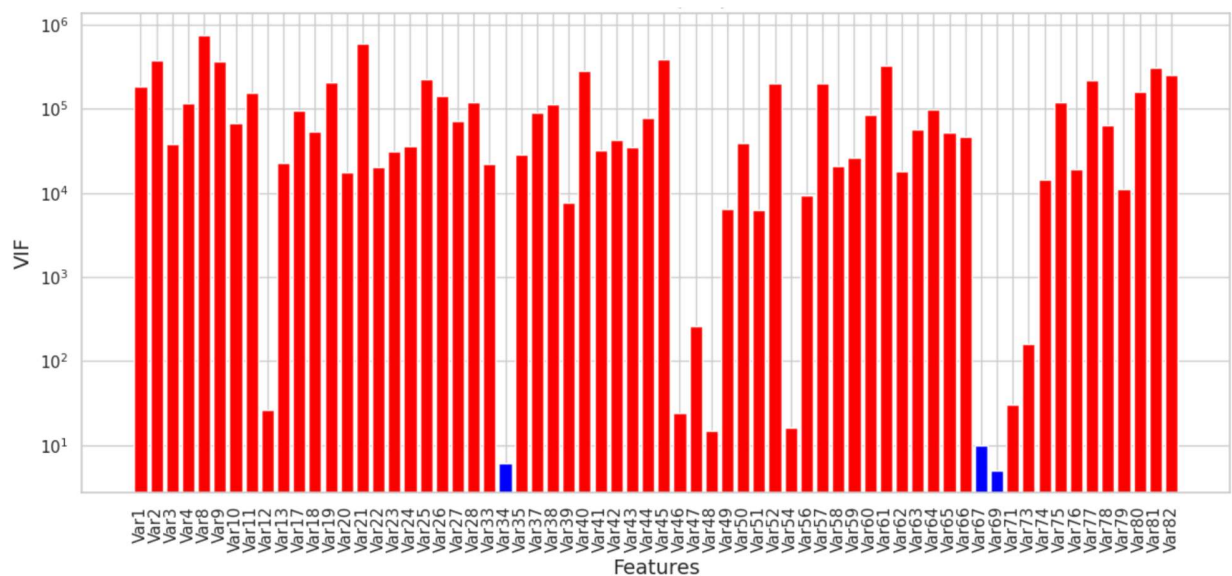


Рисунок 2.5 – Графічне представлення мультиколінеарності змінних

Оскільки стандартні процедури типу усунення мультиколінеарних змінних можуть призвести до усунення критичної більшості з них, то в даному випадку

найліпшим виходом є провести групування змінних з послідовним усуненням деяких найбільш корельованих змінних. Для цієї процедури проведемо ієрархічну кластеризацію методом Уорда, яка краще за все дозволить виділити групи змінних, які необхідно об'єднати (Рисунок 2.6).

```

from scipy.cluster.hierarchy import linkage, dendrogram, fcluster

# Налаштуємо стиль графіка
sns.set(style="dark")

# Обчислення кореляційної матриці
correlation_matrix = df.corr().abs()
np.fill_diagonal(correlation_matrix.values, 0)

# Використання агломеративної кластеризації
linked = linkage(correlation_matrix, method='ward')

# Побудова дендрограми
plt.figure(figsize=(12, 8))
dendrogram(linked, labels=correlation_matrix.columns, orientation='top', distance_sort='descending', show_leaf_counts=True)
plt.title('Dendrogram for Hierarchical Clustering')
plt.xlabel('Variables')
plt.ylabel('Distance')
plt.show()

```

Рисунок 2.6 – Побудова дендрограми змінних

Результати ієрархічного аналізу досліджуваних змінних на основі їх подібностей представлені на дендрограмі (Рисунок 2.7). Вона показує, як об'єкти або елементи групуються або ієрархічно об'єднуються в кластери на різних рівнях подібності.

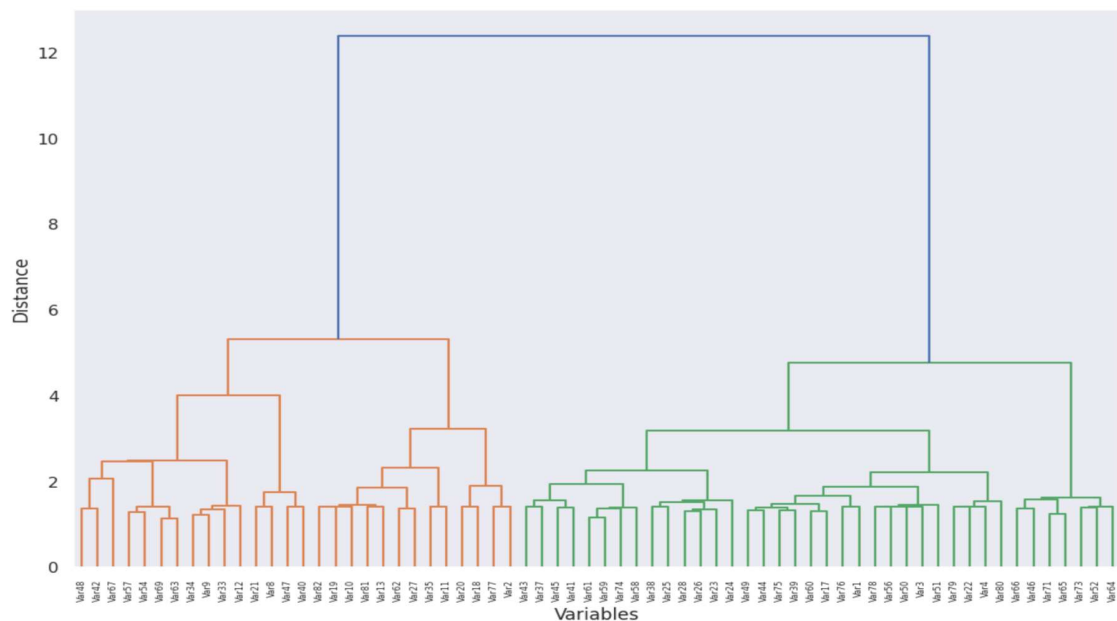


Рисунок 2.7 – Дендрограма результатів кластеризації досліджуваних змінних

На базі побудованої дендрограми було сформовано дванадцять кластерів, які дозволили провести групування і сформувати дванадцять змінних. Дану кількість було визначено шляхом поступового об'єднання показників, розрахунку VIF та видалення деяких найбільш корельованих факторів, що призводили до завищення до мультиколінеарності навіть у групі. Результати розрахованих значень VIF було візуалізовано (Рисунок 2.8). Як видно з результатів розрахунків, ми отримали цілком прийнятні значення показника фактору інфляції варіансу (менше 10), що свідчить про відсутність сильної кореляції між незалежними змінними або наявність незначної, допустимої для розрахунків кореляції. Отриманий результат забезпечить стабільність і надійність регресійної моделі. Таким чином, ми зможемо побудувати модель, яка точно відображає залежності між змінними без втрати значущості окремих факторів.

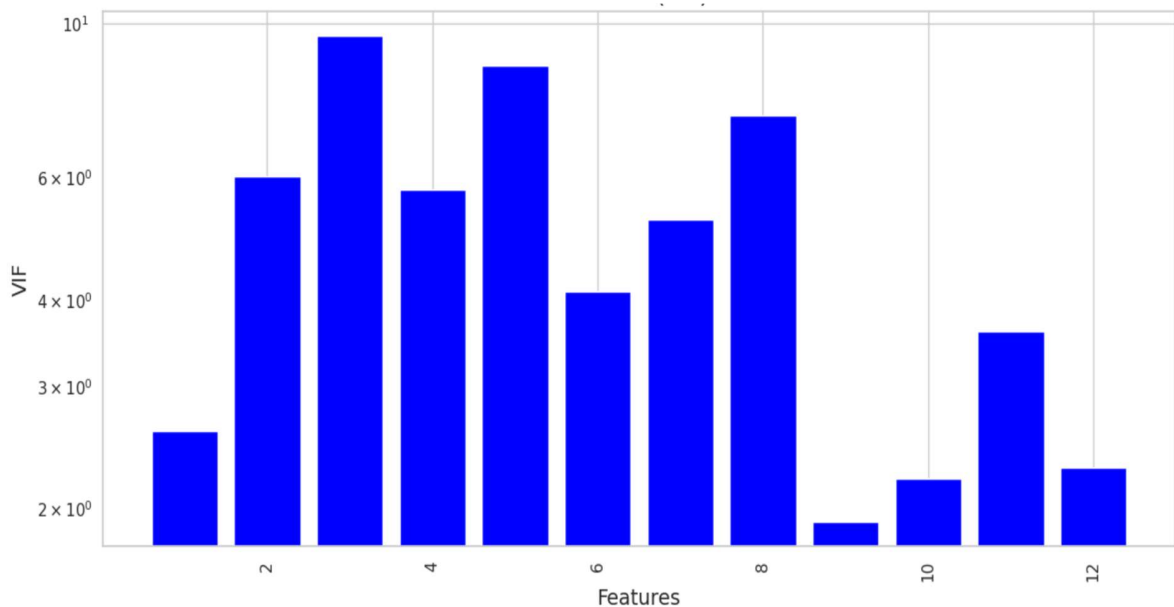


Рисунок 2.8 – Графічне представлення мультиколінеарності кластерів

2.3. Здійснення факторного аналізу

Для здійснення прихованого впливу окремих змінних та підготовки даних до моделювання, здійснимо факторний аналіз виділених у попередньому розділі згрупованих змінних (кластерів). Спочатку оцінимо, чи підходять дані для проведення даної процедури. Для цього скористаємося тестами Кайзера-Мейєра-Олкіна (КМО) та Тестом сферичності Бартлета (Рисунок 2.9).

```
# Перевірка адекватності даних для факторного аналізу
kmo_all, kmo_model = calculate_kmo(aggregated_df1)
bartlett_sphericity = calculate_bartlett_sphericity(aggregated_df1)

print(f"КМО: {kmo_model:.4f}")
print(f"Bartlett's Test: Chi-square = {bartlett_sphericity[0]:.4f}, p-value = {bartlett_sphericity[1]:.4f}")
```

Рисунок 2.9 – Перевірка адекватності даних для факторного аналізу

В результаті отримано, що КМО дорівнює 0.7959, що говорить про дуже добру адекватність, оскільки значення близько 0.8. За тестом Бартлета виявлено, що Chi-square = 1720.9165, а p-value = 0.0000, тобто нульова гіпотеза відхиляється і дані підходять для факторного аналізу.

Для того, щоб визначити, яку оптимальну кількість факторів необхідно використати у факторному аналізі, було побудовано графік власних значень або графік каменистого осипу (scree plot) (Рисунок 2.10).

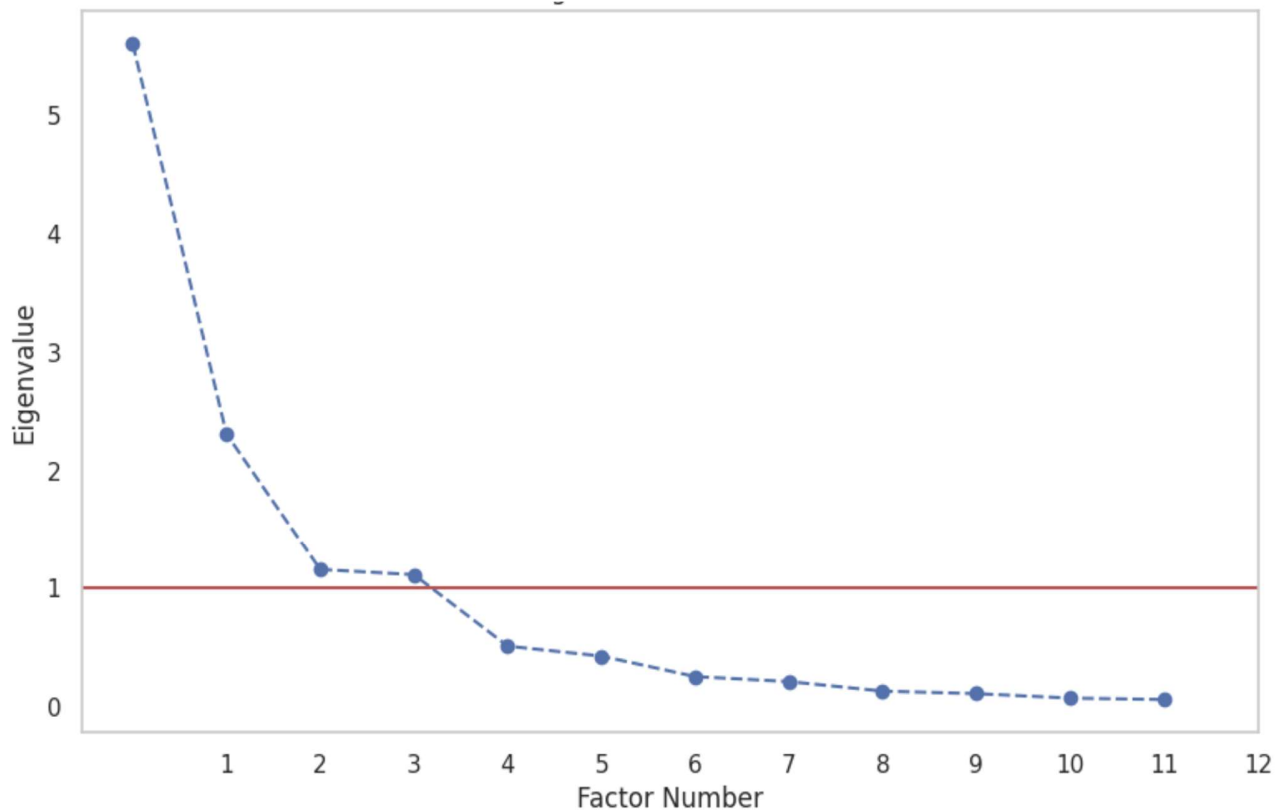


Рисунок 2.10 – Графічне зображення вибору оптимальної кількості факторів для факторного аналізу

Як бачимо з графіку, основними факторами, що пояснюють варіативність у наборі даних є: 0, 1, 2 та 3. Їх власне значення вище одиниці. Таким чином, для подальшого моделювання будемо використовувати саме їх, оскільки це дозволить нам зменшити кількість змінних, а отже знизить складність моделювання та полегшить інтерпретацію даних.

Здійснення факторного аналізу відбувалося за допомогою бібліотеки Python FactorAnalyzer та проведення обертання за методом VARIMAX (Рисунок 2.11). Це дозволило отримати чіткішу структуру факторів, зробивши зв'язки між змінними та факторами більш очевидними.

```

# Факторний аналіз
fa = FactorAnalyzer(n_factors=4, rotation='varimax')
fa.fit(aggregated_df1)

# Отримання завантажень факторів
loadings = fa.loadings_
loading_df = pd.DataFrame(loadings, index=aggregated_df1.columns)

# Візуалізація завантажень факторів
plt.figure(figsize=(12, 8))
sns.heatmap(loading_df, annot=True, cmap='coolwarm', center=0)
plt.title('Factor Loadings')
plt.xlabel('Factors')
plt.ylabel('Variables')
plt.show()

# Виведення завантажень факторів
print("Factor Loadings:")
print(loading_df)

```

Рисунок 2.11 – Реалізація факторного аналізу

Результати факторного аналізу представлені на рисунку 2.12. Можна побачити, що 0-ий фактор сформували Var 1-5 та Var 11, 1-й фактор - Var 6-9, 2-й фактор - Var 10, 3-й фактор - Var 12. Проведемо ідентифікацію цих факторів, базуючись на тих початкових змінних, які були агреговані та використані для реалізації факторного аналізу.

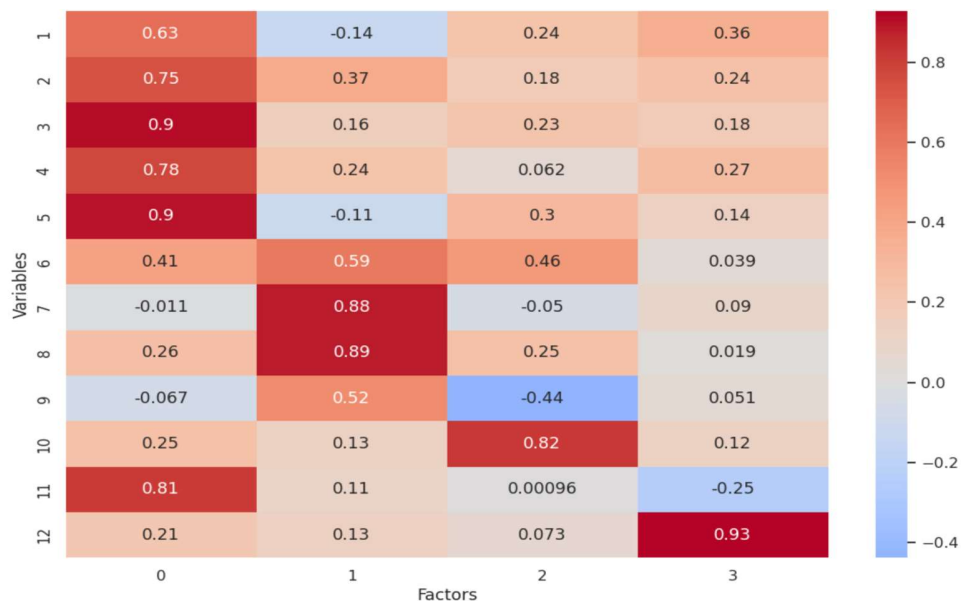


Рисунок 2.12 – Результати факторного аналізу

В результаті попереднього аналізу вхідних даних та проведених з ними маніпуляцій було визначено чотири фактори, які були сформовані на основі низки вхідних індикаторів, які характеризують цифровий розвиток країни з урахуванням різних аспектів (Додаток Г).

Перша група факторів – «Цифрова компетентність та бізнес-інновації», охоплює показники, що стосуються людського капіталу, цифрових навичок, інтеграції цифрових технологій, е-урядування, а також застосування цифрових рішень у бізнесі.

Друга група – «Цифрова інфраструктура та підключення», включає показники, пов'язані з інтернет-зв'язком, широкопasmовим доступом, індексом цін на підключення, а також відкритими даними від уряду.

Третя група – «Широкопasmове покриття та проникнення», охоплює показники, що описують покриття та використання фіксованого широкопasmового інтернету, включаючи швидкісні та ультрашвидкісні технології.

Четверта група – «Кадри та спеціалізація у сфері ІКТ», включає показники, пов'язані з випускниками ІКТ-спеціальностей, їх кількістю та вкладом в е-комерцію.

Результати факторного аналізу та формування чотирьох груп факторів дозволяють глибше зрозуміти ключові аспекти цифрового розвитку країни та їх взаємозв'язок. Визначені групи факторів забезпечують структуру для оцінки сильних і слабких сторін у сфері цифрової трансформації, таких як компетентність людського капіталу, якість цифрової інфраструктури, проникнення широкопasmового інтернету та рівень спеціалізації в ІКТ. Це, у свою чергу, створює базу для розробки ефективних стратегій і політик, спрямованих на вдосконалення цифрової екосистеми країни, оптимізацію ресурсів і підвищення конкурентоспроможності в умовах глобальної цифрової економіки.

Отримані фактори будуть використані для моделювання впливу рівня цифровізації на економічний розвиток 27 європейських країн.

2.4. Характеристика основних методів моделювання

2.4.1. Побудова OLS регресії

Побудова регресії методом OLS (ordinary least squares) є одним із найбільш популярних і фундаментальних методів моделювання в статистиці та машинному навчанні для побудови лінійної регресії. Його основна мета – знайти лінійну залежність між однією або кількома незалежними змінними та залежною змінною. Метод OLS базується на тому, щоб мінімізувати суму квадратів відхилень (помилки) між фактичними значеннями залежної змінної та прогнозованими значеннями моделі:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \varepsilon, \quad (2.1)$$

де Y – залежна змінна (цільова змінна);

β_0 – вільний член;

$\beta_1, \beta_2 \dots \beta_n$ – коефіцієнти регресії для незалежних змінних;

$X_1, X_2 \dots X_n$ – незалежні змінні;

ε – похибка моделі (залишкове відхилення);

n – кількість спостережень.

Серед основних переваг методу OLS – простота і зрозумілість, адже модель легко інтерпретувати, а також висока ефективність при дотриманні припущень. Метод є обчислювально ефективним навіть для великих обсягів даних. Водночас, метод OLS має недоліки: він є чутливим до викидів, що можуть суттєво вплинути на модель, вимагає лінійності залежності, і може страждати від проблем

мультиколінеарності, коли змінні сильно корельовані між собою, що погіршує точність оцінок.

2.4.2. Побудова Ridge та Lasso регресій

Ridge та Lasso регресії – це методи регуляризації, які застосовуються для покращення лінійних моделей у випадках, коли дані мають високий рівень мультиколінеарності, великий обсяг ознак або присутні шуми. Обидва методи базуються на тому ж принципі, що й лінійна регресія, але додають до функції втрат штрафні члени для зменшення складності моделі та уникнення перенавчання.

Ridge регресія – це метод, який зменшує величину коефіцієнтів шляхом додавання штрафного члена, пропорційного сумі квадратів коефіцієнтів, до функції втрат. Серед переваг методу є те, що він ефективно працює в умовах мультиколінеарності, зменшує варіацію моделі, що призводить до кращого узагальнення, а також не занулює коефіцієнти, тому всі ознаки залишаються в моделі:

$$J(\omega, b) = \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \lambda \|\omega\|^2, \quad (2.2)$$

де $J(\omega, b)$ – функція вартості, яку потрібно мінімізувати;

y_i – фактичне значення залежної змінної для i -го спостереження;

$\hat{y}_i = \omega x_i + b$ – прогнозоване значення;

ω – вектор коефіцієнтів для незалежних змінних;

b – зміщення (bias);

$\|\omega\|^2$ – квадрат L2-норми вектора коефіцієнтів ω ;

λ – гіперпараметр, який визначає силу регуляризації.

Lasso регресія – це метод, що додає до функції втрат штрафний член, пропорційний сумі модулів коефіцієнтів. Це дозволяє моделі автоматично обирати значущі ознаки, зануляючи деякі коефіцієнти:

$$J(\omega, b) = \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \lambda \|\omega\|_1, \quad (2.3)$$

де $J(\omega, b)$ – функція вартості, яку потрібно мінімізувати;

y_i – фактичне значення залежної змінної для i -го спостереження;

$\hat{y}_i = \omega x_i + b$ – прогнозоване значення;

ω – вектор коефіцієнтів для незалежних змінних;

b – зміщення (bias);

$\|\omega\|_1$ – L1-норма вектора коефіцієнтів ω ;

λ – гіперпараметр, який визначає силу регуляризації.

Lasso регресія має тенденцію зануляти деякі коефіцієнти повністю, що фактично виконує відбір ознак. Це робить Lasso підходящим для задач, де потрібно визначити найважливіші предиктори серед великої кількості ознак.

Ridge та Lasso регресії є потужними методами регуляризації, які допомагають вирішувати проблеми перенавчання та мультиколінеарності, особливо при роботі з великими обсягами даних або високорозмірними ознаковими просторами. Ridge зменшує величину коефіцієнтів, але залишає всі ознаки в моделі, тоді як Lasso виконує відбір ознак, зануляючи деякі з них.

2.4.3. Побудова Random Forest

Використання методу Random Forest дозволяє будувати надійні моделі, що здатні передбачати числові значення з високою точністю, навіть у складних ситуаціях, де інші моделі можуть давати слабші результати.

Модель будується на основі ансамблю дерев рішень, коли усереднюються прогнози всіх дерев для отримання кінцевого результату. Це дозволяє зменшити дисперсію прогнозів та підвищити стабільність і точність моделі:

$$\hat{y} = \frac{1}{m} \sum_{i=1}^m h_i(x), \quad (2.4)$$

де $h_i(x)$ – передбачення i -го дерева для вхідного вектора x ;

m – кількість дерев.

Для кожного дерева в ансамблі Random Forest генерує випадкову підмножину вихідних даних за допомогою бутстреп-вибірки (випадкова вибірка з заміною). Деякі зразки можуть бути використані кілька разів, а деякі – не використовуватись взагалі.

Кожне дерево навчається на своїй бутстреп-вибірці. На кожному вузлі дерева модель випадковим чином обирає підмножину ознак для розбиття. Це допомагає уникнути перенавчання та зменшити кореляцію між деревами. Древа будуються до певної глибини або до виконання критеріїв зупинки (наприклад, мінімальна кількість зразків у вузлі).

Після побудови всіх дерев модель робить прогноз для нових даних шляхом усереднення значень прогнозів від усіх дерев у ансамблі. Це забезпечує більш стійкий і точний результат, ніж використання одного дерева рішень.

Метод забезпечує високу точність навіть для складних і нелінійних залежностей у даних, при цьому є досить гнучким, оскільки може працювати з даними, що містять пропущені значення або шум, завдяки бутстреп-вибірці та випадковому вибору ознак.

2.4.4. Побудова XGBoost

XGBoost або eXtreme Gradient Boosting – це потужний ансамблевий метод, який часто використовується для регресійних задач завдяки своїй високій точності, швидкості та ефективності. Цей метод прогнозує числові значення, використовуючи ансамбль дерев рішень.

Модель $f(x)$ представляється як сума прогнозів K дерев:

$$f(x) = \sum_{k=1}^K f_k(x), \quad (2.5)$$

де $f_k(x)$ – прогноз k -го дерева;

K – загальна кількість дерев.

Замість того, щоб створювати одну складну модель, XGBoost будує послідовність дерев, кожне з яких намагається скоригувати помилки, зроблені попередніми деревами. Цей підхід дозволяє моделі досягати високої точності навіть у випадку складних і нелінійних залежностей.

Спочатку модель використовує середнє значення цільової змінної для всіх спостережень як початкове передбачення. На кожному етапі створюється нове дерево, яке намагається виправити похибки попередніх дерев. Глибина дерева та кількість вузлів обираються так, щоб мінімізувати функцію втрат. Кожне нове дерево навчається на залишкових помилках попередніх прогнозів.

XGBoost є одним із найефективніших методів для вирішення регресійних задач завдяки здатності знаходити складні залежності в даних і забезпечувати високу точність прогнозів. Цей метод особливо підходить для великих і складних наборів даних, але потребує уважного налаштування гіперпараметрів для досягнення оптимальних результатів.

2.4.5. Побудова Support Vector Regression

Support Vector Regression – це метод моделювання, який застосовується для вирішення регресійних задач за допомогою ідеї опорних векторів (Support Vector Machines). В Support Vector Regression модель намагається знайти таку функцію, яка мінімізує помилки, при цьому залишаючи значну кількість точок даних у межах заданої толерантності. Він активно використовується в задачах, де важливо, щоб модель мала високу узагальнюючу здатність й могла прогнозувати неперервні значення. У загальному вигляді SVR шукає функцію вигляду:

$$f(x) = \omega^T x + b, \quad (2.6)$$

де $\omega^T x$ – скалярний добуток між вектором ваг ω і вектором ознак (вхідних даних) x ;

b – зміщення (bias).

Ключовою ідеєю Support Vector Regression є встановлення певного діапазону помилки, у межах якого точка вважається “правильно передбаченою” (формула (2.6)). Модель намагається мінімізувати кількість точок, що знаходяться поза цим діапазоном. Якщо точка не виходить за його межі, вона не враховується в процесі оптимізації, тобто модель “ігнорує” її для побудови регресійної функції, оскільки вони не порушують точність передбачень.

$$\min \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^n (\zeta_i + \zeta_i^*), \quad (2.7)$$

де $\|\omega\|^2$ – регуляризаційний член, який намагається мінімізувати складність моделі;

C – гіперпараметр, який контролює баланс між простотою моделі та величиною похибок;

ζ_i, ζ_i^* – змінні, які представляють відхилення від ε -границі (відхилення понад ε).

Як видно з формули (2.7), модель включає окремий параметр, який контролює компроміс між узагальненням та точністю. Високе значення цього параметру призводить до того, що модель намагається максимально точно вгадати усі дані, навіть якщо це спричинить перенавчання, тоді як низьке значення параметру дозволяє моделі бути більш гладкою і менш схильною до перенавчання.

РОЗДІЛ 3. АНАЛІЗ І ВЕРИФІКАЦІЯ МОДЕЛЕЙ ВПЛИВУ ЦИФРОВИХ ТРАНСФОРМАЦІЙ НА ЕКОНОМІКУ КРАЇНИ

3.1. Аналіз моделей та їх якості для ВВП на душу населення

3.1.1. OLS, Ridge та Lasso регресії

Перед реалізацією визначених у другому розділі моделей було проведено перетворення змінної “Country” на фіктивну змінну, що дозволило врахувати панельну структуру даних в процесі моделювання (Рисунок 3.1).

```
df_encoded = pd.get_dummies(merged_df, columns=['Country'], drop_first=True)
```

Рисунок 3.1 – Перетворення змінної “Country” на фіктивну

На початку моделювання було побудовано OLS (Ordinary Least Squares) регресію (Рисунок 3.2). Основна мета даного кроку - це провести початковий аналіз, щоб побачити, чи існує зв'язок між незалежними змінними та залежною змінною ВВП на душу населення. Часовий ряд панелі є дуже маленьким, що не дозволяє здійснити перевірку автокореляції та виявити наявність чи відсутність випадкових та фіксованих ефектів, тому для подальшого моделювання було обрано інші методи машинного навчання.

```
# Додавання константи до моделі  
X = sm.add_constant(X)  
  
# Побудова Pooled OLS моделі  
pooled_ols_model = sm.OLS(y, X).fit()  
  
# Виведення результатів  
print(pooled_ols_model.summary())
```

Рисунок 3.2 – Побудова OLS регресії

Результат побудови OLS регресії представлений на рисунку 3.3.

OLS Regression Results						
Dep. Variable:	GDP per capita		R-squared:	0.439		
Model:	OLS		Adj. R-squared:	0.424		
Method:	Least Squares		F-statistic:	30.67		
Date:	Thu, 05 Dec 2024		Prob (F-statistic):	7.32e-19		
Time:	18:44:59		Log-Likelihood:	-1819.1		
No. Observations:	162		AIC:	3648.		
Df Residuals:	157		BIC:	3664.		
Df Model:	4					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
const	3.664e+04	1454.119	25.196	0.000	3.38e+04	3.95e+04
Factor1	1.354e+04	1493.864	9.065	0.000	1.06e+04	1.65e+04
Factor2	-657.6410	1495.265	-0.440	0.661	-3611.073	2295.791
Factor3	7818.5624	1630.767	4.794	0.000	4597.490	1.1e+04
Factor4	4891.1432	1507.320	3.245	0.001	1913.901	7868.385
Omnibus:	94.597		Durbin-Watson:	2.046		
Prob(Omnibus):	0.000		Jarque-Bera (JB):	458.025		
Skew:	2.234		Prob(JB):	3.48e-100		
Kurtosis:	9.921		Cond. No.	1.13		

Рисунок 3.3 – Результати побудови OLS регресії

Згідно з результатами регресійного аналізу, залежна змінна GDP per capita пояснюється чотирма факторами, з яких три є статистично значущими. R-squared становить 0.439, що свідчить про те, що модель пояснює 43.9% варіації залежної змінної. Хоча це не є дуже високим значенням, модель все ж демонструє помітний рівень пояснювальної здатності. Це очікуваний рівень залежності, який свідчить про те, що на ВВП на душу населення впливають також й інші фактори, окрім факторів цифровізації.

Factor1 (Цифрова компетентність та бізнес-інновації) має найсильніший позитивний вплив на GDP per capita, з коефіцієнтом 1.354e+04, що є статистично значущим ($P < 0.001$).

Factor3 (Широкосмугове покриття та проникнення) також має значний позитивний вплив із коефіцієнтом 7818.56 ($P < 0.001$).

Factor4 (Кадри та спеціалізація у сфері ІКТ) має помірний позитивний вплив із коефіцієнтом 4891.14 і $P = 0.001$, що робить його також статистично значущим.

Factor2 (Цифрова інфраструктура та підключення) є статистично незначущим ($P = 0.661$), що свідчить про відсутність суттєвого впливу цього чинника на GDP per capita в межах даної моделі.

Значення Durbin-Watson (2.046) вказує на відсутність серйозної автокореляції залишків, що є позитивною характеристикою моделі.

Проте тести нормальності (Omnibus, Jarque-Bera) свідчать про те, що залишки не розподілені нормально ($\text{Prob}(\text{Omnibus})$ і $\text{Prob}(\text{JB}) = 0.000$), що може вказувати на певні обмеження моделі. Це може бути наслідком того, що дані мають панельну структуру. Можна тільки припустити, що ряд країн мають або свою модель розвитку, або краще застосувати нелінійні моделі для більш ефективного моделювання.

Оскільки побудова OLS регресії з урахуванням параметру року та країни буде мати завищені оцінки (Рисунок Д.1, Додаток Д), то для врахування панельної структури та позбавлення такої проблеми, як завищення оцінок параметрів регресії, було побудовано Ridge та Lasso регресії. Код їх побудови представлений у додатку Е. Результати моделювання представлені на рисунку 3.4, де також виведено параметри якості регресій, які було розраховано для тестового набору даних.

Ridge Regression Results:		Lasso Regression Results:	
	Coefficient		Coefficient
Intercept	4.682257e+06	Intercept	1.922611e+06
const	0.000000e+00	const	0.000000e+00
Factor1	1.400189e+04	Factor1	1.976384e+03
Factor2	1.126135e+03	Factor2	6.667030e+03
Factor3	7.346449e+03	Factor3	-1.596111e+03
Factor4	4.742838e+03	Factor4	1.075342e+03
Year	-2.299246e+03	Year	-9.253053e+02
Country_BE	-2.853238e+03	Country_BE	-1.749793e+02
Country_BG	1.234871e+02	Country_BG	-4.629052e+04
Country_CY	-5.633721e+03	Country_CY	-1.214192e+04
Country_CZ	-1.354826e+04	Country_CZ	-2.315104e+04
Country_DE	6.144371e+03	Country_DE	7.392827e+02
Country_DK	2.644856e+03	Country_DK	4.337516e+03
Country_EE	-1.659879e+04	Country_EE	-3.618125e+04
Country_EL	-1.540084e+03	Country_EL	-2.614109e+04
Country_ES	-1.160332e+04	Country_ES	-3.023221e+04
Country_FI	-9.004234e+03	Country_FI	-1.055779e+04
Country_FR	6.966907e+03	Country_FR	-1.340121e+04
Country_HR	-1.620024e+04	Country_HR	-3.206915e+04
Country_HU	-8.940598e+03	Country_HU	-3.421802e+04
Country_IE	2.122763e+04	Country_IE	3.521525e+04
Country_IT	5.306548e+03	Country_IT	-1.534329e+04
Country_LT	-6.972154e+03	Country_LT	-3.983948e+04
Country_LU	5.771571e+04	Country_LU	6.245964e+04
Country_LV	-7.334157e+03	Country_LV	-4.384456e+04
Country_MT	-1.987940e+04	Country_MT	-2.013260e+04
Country_NL	-7.842533e+03	Country_NL	2.907729e+03
Country_PL	-1.850193e+03	Country_PL	-3.666412e+04
Country_PT	-9.602492e+03	Country_PT	-3.605614e+04
Country_RO	2.270216e+03	Country_RO	-4.572269e+04
Country_SE	-4.286356e+03	Country_SE	-5.949629e+03
Country_SI	-1.008612e+04	Country_SI	-2.902517e+04
Country_SK	-7.870412e+03	Country_SK	-3.206054e+04
R-squared:	0.9892	R-squared:	0.9892
Adj. R-squared:	0.9857	Adj. R-squared:	0.9857
AIC:	2094.0276	AIC:	2094.0276
BIC:	2188.4014	BIC:	2188.4014

Рисунок 3.4 – Результати Ridge та Lasso регресій для моделювання ВВП на душу населення

Побудовані моделі Ridge та Lasso показують дуже високу апроксимацію даних (значення коефіцієнту детермінації та його скориговане значення наближаються до одиниці). Інформаційні критерії Акайке (AIC) та Баєсівський (BIC) є значно меншими ніж для OLS регресії. Результати прогнозів, побудованих для тестових даних за даними моделями, представлені на рисунках 3.5 та 3.6. Візуалізація прогнозних значень підтверджує, що Lasso регресія за заданим набором вхідних даних формує більш точні прогнози, ніж Ridge.

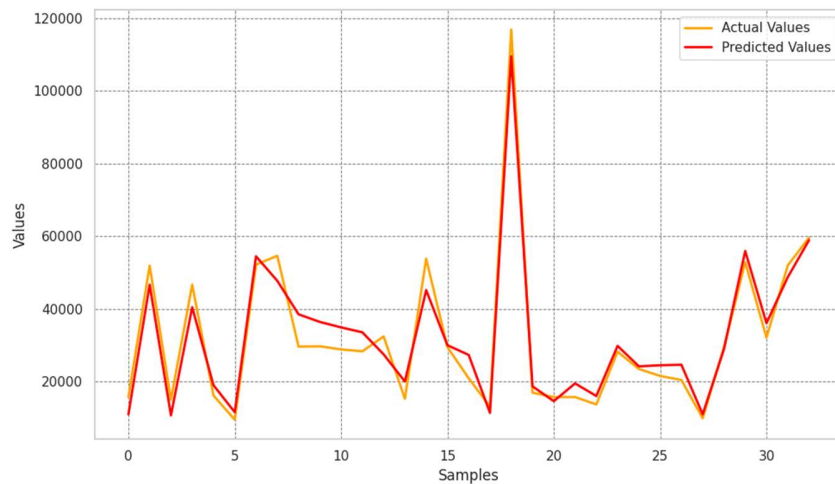


Рисунок 3.5 – Прогнозні значення ВВП на душу населення з урахуванням впливів цифрового розвитку та урахуванням структури панелі (Ridge регресія)

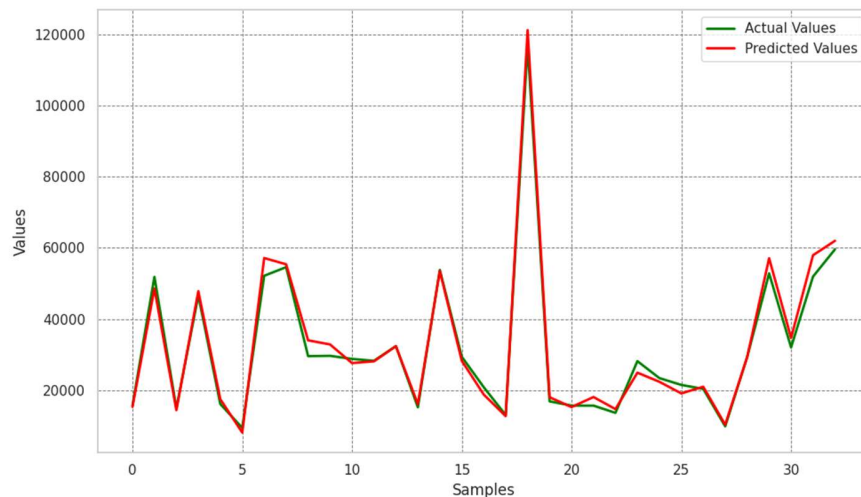


Рисунок 3.6 – Прогнозні значення ВВП на душу населення з урахуванням впливів цифрового розвитку та урахуванням структури панелі (Lasso регресія)

3.1.2. Random Forest

Модель Random Forest була побудована із застосуванням коду, наведеного в додатку Ж. На графіку 3.7, що відображає результати моделювання, видно, що ця модель добре захоплює загальну тенденцію змін ВВП на душу населення під впливом цифрових трансформацій. Хоча модель у цілому відображає загальні

тенденції, відхилення помітні для деяких спостережень і не мають чіткої прив'язки до країн із мінімальним чи максимальним рівнем ВВП. Це може свідчити про те, що Random Forest недостатньо ефективно враховує специфіку панельних даних, таких як взаємозв'язок між часовими рядами та міжгруповими відмінностями.

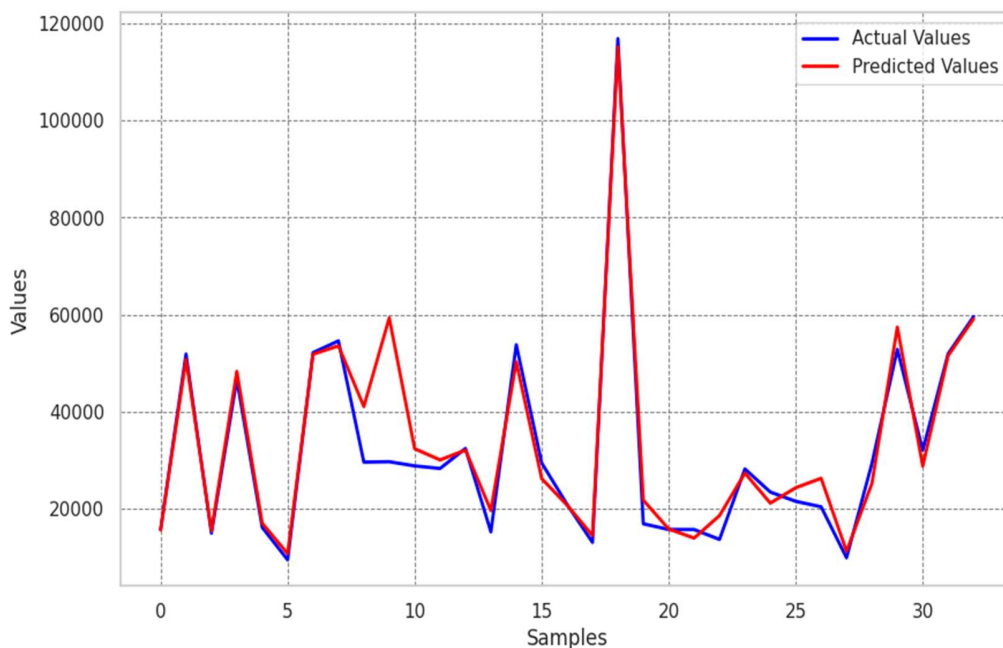


Рисунок 3.7 – Прогнозні значення ВВП на душу населення з урахуванням впливів цифрового розвитку та урахуванням структури панелі (Random Forest)

3.1.3. XGBoost

Модель XGBoost була побудована із застосуванням коду, наведеного в додатку II. Рисунок 3.8 показує значно кращі результати, демонструючи більш точне відображення прогнозних значень у порівнянні зі спостережуваними. Хоча відхилення присутні для окремих спостережень, загалом прогнози є ближчими до фактичних значень, ніж у Random Forest. Це свідчить про те, що XGBoost краще адаптується до динамічних змін і варіативності між об'єктами дослідження, що є ключовою характеристикою панельних даних.

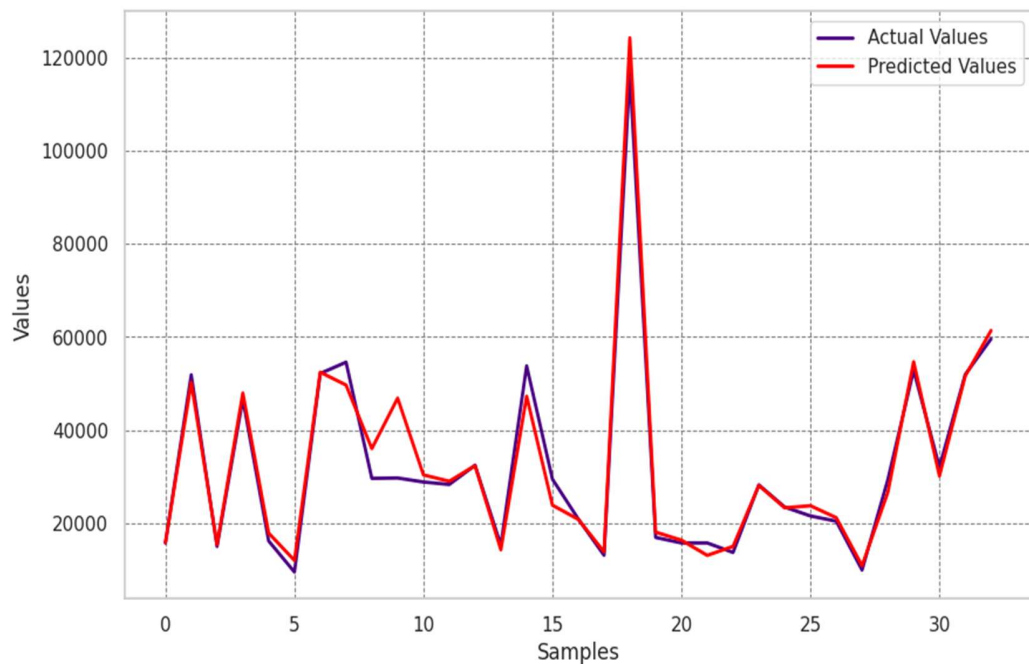


Рисунок 3.8 – Прогнозні значення ВВП на душу населення з урахуванням впливів цифрового розвитку та урахуванням структури панелі (XGBoost)

3.1.4. Support Vector Regression

Модель Support Vector Regression була побудована із застосуванням коду, наведеного в додатку К. Вона показала найкращі результати серед усіх моделей (Рисунок 3.9). Прогнозні значення майже збігаються з фактичними для більшості спостережень, а відхилення є мінімальними й рідкісними. Це вказує на високу здатність SVR моделювати зв'язок між цифровими трансформаціями та ВВП на душу населення, враховуючи як міжгрупові відмінності, так і часову варіативність. Такий результат підтверджує ефективність SVR для панельних даних у цьому контексті.

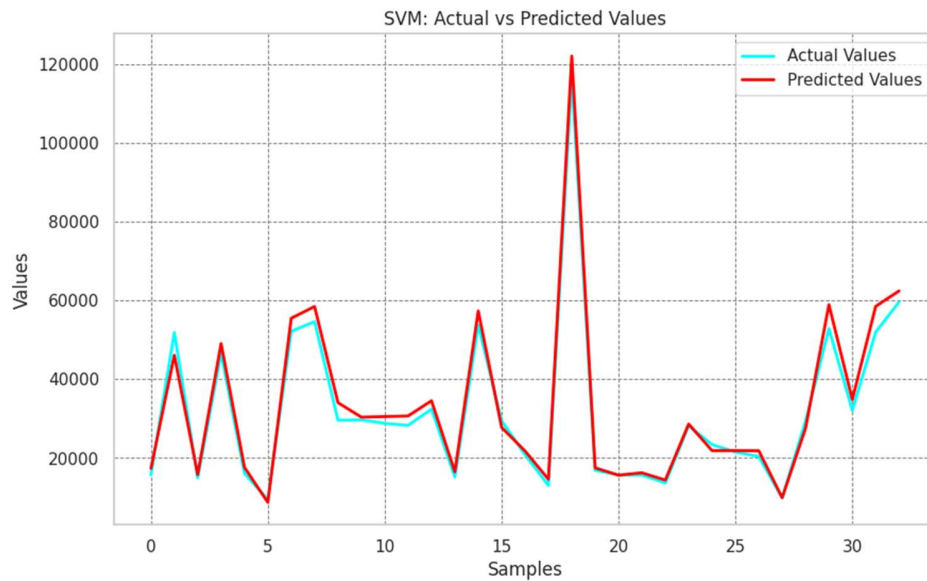


Рисунок 3.9 – Прогнозні значення ВВП на душу населення з урахуванням впливів цифрового розвитку та урахуванням структури панелі (Support Vector Regression)

3.1.5. Оцінка якості прогнозів

В таблиці 3.1 представлені результати оцінок якості прогнозів для п'яти моделей. Орієнтуючись на значення MAPE для тестових даних, можна сказати, що Lasso регресія, XGBoost та SVR дають прогнози високої якості, оскільки похибка менше 10%. Ridge регресія та Random Forest дозволили отримати прогнози гарної якості (від 10% до 20%), що також робить придатним обидві моделі для прогнозування. Аналізуючи значення R^2 , то можна сказати, що всі моделі показують високий рівень апроксимації. Якщо порівнювати значення MSE, MAE та MAD, то найнижчі відповідають моделі Lasso, хоча SVR також не далеко від неї пішла.

Таблиця 3.1 - Оцінка якості прогнозів побудованих моделей

Критерій якості	Ridge	Lasso	Random Forest	XGBoost	SVR
R ² Score	0.9892	0.9892	0.917	0.963	0.983
MSE	20372149.3937	5784563.728	37326913.805	16446042.020	7751732.741
MAE	3844.7202	1813.4772	3201.903	2383.194	2163.956
MAPE	14.7217%	6.1795%	12.961%	8.417%	6.577%
MAD	3756.1761	1174.1928	3201.903	2383.194	2163.956

Таким чином, для моделювання впливів цифрових трансформацій на ВВП на душу населення за результатами аналізу їх оцінок якості краще за все обрати моделі Lasso регресію та SVR, оскільки вони показали найкращі спроможності відобразити залежності та зробити прогноз.

3.2. Аналіз моделей та їх якості для очікуваної тривалості життя

3.2.1. OLS, Ridge та Lasso регресії

В процесі моделювання очікуваної тривалості життя при народженні було виконано аналогічні кроки, які були передбачені запропонованою методикою. В результаті було побудовано OLS регресію (Рисунок 3.10), результати якої свідчать про те, що між незалежними та залежною ознакою існує тісний зв'язок, оскільки значення коефіцієнту детермінації та його скоригованого значення близькі до 1.

Тобто можна сказати, що зміна тривалості життя при народженні буде пояснюватися цифровими трансформаціями, які відбуваються в різних країнах.

OLS Regression Results						
Dep. Variable:	Life expectancy	R-squared:	0.572			
Model:	OLS	Adj. R-squared:	0.561			
Method:	Least Squares	F-statistic:	52.46			
Date:	Thu, 05 Dec 2024	Prob (F-statistic):	5.35e-28			
Time:	18:46:48	Log-Likelihood:	-338.35			
No. Observations:	162	AIC:	686.7			
Df Residuals:	157	BIC:	702.1			
Df Model:	4					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
const	79.8226	0.156	511.967	0.000	79.515	80.131
Factor1	1.9240	0.160	12.012	0.000	1.608	2.240
Factor2	-0.9727	0.160	-6.067	0.000	-1.289	-0.656
Factor3	0.7693	0.175	4.400	0.000	0.424	1.115
Factor4	-0.2938	0.162	-1.818	0.071	-0.613	0.025
Omnibus:	10.059	Durbin-Watson:	2.346			
Prob(Omnibus):	0.007	Jarque-Bera (JB):	4.437			
Skew:	-0.114	Prob(JB):	0.109			
Kurtosis:	2.222	Cond. No.	1.13			

Рисунок 3.10 – Результати побудови OLS регресії

Згідно з результатами регресійного аналізу, значення коефіцієнта детермінації (R-squared) становить 0.572, що означає, що 57.2% варіації залежної змінної (тривалості життя) пояснюється включеними факторами. Скориговане значення R-squared становить 0.561, що враховує кількість предикторів і розмір вибірки, підтверджуючи відносно хорошу пояснювальну силу моделі. Значення p-value для F-тесту (Prob (F-statistic)) є 5.35e-28, що вказує на загальну статистичну значущість моделі.

Коефіцієнт для Factor1 «Цифрова компетентність та бізнес-інновації» становить 1.924, $p\text{-value} = 0.000$. Це вказує на позитивний і статистично значущий вплив цього фактора на тривалість життя.

Коефіцієнт для Factor2 «Цифрова інфраструктура та підключення» дорівнює -0.973, $p\text{-value} = 0.000$. Це свідчить про негативний і статистично значущий вплив, що може вказувати на структурні проблеми в інфраструктурі, які впливають на якість життя.

Коефіцієнт для Factor3 «Широкосмугове покриття та проникнення»: Коефіцієнт становить 0.769, $p\text{-value} = 0.000$, що демонструє позитивний і статистично значущий вплив, підкреслюючи важливість доступу до швидкісного інтернету для підвищення якості життя.

Коефіцієнт для Factor4 «Кадри та спеціалізація у сфері ІКТ» дорівнює -0.294, $p\text{-value} = 0.071$, що означає, що вплив цього фактора негативний, але статистично незначущий (на рівні значущості 0.05), можливо, через недостатній розвиток ІКТ-кадрів або інші взаємопов'язані фактори.

Тест на нормальність (Omnibus $p\text{-value} = 0.007$) свідчить про певні відхилення від нормальності залишків, що може вказувати на певні обмеження моделі. Durbin-Watson = 2.346 вказує на відсутність серйозної автокореляції в залишках.

Оскільки побудова OLS регресії з урахуванням параметру року та країни буде мати завищені оцінки (Рисунок Д.2, Додаток Д), то для врахування панельної структури та позбавлення такої проблеми, як завищення оцінок параметрів регресії, було побудовано Ridge та Lasso регресії.

Результати даного моделювання представлені на рисунку 3.11, де також виведено інформаційні параметри AIC та BIC, за якими можна побачити, що обидві моделі є рівнозначними. Але оцінки параметрів для Lasso регресії є наближеними до нуля, що свідчить про її властивість оцінювання даних за типом експоненційного згладжування.

Ridge Regression Results:		Lasso Regression Results:	
	Coefficient		Coefficient
Intercept	658.548814	Intercept	79.890786
const	0.000000	const	0.000000
Factor1	1.860608	Factor1	0.727133
Factor2	-0.582076	Factor2	-0.000000
Factor3	0.792211	Factor3	0.000000
Factor4	-0.108961	Factor4	-0.000000
Year	-0.286513	Year	-0.000000
Country_BE	-1.021758	Country_BE	0.000000
Country_BG	-1.966782	Country_BG	-0.000000
Country_CY	0.075843	Country_CY	0.000000
Country_CZ	-1.219817	Country_CZ	-0.000000
Country_DE	0.595228	Country_DE	0.000000
Country_DK	-0.630276	Country_DK	0.000000
Country_EE	-0.717442	Country_EE	-0.000000
Country_EL	1.797827	Country_EL	0.000000
Country_ES	2.505670	Country_ES	0.000000
Country_FI	-1.024758	Country_FI	0.000000
Country_FR	2.242881	Country_FR	0.000000
Country_HR	-2.002596	Country_HR	-0.000000
Country_HU	-2.035271	Country_HU	-0.000000
Country_IE	0.766155	Country_IE	0.000000
Country_IT	2.494384	Country_IT	0.000000
Country_LT	-2.546824	Country_LT	-0.000000
Country_LU	1.375886	Country_LU	0.000000
Country_LV	-2.242664	Country_LV	-0.000000
Country_MT	0.751362	Country_MT	0.000000
Country_NL	-1.945806	Country_NL	0.000000
Country_PL	-0.630944	Country_PL	-0.000000
Country_PT	1.759676	Country_PT	0.000000
Country_RO	0.081045	Country_RO	-0.000000
Country_SE	0.468861	Country_SE	0.000000
Country_SI	0.928766	Country_SI	0.000000
Country_SK	-1.746629	Country_SK	-0.000000
R-squared:	0.9662	R-squared:	0.9662
Adj. R-squared:	0.9553	Adj. R-squared:	0.9553
AIC:	-98.4013	AIC:	-98.4013
BIC:	-4.0275	BIC:	-4.0275

Рисунок 3.11 – Результати Ridge та Lasso регресій для моделювання очікуваної тривалості життя при народженні

Побудовані моделі Ridge та Lasso показують дуже високу апроксимацію даних (значення коефіцієнту детермінації та його скориговане значення наближаються до одиниці). Інформаційні критерії Акайке та Баєсівський є найнижчими ніж для OLS регресії. Результати прогнозів, побудованих для тестових даних за даними моделями, представлені на рисунках 3.12 та 3.13. Візуалізація прогнозних значень підтверджує, що Ridge регресія за заданим

набором вхідних даних формує більш точні прогнози, ніж Lasso, яка демонструє ряд, що відповідає експоненційному згладжуванню. Оскільки ми маємо обмеженість ряду за часом, то це обмеження не дає можливості провести дослідження на наявність або відсутність авторегресійних чи ковзного середнього процесів. Тому в даному випадку віддаємо перевагу Ridge регресії, точність якої буде оцінена далі.

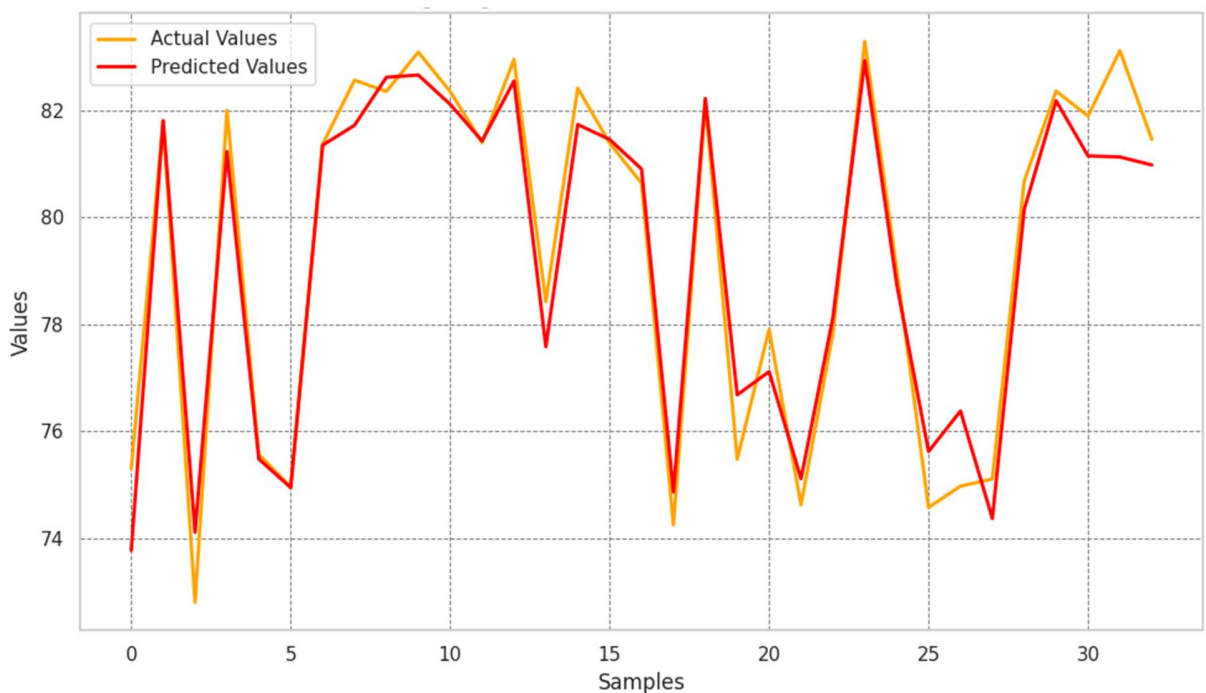


Рисунок 3.12 – Прогнозні значення очікуваної тривалості життя при народженні з урахуванням впливів цифрового розвитку та урахуванням структури панелі (Ridge регресія)

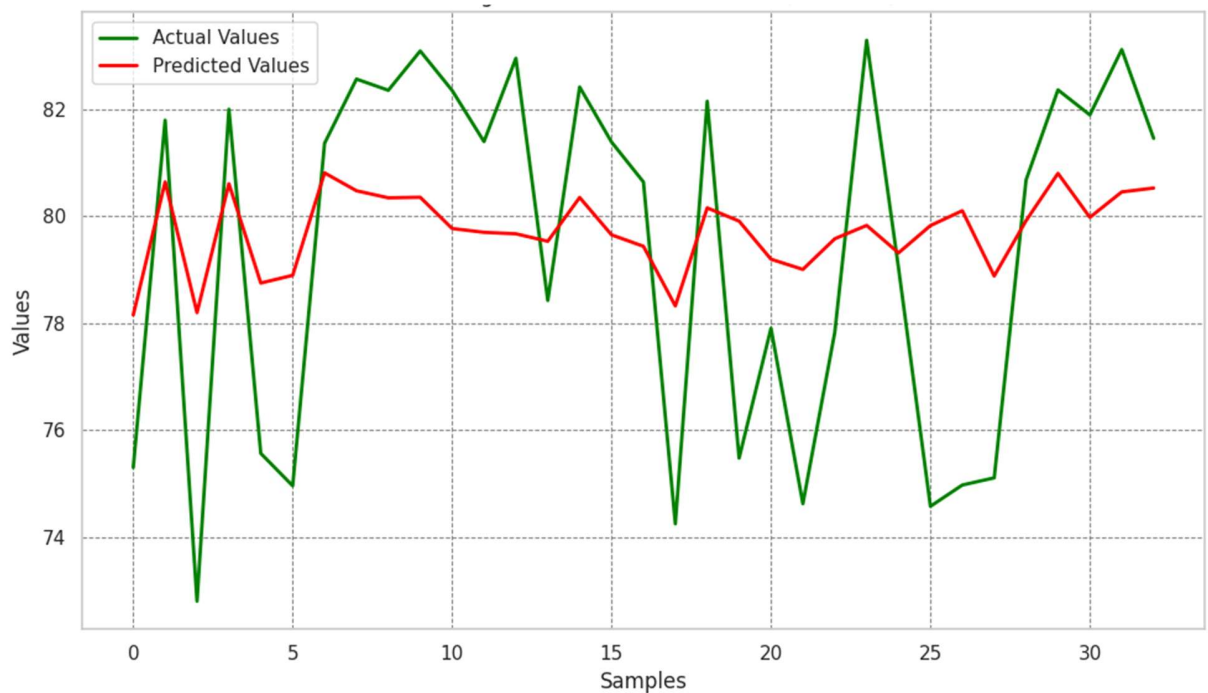


Рисунок 3.13 – Прогнозні значення очікуваної тривалості життя при народженні з урахуванням впливів цифрового розвитку та урахуванням структури панелі (Lasso регресія)

3.2.2. Random Forest

На графіку 3.14 результатів моделювання очікуваної тривалості життя при народженні за допомогою Random Forest видно значні відхилення між фактичними та прогнозними значеннями для деяких спостережень. Особливо помітні розбіжності для показників із високими та низькими значеннями, що свідчить про чутливість моделі до крайніх значень. Це може вказувати на те, що Random Forest не повною мірою враховує складну природу зв'язків між змінними в панельних даних, що впливає на точність прогнозів.

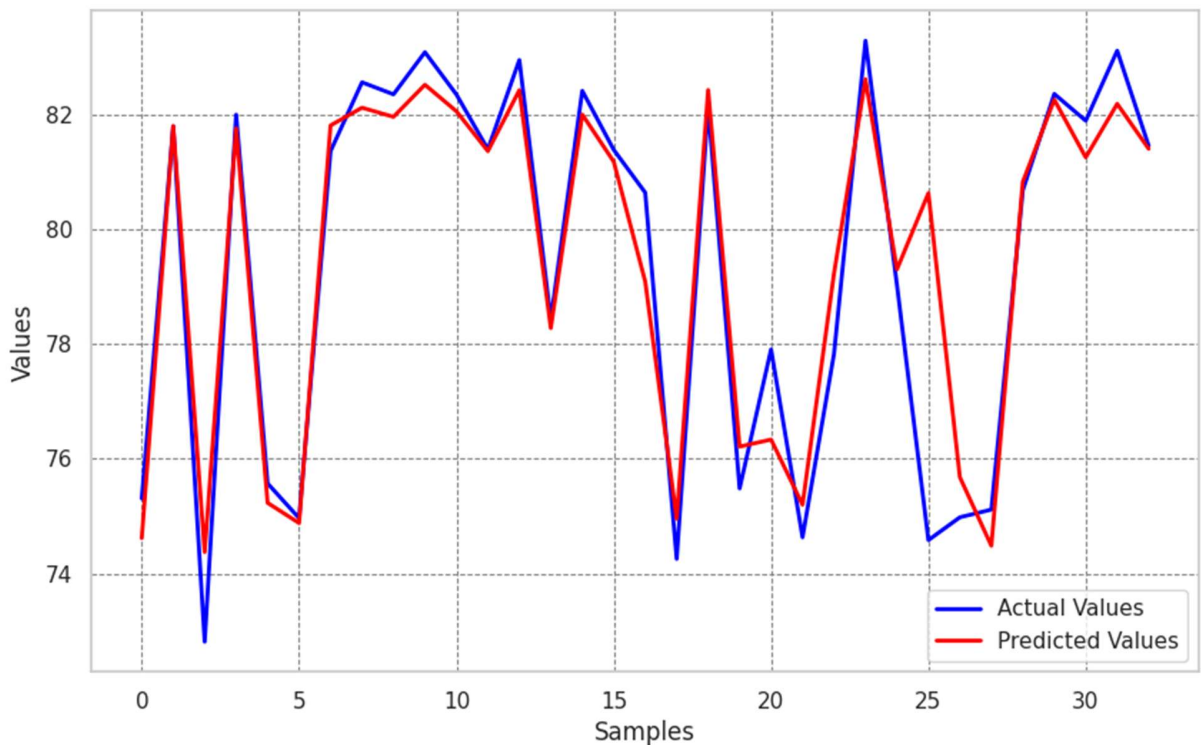


Рисунок 3.14 – Прогнозні значення очікуваної тривалості життя при народженні з урахуванням впливів цифрового розвитку та урахуванням структури панелі (Random Forest)

3.2.3. XGBoost

Графік для XGBoost (Рисунок 3.15) демонструє гіршу відповідність між прогнозами та фактичними значеннями в порівнянні з Random Forest, оскільки спостерігається більше відхилень. Вони також найбільш помітні для показників із крайніми значеннями. В середньому модель гірше прогнозує тривалість життя. Це свідчить про те, що XGBoost все ж таки гірше адаптується до варіативності даних і враховує складні залежності та для екстремальних спостережень її точність все ж обмежена.

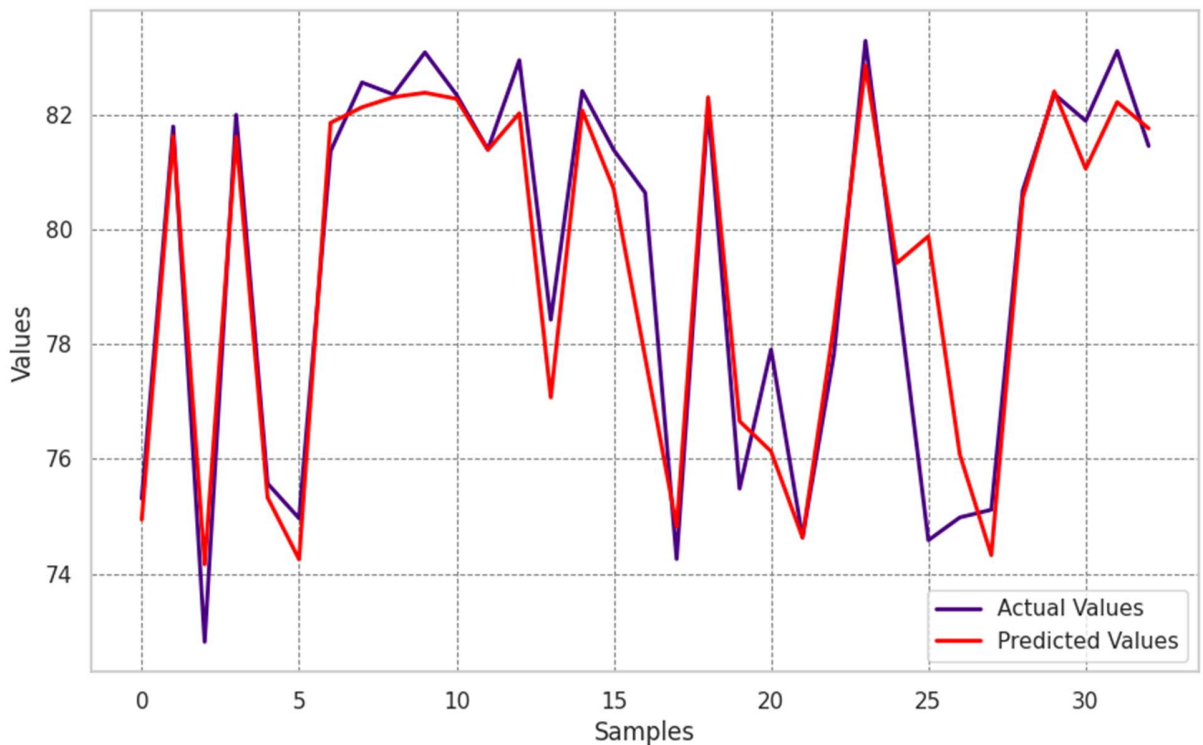


Рисунок 3.15 – Прогнозні значення очікуваної тривалості життя при народженні з урахуванням впливів цифрового розвитку та урахуванням структури панелі (XGBoost)

3.2.4. Support Vector Regression

Support Vector Regression продемонструвала найвищу точність серед моделей (Рисунок 3.16). На графіку відхилення між фактичними та прогнозними значеннями є мінімальними навіть для спостережень з крайніми значеннями тривалості життя. Це свідчить про високу здатність SVR враховувати як внутрішню варіативність панельних даних, так і вплив крайніх значень на залежність між змінними. Модель успішно моделює як загальні тенденції, так і специфіку окремих спостережень.

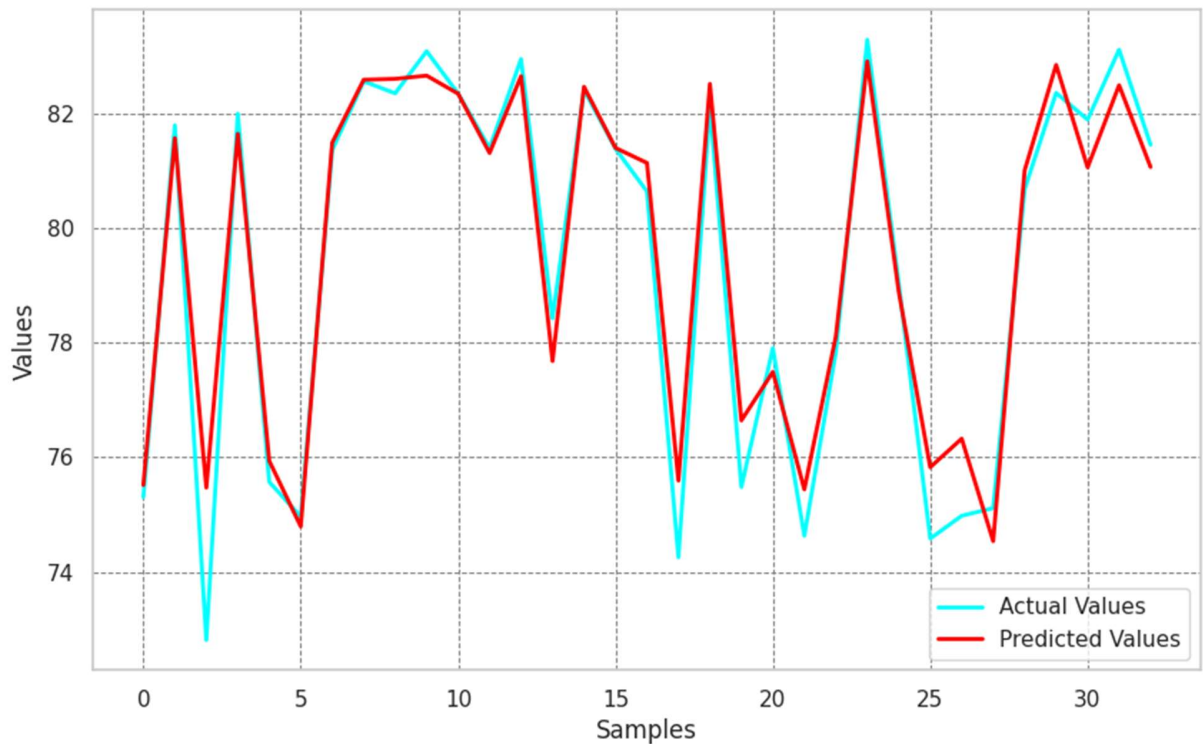


Рисунок 3.16 – Прогнозні значення очікуваної тривалості життя при народженні з урахуванням впливів цифрового розвитку та урахуванням структури панелі (Support Vector Regression)

3.2.5. Оцінка якості прогнозів

В таблиці 3.2 представлені результати оцінок якості прогнозів для очікуваної тривалості життя при народженні за п'ятьма моделями. Орієнтуючись на значення MAPE для тестових даних, можна сказати, що всі моделі дають прогнози високої якості, оскільки похибка менше 10%. Тобто всі моделі можна використовувати для прогнозування очікуваної тривалості життя при народженні під впливом цифрових трансформацій. Значення R^2 демонструє високий рівень апроксимації для моделей Ridge, Lasso та SVR. Порівняння значень MSE, MAE та MAD свідчить, що моделі SVR та Ridge мають найнижчі похибки прогнозів, тому вони є найкращими.

Таблиця 3.2 - Оцінка якості прогнозів побудованих моделей

Критерій якості	Ridge	Lasso	Random Forest	XGBoost	SVR
R ² Score	0.9662	0.9662	0.860	0.860	0.951
MSE	0.5662	8.2061	1.578	1.579	0.553
MAE	0.5755	2.5003	0.708	0.773	0.522
MAPE	0.7351%	3.2059%	0.914%	0.994%	0.677%
MAD	0.4760	2.0625	0.708	0.773	0.522

Таким чином, для моделювання впливів цифрових трансформацій на очікувану тривалість життя при народженні за результатами аналізу їх оцінок якості краще за все обрати моделі Ridge регресію та SVR, оскільки вони показали найкращі спроможності відобразити залежності та зробити прогноз. Але попри їх високу здатність прогнозувати, треба бути обережними для країн з високим та низьким значеннями, що значно відрізняються від середнього.

ВИСНОВКИ

У процесі виконання магістерської роботи на тему «Математичне моделювання впливу цифрової трансформації на економічне зростання країни» було досягнуто поставленої мети – розроблено методичку математичного моделювання для оцінки впливу цифрових трансформацій на окремі показники економічного розвитку країн, зокрема ВВП на душу населення та очікувану тривалість життя.

Перш за все було проаналізовано сутність, значення та ключові аспекти цифрової трансформації у сучасній економіці, визначено основні виклики та бар'єри, пов'язані з впровадженням цифрових технологій, зокрема нерівномірний доступ до інфраструктури, кіберзагрози та низький рівень цифрової грамотності.

Проведено попередній аналіз даних щодо цифрового розвитку 27 європейських країн та використано сучасні статистичні методи (аналіз мультиколінеарності, факторний аналіз), що дозволило структурувати змінні та зменшити їх кількість, зберігаючи основні фактори впливу.

Побудовано кілька типів моделей (OLS-регресія, Ridge, Lasso, Random Forest, XGBoost, Support Vector Regression) для аналізу впливу цифрової трансформації на ВВП на душу населення та очікувану тривалість життя. Визначено, що найбільший вплив на економічний розвиток мають показники цифрової компетентності, бізнес-інновацій та широкосмугового покриття.

Проведено оцінку якості моделей, в результаті було з'ясовано, що при моделюванні впливів цифрових трансформацій на ВВП на душу населення найточніші результати показали моделі Lasso-регресії та Support Vector Regression; на очікувану тривалість життя – моделі Ridge-регресії та Support Vector Regression. Ці моделі дозволяють прогнозувати вплив цифрових трансформацій з мінімальними похибками.

Розроблено рекомендації щодо вдосконалення цифрової політики, зокрема в частині розвитку інфраструктури, підвищення цифрової грамотності та підтримки інновацій, а саме: для прискорення економічного зростання країн рекомендовано зосередитись на розвитку цифрових навичок населення, розширенні доступу до високошвидкісного інтернету та стимулюванні бізнес-інновацій.

Державам, що розвиваються, запропоновано активніше залучати інвестиції в цифрову інфраструктуру, зменшувати цифрову нерівність та підтримувати стартапи у сфері ІТ.

Таким чином, результати роботи підтверджують, що цифрова трансформація є одним із ключових драйверів економічного зростання сучасної економіки. Запропоновані математичні моделі можуть бути використані як інструмент для оцінки впливу цифрових ініціатив та прийняття обґрунтованих управлінських рішень у сфері цифрової політики.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Бондар А.О. Сутність і пріоритетність цифрової трансформації для міжнародного бізнесу [Електронний ресурс] / Бондар А.О. // Актуальні проблеми розвитку економіки регіону. – 2023. – випуск 19. – С. 96-108 Режим доступу: <https://journals.pnu.edu.ua/index.php/aprde/issue/view/326/120>
2. Офіційні сторінки Міністерства цифрової трансформації України – Режим доступу: <https://thedigital.gov.ua/>
3. Якушко І. В. Рейтинговий аналіз розвитку цифрової економіки в Україні – Проблеми економіки, №4 (54), 2022 – Режим електронного доступу https://www.problecon.com/export_pdf/problems-of-economy-2022-4_0-pages-87_93.pdf
4. 6 hard truths CEOs must face. How to leap forward with courage and conviction in the generative AI era [Electronic resource] – Mode of access: <https://www.ibm.com/thought-leadership/institute-business-value/en-us/c-suite-study/ceo>
5. Abhishek Jain Unveiling the power of heatmaps in data visualization [Electronic resource] – Mode of access: <https://medium.com/@abhishekjainindore24/unveiling-the-power-of-heatmaps-in-data-visualization-8b370efca935>
6. Alex J. Smola, Bernhard Scholkopf A tutorial on support vector regression [Electronic resource] – Mode of access: <https://i2pc.es/coss/Docencia/SignalProcessingReviews/Smola2004.pdf>
7. Bagging vs Boosting vs Stacking in Machine Learning [Electronic resource] – Mode of access: <https://medium.com/grabngoinfo/bagging-vs-boosting-vs-stacking-in-machine-learning-65fe4d1684c0>

8. Ceyhun Elgin, Oguz Oztunali Institutions, Informal Economy, and Economic Development [Electronic resource] – Mode of access: <https://www.tandfonline.com/doi/abs/10.2753/REE1540-496X500409>
9. DESI dashboard for the Digital Decade [Electronic resource] – Mode of access: <https://digital-decade-desi.digital-strategy.ec.europa.eu/datasets/desi/indicators>
10. Eoghan Keany What makes «XGBoost» so extreme? [Electronic resource] – Mode of access: <https://medium.com/analytics-vidhya/what-makes-xgboost-so-extreme-e1544a4433bb>
11. How to create a seaborn correlation heatmap in Python? [Electronic resource] – Mode of access: <https://www.geeksforgeeks.org/how-to-create-a-seaborn-correlation-heatmap-in-python/>
12. Jason Brownlee A gentle introduction to the gradient boosting algorithm for machine learning: [Electronic resource] – Mode of access: <https://machinelearningmastery.com/gentle-introduction-gradient-boosting-algorithm-machine-learning/>
13. Julianna Delua Supervised versus unsupervised learning: What's the difference? [Electronic resource] – Mode of access: <https://www.ibm.com/think/topics/supervised-vs-unsupervised-learning>
14. Machine Learning Methods and Applications to Brain Disorders / A. Mechelli, S. Vieira – 2019 by Academic Press – 393p.
15. Maria Webb, Alexandra Pankratyeva 150+ Essential Artificial Intelligence Statistics for 2024: Who's Using AI & How? [Electronic resource] – Mode of access: <https://www.techopedia.com/artificial-intelligence-statistics>
16. Mariette Awad, Rahul Khanna Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers / M. Awad, R. Khanna – 2015 by Apress Media – 263p.
17. Martin Fritz, Max Koch Economic development and prosperity patterns around the world: Structural challenges for a global steady-state economy [Electronic resource] –

Mode of access:
<https://www.sciencedirect.com/science/article/abs/pii/S0959378016300188>

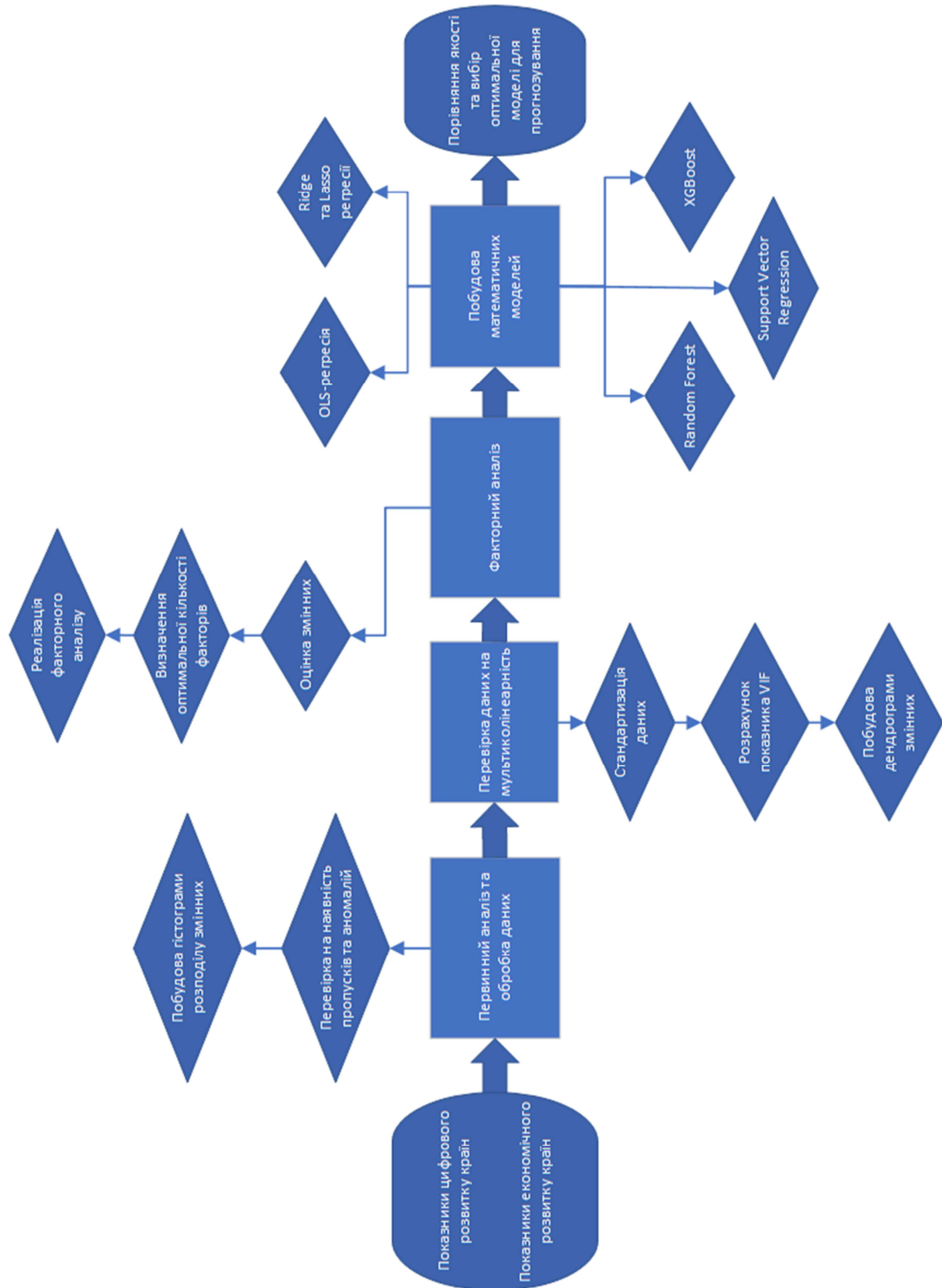
18. Mary K. Pratt What is digital transformation? Ongoing reinvention [Electronic resource] – Mode of access: <https://www.cio.com/article/230425/what-is-digital-transformation-a-necessary-disruption.html>
19. Michel Ferrari, Chandi Fernando Human Growth and Development [Electronic resource] – Mode of access: <https://www.sciencedirect.com/science/article/abs/pii/B0123693985005211?via%3Dihub>
20. Pillars of Resilient Digital Transformation: How CIOs Are Driving Organizational Agility By Harvard Business Review Analytic Services [Electronic resource] – Mode of access: <https://enterpriseproject.com/pillars-resilient-digital-transformation?intcmp=7016000000h0aXAAQ>
21. Random forest regression – how it helps in predictive analytics? [Electronic resource] – Mode of access: <https://medium.com/@byanalytixlabs/random-forest-regression-how-it-helps-in-predictive-analytics-01c31897c1d4>
22. Ridge regression vs lasso regression [Electronic resource] – Mode of access: <https://www.geeksforgeeks.org/ridge-regression-vs-lasso-regression/>
23. Sara Brown Machine learning, explained [Electronic resource] – Mode of access: <https://mitsloan.mit.edu/ideas-made-to-matter/machine-learning-explained>
24. Support vector regression: unleashing the power of non-linear predictive modeling [Electronic resource] – Mode of access: <https://medium.com/@vk.viswa/support-vector-regression-unleashing-the-power-of-non-linear-predictive-modeling-d4495836884>
25. The state of AI in early 2024: Gen AI adoption spikes and starts to generate value [Electronic resource] / McKinsey. – Mode of access: <https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai>
26. Tianqi Chen, Carlos Guestrin XGBoost: a scalable tree boosting system [Electronic resource] – Mode of access: <https://arxiv.org/pdf/1603.02754>

27. Time series analysis (TSA) in Python – linear models to GARCH [Electronic resource] – Mode of access: <https://www.blackarbs.com/blog/time-series-analysis-in-python-linear-models-to-garch/11/1/2016>
28. Understanding ridge regression vs. lasso regression: a mathematical comparison [Electronic resource] – Mode of access: <https://medium.com/@technicalpanchayat18/understanding-ridge-regression-vs-lasso-regression-a-mathematical-comparison-e4bba22816a7>
29. Variance inflation factor (VIF) [Electronic resource] – Mode of access: <https://www.investopedia.com/terms/v/variance-inflation-factor.asp>
30. Wei Zhang, Siqi Zhao Study on the effect of digital economy on high-quality economic development in China [Electronic resource] – Mode of access: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0257365>
31. What is digital transformation? [Electronic resource] – Mode of access: <https://enterprisersproject.com/what-is-digital-transformation>
32. World Development Indicators [Electronic resource] – Mode of access: <https://databank.worldbank.org/source/world-development-indicators>
33. XGBoost Documentation [Electronic resource] – Mode of access: <https://xgboost.readthedocs.io/en/stable/>
34. XGBoost for regression [Electronic resource] – Mode of access: https://www.geeksforgeeks.org/xgboost-for-regression/?ref=header_outind
35. Yarovenko H., Ohol D. (2024). MODELING THE IMPACT OF DIGITAL TRANSFORMATION ON THE ECONOMIC DEVELOPMENT OF THE COUNTRY. In Grundlagen der modernen wissenschaftlichen Forschungder Sammlung wissenschaftlicher Arbeiten «ΛΟΓΟΣ» zu den Materialien der Vinternationalen wissenschaftlich-praktischen Konferenz, Zürich, 13 December, 2024. Zürich-Vinnytsia: BOLESWA Publishers & Europäische Wissenschaftsplattform, 2024. - In press. [Electronic resource] – Mode of access: <https://archive.logos-science.com/index.php/conference-proceedings/issue/view/30>

ДОДАТКИ

Додаток А

Концептуальна схема методики проведених розрахунків



Додаток Б

Відповідник змінних, що використовуються у моделюванні індикаторам DESI

Змінна	Індикатор DESI	Значення індикатору	Од.вимір у
1	2	3	4
Var1	desi_desi_hc	Human capital	%
Var2	desi_desi_conn	Connectivity	%
Var3	desi_desi_idt	Integration of Digital Technology	%
Var4	desi_desi_dps	Digital Public Services	%
Var5	desi_5g_5g_pb	5G spectrum, 5G pioneer bands	%
Var6	desi_5gc_total_pophh	Overall 5G coverage	%
Var7	desi_bbpi_desi_c_bbpi	Broadband price index	score
Var8	desi_c_bbpi_desi_bbpi	Broadband price index	%
Var9	desi_c_fbnc_desi_fbnc	Fixed broadband coverage	%
Var10	desi_c_fbnc_desi_vhnc	Very high capacity network coverage	%
Var11	desi_c_fbnc_desi_fttpc	Fibre to the premises coverage	%
Var12	desi_c_fbbtc_desi_fbbtc	Fixed broadband take-up	%
Var13	desi_c_fbbtc_desi_ufbbtu	Ultrafast broadband take-up	%
Var14	desi_c_fbbtc_desi_gbpstc	Gigabit broadband subscription	%
Var15	desi_c_mbb_desi_5g	5G coverage	%
Var16	desi_c_mbb_desi_5gc	5G connectivity	%
Var17	desi_c_mbb_desi_mbbtu	Mobile Broadband Take-up	%
Var18	desi_conn_desi_c_fbbtc	Fixed broadband take-up	%
Var19	desi_conn_desi_c_fbnc	Fixed broadband coverage	%
Var20	desi_conn_desi_c_mbb	Mobile broadband	%
Var21	desi_conn_desi_c_bbpi	Broadband price index	%
Var22	desi_dps_desi_dps_egov	Digital public services, eGovernment	%
Var23	desi_dps_egov_desi_dps_egovuser	eGovernment users	%
Var24	desi_dps_egov_desi_dps_prefform	Pre-filled forms	%
Var25	desi_dps_egov_desi_dps_pscit	Public services citizen interaction technologies	%
Var26	desi_dps_egov_desi_dps_psbiz	Public services business interaction technologies	%
Var27	desi_dps_egov_desi_dps_opendata	Open data from government	%
Var28	desi_dps_egovuser_ind_total	eGovernment users (last 12 months)	%
Var29	desi_dps_opendata_total	Open data	score
Var30	desi_dps_prefform_all_egov_le	Pre-filled forms (all life events)	score
Var31	desi_dps_psbiz_all_egov_le	Public services business interaction technologies (all life events)	score
Var32	desi_dps_pscit_all_egov_le	Public services citizen interaction technologies (all life events)	score
Var33	desi_fbnc_total_pophh	Fast broadband (NGA) coverage (total)	%
Var34	desi_fbbtc_hh_total	Overall fixed broadband take-up (all households)	%
Var35	desi_fttpc_total_pophh	Fibre to the premises (FTTP) coverage (total)	%
Var36	desi_gbpstc_hh_total	At least 1 Gbps broadband take-up (all households)	%
Var37	desi_hc_desi_hc_bsu	Internet User Skills	%
Var38	desi_hc_desi_hc_asd	Advanced Skills and Development	%

Продовження додатку Б

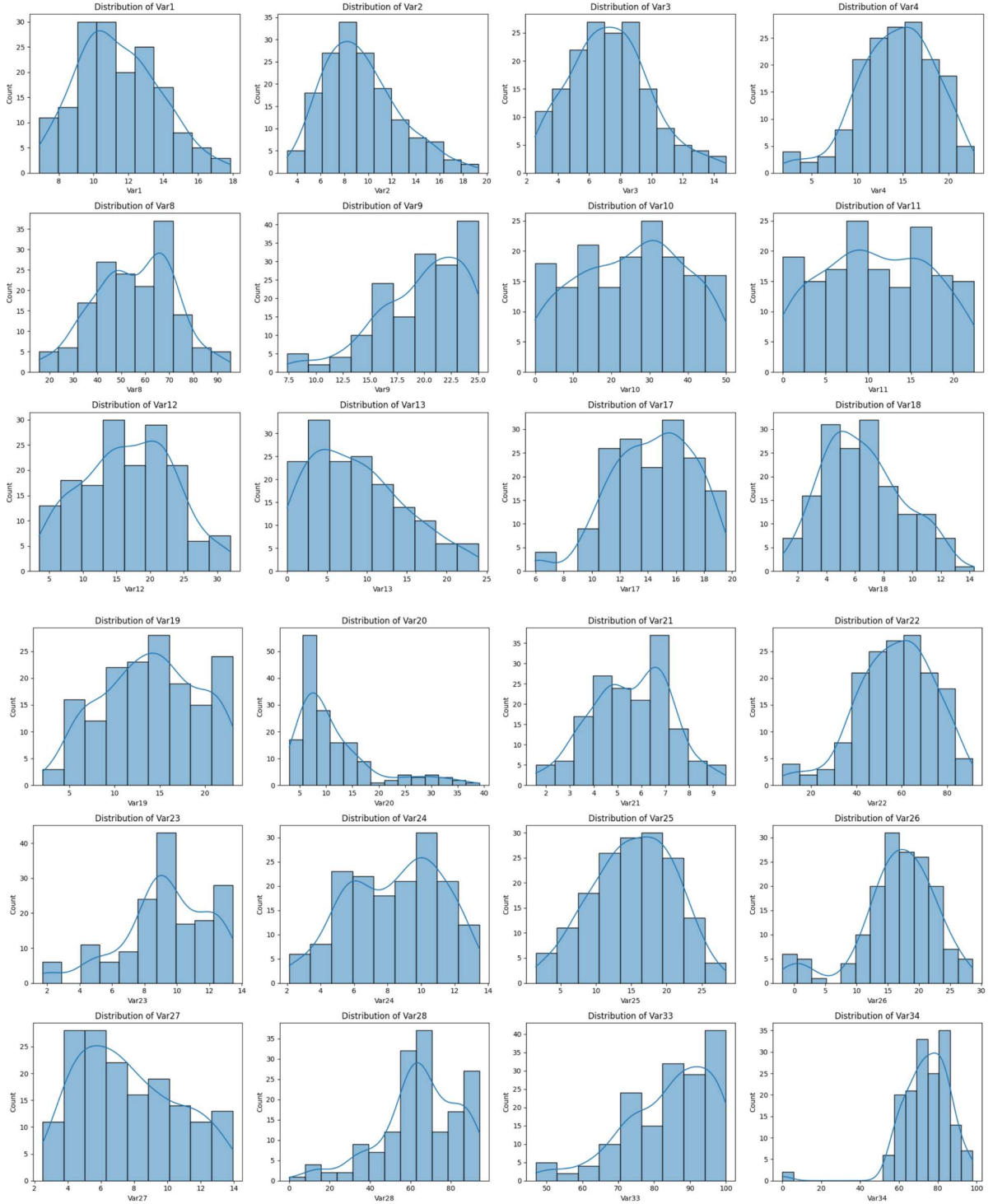
1	2	3	4
Var39	desi hc asd desi hc ictspec	ICT specialists	%
Var40	desi hc asd desi hc fictspect	Female ICT specialists	%
Var41	desi hc asd desi hc entict	Enterprises providing ICT training	%
Var42	desi hc asd desi hc ictg	ICT graduates	%
Var43	desi hc bsu desi hc bds	Internet user skills (at least basic digital skills)	%
Var44	desi hc bsu desi hc abds	Internet user skills (above basic digital skills)	%
Var45	desi hc bsu desi hc abss	Internet user skills (at least basic digital content creation skills)	%
Var46	desi hc entict ent all xfin	Enterprises providing ICT training (all enterprises (10 persons employed or more))	%
Var47	desi hc fictspect total	Female ICT specialists (total)	%
Var48	desi hc ictg total	ICT graduates (total)	%
Var49	desi hc ictspec total	ICT specialists (total)	%
Var50	desi idt desi idt di	Digital intensity	%
Var51	desi idt desi idt dtb	Digital technologies for businesses	%
Var52	desi idt desi idt ecomm	e-Commerce	%
Var53	desi idt ai ent all xfin	Artificial intelligence	%
Var54	desi idt bigdat ent all xfin	Big data	%
Var55	desi idt cloud ent all xfin	Cloud	%
Var56	desi idt di desi idt smedi	SMEs with at least a basic level of digital intensity	%
Var57	desi idt dtb desi idt eis	Digital technologies for businesses (Electronic information sharing)	%
Var58	desi idt dtb desi idt socmed	Digital technologies for businesses (Social media)	%
Var59	desi idt dtb desi idt bigdat	Digital technologies for businesses (Big data)	%
Var60	desi idt dtb desi idt cloud	Digital technologies for businesses (Cloud)	%
Var61	desi idt dtb desi idt ai	Digital technologies for businesses (Artificial intelligence)	%
Var62	desi idt dtb desi idt ictsust	Digital technologies for businesses (ICT for environmental sustainability)	%
Var63	desi idt dtb desi idt einv	Digital technologies for businesses (e-Invoices)	%
Var64	desi idt ecomm desi idt smeso	e-Commerce (SMEs selling online)	%
Var65	desi idt ecomm desi idt ecomturn	e-Commerce (turnover)	%
Var66	desi idt ecomm desi idt sellcb	e-Commerce (Selling online cross-border)	%
Var67	desi idt ecomturn ent sm xfin	e-Commerce turnover (Small and medium enterprises (10-249 persons employed))	%
Var68	desi idt einv ent all xfin	e-Invoices (All enterprises (10 persons employed or more))	%
Var69	desi idt eis ent all xfin	Electronic information sharing (All enterprises (10 persons employed or more))	%
Var70	desi idt ictsust ent all xfin	ICT for environmental sustainability (All enterprises (10 persons employed or more))	%
Var71	desi idt sellcb ent sm xfin	Selling online cross-border (Small and medium enterprises (10-249 persons employed))	%
Var72	desi idt smedi ent sm xfin	SMEs with at least a basic level of digital intensity (Small and medium enterprises (10-249 persons employed))	%

Продовження додатку Б

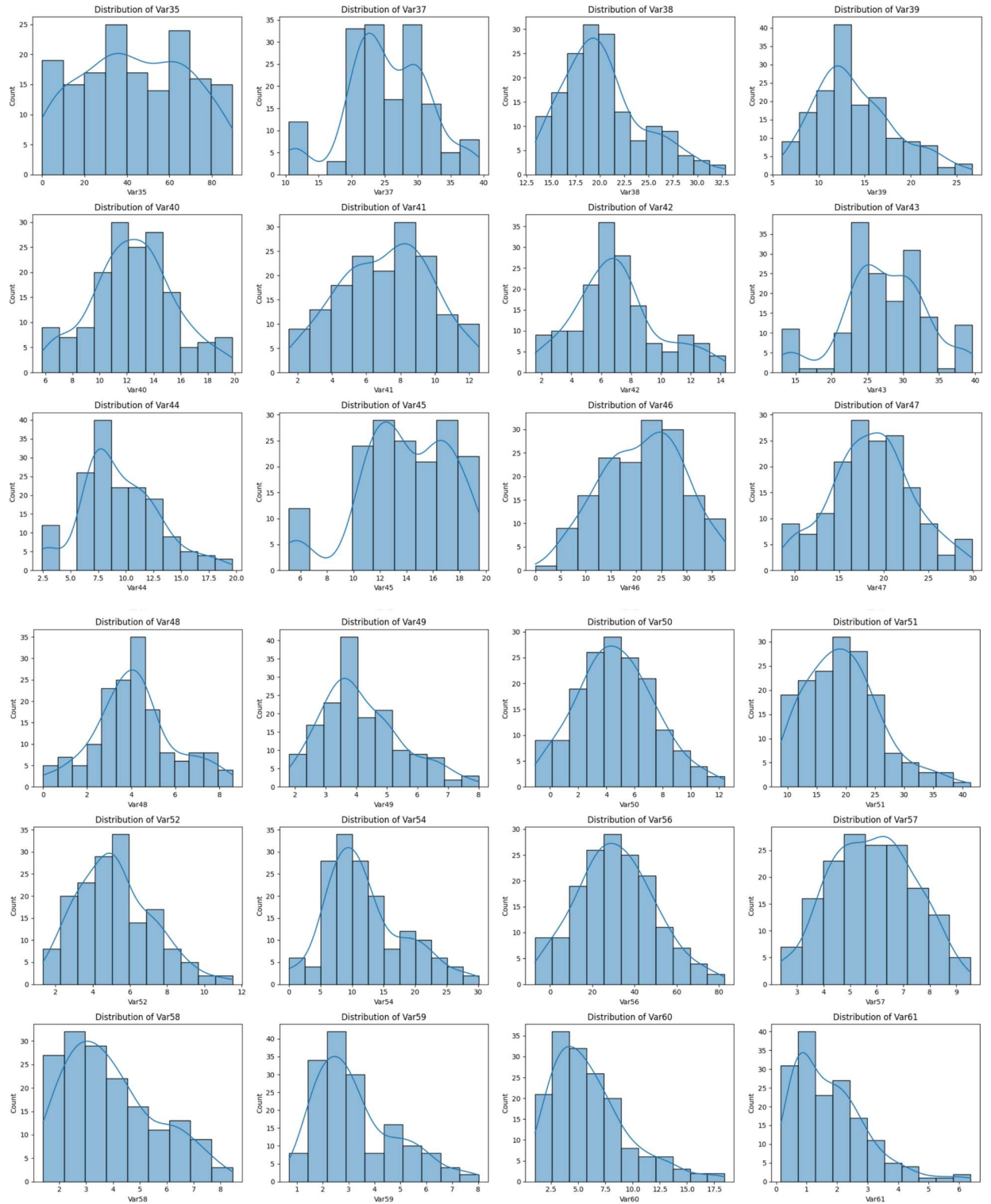
1	2	3	4
Var73	desi_idt_smeso_ent_sm_xfin	SMEs selling online (Small and medium enterprises (10-249 persons employed))	%
Var74	desi_idt_socmed_ent_all_xfin	Social media (Small and medium enterprises (10-249 persons employed))	%
Var75	desi_mbbtu_ind_total	Mobile broadband take-up (all individuals)	%
Var76	desi_sliders_desi_hc	DESI (human capital)	%
Var77	desi_sliders_desi_conn	DESI (connectivity)	%
Var78	desi_sliders_desi_idt	DESI (integration of digital technology)	%
Var79	desi_sliders_desi_dps	DESI (digital public services)	%
Var80	desi_total_desi_total	DESI	%
Var81	desi_ufbbtu_hh_total	At least 100 Mbps fixed broadband take-up (all households)	%
Var82	desi_vhenc_total_pophh	Fixed very high capacity network (VHCN) coverage (total)	%
Var83	desi_hc_abds_ind_total	Above basic digital skills	%
Var84	desi_hc_abss_ind_total	At least basic digital content creation skills	%
Var85	desi_hc_bds_ind_total	At least basic digital skills	%

Додаток В

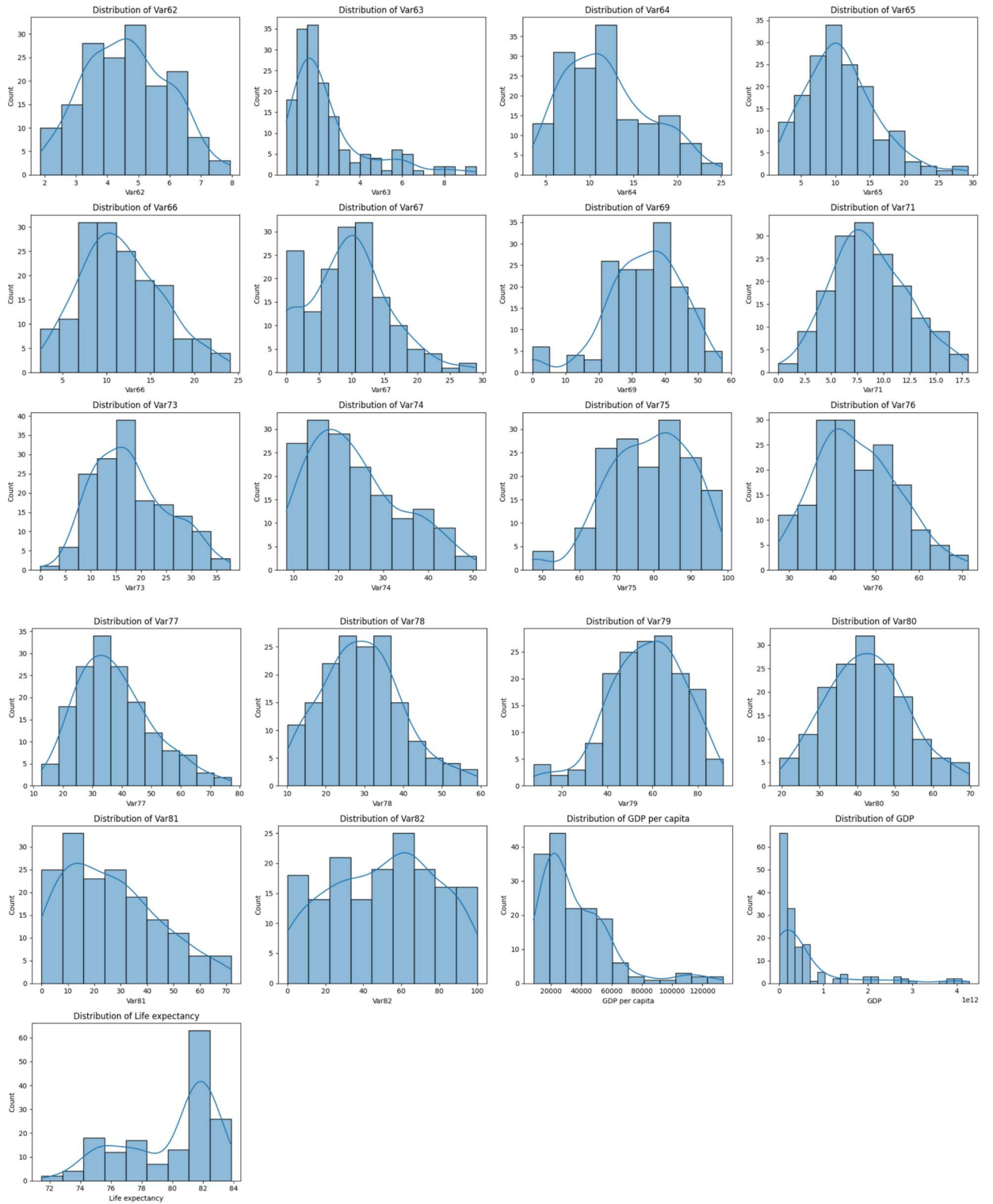
Гістограми розподілу вхідних даних



Продовження додатку В



Продовження додатку В



Додаток Г

Визначені чотири фактори, які були сформовані на основі факторного аналізу
вхідних індикаторів

Цифрова компетентність та бізнес-інновації		Цифрова інфраструктура та підключення	
0	Human capital	1	Connectivity
0	Integration of Digital Technology	1	Broadband price index
0	Digital Public Services	1	Very high capacity network coverage
0	Mobile Broadband Take-up	1	Fibre to the premises coverage
0	Digital public services, eGovernment	1	Ultrafast broadband take-up
0	eGovernment users	1	Fixed broadband take-up
0	Pre-filled forms	1	Fixed broadband coverage
0	Public services citizen interaction technologies	1	Mobile broadband
0	Public services business interaction technologies	1	Broadband price index
0	eGovernment users (last 12 months)	1	Open data from government
0	Internet User Skills	1	Fibre to the premises (FTTP) coverage (total)
0	Advanced Skills and Development	1	Female ICT specialists
0	ICT specialists	1	Female ICT specialists (total)
0	Enterprises providing ICT training	1	Digital technologies for businesses (ICT for environmental sustainability)
0	Internet user skills (at least basic digital skills)	1	DESI (connectivity)
0	Internet user skills (above basic digital skills)	1	At least 100 Mbps fixed broadband take-up (all households)
0	Internet user skills (at least basic digital content creation skills)	1	Fixed very high capacity network (VHCN) coverage (total)
0	Enterprises providing ICT training (all enterprises (10 persons employed or more))		
0	ICT specialists (total)		
0	Digital intensity	2	Fixed broadband coverage
0	Digital technologies for businesses	2	Fixed broadband take-up
0	e-Commerce	2	Fast broadband (NGA) coverage (total)
0	Big data	2	Overall fixed broadband take-up (all households)
0	SMEs with at least a basic level of digital intensity		
0	Digital technologies for businesses (Electronic information sharing)		
0	Digital technologies for businesses (Social media)	3	ICT graduates
0	Digital technologies for businesses (Big data)	3	ICT graduates (total)
0	Digital technologies for businesses (Cloud)	3	e-Commerce turnover (Small and medium enterprises (10-249 persons employed))
0	Digital technologies for businesses (Artificial intelligence)		
0	Digital technologies for businesses (e-Invoices)		
0	e-Commerce (SMEs selling online)		
0	e-Commerce (turnover)		
0	e-Commerce (Selling online cross-border)		
0	Electronic information sharing (All enterprises (10 persons employed or more))		
0	Selling online cross-border (Small and medium enterprises (10-249 persons employed))		
0	SMEs selling online (Small and medium enterprises (10-249 persons employed))		
0	Social media (Small and medium enterprises (10-249 persons employed))		
0	Mobile broadband take-up (all individuals)		
0	DESI (human capital)		
0	DESI (integration of digital technology)		
0	DESI (digital public services)		
0	DESI		

Додаток Д
Результати побудови OLS регресій для панельних даних

OLS Regression Results						
=====						
Dep. Variable:	GDP per capita	R-squared:	0.989			
Model:	OLS	Adj. R-squared:	0.987			
Method:	Least Squares	F-statistic:	391.5			
Date:	Thu, 05 Dec 2024	Prob (F-statistic):	1.08e-113			
Time:	19:37:01	Log-Likelihood:	-1497.6			
No. Observations:	162	AIC:	3059.			
Df Residuals:	130	BIC:	3158.			
Df Model:	31					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975]

const	9.304e+05	1.3e+06	0.717	0.475	-1.64e+06	3.5e+06
Factor1	1902.0763	1734.100	1.097	0.275	-1528.634	5332.786
Factor2	5249.7906	1185.985	4.427	0.000	2903.461	7596.120
Factor3	-1684.1837	1127.056	-1.494	0.138	-3913.930	545.562
Factor4	889.1779	856.682	1.038	0.301	-805.665	2584.021
Year	-434.3445	642.296	-0.676	0.500	-1705.050	836.361
Country_BE	-399.1413	2178.188	-0.183	0.855	-4708.425	3910.142
Country_BG	-4.469e+04	4048.640	-11.039	0.000	-5.27e+04	-3.67e+04
Country_CY	-1.358e+04	2337.972	-5.806	0.000	-1.82e+04	-8949.694
Country_CZ	-2.333e+04	1959.639	-11.903	0.000	-2.72e+04	-1.94e+04
Country_DE	567.3206	2022.071	0.281	0.779	-3433.106	4567.747
Country_DK	5418.5348	3251.095	1.667	0.098	-1013.369	1.19e+04
Country_EE	-3.436e+04	2712.057	-12.671	0.000	-3.97e+04	-2.9e+04
Country_EL	-2.684e+04	2888.967	-9.290	0.000	-3.26e+04	-2.11e+04
Country_ES	-2.824e+04	2581.664	-10.938	0.000	-3.33e+04	-2.31e+04
Country_FI	-9989.9518	3466.403	-2.882	0.005	-1.68e+04	-3132.089
Country_FR	-1.33e+04	2034.116	-6.539	0.000	-1.73e+04	-9276.391
Country_HR	-3.281e+04	1798.248	-18.246	0.000	-3.64e+04	-2.93e+04
Country_HU	-3.343e+04	2758.784	-12.118	0.000	-3.89e+04	-2.8e+04
Country_IE	3.538e+04	3483.349	10.156	0.000	2.85e+04	4.23e+04
Country_IT	-1.563e+04	2489.299	-6.277	0.000	-2.06e+04	-1.07e+04
Country_LT	-3.86e+04	2692.087	-14.339	0.000	-4.39e+04	-3.33e+04
Country_LU	6.363e+04	2855.200	22.285	0.000	5.8e+04	6.93e+04
Country_LV	-4.123e+04	3144.560	-13.111	0.000	-4.74e+04	-3.5e+04
Country_MT	-2.005e+04	2903.701	-6.906	0.000	-2.58e+04	-1.43e+04
Country_NL	2514.7654	3509.756	0.717	0.475	-4428.868	9458.399
Country_PL	-3.601e+04	3182.305	-11.316	0.000	-4.23e+04	-2.97e+04
Country_PT	-3.411e+04	2606.249	-13.087	0.000	-3.93e+04	-2.9e+04
Country_RO	-4.316e+04	4569.554	-9.446	0.000	-5.22e+04	-3.41e+04
Country_SE	-5584.1736	3452.692	-1.617	0.108	-1.24e+04	1246.565
Country_SI	-2.786e+04	2082.758	-13.377	0.000	-3.2e+04	-2.37e+04
Country_SK	-3.164e+04	2219.232	-14.256	0.000	-3.6e+04	-2.72e+04
=====						
Omnibus:	40.845	Durbin-Watson:	1.804			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	373.644			
Skew:	0.516	Prob(JB):	7.32e-82			
Kurtosis:	10.368	Cond. No.	1.19e+07			
=====						

Рисунок Д.1 - Результати побудови OLS регресії для панельних даних «ВВП на душу населення»

OLS Regression Results						
Dep. Variable:	Life expectancy	R-squared:	0.968			
Model:	OLS	Adj. R-squared:	0.960			
Method:	Least Squares	F-statistic:	126.2			
Date:	Thu, 24 Oct 2024	Prob (F-statistic):	1.69e-82			
Time:	16:23:30	Log-Likelihood:	-128.68			
No. Observations:	162	AIC:	321.4			
Df Residuals:	130	BIC:	420.2			
Df Model:	31					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
const	784.1886	277.539	2.826	0.005	235.112	1333.265
Factor1	0.8141	0.371	2.195	0.030	0.080	1.548
Factor2	0.3327	0.254	1.311	0.192	-0.169	0.835
Factor3	0.1853	0.241	0.769	0.443	-0.292	0.662
Factor4	0.0385	0.183	0.210	0.834	-0.324	0.401
Year	-0.3480	0.137	-2.533	0.013	-0.620	-0.076
Country_BE	-0.3721	0.466	-0.799	0.426	-1.294	0.550
Country_BG	-6.3000	0.866	-7.275	0.000	-8.013	-4.587
Country_CY	0.4171	0.500	0.834	0.406	-0.572	1.406
Country_CZ	-2.3310	0.419	-5.561	0.000	-3.160	-1.502
Country_DE	-0.0874	0.433	-0.202	0.840	-0.943	0.768
Country_DK	-1.2641	0.695	-1.818	0.071	-2.640	0.112
Country_EE	-3.5230	0.580	-6.073	0.000	-4.671	-2.375
Country_EL	0.6787	0.618	1.098	0.274	-0.544	1.901
Country_ES	1.2045	0.552	2.181	0.031	0.112	2.297
Country_FI	-1.2367	0.741	-1.668	0.098	-2.704	0.230
Country_FR	1.0606	0.435	2.438	0.016	0.200	1.921
Country_HR	-3.3172	0.385	-8.624	0.000	-4.078	-2.556
Country_HU	-4.9457	0.590	-8.381	0.000	-6.113	-3.778
Country_IE	0.6378	0.745	0.856	0.394	-0.836	2.112
Country_IT	1.8226	0.532	3.423	0.001	0.769	2.876
Country_LT	-6.1980	0.576	-10.764	0.000	-7.337	-5.059
Country_LU	0.3572	0.611	0.585	0.560	-0.851	1.565
Country_LV	-6.6752	0.673	-9.925	0.000	-8.006	-5.345
Country_MT	0.5526	0.621	0.890	0.375	-0.676	1.781
Country_NL	-1.0646	0.751	-1.418	0.159	-2.550	0.421
Country_PL	-3.4767	0.681	-5.108	0.000	-4.823	-2.130
Country_PT	-0.3052	0.557	-0.548	0.585	-1.408	0.798
Country_RO	-5.2487	0.977	-5.370	0.000	-7.182	-3.315
Country_SE	0.0017	0.738	0.002	0.998	-1.459	1.463
Country_SI	-0.5161	0.445	-1.158	0.249	-1.397	0.365
Country_SK	-4.1219	0.475	-8.684	0.000	-5.061	-3.183
Omnibus:	39.535	Durbin-Watson:	1.453			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	78.349			
Skew:	-1.114	Prob(JB):	9.70e-18			
Kurtosis:	5.577	Cond. No.	1.19e+07			

Рисунок Д.2 - Результати побудови OLS регресії для панельних даних
«Очікувана тривалість життя»

Додаток Е

Побудова Ridge та Lasso регресій

```

from sklearn.model_selection import train_test_split
from sklearn.linear_model import Ridge, Lasso
from sklearn.metrics import mean_squared_error

# Розділення даних на тренувальні та тестові
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Додаємо константу для стат. моделі
X_train_with_const = sm.add_constant(X_train)
X_test_with_const = sm.add_constant(X_test)

# Функція для підрахунку AIC і BIC
def calculate_aic_bic(model, X, y):
    residuals = y - model.predict(X)
    mse = mean_squared_error(y, model.predict(X))
    n = len(y)
    aic = n * np.log(mse) + 2 * (X.shape[1] + 1)
    bic = n * np.log(mse) + np.log(n) * (X.shape[1] + 1)
    return aic, bic

# Ridge регресія
ridge_model = Ridge(alpha=1.0)
ridge_model.fit(X_train, y_train)

# Підсумок Ridge
ridge_coef = np.append(ridge_model.intercept_, ridge_model.coef_)
ridge_summary = pd.DataFrame({'Coefficient': ridge_coef}, index=['Intercept'] + list(X_train.columns))
ridge_model_sm = sm.OLS(y_train, X_train_with_const).fit()

ridge_aic, ridge_bic = calculate_aic_bic(ridge_model_sm, X_train_with_const, y_train)
ridge_predictions = ridge_model.predict(X_test)
ridge_mse = mean_squared_error(y_test, ridge_predictions)
ridge_mae = np.mean(np.abs(y_test - ridge_predictions))
ridge_mape = np.mean(np.abs((y_test - ridge_predictions) / y_test)) * 100
ridge_mad = np.median(np.abs(y_test - ridge_predictions))

# Lasso регресія
lasso_model = Lasso(alpha=1.0)
lasso_model.fit(X_train, y_train)

# Підсумок Lasso
lasso_coef = np.append(lasso_model.intercept_, lasso_model.coef_)
lasso_summary = pd.DataFrame({'Coefficient': lasso_coef}, index=['Intercept'] + list(X_train.columns))
lasso_model_sm = sm.OLS(y_train, X_train_with_const).fit()
lasso_aic, lasso_bic = calculate_aic_bic(lasso_model_sm, X_train_with_const, y_train)
lasso_predictions = lasso_model.predict(X_test)
lasso_mse = mean_squared_error(y_test, lasso_predictions)
lasso_mae = np.mean(np.abs(y_test - lasso_predictions))
lasso_mape = np.mean(np.abs((y_test - lasso_predictions) / y_test)) * 100

# Вивід результатів
print("Ridge Regression Results:")
print(ridge_summary)
print(f"R-squared: {ridge_model_sm.rsquared:.4f}")
print(f"Adj. R-squared: {ridge_model_sm.rsquared_adj:.4f}")
print(f"AIC: {ridge_aic:.4f}")
print(f"BIC: {ridge_bic:.4f}")
print(f"Ridge MSE: {ridge_mse:.4f}")
print(f"Ridge MAE: {ridge_mae:.4f}")
print(f"Ridge MAPE: {ridge_mape:.4f}%")
print(f"Ridge MAD: {ridge_mad:.4f}")

print("\nLasso Regression Results:")
print(lasso_summary)
print(f"R-squared: {lasso_model_sm.rsquared:.4f}")
print(f"Adj. R-squared: {lasso_model_sm.rsquared_adj:.4f}")
print(f"AIC: {lasso_aic:.4f}")
print(f"BIC: {lasso_bic:.4f}")
print(f"Lasso MSE: {lasso_mse:.4f}")
print(f"Lasso MAE: {lasso_mae:.4f}")
print(f"Lasso MAPE: {lasso_mape:.4f}%")
print(f"Lasso MAD: {lasso_mad:.4f}")

```

Додаток Ж

Побудова моделі Random Forest

```
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestRegressor
from sklearn.metrics import mean_squared_error, r2_score, mean_absolute_error

# Розділення даних на тренувальні та тестові
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Навчання моделі
model = RandomForestRegressor(n_estimators=100, random_state=42)
model.fit(X_train, y_train)

# Прогнозування
y_pred = model.predict(X_test)

# Оцінка моделі
mse = mean_squared_error(y_test, y_pred)
r2 = r2_score(y_test, y_pred)
mae = mean_absolute_error(y_test, y_pred)
mape = np.mean(np.abs((y_test - y_pred) / y_test)) * 100
mad = np.mean(np.abs(y_test - y_pred))

# Виведення метрик
print(f'R2 Score: {r2:.3f}')
print(f'Mean Squared Error (MSE): {mse:.3f}')
print(f'Mean Absolute Error (MAE): {mae:.3f}')
print(f'Mean Absolute Percentage Error (MAPE): {mape:.3f}%')
print(f'Mean Absolute Deviation (MAD): {mad:.3f}')

# Побудова графіку фактичних та прогнозованих значень
plt.figure(figsize=(10, 6))
plt.plot(y_test.reset_index(drop=True), label='Actual Values', color='blue', linewidth=2)
plt.plot(y_pred, label='Predicted Values', color='red', linewidth=2)
plt.title('Random Forest: Actual vs Predicted')
plt.xlabel('Samples')
plt.ylabel('Values')
plt.legend()
plt.grid(visible=True, color='gray', linestyle='--', linewidth=0.7)
plt.show()
```

Додаток И

Побудова моделі XGBoost

```
from xgboost import XGBRegressor
from sklearn.preprocessing import MinMaxScaler

# Ініціалізація та навчання моделі XGBoost
model = XGBRegressor(n_estimators=500, learning_rate=0.1, max_depth=4, random_state=42)
model.fit(X_train, y_train)
# Прогнозування
y_pred = model.predict(X_test)
# Оцінка моделі
mse = mean_squared_error(y_test, y_pred)
r2 = r2_score(y_test, y_pred)
mae = mean_absolute_error(y_test, y_pred)
mape = np.mean(np.abs((y_test - y_pred) / y_test)) * 100
mad = np.mean(np.abs(y_test - y_pred))

# Виведення метрик
print(f'R2 Score: {r2:.3f}')
print(f'Mean Squared Error (MSE): {mse:.3f}')
print(f'Mean Absolute Error (MAE): {mae:.3f}')
print(f'Mean Absolute Percentage Error (MAPE): {mape:.3f}%')
print(f'Mean Absolute Deviation (MAD): {mad:.3f}')

# Побудова графіку фактичних та прогнозованих значень
plt.figure(figsize=(10, 6))
plt.plot(y_test.reset_index(drop=True), label='Actual Values', color='indigo', linewidth=2)
plt.plot(y_pred, label='Predicted Values', color='red', linewidth=2)
plt.title('XGBoost: Actual vs Predicted Values')
plt.xlabel('Samples')
plt.ylabel('Values')
plt.legend()
plt.grid(visible=True, color='gray', linestyle='--', linewidth=0.7)
plt.show()
```


Додаток К

Побудова Support Vector Regression

```

from sklearn.svm import SVR

#Побудова SVM моделі
scaler_X = StandardScaler()
X_scaled = scaler_X.fit_transform(X)

# Нормалізація цільової змінної (опціонально)
scaler_y = StandardScaler()
y_scaled = scaler_y.fit_transform(y.values.reshape(-1, 1)).flatten()

# Розділення на тренувальні та тестові дані
X_train, X_test, y_train, y_test = train_test_split(X_scaled, y_scaled, test_size=0.2, random_state=42)

# Прогнозування
from sklearn.pipeline import make_pipeline
from sklearn.model_selection import GridSearchCV

pipeline = make_pipeline(StandardScaler(), SVR(kernel='rbf'))
# Параметри для підбору
param_grid = {
    'svr_C': [0.1, 1, 10, 100], # Варіації параметра регуляризації
    'svr_epsilon': [0.01, 0.1, 0.5, 1.0], # Підбираємо епсилон
    'svr_gamma': ['scale', 'auto', 0.01, 0.1, 1] # Підбираємо гамма
}

# Налаштовуємо GridSearchCV
grid_search = GridSearchCV(pipeline, param_grid, cv=5, scoring='neg_mean_squared_error', verbose=1)

# Тренуємо модель на тренувальних даних
grid_search.fit(X_train, y_train)

# Отримуємо найкращі параметри
best_params = grid_search.best_params_
print(f"Best parameters: {best_params}")

# Робимо прогноз на тестових даних
y_pred = grid_search.predict(X_test)

# Оцінка моделі
r2 = r2_score(y_test, y_pred)
mse = mean_squared_error(y_test, y_pred)
mae = mean_absolute_error(y_test, y_pred)
mape = np.mean(np.abs((y_test - y_pred) / y_test)) * 100
mad = np.mean(np.abs(y_test - y_pred))

# Якщо ви нормалізували цільову змінну, можна зробити зворотнє перетворення:
y_test_original = scaler_y.inverse_transform(y_test.reshape(-1, 1)).flatten()
y_pred_original = scaler_y.inverse_transform(y_pred.reshape(-1, 1)).flatten()

# Виведення результатів у вихідній шкалі
r2_original = r2_score(y_test_original, y_pred_original)
mse_original = mean_squared_error(y_test_original, y_pred_original)
mae_original = mean_absolute_error(y_test_original, y_pred_original)
mape_original = np.mean(np.abs((y_test_original - y_pred_original) / y_test_original)) * 100
mad_original = np.mean(np.abs(y_test_original - y_pred_original))

print(f'R² Score (Original Scale): {r2_original:.3f}')
print(f'MSE (Original Scale): {mse_original:.3f}')
print(f'MAE (Original Scale): {mae_original:.3f}')
print(f'MAPE (Original Scale): {mape_original:.3f}%',)
print(f'MAD (Original Scale): {mad_original:.3f}')

```

Продовження додатку К

```
# Візуалізація результатів
plt.figure(figsize=(10, 6))
plt.plot(y_test_original, label='Actual Values', color='aqua', linewidth=2)
plt.plot(y_pred_original, label='Predicted Values', color='red', linewidth=2)
plt.title('SVM: Actual vs Predicted Values')
plt.xlabel('Samples')
plt.ylabel('Values')
plt.legend()
plt.grid(visible=True, color='gray', linestyle='--', linewidth=0.7)
plt.show()
```