




“Data science methods for comprehensive assessment of regional economic development”

AUTHORS	Liubov Chagovets  https://orcid.org/0000-0003-4064-9712 Svitlana Prokopovych  https://orcid.org/0000-0002-6333-2139 Viktor Kholod
ARTICLE INFO	Liubov Chagovets, Svitlana Prokopovych and Viktor Kholod (2020). Data science methods for comprehensive assessment of regional economic development. <i>Development Management</i> , 18(2), 43-56. doi: 10.21511/dm.18(2).2020.05
DOI	http://dx.doi.org/10.21511/dm.18(2).2020.05
RELEASED ON	Tuesday, 29 September 2020
RECEIVED ON	Wednesday, 29 April 2020
ACCEPTED ON	Thursday, 03 September 2020
LICENSE	 This work is licensed under a Creative Commons Attribution 4.0 International License
JOURNAL	"Development Management"
ISSN PRINT	2413-9610
ISSN ONLINE	2663-2365
FOUNDER	Simon Kuznets Kharkiv National University of Economics



NUMBER OF REFERENCES

34



NUMBER OF FIGURES

7



NUMBER OF TABLES

3

Liubov Chagovets (Ukraine), Svitlana Prokopovych (Ukraine),
Viktor Kholod (Ukraine)

DATA SCIENCE METHODS FOR COMPREHENSIVE ASSESSMENT OF REGIONAL ECONOMIC DEVELOPMENT

Abstract

The paper deals with the assessment of the socio-economic development of Ukrainian regions using Data Science methods and multidimensional analysis, including taxonomy, n-dimensional classification, and ensemble decision trees methods. The methodological bases of economic regions development by the economic and mathematical modeling methods were investigated. The necessity of improving and further developing estimation models of the regional economic development using business analytics tools and multidimensional scaling methods was investigated.

The ensemble decision trees methods was applied for the classification model of economic development of the Ukrainian regions according to the conceptual base of the research on regional economic development. It will increase the quality level of administrative decisions making on regional de-velopment asymmetry equalization. It is determined that in Ukraine, there is a significant imbalance of regions clusters with high and low economic level. Here was investigated the relationship between the two groups of economic development indicators – the indicators of the economic development regional performance and the group of economic potential. The results of the classification model allow identifying the set of indicators that have significant impact on the overall economic development. The developed ensemble model allows carrying out qualitative recognition and prediction of the state probability of economic development. It will improve the quality of decision making processes on equalization of regional development asymmetry.

The further research gives the possibility to develop the system of levers directions of regional development imbalance equalization, to determine priority vectors of sustainable development of both the regions and the country.

Keywords

region, model, boosting, bagging, tree ensembles, regional economic development, modeling

JEL Classification

C38, O18, R11

Л. О. Чаговець (Україна), С. В. Прокопович (Україна), В. А. Холод (Україна)

ЗАСТОСУВАННЯ МЕТОДІВ DATA SCIENCE У КОМПЛЕКСНОМУ ОЦІНЮВАННІ ЕКОНОМІЧНОГО РОЗВИТКУ РЕГІОНІВ

Анотація

У роботі розглянуто питання оцінювання економічного розвитку регіонів України методами Data Science та багатовимірного аналізу, зокрема методами таксономії, n-вимірної класифікації та ансамблевих дерев рішень. Досліджено методологічні основи використання інструментарію економіко-математичного моделювання економічного розвитку регіонів. Встановлено доцільність удосконалення та подальший розвиток моделей оцінювання економічного розвитку регіонів на базі сучасного інструментарію дослідження багатовимірних об'єктів.

Проведене дослідження з питань економічного розвитку регіонів дозволило запропонувати концептуальний базис та побудувати модель класифікації регіонів України методами ансамблевих дерев рішень, що дає змогу підвищити якість формування та ухвалення управлінських рішень з вирівнювання асиметричності регіонального розвитку. Визначено, що спостерігається суттєва незбалансованість кластерів регіонів з високим та низьким рівнем економічного розвитку регіонів України. Виявлено тісний зв'язок між двома групи показників економічного розвитку – показниками результативності економічного розвитку регіонів та їх економічним потенціалом. Розроблена модель класифікації дозволяє здійснити якісне розпізнавання та прогнозування ймовірності стану економічного розвитку регіонів, що сприяє підвищенню якості формування та ухвалення управлінських рішень з вирівнювання асиметричності регіонального розвитку. Перспективами подальших досліджень є можливість



S. KUZNETS KHNUe



Founder

Simon Kuznets Kharkiv National
University of Economics, Nauky
avenue, 9-A, Kharkiv, 61166,
Ukraine

<http://www.hneu.edu.ua/>

Received on: 29th of April, 2020

Accepted on: 3rd of September, 2020

Published on: 29th of
September, 2020

© Liubov Chagovets,
Svitlana Prokopovych,
Viktor Kholod, 2020

Liubov Chagovets, Associated
Professor, Faculty of Economic
Informatics, Simon Kuznets
Kharkiv University of Economic,
Ukraine.

Svitlana Prokopovych, Associated
Professor, Faculty of Economic
Informatics, Simon Kuznets
Kharkiv University of Economic,
Ukraine.

Viktor Kholod, Data Scientist,
VisiQuate, Inc., Ukraine.



This is an Open Access article,
distributed under the terms of the
[Creative Commons Attribution 4.0
International license](https://creativecommons.org/licenses/by/4.0/), which permits
unrestricted re-use, distribution,
and reproduction in any medium,
provided the original work is
properly cited.

на базі сформованого комплексу моделей розробки системи важелів вирівнювання дисбалансу регіонального розвитку, визначення пріоритетних векторів стійкого розвитку як окремих регіонів, так і країни в цілому.

Ключові слова регіон, модель, бустинг, беггінг, ансамблі дерев, економічний розвиток регіонів, моделювання

Класифікація JEL C38, O18, R11

ВСТУП

Сучасні умови функціонування та розвитку вітчизняної економіки в цілому та окремих економічних суб'єктів формуються під дією значної кількості чинників, які мають характер невизначеності та стохастичності, а також суперечності апріорної інформації для управлінських рішень. Це дає змогу стверджувати, що основною передумовою досягнення високого рівня розвитку регіону є формування бази ефективних стратегій та комплексу важелів вирівнювання та перерозподілу економічних ресурсів, що стає можливим лише за розробки відповідного статистично адекватного модельного комплексу.

Рівень розвитку регіонів є складним та багатоплановим поняттям, яке об'єднує широкий спектр економічних відносин, пов'язаних з умовами та станом життєдіяльності людини у суспільстві. Рівень соціального-економічного розвитку регіонів впродовж тривалого часу залишається однією з основних економічних категорій, що характеризує не лише матеріальний добробут окремої людини, а й визначає узагальнений результат діяльності економіки країни за певний період. Невизначеність зовнішнього середовища істотно впливає на збільшення незбалансованості розвитку регіонів, зростання втрат в процесі прийняття управлінських рішень, веде до загального кризового економічного, соціального і фінансового стану регіонів. Динамізм розвитку таких процесів визначає особливості формування сценаріїв управління економічним розвитком регіонів та свідчить про важливість і необхідність розробки концептуальних основ і побудови комплексу моделей вирівнювання диспропорцій розвитку територій. Таким чином, процес прийняття рішень під час управління економічним розвитком регіону в умовах неповної інформації має враховувати фактори нерівномірності і асиметричності територіального розвитку.

1. ЛІТЕРАТУРНИЙ ОГЛЯД

Висвітленню проблем розвитку регіонів, оцінювання економічної ситуації у регіонах України, дослідження нерівномірності їх функціонування та розвитку значну увагу приділено у роботах багатьох вітчизняних та закордонних вчених, серед яких: Гребенкін [7], Глушенко [6], Самаріна [24], Клебанова та Гур'янова [13], Гавкалова [8], Романюк [22], Архипенко [2], Гейман [14], Бойко [3] та багатьох інших. У роботах цих авторів розглядається широке коло питань з визначення чинників, показників, що характеризують або впливають на рівень розвитку регіонів та рівень життя населення в них.

Рівень економічного розвитку (ЕР) регіону зазвичай ототожнюється з широким колом синонімічних і близьких за значенням термінів: промислова, інвестиційна привабливість регіону, його геополітичне розміщення тощо; економічний потенціал і ефективність його використання; рівень розвитку технологій; народний добробут, матеріальне становище населення. Разом з цим, поряд зі значними доробками у галузі регіонального розвитку, на сьогодні залишаються дискусійними питання методологічного та методичного характеру зі створення єдиної системи оцінювання показників рівня розвитку, так і залишаються невирішеними проблеми щодо розрахунку окремих індикаторів та побудови моделей розпізнавання та прогнозування стану розвитку регіонів.

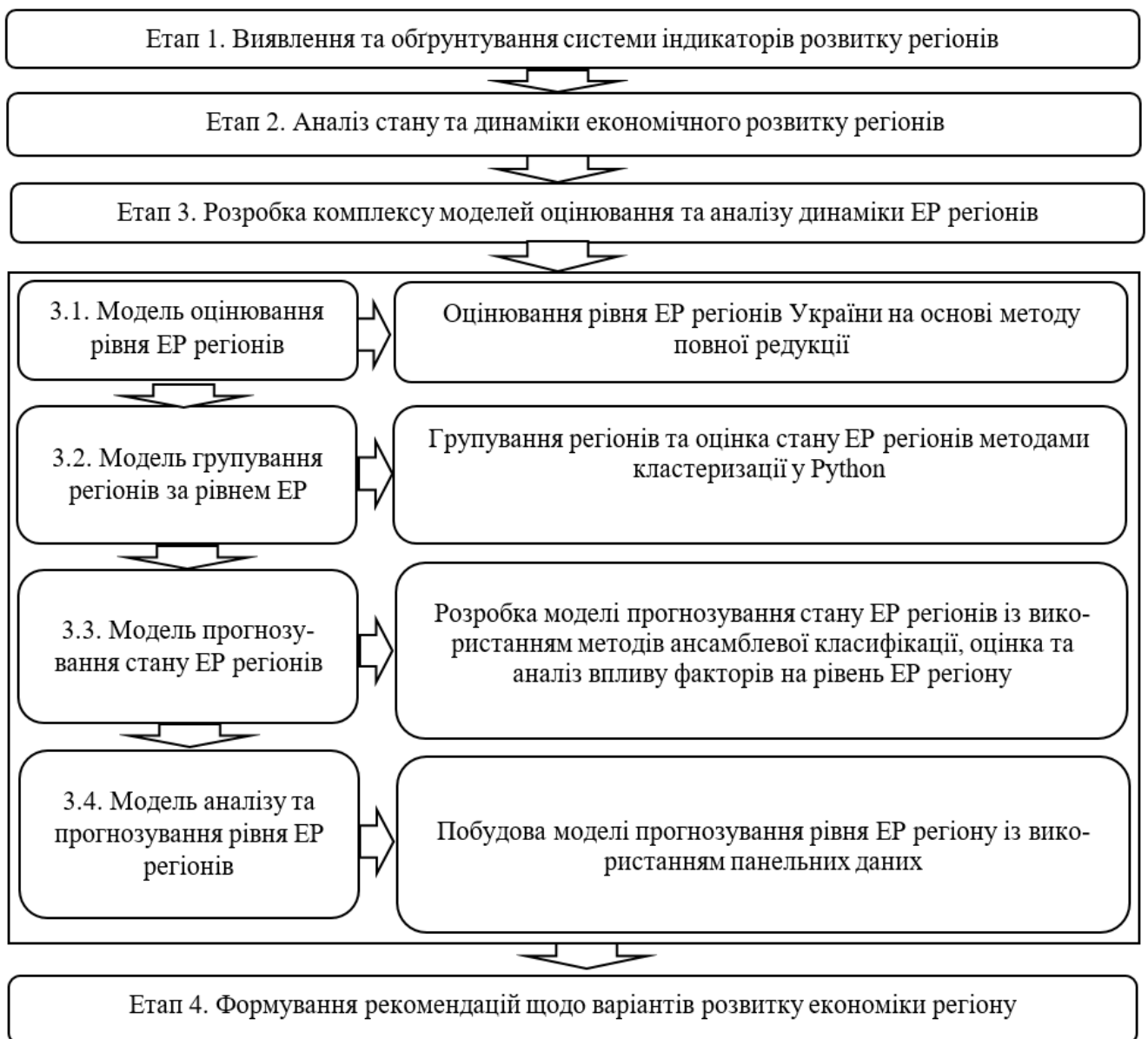
Таким чином, з урахуванням накопиченого позитивного досвіду, стає доцільним удосконалення та подальший розвиток моделей оцінювання економічного розвитку регіонів на базі сучасного інструментарію бізнес-аналітики та методів дослідження багатовимірних об'єктів, що надає можливість покращити якість управлінських рішень з розробки стратегій стійкого регіонального розвитку та формування системи важелів для вирівнювання регіональних диспропорцій.

2. МЕТА ДОСЛІДЖЕННЯ

Метою дослідження є побудова комплексної моделі оцінювання економічного розвитку регіонів України методами Data Science та багатовимірною аналізу, що дозволить підвищити якість формування та ухвалення управлінських рішень з вирівнювання асиметричності регіонального розвитку.

3. МЕТОДИ

З використанням сучасного інструментарію моделювання можна визначити рівень економічного розвитку регіонів України в контексті взаємодії регіонів, що надає можливість подальшого дослідження процесів регіональної конвергенції. Для якісного аналізу рівня економічного розвитку регіонів України та його оцінки, а також для вирішення зазначених проблем була запропонована наступна концептуальна схема моделювання рівня економічного розвитку регіонів (Рисунок 1).



Джерело: Розробка авторів.

Рисунок 1. Концептуальна схема моделювання рівня ЕР регіонів

Під час розробки концептуального базису оцінювання регіонального розвитку було використано методи синтезу та аналізу. Для побудови системи інформаційних індикаторів дослідження було використано спеціалізовані методи Data Science та багатовимірний статистичний аналіз – методи дерев класифікації та повної редукції даних. Зокрема, для розробки моделей класифікації стану економічного розвитку регіонів було використано методи ансамблів дерев класифікації, метод випадкового лісу (Random Forest) з можливістю посилення ансамблю алгоритмами беггінгу та бустингу.

Розглянемо головні етапи концептуальної схеми дослідження та процесу побудови моделей.

На першому етапі формується первісна система індикаторів, що характеризує економічний розвиток регіонів.

На другому етапі проводиться аналіз сучасного стану та динаміки економічного розвитку регіонів за обраною системою індикаторів.

На третьому етапі будується комплекс моделей для оцінювання та аналізу динаміки ЕР регіонів. Виявляється взаємозв'язок між вихідними показниками, що характеризують ЕР, та розраховується таксономічний індекс ЕР на базі методів повної редукції. Досліджується рівень економічного розвитку регіону в динаміці на основі історичних даних. Далі будується модель класифікації рівнів економічного розвитку регіонів та ансамблева модель прогнозування ймовірності приналежності регіону до класу ЕР. Для аналізу впливу вихідних показників на рівень ЕР регіону, визначення індивідуальних особливостей та ефектів ЕР регіонів будується модель на панельних даних. В останньому блоці третього етапу, виконується прогнозування показників-індикаторів оцінки ЕР, що надає можливість отримати прогноз загального рівня та стану ЕР вибраного регіону.

Завершальний етап концептуальної схеми моделювання рівня економічного розвитку регіонів полягає в аналізі та оцінюванні отриманих результатів попередніх етапів для формування комплексних висновків щодо стану та тенденції змін рівня ЕР регіонів України під впливом процесів євроінтеграції та структурних економічних зрушень пов'язаних з кризовою ситуацією, що склалась в Україні.

У рамках даної роботи розглянемо більш детально моделі класифікації. Для розробки даних моделей на базі моделей беггінгу та бустингу використано методи ансамблів дерев класифікації Random Forest. Особливість цих методів полягає у наступному. Класичні алгоритми побудови дерев мають суттєві недоліки для вирішення задач класифікації та розпізнавання [20, 27, 28, 30], оскільки зі зміною вхідних даних, створене дерево рішень стає чутливим до них та змінює конфігурацію. У такому випадку слід створити надійну модель (або зменшити дисперсію помилки) за допомогою процедури пакетування або «мішку» – який передбачає розробку різних моделей шляхом перекомпонування даних, таким чином отримана модель має вищу статистичну значущість та надійність, утворюється ансамбль з дерев.

Як справедливо зазначено у [28] усереднення результатів спостережень може дати більш стійку і надійну оцінку, оскільки послаблюється вплив випадкових флуктуацій в окремому вимірі. За схожим принципом відбувався розвиток алгоритмів комбінування моделей, в результаті чого побудова їх ансамблів виявилась одним з найбільш потужних методів машинного навчання, що нерідко за якістю прогнозів є кращим за інші методи. Одним з рішень, що забезпечують необхідну різноманітність моделей, є їх повторне навчання на вибірках, випадково вибраних з генеральної сукупності, або інших підмножинах даних, що складає базис для побудови ансамблів дерев. Існує два типи моделей на основі ансамблів дерев – це беггінг (англ. Bagging) та бустинг (англ. Boosting). На основі беггінгу працює алгоритм Random Forest – це метаоцінювач, який підходить до ряду класифікаторів дерев рішень на різних підвибірках даних і використовує усереднення для підвищення точності прогнозування та контролю [26]. Розмір вибірки завжди такий самий, як розмір вхідної, але спостереження витягуються із заміною. Тобто за алгоритмом будується багато дерев рішень, розділяються вхідні дані на невеликі вибірки, на яких будує прогноз, а кінцевий результат визначається методом середнього арифметичного.

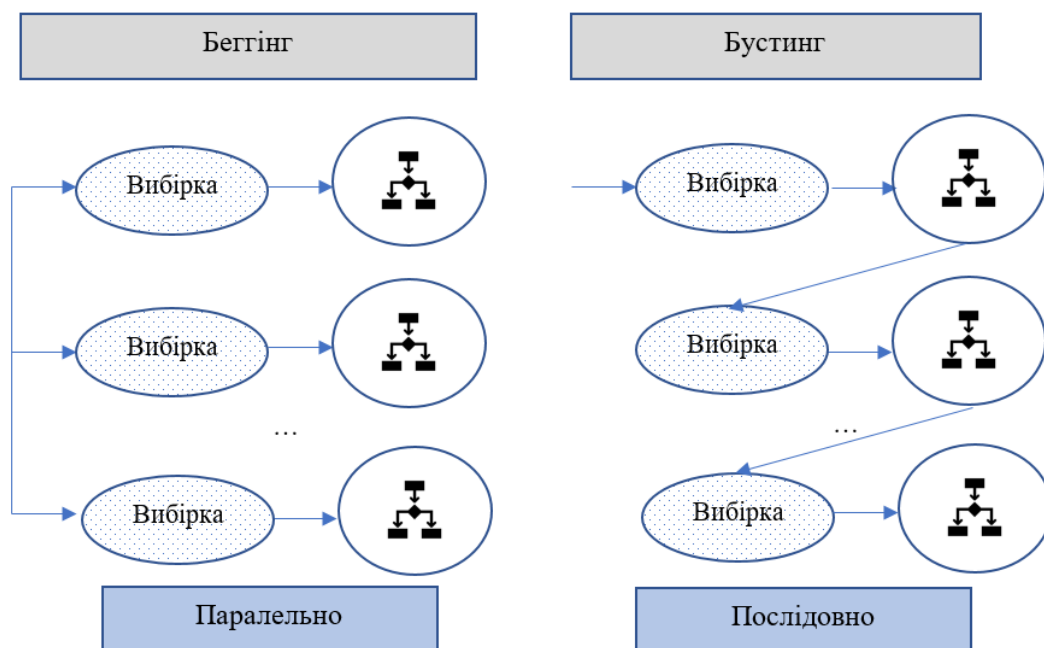
Дослідження основних алгоритмів бустингу дозволило визначити, що на основі бустингу працює XGB алгоритм (метод градієнтного спуску), який, на відміну від бегінгу, навчається на своїх помилках [34]. Алгоритми схожі, але, якщо в Random Forest всі дерева працюють паралельно, то в бустингу дані помилок першого дерева використовуються під час побудови наступних конфігурацій дерев. Бустинг зменшує дисперсію та упередженість моделі, оскільки використовується набір моделей (ансамбль). Одним з алгоритмів бустингу є Adaboost – алгоритм, за яким для кожної наступної моделі здійснюється перевірка та відсіювання спостережень-викидів, помилкових висновків попередньої моделі. Інший алгоритм – градієнтний бустинг передбачає, що виконують тренування кожної наступної моделі за допомогою залишків (різниця між передбачуваними та реальними значеннями). У цих ансамблях базова модель є достатньо слабкою, оскільки в подальшому проводиться додатковий аналіз залишків, на яких можна удосконалювати подальші конфігурації моделі (Рисунок 2) [21].

Для реалізації алгоритму збільшення градієнтів у дослідженні було використано оптимізовану бібліотеку Python – XGBoost. Вона є високоефективною, гнучкою та портативною, реалізує алгоритми машинного навчання в рамках градієнтного бустингу [21, 34]. Алгоритми, реалізовані у XGBoost, забезпечують паралельну побудову дерев, вони також відомі як GBDT або GBM, мають високу швидкість та точність роботи з даними. Цей же код працює у великих розподілених середовищах (Hadoop, SGE, MPI) і може працювати з великими даними (Big Data). Вищезазначене обумовило вибір саме алгоритму бустингу для подальшої розробки моделі класифікації.

4. РЕЗУЛЬТАТИ

Відповідно до запропонованої концептуальної схеми перший етап дослідження полягає у формуванні масивів вихідних даних, що в подальшому були використані в якості показників аналізу ЕР регіонів. Попередній монографічний аналіз [1-3, 5-11, 13-19, 22-24, 31, 32] дозволив визначити первісну систему оціночних індикаторів ЕР, які можна поділити на дві групи.

Перша група – це показники потенціалу ЕР регіонів: ЕАН – чисельність економічно активного населення (тис осіб) та КІ – капітальні інвестиції (млн грн). Саме ці два фактори є основним джерелом та рушійною силою економічного розвитку.



Джерело: Розробка авторів за джерелом [29].

Рисунок 2. Схема роботи алгоритмів бустингу та бегінгу

Друга група – це показники, що відображають результативність ЕР регіонів або ефективність використання економічного потенціалу. У свою чергу ці показники діляться на:

- стимулятори: СЗП – середня заробітна плата (грн), ВРП – валовий регіональний продукт (ВРП) у розрахунку на одну особу (грн), ІПП – індекс промислової продукції (відсотків до попереднього періоду), СЕІ – сальдо експорту-імпорту, (млн доларів);
- дестимулятори: БН – чисельність безробітного населення (за методологією Міжнародної організації праці [4, 12, 29, 33], тис осіб), ВЗР – обсяги викидів забруднюючих речовин (тис т).

Збільшення показників-стимуляторів демонструє зростання рівня ЕР, тоді як збільшення показників-дестимуляторів свідчить про неефективне використання економічного потенціалу регіону.

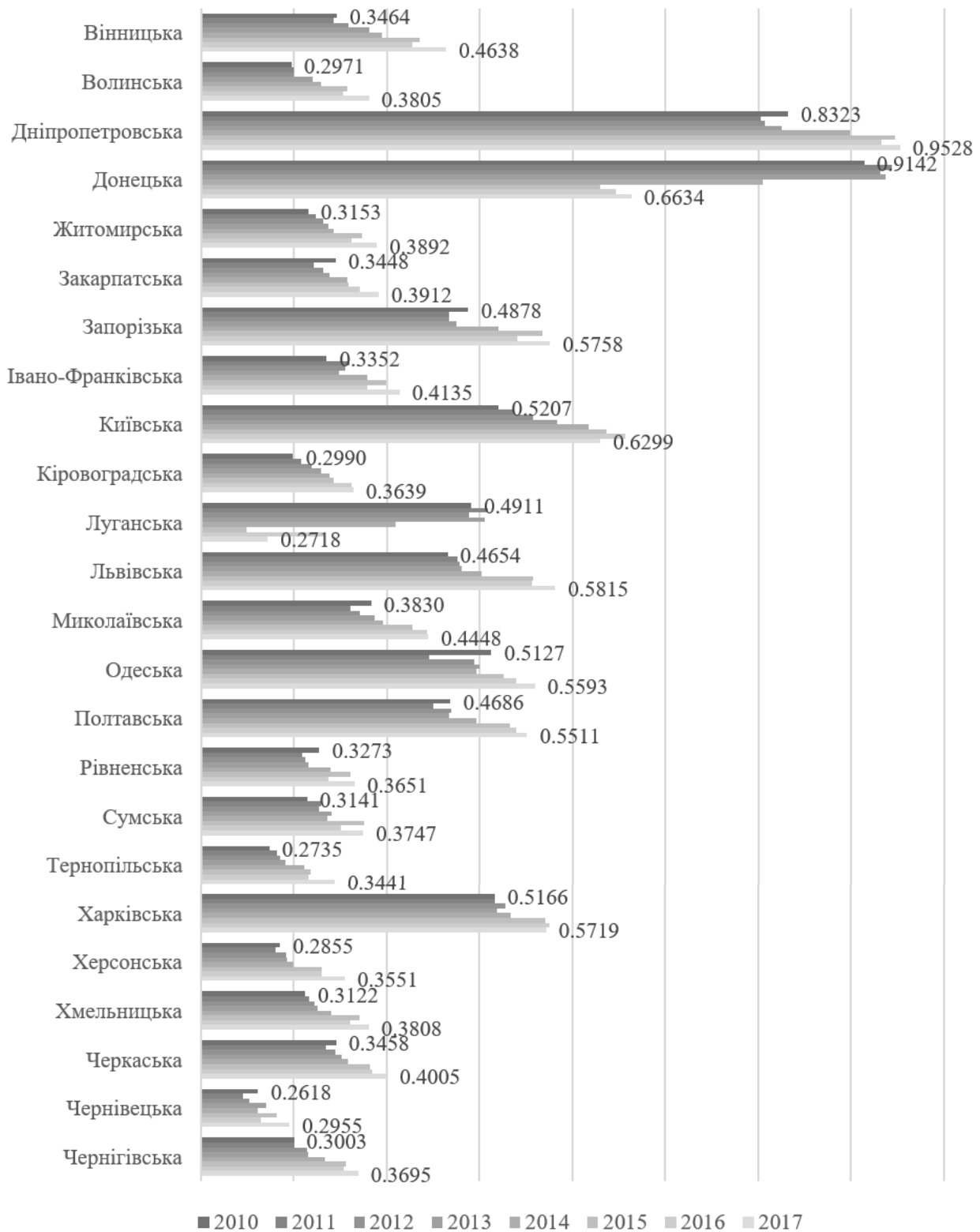
Вихідні дані були зібрані за період з 2010 по 2017 роки [30]. За вихідними даними був розрахований загальний рівень ЕР на базі методу повної редукції – таксономічного показника рівня розвитку для кожного регіону за кожним роком. Результати розрахунків (Рисунок 3) дозволили дійти висновку, що рівень ЕР є не дуже високим майже для всіх регіонів. Середнє значення для Харківського регіону складає 54.17%. Найвищий результат показав Дніпропетровський регіон, його середнє значення за 8 років досягло 89.5%.

Також високий показник має Донецький регіон (80.1%). До регіонів з найнижчим рівнем ЕР увійшли: Чернівецький, Херсонський, Волинський, Кіровоградський та Чернігівський регіони, значення їхнього показника в середньому знаходиться в межах 30%. Можемо спостерігати, що регіони України за рівнем ЕР мають досить високий розкид, що вказує на відсутність спільної динаміки розвитку, та можемо спостерігати тенденцію до їх дивергенції. Загалом, у сформованій просторово-динамічній вибірці 102 спостереження (регіони) належало до групи з високим рівнем економічного розвитку (середнє значення рівня – 0.5158), та 90 регіонів – до групи з низьким рівнем розвитку (0.3449). Проте слід зазначити, що в нашому випадку регіони розділяються скоріше на клас з низьким та середнім рівнем розвитку, оскільки середнє значення для першого кластеру складає 0.51, а це не дуже високий результат. Як зазначалося вище, тільки Дніпропетровська та Донецька області мають середні значення рівня ЕР за 8 років вище ніж 0.75%. Також слід звернути увагу, що рівень ЕР Донецького та Луганського регіонів суттєво зменшився з 2014 року, що викликано об'єктивними обставинами іноземної військової агресії на Сході нашої держави.

Побудова матриці кореляційних зв'язків між показниками груп дозволила визначити, що найменший вплив на рівень ЕР мають СЗП (–1%), ІПП (–2%) та ВРП (5%) (Рисунок 4).

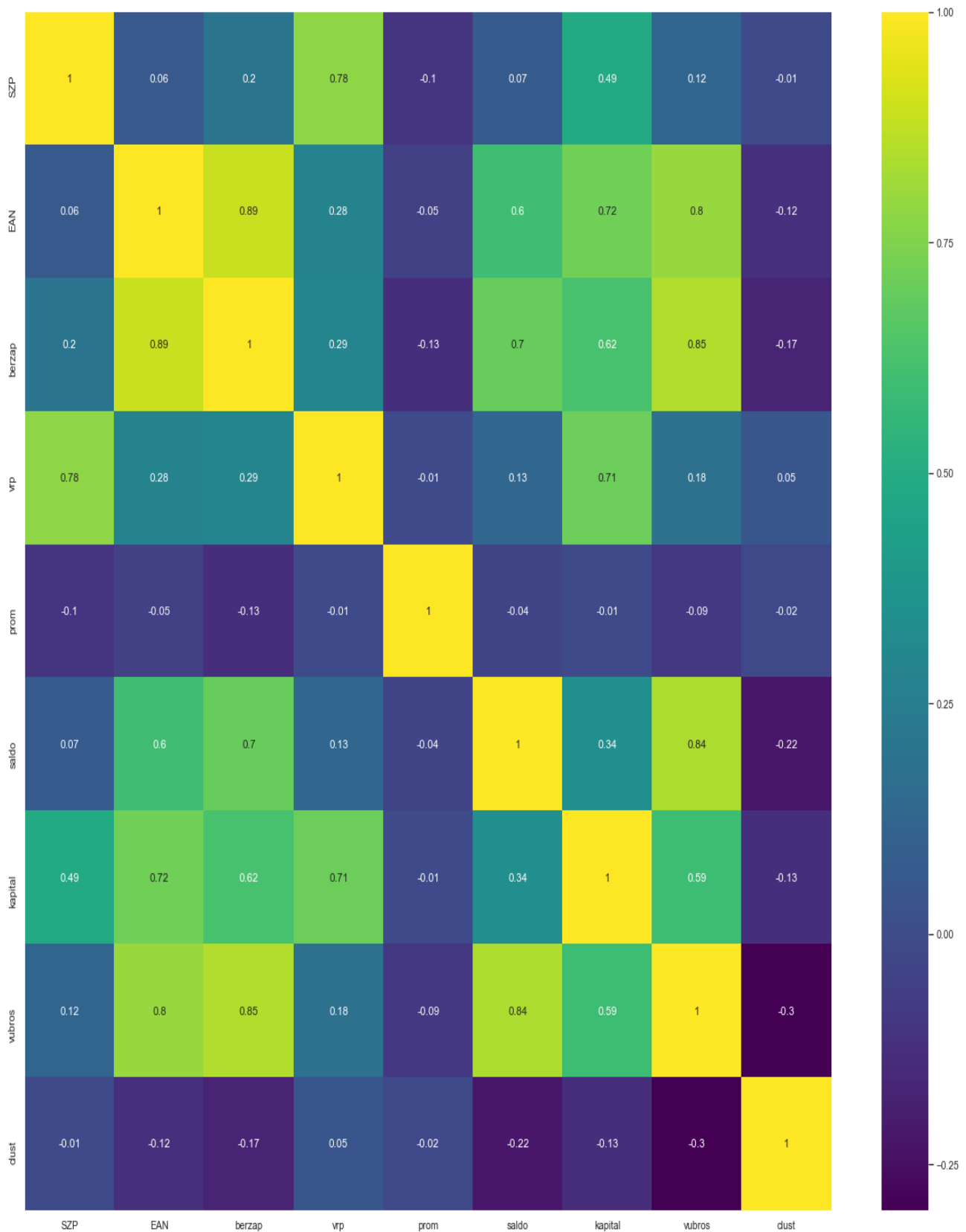
Найбільший вплив має показник викидів забруднюючих речовин в атмосферу (30%), сальдо експорту – імпорту (22%) та чисельність безробітного населення (17%). Слід також зазначити, що між показниками спостерігається мультиколінеарність. Так, наприклад, між чисельністю економічно активного населення та чисельністю безробітного населення зв'язок становить 89%, між СЕІ та ВЗР – 84%, між ВРП та СЗП – 78%, між сальдо експорту-імпорту та БН – 70%. Тобто можна прослідкувати тісний зв'язок між двома групами показників ЕР – показниками результативності економічного розвитку регіонів та їх економічним потенціалом.

На наступному етапі дослідження було здійснено побудову моделі прогнозування стану ЕР регіонів методами ансамблів дерев класифікації. Для побудови класифікатора було вибрано XGBoost (xgboost) алгоритм. Алгоритм був імплементован мовою Python, версія 3.7. Початковий розподіл вибірки на тестову та навчальну було виконано у пропорції 1:10, тобто з 192 спостережень до тестової вибірки потрапило 20, а до навчальної – 172. Навчання було проведено з використанням методу крос-валідації, за яким випадковим чином для 5 вибірок було сформовано навчальну та тестову пропорційно по 30% спостережень. Результати навчання – це середнє значення моделі на 5 вибірках (фолдах) (Рисунок 5).



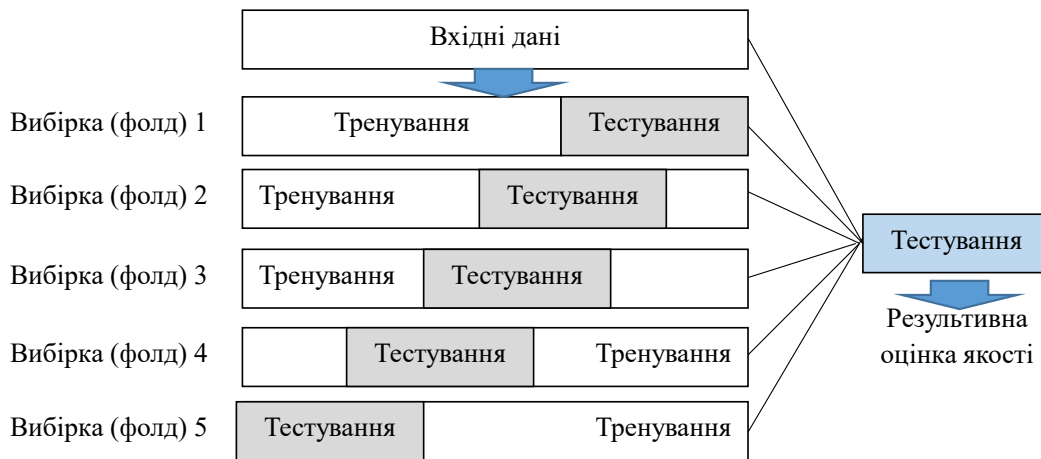
Джерело: Розробка авторів.

Рисунок 3. Результати розрахунку рівня ER за регіонами



Джерело: Розробка авторів.

Рисунок 4. Кореляційна матриця індикаторів



Джерело: Розробка авторів за джерелом [34].

Рисунок 5. Схема крос-валідації

Оцінку якості моделі було перевірено за 4 критеріями (метриками):

- 1) *accuracy* – якість класифікації або значущість моделі;
- 2) $f1_{score}$ – середньозважений рівень точності класифікації (збільшення рівня точності відповідає зростанню показника, який прямує до 1):

$$f1_{score} = \frac{2 \cdot precision \cdot recall}{precision + recall},$$

$$recall = \frac{t_p}{t_p + f_n} \cdot precision = \frac{t_p}{t_p + f_p},$$

де t_p – кількість вірно класифікованих спостережень справжніх позитивних результатів, f_n – кількість невірно класифікованих помилок, f_p – кількість вірно класифікованих помилок, *recall* – рівень вірно класифікованих спостережень, *precision* – рівень вірно класифікованих помилкових спостережень.

Для класифікатора використано такі параметри: {'max_depth': 50, 'min_child_weight': 5, 'n_estimators': 250, 'learning_rate': 0.1}. Де *max_depth* – максимальна глибина дерева; *min_child_weight* – мінімальна кількість листків на дереві; *n_estimators* – кількість дерев; *learning_rate* – крок (швидкість) навчання. Результати крос-валідації наведено у Таблиці 1.

Таблиця 1. Результати крос-валідації класифікатора на навчальній виборці

Джерело: Розробка авторів.

№ дерева	Метрика			
	<i>accuracy</i>	<i>f1_score</i>	<i>precision</i>	<i>recall</i>
1	0.846154	0.846154	0.846154	0.846154
2	0.846154	0.845696	0.847768	0.846154
3	0.788462	0.788853	0.803915	0.788462
4	0.807692	0.807120	0.808961	0.807692
5	0.846154	0.846154	0.846154	0.846154
Середнє	0.826923	0.826795	0.830590	0.826923

Як видно з Таблиці 1 усі значення за метриками перевищують 0.78, а більшість з них перевищують 0.84, що свідчить про досить високу якість моделі. Результати обчислення міри впливу показників на процес класифікації дозволив дійти висновку, що показник обсягу капітальних інвестицій має ступінь значущості для класифікації – 20%, чисельність економічно активного населення – 18%, індекс промислової продукції – 13%, викиди забруднюючих речовин та сальдо експорту-імпорту – 12.5%, чисельність безробітного населення – 11.5% (Рисунок 6).

Кумулятивний вплив цих показників на загальний стан ЕР становить 75%, а показники ВРП та СЗП мають найменший внесок у процесі класифікації. Побудовану модель було перевірено на тестовій вибірці з 20 спостережень (Таблиця 2).

Таблиця 2. Результати валідації моделі на тестовій вибірці

Джерело: Розробка авторів.

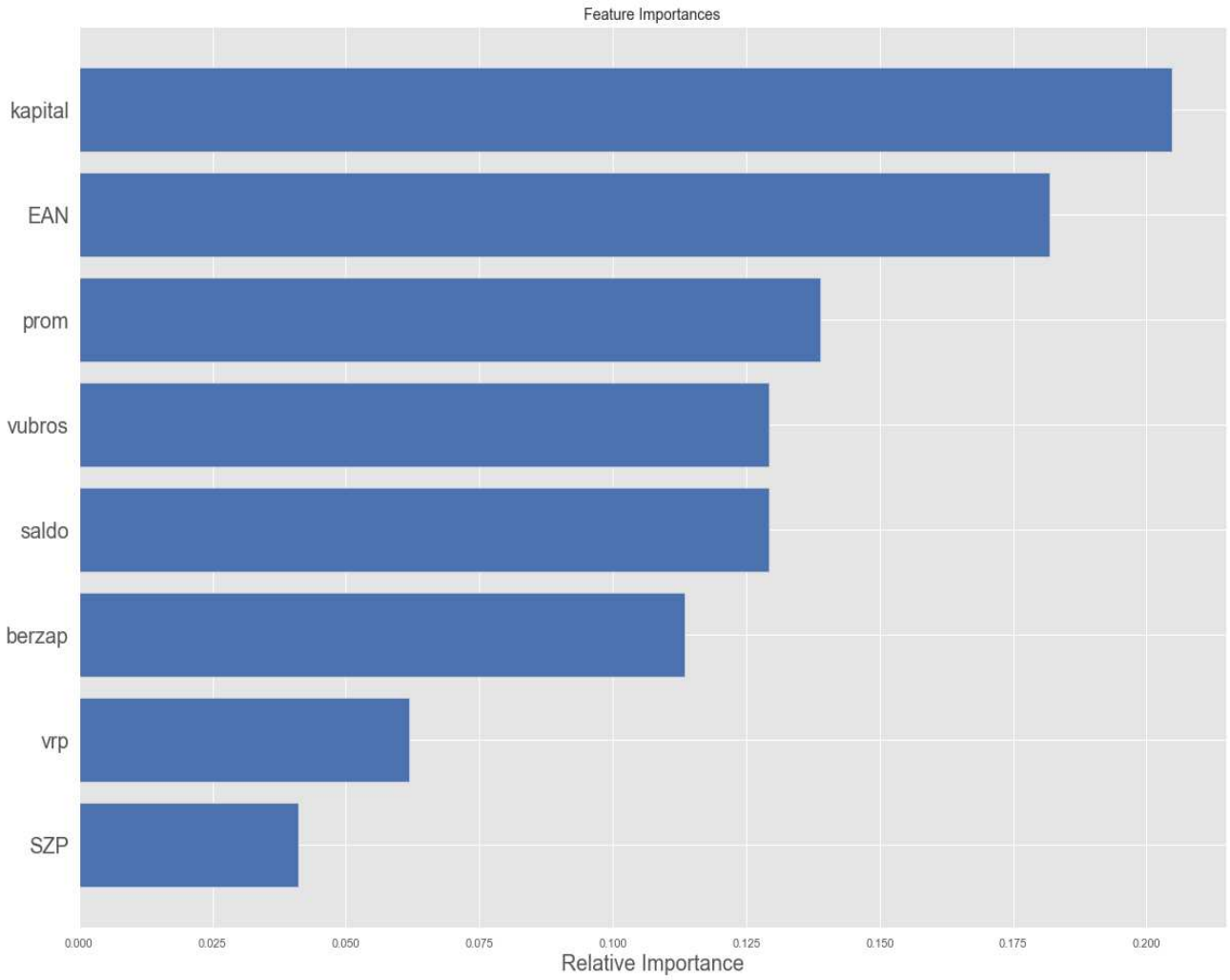
Регіон	Рік	Ймовірність приналежності	Фактичний клас	Прогнозний клас
Полтавський	2014	0.852365	1	1
Дніпровський	2013	0.980530	1	1
Херсонський	2016	0.013524	0	0
Волинський	2014	0.282792	0	0
Запорізький	2015	0.727431	1	1
Чернігівський	2012	0.273373	0	0
Харківський	2010	0.695317	1	1
Одеський	2016	0.927397	1	1
Волинський	2016	0.178018	0	0
Івано-Франківський	2010	0.334574	1	0
Закарпатський	2010	0.558368	1	1
Херсонський	2015	0.075045	0	0
Полтавський	2016	0.879024	1	1
Дніпровський	2017	0.984242	1	1
Сумський	2016	0.140028	0	0
Чернігівський	2016	0.111619	0	0
Одеський	2011	0.788238	1	1
Харківський	2012	0.955387	1	1
Харківський	2015	0.917986	1	1
Дніпровський	2015	0.964841	1	1

Під час валідації моделі помилка класифікації спостерігалась лише для спостереження, що відповідає Івано-Франківському регіону. 19 з 20 спостережень вибірки було класифіковано правильно. Вибірку для тестування була сформовано випадковим чином з таким розподілом регіонів: 13 регіонів кластеру з високим рівнем економічного розвитку та 8 регіонів з низьким рівнем розвитку (Таблиця 3).

Таблиця 3. Результати основних метрик валідації моделі

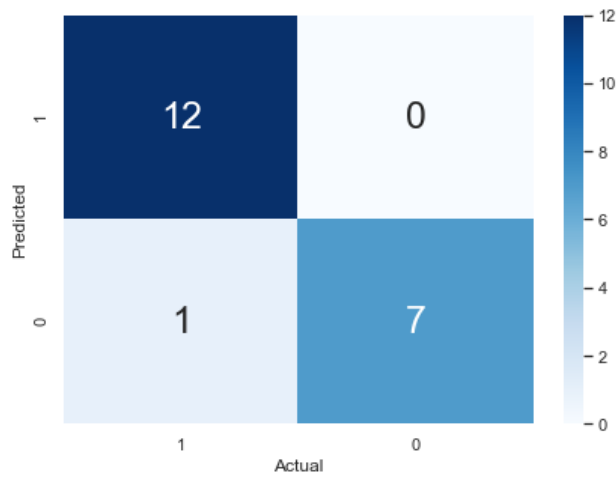
Джерело: Розробка авторів.

Метрика	Значення
accuracy	0.95
f1_score	0.95
precision	1
recall	0.92



Джерело: Розробка авторів.

Рисунок 6. Графік впливу факторів



Джерело: Розробка авторів.

Рисунок 7. Перехресна матриця класифікації на тестовій вибірці

З метою додаткового аналізу результатів класифікації, було побудовано перехресну матрицю валідації, яка дозволяє дійти висновку про достатньо високий рівень якості моделі, оскільки кількість невірної класифікованих об'єктів становить 1, тобто 5% (Рисунок 7).

Якість класифікації на тестовій вибірці є високою та модель є статистично значущою на рівні 95%, про що вказує досить високе значення метрик для даної моделі, зокрема, точності (accuracy).

ВИСНОВКИ

Таким чином, здобуті результати дослідження дозволили дійти низки висновків. Так, у результаті побудови загального рівня EP можна визначити, що регіони України мають суттєву різницю у динаміці розвитку, про що свідчить великий розмах інтегрального показника. Лідерами за значенням показника стали Дніпропетровський та Донецький регіон, найнижче значення у Чернівецького, Херсонського та Волинського регіону. Слід також зазначити, що протягом останніх 8 років, майже у всіх регіонів спостерігається позитивна динаміка рівня EP, проте існує велика частка регіонів, де рівень EP залишається низькими та дуже низьким. Найбільший розмах показника у динамічній вибірці мають Донецький та Луганський регіони, рівень розвитку яких стрімко впав у 2015 році, але спостерігається незначне покращення з 2016 року.

Було визначено, що в Україні спостерігається суттєва незбалансованість кластерів за часткою регіонів, із перевагою кластеру з низьким станом економічного розвитку. Крім того, слід відмітити тісний зв'язок між двома групи показників EP – показниками результативності економічного розвитку регіонів та їх економічним потенціалом, що має особливе значення для визначення подальшої розробки прогнозних характеристик стану EP.

Таким чином, побудована комплексна модель оцінювання економічного розвитку регіонів України методами Data Science та багатовимірного аналізу, що містить моделі оцінювання рівня EP регіонів, групування регіонів за рівнем EP, прогнозування стану EP регіонів та аналізу та прогнозування рівня EP регіонів дозволяє підвищити якість формування та ухвалення управлінських рішень з вирівнювання асиметричності регіонального розвитку.

Перспективами подальших досліджень на базі сформованого комплексу моделей є можливість розробки стратегії вирівнювання дисбалансу регіонального розвитку, визначення пріоритетних векторів стійкого розвитку та досягнення належного економічного стану як окремих регіонів, так і країни в цілому.

AUTHORS CONTRIBUTIONS

Conceptualization: Liubov Chagovets, Svitlana Prokopovych, Viktor Kholod.

Data curation: Viktor Kholod.

Formal analysis: Viktor Kholod.

Funding acquisition: Liubov Chagovets, Svitlana Prokopovych.

Methodology: Liubov Chagovets, Svitlana Prokopovych.

Project administration: Svitlana Prokopovych.

Software: Viktor Kholod.

Supervision: Svitlana Prokopovych.

Visualization: Liubov Chagovets, Viktor Kholod.

Writing – original draft: Liubov Chagovets, Svitlana Prokopovych.

Writing – review & editing: Liubov Chagovets, Svitlana Prokopovych.

СПИСОК ЛІТЕРАТУРИ

1. Agishev, R., & Karaeva, L. (2009). Problemy formirovaniya sistemy kompleksnykh pokazateley sotsialno-ekonomicheskogo razvitiya munitsipalnykh obrazovaniy [Problems of forming a system of complex indicators of socio-economic development of municipalities]. *Regionology*, 2. (In Russian). Retrieved from <http://regionsar.ru/node/316>
2. Arkhyenko, I. (2019). Foreign experience of formation and implementation of regional economic policy. *Collection of scientific works «Public Administration and Local Government»*, 2(41), 32-42. (In Ukrainian). Retrieved from [http://www.dridu.dp.ua/vidavnic-tvo/2019/2019_02\(41\)/6.pdf](http://www.dridu.dp.ua/vidavnic-tvo/2019/2019_02(41)/6.pdf)
3. Boyko, A. (2014). Convergence and unevenness in regional development of Ukraine: risks, trends and prospects. *Economics and region*, 1, 72-78. (In Ukrainian). Retrieved from http://nbuv.gov.ua/UJRN/econrig_2014_1_13
4. Communications Development Incorporated (2015). *Human Development Report 2015. Work for human development* (288 p.). Retrieved from http://hdr.undp.org/sites/default/files/2015_human_development_report.pdf
5. Dutta, S., Lanvin, B., & Wunsch-Vincent, S. (Eds.) (2016). *The Global Innovation Index 2016: Winning with Global Innovation*. Retrieved from https://www.wipo.int/edocs/pubdocs/en/wipo_pub_gii_2016.pdf
6. Glushchenko, K. (2010). Metody analiza mezhhregionalnogo neravenstva po dokhodam [Methods for analyzing interregional income inequality]. *Region: Economics and Sociology*, 1, 54-87. (In Russian). Retrieved from https://www.sibran.ru/journals/issue.php?ID=120715&ARTICLE_ID=130031
7. Grebenkin, A. (2005). Asimetriya razvitiya regionov: faktory vozniknoveniya i regulirovaniye [Asimetriya razvitiya regionov: faktory vozniknoveniya i regulirovanie]. *Nalogi, Investitsii, Kapital - Taxes, Investments, Capital*, 1-3, 182-189. (In Russian)
8. Havkalova, N. (Ed.) (2015). *Zabezpechennia rehionalnogo rozvytku ekolohichnogo suspilstva zasobom stvorennia systemnoho bazysu syntezyvanoho kapitalu [Ensuring the regional development of ecological society by means of creating a systemic basis of synthesized capital]* (304 p.). Kharkiv: HNEU. (In Ukrainian). Retrieved from <http://www.repository.hneu.edu.ua/handle/123456789/13408>
9. Helliwell, J., Layard, R., Sachs, J., & De Neve, J. (Eds.) (2020). *World Happiness Report 2020*. Retrieved from <http://worldhappiness.report>
10. Hnativ, B., Tokarchuk, R., Kostrobij, P., & Tokarchuk, M. (2011). Mathematical modeling of economic processes by methods of nonequilibrium statistical mechanics. *Journal of Lviv Polytechnic National University. Series of Physical and mathematical sciences*, 696, 93-100. (In Ukrainian)
11. Hrechana, S., Kharytonova, O., & Klius, I. (2012). Research disproportions in development economy region. *Efektivna ekonomika - Efficient economy*, 11. (In Ukrainian). Retrieved from <http://www.economy.nayka.com.ua/?op=1&z=1700>
12. ILO (2013). *Resolution concerning statistics of work, employment and labour underutilization*. Retrieved from http://www.ilo.ch/wcmsp5/groups/public/---dgreports/---stat/documents/normativeinstrument/wcms_230304.pdf
13. Klebanova, T., & Kizima, N. (Eds.) (2011). *Modeli otsenki neravnomernosti i tsiklicheskoj dinamiki razvitiya territoriy [Models for assessing the unevenness and cyclical dynamics of the development of territories]* (352 p.). Kharkov: INZHEK. (In Russian)
14. Kozryieva, O., & Heiman, O. (2015). An Analysis of Inequality in the Socio-Economic Development of the Regions of Ukraine. *Business Inform*, 12, 93-104. (In Ukrainian). Retrieved from http://nbuv.gov.ua/UJRN/binf_2015_12_16
15. Makarova, O. (2015). *Sotsialna polityka v Ukraini [Social policy in Ukraine]* (244 p.). Kyiv: Ptoukha Institute for Demography and Social Studies of the National Academy of Sciences of Ukraine. (In Ukrainian). Retrieved from https://idss.org.ua/monografii/2015_Makarova.pdf
16. Mokij, A., & Daczko O. (2018). *Implementatsiia instytutsionalnogo bazysu rehionalnoi polityky YeS v Ukraini [Implementation of the institutional basis of EU regional policy in Ukraine]* (22 p.). Retrieved from https://niss.gov.ua/sites/default/files/2019-01/1_Zapiska-Mok_y-Datsko-pogodzhena-41230.pdf
17. Muzychuk, V. (2011). *Indeks rozvytku liudskoho potentsialu yak indyikator efektyvnosti sotsialnoi oriientsatsii ekonomiky [Human resources development index as indicator of the economies social orientation efficiency]*. Retrieved from http://nbuv.gov.ua/j-pdf/vkpnu-en_2011_4_146.pdf
18. NISS (2009). *Instytutsiini zasady derzhavnogo rehuliuвання rozvytku depresyvnnykh terytorii [Institutional framework of state regulation of depressed areas]*. Retrieved from <http://www.niss.gov.ua/Monitor/june2009/9.html>
19. Pilko, A., & Harda, T. (2018). The models of assessment and analysis of the regional development asymmetry. *Economics of Development*, 2(86), 24-35. (In Ukrainian). Retrieved from http://www.ed.ksue.edu.ua/ER/knt/eu182_86/e182pil.pdf
20. Ponomarenko, V., & Malyarets, L. (2009). *Multivariate analysis of the socio-economic systems: study guide* (384 p.). Kharkiv: KHNEU. (In Ukrainian). Retrieved from <http://www.repository.hneu.edu.ua/handle/123456789/3113>
21. QuantDare (2016). *What is the difference between Bagging and Boosting?* Retrieved from <https://quantdare.com/what-is-the-difference-between-bagging-and-boosting>
22. Romanyuk, S. (2013). *Rozvytok rehioniv u vidkrytii ekonomitsi: teoriia, polityka, praktyka [Development of regions in an open economy: theory, policy, practice]* (375 p.). Kyiv: NADU. (In Ukrainian)
23. Rudenko, L. (2005). Kontsepsiia zbalansovanosti rozvytku i yakist zhyttia naselennia [The concept of balanced development and quality of life]. *Ukrainian Geographic Journal*, 2, 44-50. (In Ukrainian)
24. Samarina, V. (2008). Osobennosti otsenki neravnomernosti sotsialno-ekonomicheskogo razvitiya regionov [Features of assessing the uneven socio-economic development of regions]. *Problems of modern economics*, 1(25), 300-303. (In Russian). Retrieved from <http://www.m-economy.ru/art.php?nArtId=1820>
25. Savchuk, T., & Luzhenetskyi, R. (2011). Vykorystannia klasternoho analizu dlia vyrishennia zadach tsilovoho marketynhu [Using cluster analysis to solve targeted marketing problems]. *Measuring and computing devices in technological processes*, 2, 144-148. (In Ukrainian). Retrieved from <https://journals.khnu.km.ua/index.php/MeasComp/article/view/1836>
26. Scikit learn (n.d.). *A random forest classifier*. Retrieved from <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html>
27. Scikit Learn (n.d.). *Density-Based Spatial Clustering of Applications with Noise*. Retrieved from <https://scikit-learn.org/stable/modules/generated/sklearn.cluster.DBSCAN.html>

28. Shitikov, V., & Mastickij, S. (2017). *Klassifikatsiya, regressiya i drugiye algoritmy Data Mining s ispolzovaniyem R [Classification, Regression, and Other Data Mining Algorithms Using R]*. (In Russian)
29. Social Progress Index (n.d.). *Official web-site*. Retrieved from <http://www.socialprogressindex.com>
30. State Statistics Service of Ukraine (n.d.). *Official web-site*. Retrieved from <http://www.ukrstat.gov.ua>
31. Tomareva-Patlakhova, V. (2013). Model of regional development in the context of economic reforms. *State and regions. Series: Economics and Business*, 1(70), 76-80. (In Ukrainian). Retrieved from http://nbuv.gov.ua/UJRN/drep_2013_1_17
32. Vershygora, Yu., & Vershygora, V. (2017). Unevenness of development of the regions of Ukraine and ways of its overcoming. *International Humanitarian University Herald. Economics and Management*, 2(23), 25-28. (In Ukrainian). Retrieved from <http://www.vestnik-econom.mgu.od.ua/journal/2017/23-2-2017/7.pdf>
33. World Intellectual Property Organization (n.d.). *Official web-site*. Retrieved from <http://www.wipo.int>
34. XGBoost (n.d.). *XGBoost Documentation*. Retrieved from <https://xgboost.readthedocs.io/en/latest>