УДК 004.021: 330.341

Міністерство освіти і науки України
Харківський національний університет радіоелектроніки

Факультет *Комп'ютерних наук*
(повна назва)
Кафедра *Системотехніки*
(повна назва)

АТЕСТАЦІЙНА РОБОТА
Пояснювальна записка

Рівень вищої освіти другий (магістерський)

ГЮИК.506900.007 ПЗ

Дослідження методів багатовимірного аналізу даних для оцінки рівня соціально-економічного розвитку регіонів

(тема)

Виконав:
Студент 2 курсу, групи *ІТПм-19-1*
Спеціальність *122 – Комп'ютерні науки*
(код і повна назва напряму)
Тип програми *освітньо-професійна*
(освітньо-професійна або освітньо-наукова)
Освітня програма *Інформаційні технології проектування*
(повна назва освітньої програми)
Ісакій К.Г.
(прізвище, ініціали)

Керівник доцент каф. *Коваленко А.І*
(посада, прізвище, ініціали)

Допускається до захисту
Зав. кафедри _____ *Гребеннік І. В.*
(підпис) (прізвище, ініціали)

2020 р.

Атестаційна робота не містить відомостей заборонених до відкритого опублікування.

Атестаційна робота виконана у відповідності до стандартів, що діють в Україні.

Попередній захист проведений «18» грудня 2020 р.

Керівник атестаційної роботи  доц. Коваленко А.І.

Харківський національний університет радіоелектроніки

Факультет *Комп'ютерних наук*
(повна назва)

Кафедра *Системотехніки*
(повна назва)

Рівень вищої освіти другий (магістерський)

Спеціальність *122 – Комп'ютерні науки*
(код і повна назва)

Тип програми *освітньо-професійна*
(освітньо-професійна або освітньо-наукова)

Освітня програма *Інформаційні технології проектування*
(повна назва)

ЗАТВЕРДЖУЮ:
Зав. кафедри _____
(підпис)
«____»_____ 20___ р.

ЗАВДАННЯ

НА АТЕСТАЦІЙНУ РОБОТУ

студентові Ісакію Костянтину Григоровичу
(прізвище, ім'я, по батькові)

1. Тема роботи *Дослідження методів багатовимірного аналізу даних для оцінки рівня соціально-економічного розвитку регіонів*
затверджена наказом по університету від «02» листопада 2020 р. № 1517Ст
2. Термін подання студентом роботи (проекту) 18 грудня 2020 р.
3. Вихідні дані до роботи (проекту) Функція: визначення соціально-економічного рейтингу регіонів за таксономічним показником, який визначається на основі статистичних даних з подальшим визначенням кластерів та прогнозів регіональних показників. Дані Держкомстату: середня заробітна плата, економічно активне населення, безробітне населення, валовий регіональний продукт у розрахунку на одну особу, індекси промислової продукції, сальдо (експорт-імпорт), капітальні інвестиції, обсяги викидів забруднюючих речовин у період з 2012 по 2019 роки.
4. Зміст пояснювальної записки (перелік питань, що потрібно розробити) Вступ, Аналіз предметної галузі та постановка задачі, Актуальність теми, Огляд і аналіз сучасного стану проблеми, Постановка задачі, Теоретико-методологічні основи аналізу рівня соціально-економічного розвитку регіонів, Розробка концептуальної схеми моделювання рівня соціально-економічного розвитку регіонів з визначенням системи

індикаторів, Вибір методів багатовимірної статистики згідно концептуальної схеми моделювання рівня соціально-економічного розвитку регіонів. Висновки за розділом 2, Аналіз та експеременталька оцінка рівня соціально-економічного розвитку регіонів, Визначення послідовності використання розроблених програмних застосувань для експериментальної оцінки рівня соціально-економічного розвитку регіонів, Експериментальна реалізація методу таксономії та моделі кластерного аналізу рівня соціально-економічного розвитку регіонів України, Експериментальна побудова прогнозів регіональних показників соціально-економічного розвитку, Висновки за розділом 3, Висновки, Перелік джерел посилань.

5. Перелік графічного матеріалу (з точним зазначенням обов'язкових креслеників, плакатів): Таксономічний показник , Результати таксономічного показника за регіонами, Метод головних компонент для 2012 року, Метод головних компонент для 2019 року, Варіативність кожної компоненти методу головних компонент, Результати розподілу кластерів та середнє значення таксономічного показника, Матриця кореляційного аналізу, Результати моделі на панельних даних, Вплив факторів (feature_importances_), Прогноз сальдо на 2020 - 2021 роки, Прогноз капітальних інвестицій на 2020 - 2021 роки, Прогноз індексу промисловості на 2020 - 2021 роки, Прогноз валового регіонального на 2020 - 2021 роки, Прогноз шкідливих викидів на 2020 - 2021 роки, Динаміка індексу таксономії Харківського регіону з прогнозними на 2020-2021 роки значеннями, Результати прогнозування потрапляння Харківського регіону до кластеру регіонів з високим рівнем розвитку.

6. Консультанти розділів роботи (проекту)

Найменування розділу	Консультант (посада, прізвище, ім'я, по батькові)	Позначка консультанта про виконання розділу	
		підпис	дата
Спец. частина	Доц. Коваленко А.І.		24.12.2020

КАЛЕНДАРНИЙ ПЛАН

№	Назва етапів роботи	Терміни виконання етапів роботи	Примітка
1.	Отримання завдання на атестаційне проектування	02.11.2020	
2.	Аналіз завдання та пошук літератури з теми роботи	05.11.2020	
3.	Опрацювання літератури та аналіз об'єкту дослідження	10.11.2020	
4.	Вибір апаратного набору та його збірка	22.12.2020	
5.	Розробка алгоритмів	24.12.2020	
6.	Аналіз отриманих результатів	27.12.2020	
7.	Оформлення пояснювальної записки та документації	01.12.2020	
8.	Оформлення презентаційних матеріалів	15.12.2020	
9.	Представлення на рецензування	15.12.2020	
10.	Представлення атестаційної роботи	18.12.2020	

Дата видачі завдання *02 листопада 2020 р.*

Студент _____ *Ісакій К.Г*

(підпис)

Керівник роботи _____ доцент каф. *Коваленко А.І.*

(підпис) (посада, прізвище, ініціали)

РЕФЕРАТ

Пояснювальна записка до атестаційної роботи: 64 с., 15 рис., 12 табл., 3 додатки, 30 джерел інформації.

СОЦІАЛЬНО-ЕКОНОМІЧНИЙ РОЗВИТОК, МЕТОД, АНАЛІЗ, ІНДЕКС, БАГАТОВИМІРНА СТАТИСТИКА.

Об'єктом дослідження даної роботи є процес оцінювання та визначення рівня соціально-економічного розвитку регіонів.

Предметом дослідження даної роботи є методи багатовимірного аналізу статистичних даних, які можуть використовуватися для оцінювання та визначення рівня соціально-економічного розвитку регіонів.

Методи дослідження: системний підхід для аналізу рівня соціально-економічного розвитку регіонів; метод таксономічного показника (методом Хельвіга) метод кластеризації (методом DBSCAN), метод XGB класифікації (з використанням бустингу) та методи статистичного та кореляційного аналізу.

У роботі запропонована теоретико-методологічна концептуальна модель з визначеними методами для проведення аналізу рівня соціально-економічного розвитку регіонів. Модель дозволяє на основі побудови таксономічного показника рівня розвитку регіонів України кластеризувати регіони України за рівнем соціально-економічного розвитку та його прогнозування на майбутні періоди.

Галузь застосування – департаменти та відділи державної влади, відділи підприємств, організацій та компаній, що займаються оцінкою і прогнозом розвитку різних галузей економіки.

ABSTRACT

Attestation work: 64 p., 15 pic., 12 tables, 30 sources, 3 applications.

SOCIO-ECONOMIC DEVELOPMENT, METHOD, ANALYSIS, INDEX, MULTIDIMENSIONAL STATISTICS.

The object of study of this work is the process of assessing and determining the level of socio-economic development of regions.

The subject of research of this work are methods of multidimensional analysis of statistical data that can be used to assess and determine the level of socio-economic development of regions.

Research methods: a systematic approach to analyze the level of socio-economic development of regions; taxonomic index method (Helwig method) clustering method (DBSCAN method), classification (boosting method) and statistical and correlation analysis.

The paper proposes a theoretical and methodological conceptual model with certain methods for analyzing the level of socio-economic development of regions. The model allows based on the construction of a taxonomic indicator of the level of development of the regions of Ukraine to cluster the regions of Ukraine by the level of socio-economic development and forecasting it for future periods.

Field of application - departments and divisions of state power, divisions of enterprises, organizations and companies engaged in the assessment and forecast of the development of institutions, enterprises and industries.

ЗМІСТ

ВСТУП

Рівень розвитку регіонів – складне та багатопланове поняття, яке об’єднує широкий спектр соціально-економічних відносин, пов’язаних з умовами та станом життєдіяльності людини у суспільстві. Рівень соціального-економічного розвитку регіонів впродовж тривалого часу залишається однією з основних соціально-економічних категорій, що характеризує не лише матеріальний добробут окремої людини, а й визначає узагальнений результат діяльності економіки країни за певний період.

Рівень соціально-економічного розвитку регіонів можна також оцінювати, як рівень життя населення у цих регіонах, як на макро-рівні (становище України на світовій арені), та і на мікро-рівні, тобто оцінювати становище регіонів між собою в розрізі держави. Проте, якщо Україна широко висвітлена в різних соціально-економічних рейтингах, що ведуться міжнародними інститутами, то регіони все ще залишаються малодослідженими. Недослідженим залишається вплив регіонів один на один, оскільки великі кількість індикаторів рівня життя населення доступна в регіональному розрізі, постає необхідність дослідити вплив розвитку одного регіону на інший.

Об'єктом дослідження даної роботи є процес оцінювання та визначення рівня соціально-економічного розвитку регіонів.

Предметом дослідження даної роботи є методи багатовимірного аналізу статистичних даних, які можуть використовуватися для оцінювання та визначення рівня соціально-економічного розвитку регіонів.

Мета роботи полягає в якісному аналізі та оцінці рівня життя населення в регіонах України під впливом структурних економічних змін за допомогою багатовимірних методів аналізу даних.

Для досягнення поставленої мети необхідно вирішити такі завдання:

– проаналізувати сучасні підходи для оцінювання рівня розвитку регіонів;

– побудувати концептуальну схему моделювання рівня регіонів для проведення соціально-економічного аналізу рівня розвитку регіонів;

– обґрунтувати вибір методів багатовимірного аналізу даних для визначення соціально-економічного аналізу рівня розвитку регіонів;

– обґрунтувати вибір методів багатовимірного аналізу даних для прогнозування рівня розвитку регіонів;

– визначити регіони України в контексті їхньої кластеризації;

– визначити таксономічні показники регіонів України;

– провести кластеризацію регіонів України за рівнем соціально-економічного розвитку;

– проаналізувати стан і тенденції зміни рівня розвитку регіонів України;

– побудувати прогнозну модель рівня розвитку регіонів України.

Методи дослідження: системний підхід для аналізу рівня соціально-економічного розвитку регіонів; метод таксономічного показника (методом Хельвіга) метод кластеризації (методом DBSCAN), метод XGB класифікації (з використанням бустингу) та методи статистичного та кореляційного аналізу.

Практичне значення результатів роботи полягає в тому, що розроблена концептуальну схема моделювання та обрані методи багатовимірного аналізу даних дозволяють проводити розрахунок і аналіз рівня соціально-економічного розвитку регіонів України.

Основні результати за темою магістерської роботи пройшли апробацію:

– на хакатоні «Open Data Campus» (м. Харків), що відбувся 25-26 січня 2020 р. у рамках відкритого проекту USAID / UK «Прозорість та підзвітність у державному управлінні та послугах» під егідою Міністерства цифрової трансформації України [1-2];

– на міжнародній науково-практичній конференції «World science: problems, prospects and innovations» (м.Торонто, Канада) [3].

1 АНАЛІЗ ПРЕДМЕТНОЇ ГАЛУЗІ ТА ПОСТАНОВКА ЗАДАЧІ

1.1 Актуальність теми

Економіка України розвивається нерівномірно, різні регіони історично мають різну спрямованість у виробництві ВРП. Розрізняють індустріально спрямовані регіони, аграрно направлені, а також мають місце депресивні регіони, які втратили ті галузі, на яких спеціалізувалися. Виникає необхідність провести аналіз і порівняльну оцінку для виявлення макроекономічних процесів, які мають місце в економіці України в цілому.

Одним з найбільш дієвих інструментів розподілу досліджуваної сутності є кластерний аналіз. Його перевага полягає в тому, що він поєднує об'єкти дослідження в однорідні за кількома показниками групи (кластери). Кластерний аналіз є найбільш універсальним інструментом, який можна використовувати в регіональному моделюванні. З його допомогою, можливо аналізувати дані про подібності об'єктів, інформація про які отримана різними способами і для будь-яких шкал виміру. Результати аналізу подаються в зручній наочній формі, яка полегшує прийняття рішень про оптимальне число факторів і взаємозв'язок різних кластерів.

1.2 Огляд і аналіз сучасного стану проблеми

Регіони являються невід'ємними частинами не тільки адміністративно-територіального устрою, а й всієї економічної системи країни. Регіон, представляє собою агломерацію виробничих потужностей, характер яких визначений природно-географічними, соціально-економічними, демографічними, історичними та іншими особливостями, виступає центральною ланкою економічного розвитку в масштабі всієї держави. Розвиваючись, відроджуючись після депресивного стану, долаючи кризи і прогресивно функціонуючи, кожен окремо взятий регіон вносить свою лепту в розвиток економіки всієї країни. Тому проблеми стійкості розвитку регіонів не

перестають бути актуальними і в даний час, коли глибокі трансформації в економіці роблять її найбільш вразливою і чутливою до впливу різних чинників.

Напрямок соціально-економічного розвитку регіонів завжди було і буде виділено серед всіх державних інтересів як один з найбільш актуальних і пріоритетних напрямів, так як кожен регіон, будучи частиною єдиної держави, вносить свою лепту в його історію, внутрішню політичну, економічну, культурну життя, досягнення певних результатів на міжнародній арені. А організація господарської діяльності регіонів за допомогою взаємодії виробничих потужностей своїх територій становить єдиний господарський комплекс країни. Саме активізація економічного життя в регіонах визначає тенденції економічного зростання всієї країни [10].

Світовий досвід свідчить про те, що держава є стійкою і життєздатною лише в тих випадках, коли його суб'єкти (адміністративно-територіальні утворення) є політично стабільними, економічно і соціально життєдіяльними. Таким чином, є очевидним, що дослідження проблем, пов'язаних з розробкою концепцій соціально-економічного розвитку регіонів, спрямованих на оптимізацію використання наявних в їх межах ресурсів і вибір основних пріоритетів у розвитку кожного з регіонів України, є найбільш актуальним і в даний час.

Розвиток є рушійною силою прогресу, його матеріальною основою. У широкому, філософському сенсі розвиток означає процес удосконалення тих чи інших елементів суспільних відносин або матеріально-речових елементів суспільства, перехід до принципово новим якісним характеристикам. Серцевиною розвитку є економічний розвиток, яке включає кількісне збільшення населення і багатства, поява якісно нових капітальних благ і цінностей, явищ і процесів, глибоку модернізацію та розбудову всієї економічної і соціальної системи.

Економічний розвиток – важлива мета будь-якого суспільства, держави. Незалежно від ідеології і вже досягнутого рівня добробуту, всі країни мають на меті розвитку економіки і підвищення життєвого рівня. Термін «розвиток» часто вживається в таких поєднаннях: економічний розвиток, соціально-економічний розвиток, розвиток економіки міста, регіону, країни. У кожному разі під розвитком

мається на увазі будь-який прогресивний зміна, перш за все, в економічній сфері. Якщо зміна кількісне, кажуть про економічне зростання. При якісній зміні мова може йти про структурні зміни або про зміну змісту розвитку, або про придбання економічною системою нових характеристик [11].

Подолання кризи в будь-якій сфері життя безпосередньо пов'язане з рівнем економічної активності. Соціальний розвиток визначається ресурсними можливостями, які, в свою чергу, залежать від ступеня економічного розвитку. Тому, тільки розвиваючи економічну активність, можна здійснити ті чи інші прориви в житті місцевої громади та підняти рівень добробуту населення, який, в кінцевому рахунку, завжди визначає ступінь успіху функціонування.

Економічний розвиток регіону створює базу і є важливим джерелом підвищення якості життя населення, так як в ході взаємозалежних процесів економічного розвитку створюються передумови для підвищення доходів громадян, рівня освіти та медичного обслуговування, а також створюються умови, які сприяють зростанню самоповаги людей в результаті формування соціальної, політичної, економічної та інституційної систем, орієнтованих на підвищення престижності особистості в суспільстві. Тобто економічний розвиток, як правило, створює передумови для розвитку соціальної сфери, тому поряд з чисто економічними характеристиками нерідко розглядають соціальні показники розвитку, які стали не менш важливими в оцінці ступеня розвитку будь-якого регіону.

Прогрес у розвитку регіону слід розглядати і з точки зору економіки, і з точки зору соціальної сфери. Позитивна динаміка економічних показників у відриві від прогресивних змін в соціальній сфері призведе до перемоги застарілого принципу «виробництво заради виробництва». Тому поняття «економічний розвиток», як правило, передбачає, визначає і перетікає в поняття «соціально-економічний розвиток», не змінюючи суті дослідження, а, навпаки, більш детально розкриваючи проблему.

Соціальний аспект у розвитку полягає в підпорядкуванні цілей економічного зростання першочергових завдань соціального розвитку. Економічний аспект полягає в найбільшу відповідність кінцевих результатів економічного розвитку досягненню

сукупності цілей соціального розвитку. Економічна складова соціально-економічної ефективності є матеріальною основою для поліпшення якості життя - підвищення добробуту суспільства.

У монографії Л. М. Кузьменко зазначено, що: «Розвиток регіону - це комплексний процес зміни економічної, соціальної, екологічної, політичної, духовної сфер (цей перелік можна продовжувати), які призводять до якісних перетворень в напрямку поліпшення умов життєдіяльності людини» [29].

У праці Т.С. Максимової зазначено, що: «Регіональний розвиток - це складний і комплексний процес, який має: а) своє утримання - процес виробництва і відтворення; б) матеріально-речові носії - фактори економічного зростання; в) кількісні та якісні показники, що характеризують соціально-економічний результат як суспільне багатство в різних його формах» [28].

Соціально-економічний розвиток регіонів в значній мірі залежить від обсягу та ефективності використання економічного потенціалу їх території, який є основою існування і життєдіяльності регіону.

Узагальнюючи підходи різних вчених до визначення економічного потенціалу, можна зробити висновок, що економічний потенціал регіону визначається його природними ресурсами, засобами виробництва, трудовими і науково-технічним потенціалом.

У монографії В.Н. Василенко зазначено, що: «Економічна та соціальна життєдіяльність адміністративно-територіальних утворень забезпечується наявністю природних, матеріальних і трудових ресурсів (наявних в певних межах), можливостями ефективного їх використання для підтримки (або підйому) якомога більш високого рівня якості життя населення, яке проживає на даній території.

Однією з найважливіших рис регіонального господарства є його тісний зв'язок з природно-географічним фактором. Кожен регіон має певні природні ресурси, багато в чому визначають його господарську орієнтацію. Наявність ресурсів, їх обсяг, якість землі, природно-кліматичні умови, місце розташування грають ключову роль в економіці регіонів. Обмеженість корисних копалин, збільшення витрат на їх придбання і транспортування ставлять у вигідне становище ті регіони, які володіють

багатими природними ресурсами» [27].

Проблеми впливу природних умов і ресурсів на економічний розвиток регіонів перебували в центрі уваги вчених вже тоді, коли під керівництвом академіка В.І. Вернадського в Україні тільки відбувалося становлення і розвиток науки про розміщення продуктивних сил і регіональної економіки. Близькими до природних ресурсів за значенням і змістом є екологічні умови, поліпшення екологічної ситуації – це неодмінна умова збереження навколишньої природи і природи самої людини.

У монографії З.С. Варналій зазначає: «У складі природно-ресурсного потенціалу необхідно виділяти екологічний потенціал, оскільки вже сьогодні, і особливо в перспективі, екологічна ситуація в регіонах буде визначати можливості розвитку і розміщення продуктивних сил і рівень сприятливості територій для проживання і життєдіяльності людей» [26].

Поряд з природними ресурсами і екологічними умовами значну роль у розвитку регіону відводять населенню, оскільки людина є основним творцем суспільного багатства, а чисельність населення, кваліфікація його працездатної частини є фактором, що обумовлює економічний розвиток.

Розвиток також неможливий і без засобів виробництва, виробничої і соціальної інфраструктури. Тому важливе значення для розвитку як регіону, так і країни має виробничий і науково-технічний потенціал, який знаходить відображення в сукупної можливості галузей народного господарства виробляти промислову і сільськогосподарську продукцію, здійснювати інвестиційну та інноваційну діяльність.

Велике значення кожної зі складових економічного потенціалу регіону. А це означає, що економічний потенціал - це сукупність засобів і умов для ведення господарської діяльності, тобто можливість і здатність створювати, творити, розвиватися.

Зазвичай економічний розвиток пов'язують з економічним зростанням, і на те є причини, однак, як вказано у статті Л.М. Кузьменко [29]: «Відмінність між «зростанням» і «розвитком» полягає в тому, що коли щось «росте», воно стає більше кількісно, а коли щось розвивається, воно стає якісно іншим».

В даний час пріоритетність якості розвитку над кількісними економічними показниками стає найбільш актуальним питанням, так як у різнобічної господарської діяльності, покладеної в основу економічного розвитку суспільства, є й інша сторона - екологічні наслідки, які можуть звести нанівець усі досягнення економіки, особливо з урахуванням останніх екологічних катаклізмів.

Так, наприклад, в праці П. Самуельсона вказано, що не можна прагнути лише до кількісного зростання виробництва (нехай навіть в розрахунку на душу населення). Необхідно також внутрішній розвиток всієї соціально-економічної структури суспільства. В даний час багато як зарубіжні, так і вітчизняні автори відзначають, що безконтрольний кількісне зростання може створити загрозу існування для всього людства, оскільки він породжує екологічну, енергетичну, сировинну та деякі інші глобальні проблеми [30].

Сьогодні сучасній економіці України притаманне високий вміст енергоємних і матеріаломістких технологій виробництва, а також здійснення господарської діяльності без урахування екологічних вимог. Принцип господарювання, заснований на екстенсивному використанні природних ресурсів і витратному механізмі виробничих процесів, був характерний для України багато десятиліть і зумовив нинішній рівень природотехногенной безпеки.

У всіх без винятку областях України орні землі мають деформовану ґрунтову структуру, катастрофічно знижується родючість земель: за період 2001 - 2003 рр. дефіцит гумусу в ґрунті склав 50%, частка забруднених важкими металами ґрунтів наближається до 40%, в тому числі надмірно забруднених - 13% від загальної земельної території країни і продовжує зростати.

Україна належить до регіонів недостатньо забезпеченим прісною водою, проте інтенсивність використання водних ресурсів набагато перевищує процес їх відновлення в біосфері. Протягом 2000 року в водні об'єкти України надійшло 8246 млн т забруднюючих речовин. У басейнах річок Дніпро, Дністер, Західний Буг, Південний Буг, Сіверський Донець, а також в Київському, Канівському, Кременчуцькому та Дніпродзержинському водосховищах, водах Північно-Кримського каналу виявлено високий вміст важких металів, сульфатів, сполуки

азоту, фенолів. Є серйозні проблеми, пов'язані зі зміною гідрологічного режиму малих річок, якісним станом вод Чорного і Азовського морів.

Не відповідає еколого-економічним вимогам і стан лісів України. Практично всі лісові масиви нашої країни знаходяться в зоні негативного промислового впливу. Ліси в різного ступеня забруднені радіонуклідами на площі 3,5 млн га, вилучено з експлуатації 200 тис. Га. Внаслідок цього в останні роки не добирається приблизно 1 млн м3 деревини щорічно, значно зменшилися обсяги грибів, ягід, лікарської сировини.

І це далеко не повний список проблем, обумовлених нераціональним способом господарювання і є серйозною перешкодою на шляху стійкого економічного розвитку. До того ж у всьому світі головним показником стійкості соціально-економічного розвитку є підвищення рівня якості життя населення, а це, перш за все, здоров'я і збільшення тривалості життя людей, багато в чому залежать від навколишнього природного середовища. Прогнозуючи і чекаючи результати економічної діяльності, необхідно в першу чергу враховувати ту шкоду, яку в результаті деградації природного середовища може бути заподіяна людині, а в його особі - трудових ресурсів, трудового потенціалу суспільства.

Ніякі економічні вигоди не можуть бути виправданням збільшення захворюваності, інвалідності та смертності населення, погіршення його фізичного і психічного здоров'я.

Тому сутність економічного розвитку сьогодні має полягати у вирішенні основного протиріччя економіки - між можливостями природного середовища та зростаючими потребами суспільства. Саме екологічний імператив як принцип розвитку економіки повинен надати соціально-економічного розвитку господарюючих суб'єктів будь-якого рівня (регіон, держава, світове співтовариство) нову якість. І починання втілювати в життя цю концепцію, визнану актуальною в усьому світі і увійшла в процес соціально-економічних перетворень під назвою «сталий розвиток», необхідно саме на регіональному рівні.

Сьогодні регіони України в своєму соціально-економічному розвитку повинні забезпечувати економічне зростання за допомогою такої господарської діяльності,

яка і за своєю структурою, і за своїм змістом не суперечила б принципам сталого розвитку, сприяла створенню сприятливих умов для життя і соціального благополуччя населення, рівень якого визначається не тільки економічними досягненнями, а й ступенем екологічної безпеки.

1.3 Постановка задачі

Мета атестаційної роботи полягає в якісному аналізі та оцінці рівня життя населення в регіонах України під впливом структурних економічних змін в країні на основі використання багатомірних методів аналізу.

Для досягнення визначеної мети необхідно:

– проаналізувати сучасні підходи задля подальшої оцінки рівня розвитку регіонів;

– побудувати концептуальну схему моделювання рівня регіонів для проведення соціально-економічного аналізу рівня розвитку регіонів;

– обґрунтувати вибір методів багатовимірного аналізу даних для визначення соціально-економічного аналізу рівня розвитку регіонів;

– обґрунтувати вибір методів багатовимірного аналізу даних для прогнозування рівня розвитку регіонів;

– визначити регіони України в контексті їхньої кластеризації;

– визначити таксономічні показники регіонів України;

– провести кластеризацію регіонів України за рівнем соціально-економічного розвитку;

– проаналізувати стан і тенденції зміни рівня розвитку регіонів України;

– побудувати прогнозну модель рівня розвитку регіонів України.

2 ТЕОРЕТИКО-МЕТОДОЛОГІЧНІ ОСНОВИ АНАЛІЗУ РІВНЯ СОЦІАЛЬНО-ЕКОНОМІЧНОГО РОЗВИТКУ РЕГІОНІВ

2.1 Розробка концептуальної схеми моделювання рівня соціально-економічного розвитку регіонів з визначенням системи індикаторів

Рівень-розвитку регіону – це індикатор, за яким оцінюють успішність країни чи окремого її регіону. Досягнення високого рівня розвитку, аналогічного рівню в країнах Європи є можливим для України за умови впровадження європейських соціальних стандартів у свою практику.

Подолання відставання України від країн Європейського Союзу зазначає необхідність послідовного впровадження принципів соціальної ринкової економіки, яка характеризується розвиненими ринковими відносинами, високим рівнем економічного розвитку, політичною демократією, гарантованим доступом до системи освіти та охорони здоров'я, розвиненою системою соціального захисту.

Отже, для якісного аналізу рівня соціально-економічного розвитку регіонів України та його оцінки, а також для вирішення зазначених проблем запропонована і використовується концептуальна схема моделювання рівня соціально-економічного розвитку регіонів, що подана на рис. 2.1.

Розглянемо докладніше основні етапи побудованої моделі, методи, що застосовуються на відповідному етапі та відібрані показники, на основі яких проводилися розрахунки. Перший етап дослідження полягає у формуванні масивів вихідних даних, що в подальшому використовуються в якості показників аналізу рівня життя населення та системи показників за блоками для оцінки рівня соціально-економічного розвитку регіонів.

Основним методом для обробки інформації є метод синтезу та аналізу інформації, що базується на аналізу категорійного базису та аналізу сучасних підходів до оцінки рівня життя населення (рис. 2.1).

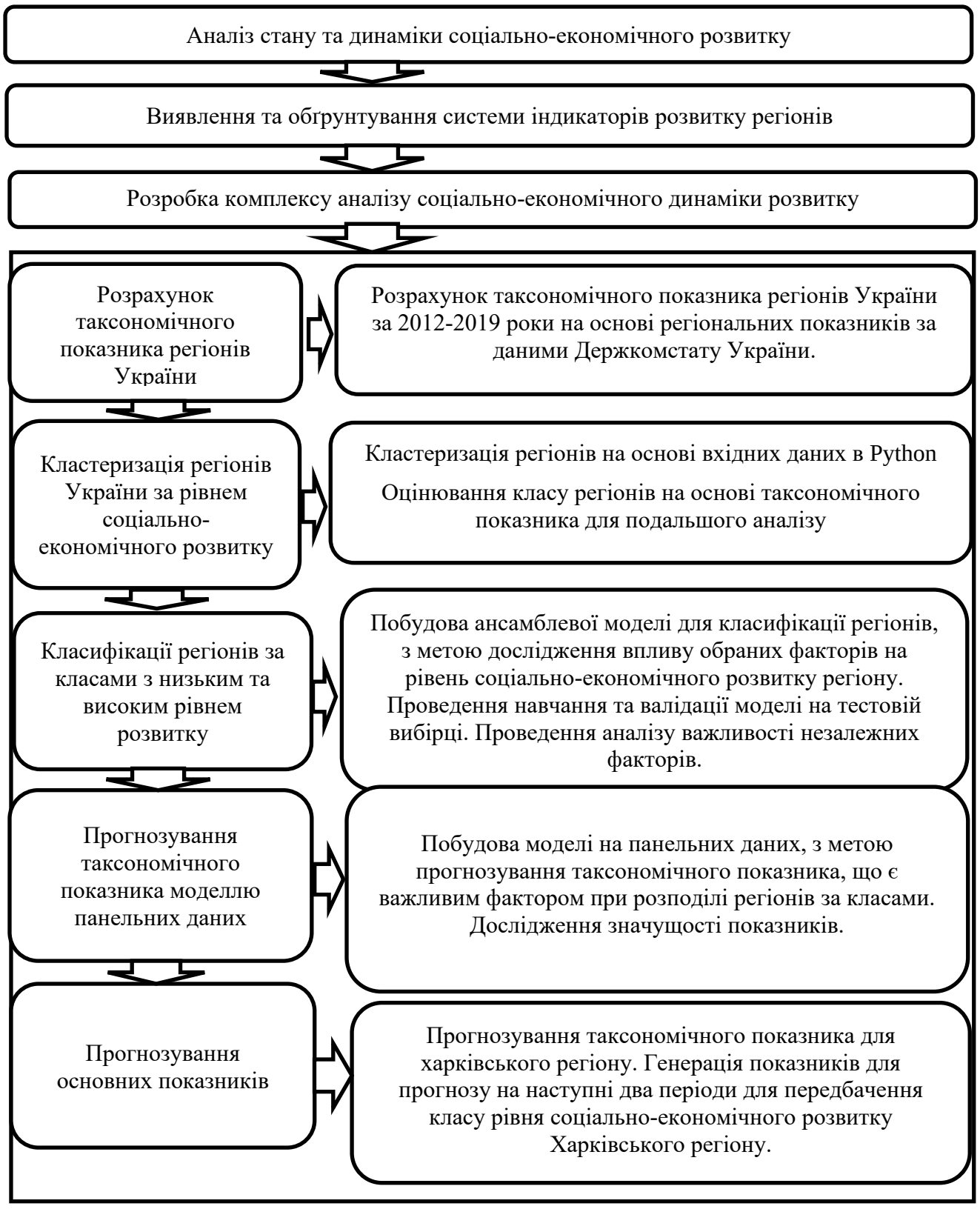

Рис. 2.1. Концептуальна схема моделювання рівня соціально-економічного розвитку регіонів

Другий етап концептуальної схеми дослідження (рис. 2.1) розкриває процес побудови моделей аналізу соціально-економічного рівня розвитку регіонів на основі методів багатовимірної статистики. Як зазначено в пп. 1.2, об’єкт дослідження немає прямого індикатора, що відображав би рівень розвитку регіонів, широкий спектр показників з економічної, соціальної, політичної, екологічної площин в цілому транслює реальний стан розвитку.

В даному дослідженні для використання методів багатовимірної статистики, а саме – методів канонічного, кластерного аналізів, масив вихідних даних був сформований з таких регіональних показників: середня заробітна плата, економічно активне населення, безробітне населення (за методологією МОП), валовий регіональний продукт у розрахунку на одну особу, індекс промислової продукції, сальдо (експорт-імпорт), капітальні інвестиції та обсяги викидів забруднюючих речовин у період с 2012 по 2019 роки.

Отже, вдалося задовольнити потребу в адекватному інструментарію та статистичній інформації, за допомогою яких можна визначити рівень соціально-економічного розвитку регіонів України в контексті впливу регіонів одного на інший, спостереженні процесів конвергенції, в контексті низького рівня життя в Україні, та аналізі вирішення цих проблем.

Для вибору методів багатовимірної статистики згідно концептуальної схеми моделювання (рис. 2.1) вирішувалися такі завдання:

– вибір методу будування ансамблевої моделі, що дозволяє отримати ймовірність потрапляння регіону в той чи інший клас;

– вибір методу для розрахування індексу, що визначає рівень розвитку регіону в той чи інший рік;

– будування регресійної моделі на панельних даних, щоб в подальшому прогнозувати таксономічних індекс для регіону;

– вибір методу кластеризації регіонів для визначення двох класів – класу регіонів з високим та низьким рівнем розвитку.

– вибір методу прогнозу показників розвитку регіонів на наступні періоди для Харківського регіону;

– проведення експериментального моделювання рівня соціально-економічного розвитку регіонів для формування комплексних висновків щодо стану та тенденції змін рівня соціально-економічного розвитку регіонів України під впливом процесів структурних економічних змін в Україні.

2.2 Вибір методів багатовимірної статистики згідно концептуальної схеми моделювання рівня соціально-економічного розвитку регіонів

Згідно до концептуальної схеми (рис. 2.1), на першому етапі здійснювалася розробка системи індикаторів соціально-економічного розвитку регіонів. Наявність великої кількості факторів, які визначені у пп. 1.3, визначає необхідність використання методів редукції з метою скорочення інформаційного простору ознак.

Редукція полягає в зведенні складного до простого і вищого до нижчого, ігноруючи специфіку вищих рівнів. Завдання полягало у тому, щоб представити вихідну інформацію, яка задана у вигляді декількох критеріїв опису, у множині меншої розмірності, з додатковою задачею – по можливості, мінімізувати втрати інформації.

Методи редукції сукупності ознак діляться на дві групи [5]:

– методи неповної редукції;

– методи повної редукції.

Повна редукція полягає у побудові синтетичних значень у вигляді деякої функції $f(y_1, y_2, \dots, y_n)$, яка відображає вплив усіх ознак, що дозволяє таким чином впорядковати усі об’єкти.

Неповна редукція приводить до отримання діагностичних ознак, які є частиною вихідних ознак. При цьому первісний набір q ознак $y = (y_1, y_2, \dots, y_q)$ замінюється набором s діагностичних ознак $x = (x_1, x_2, \dots, y_s), (s < q)$. Ця група методів дозволяє виключити з первісної системи ознак ті, які дублюють інформацію, а також забезпечують вибір ознак, що найбільш повно відображають стан досліджуваних процесів.

До складу методів неповної редукції входять [6]:

– метод головних компонент;

– факторний аналіз;

– метод центру ваги.

Повна редукція представлена таксономічним показником рівня розвитку.

Метод неповної редукції – центру ваги використовується для знаходження діагностичних ознак (репрезентантів), тобто ознак, що передають найістотніші особливості численного набору вихідних ознак.

Важливою умовою застосування методів неповної редукції є відповідність діагностичних ознак властивостям:

– ознаки некорельовані або слабко корельовані між собою (коефіцієнт кореляції менше ніж «0,1» або «0,3» відповідно);

– сильно корельовані з ознаками, що не входять у діагностичний набір (коефіцієнт кореляції більш ніж «0,5»).

Алгоритм методу центру ваги включає такі основні кроки:

Крок 1. На першому кроці алгоритму формуються матриці вихідних даних за кожною групою показників стану об'єкта дослідження $Y_1, Y_2, \dots, Y_q$, де q – кількість груп показників. Для $k-$ї групи показників структура матриці може бути визначена у такий спосіб:

$$Y_k = (y_{ij})_k, i = [1; m], j = [1; n], \tag{2.1}$$

де y_{ij} – значення i-го показника в j-ому досліджуваному періоді;

m – кількість показників, що входять у k-у групу;

n – кількість досліджуваних періодів.

Крок 2. Оскільки показники можуть бути виражені в абсолютних і відносних величинах, а також мати різні одиниці виміру, то на другому кроці здійснюється процедура їх стандартизації за формулою:

$$z_{ij} = \frac{y_{ij} - \overline{y_i}}{S_i}, \tag{2.2}$$

де z_{ij} – стандартизоване значення i-го показника в j-ому досліджуваному періоді;

$\overline{y_i}$ – середнє арифметичне значення i-го показника;

S_i – стандартне відхилення i-го показника.

Результатом цього кроку є набір матриць стандартизованих значень показників кожної групи $Z_1, Z_2, \dots, Z_q$.

Крок 3. Описані вище обчислювані процедури є основою для розрахунку матриць відстаней $\mathrm{P}_1, \mathrm{P}_2, \dots, \mathrm{P}_q$, елементи яких відображають ступінь близькості показників усередині кожної групи. Як міра відстані використовується Евклідова відстань, що визначається формулою:

$$\rho(z_i, z_j) = \sqrt{\sum_{t=1}^{n} (z_{it} - z_{jt})^2}, \tag{2.3}$$

де $\rho(z_i, z_j)$ – відстань між і-им та j-им показником групи;

z_{it}, z_{jt} – стандартизовані значення i-го й j-го показників групи у періоді t.

Крок 4. На четвертому кроці здійснюється вибір так званих показників-репрезентантів груп, які несуть у собі найбільш значну інформацію. Спосіб вибору репрезентантів за методом центру ваги залежить від розміру групи. Вирізняють групи з одного, двох та більшою кількістю елементів:

– у групах з одного елементу утворюючі їх показники мають властивості, які сильно відрізняються від показників інших груп, тому вони належать до числа показників-еталонів (репрезентантів);

– у групах, де кількість показників більша двох, розраховується сума відстаней кожного показника до інших показників групи:

$$\rho_i = \sum_{\substack{j=1 \\ j \neq i}}^{m} \rho(z_i, z_j), \tag{2.4}$$

де m – число показників групи.

До складу показників-репрезентантів входить показник з найменшою сумою відстаней: $\rho_s = min_i \rho_i$;

У групах де кількість показників дорівнює двом, визначається сума відстаней показників, що входять у групу, від показників репрезентантів, обраних за описаними вище правилами. До репрезентантів належить той показник, у якого сума відстаней від відособлених елементів і елементів-репрезентантів, виділених із груп із числом більше двох максимальна: $\rho_s = max_i \rho_i$.

Отже, результатом 4-го кроку є набір показників-репрезентантів $x = (x_1, x_2, \dots, x_k)$, що описують найбільш важливі аспекти стану об'єкта дослідження.

Описаний вище метод обраний для виділення групи показників-репрезентантів, що чинять вплив на функціонування. На основі виділених показників здійснюється побудова інтегрального показника до та після скорочення сукупності інформаційного простору з метою перевірки адекватності побудованої моделі.

Перейдемо до розгляду теоретичних умов виділення інтегральних показників, що відображають рівень розвитку об'єкту.

Для зіставлення об'єктів, які характеризуються великою кількістю ознак, найчастіше застосовуються таксономічні процедури. Одним з методів дослідження багатомірних об'єктів є таксономічний показник рівня розвитку, запропонований Хельвігом [7]. Цей показник є синтетичною величною, «рівнодіючою» всіх ознак, що характеризують об'єкти, і дозволяє лінійно впорядкувати елементи досліджуваної сукупності.

Першим кроком процесу побудови таксономічного показника рівня розвитку є визначення елементів матриці спостережень, що може бути представлена в такий спосіб:

$$X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1j} & \dots & x_{1m} \\ x_{21} & x_{22} & \dots & x_{2j} & \dots & x_{21} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ x_{i1} & x_{i1} & \dots & x_{ij} & \dots & x_{im} \\ x_{\omega 1} & x_{\omega 2} & \dots & x_{\omega j} & \dots & x_{\omega m} \end{bmatrix}, \quad (2.5)$$

де ω – кількість досліджуваних об’єктів,

m – кількість ознак,

x_{ij} – значення j-ї ознаки для і-го об'єкта.

Оскільки ознаки, що включаються в матрицю спостережень, неоднорідні, проводиться стандартизація їх значень за формулою (2.2).

Наступний крок у розглянутій процедурі полягає в диференціації ознак матриці спостережень. Усі зміні поділяються на стимулятори й дестимулятори. Підставою поділу ознак на дві групи служить характер впливу кожного з них на рівень розвитку досліджуваних об’єктів. Ознаки, що позитивно, стимулююче впливають на рівень розвитку об’єктів, називаються стимуляторами, на відміну від ознак – де-стимуляторів.

Поділ ознак на стимулятори й де-стимулятори є основою для побудови так званого еталону розвитку, що є точкою з координатами:

$$P_0 = (z_{01}, z_{02}, \dots, z_{0s}), \quad (2.6)$$

де $z_{0s} - max_r z_{rs}$, якщо $S \in I$;

$z_{0s} - min_r z_{rs}$, якщо $S \ni I, (s = 1, \dots, m)$;

I – безліч стимуляторів;

z_s – стандартизоване значення ознаки s для об’єкта r.

Відстань між окремими точками-одиницями і точкою P_0, що є еталоном розвитку, позначається c_{i0} й розраховується у такий спосіб:

$$c_{i0} = \sqrt{\sum_{j=1}^{m} (Z_{ij} - Z_{oj})^{2}}. \quad (2.7)$$

Отримані відстані служать вихідними величинами, що використовуються при розрахунку показника рівня розвитку:

$$d_i^* = 1 - \frac{c_{i0}}{c_0}, \quad (2.8)$$

$$c_0 = \bar{c_0} + 2 * S_0; \quad (2.9)$$

$$\bar{c_0} = \frac{1}{w} \sum_{i=1}^{w} c_{i0}; \quad (2.10)$$

$$S_0 = \sqrt{\frac{1}{w} \sum_{i=1}^{w} (c_{i0} - \bar{c_0})^2}. \quad (2.11)$$

Інтерпретація показника рівня розвитку така: чим ближче значення показника рівня розвитку до одиниці, тим на більш високому рівні розвитку перебуває об’єкт.

На основі методів редукції та показника рівня розвитку мають виділятися основні ознаки-репрезентанти, що чинять вплив на формування рівня розвитку регіонів. Якість мають оцінюватися на базі коефіцієнта кореляції між інтегральними показниками розрахованими на базі первісної та скороченої системі ознак [20-21].

Отже, на основі вище розглянутих алгоритмів та принципів кластеризації здійснюється розбиття країн на групи та визначаються дані для дослідження до певної групи країн. Після того як країни (регіони країн) обрані, необхідно здійснити їх класифікацію з урахуванням виділеної за допомогою методів редукції ознак-репрезентантів. Для обрання методів має бути побудована ансамблева модель та модель панельних даних.

Панельні дані – прогнозовані просторові вибірки, де кожен об'єкт спостерігається багаторазово протягом відрізка часу [8].

Традиційно лінійна залежність панельних даних для і-го об'єкта обраної генеральної сукупності має вигляд:

$$y_{it} = a_i + x'_{it}\beta + \varepsilon_{it}. \quad (2.14)$$

Параметри α_i є адаптивними константами, які підсумовують ефекти, характерні для конкретного об'єкта спостереження і періоду часу, а значить визначають середнє місце розташування y_{it}, якщо регресори зафіксовані на рівні $x_{it} = 0$. Вільні коефіцієнти прийнято називати ефектами

Найбільш прості припущення можуть мати такий вигляд:

$$y_{it} = x'_{it}\beta + \varepsilon_{it}. \quad (2.15)$$

Модель 2.13 не передбачає ніяких ефектів, характерних для окремих об'єктів спостереження або моментів часу. В такому випадку мають на увазі, що дані об'єднані [9]. Параметри моделі оцінюються за допомогою МНК за всіма спостереженнями, не враховуючи специфіку панельних даних.

У моделях панельних даних відхилення поділяються на кілька компонент. Виділяють моделі з одно- і двокомпонентною помилкою. Найбільш поширені моделі з однокомпонентною складовою помилки. У свою чергу одно- і двовимірні шоки можуть представляти фіксовані і випадкові ефекти. У першому випадку досліджують вплив специфічних для кожного об'єкту факторів, виражених у значеннях констант α_i при відсутності загального параметра розташування.

Залежно від припущення стосовно характеру величини α_i розглядаються дві моделі: модель з фіксованими і модель з випадковими ефектами. Модель з фіксованими ефектами має вигляд:

$$y_{it} = a_i + x'_{it}\beta + \varepsilon_{it}. \quad (2.16)$$

При цьому необхідно, щоб виконувалися наступні умови:

– помилки некорельовані між собою за параметрами i, і t, тобто:

$$E(\varepsilon_{it}) = 0, V(\varepsilon_{it}) = \sigma_{\varepsilon}^2, \qquad (2.17)$$

– помилки є коррельованими з регресорами.

Модель також може розглядатися як модель з індивідуальними фіктивними змінними, тобто для кожного об'єкта спостереження вводиться змінна, що має індивідуальний характер [12]. Припускаючи наявність одних і тих же параметрів для всіх об'єктів спостереження в усі моменти часу, можна досліджувати наявність гетерогенності між об'єктами спостереження з інваріантним по відношенню до часу, але специфічним для кожного об'єкта спостереження параметром місцеположення.

Якщо ввести такі фіктивні змінні:

$$d_{ij} = \begin{cases} 1, i = j \\ 0, i \neq j \end{cases}. \qquad (2.18)$$

Модель можна представити у вигляді стандартної моделі:

$$y_{it} = \sum_{j=1}^{n} a_j d_{ij} + x_{it}'\beta + \varepsilon_{it}. \qquad (2.19)$$

Основним недоліком моделі з фіксованими ефектами є те, що у моделі необхідно оцінювати велику кількість параметрів, що веде до втрати ступенів свободи. Велика кількість фіктивних змінних, що включаються, ускладнює проблеми колінеарності. Виникає ситуація, коли присутність сильного впливу будь-якої змінної в реальних змінних оцінюється як слабка, тоді як і фіктивні змінні, звертають на себе більшу частину такого взаємозв'язку. Тому для оцінювання параметрів моделі має використовуватись внутрішнє групове перетворення [15].

Модель з випадковими ефектами адаптована до структур панельних даних, що сприяє усуненню деяких недоліків моделі з фіксованими ефектами, особливо для проблеми числа оцінюваних параметрів. Свою сутність випадкові ефекти висловлюють у тому, що ефекти u_i описуючі гетерогенність, є випадковими змінними в сенсі випадковості вибірки з генеральної сукупності, так як кожен об'єкт спостереження має специфічний, чи не залежний від часу, ефект. Тобто, вибірка, що включає досліджувані об'єкти, розглядається як випадкова з деякої генеральної сукупності. Як і моделі з фіксованими ефектами, випадкові ефекти відображають наявність деяких індивідуальних характеристик об'єктів, інваріантних в часі, які складно спостерігати [13]. Однак значення цих характеристик вводяться до складу помилки.

Рівняння моделі з випадковими ефектами має вигляд:

$$y_{it} = \mu + x'_{it}\beta + u_i + \varepsilon_{it}, \qquad (2.20)$$

де u_i – випадкова помилка, інваріантна в часі для кожного об'єкту;

μ – константа.

У моделі робляться таки припущення, що виконуються за такими умовами:

– помилки ε_{it} не корелюють між собою $E(u_{it}) = 0, V(u_{it}) = \delta_u^2$;

– помилки ε_{it} не корелюють з регресорами x_{js} при всіх i, t, j, s;

– помилки u_i некоррельовані $E(u_{it}) = 0, V(u_{it}) = \delta_u^2$;

– помилки u_i некоррельовані з регресорами $\mathrm{x_{js}}$ при всіх i, t, j, s;

– помилки u_i та ε_{it} некоррельовані при всіх i, t, j, s;

Отже, в моделі передбачається наявність однакових параметрів для всіх об'єктів спостереження в усі моменти часу, проте досліджується ефект гетерогенності об'єктів спостереження за допомогою введення постійного за часом, але індивідуального для кожного об'єкта спостереження доданка помилки u_i , яка є незалежною від решти помилки.

Оскільки у моделях панельних даних можуть бути індивідуальні розбіжності через один з об'єктів у досліджуваній вибірці, то виявлення наявності таких розбіжностей і відповідного виду специфікації моделі здійснюють за допомогою ряду статистичних тестів.

Для прийняття рішення, розглядалися три основні специфікації моделі панельних даних: проста модель панельних даних, регресія з фіксованим ефектом і регресія з випадковими індивідуальними ефектами. Щоб виявити, яку з моделей слід застосовувати, проводилося попарне порівняння оцінюваних моделей за допомогою F-тесту, тесту Бреуша-Пагана і тесту Хаусмана [14].

Статистична перевірка параметрів проводилася так. Для перевірки статистичної значущості параметрів перетину у моделях формувалася нульова гіпотеза H_0 така, що H_0: $\mu_i = \mu_j$ для будь-яких i, j, що відповідає моделі з одним і тим же параметром μ для всіх об'єктів вибірки, тобто об'єднаної моделі. Альтернативна гіпотеза полягала у тому, що H_1: $\mu_i \neq \mu_{j}$ j хоча б для однієї пари i, j, що відповідає моделі з фіксованими ефектами. Дана гіпотеза перевірялася за допомогою F-тесту:

$$F = \frac{R_{FE}^2 - R_{pool}^2}{1 - R_{FE}^2} * \frac{nT - n - d}{n - 1} = \\ = \frac{Q_{pool} - Q_{FE}}{Q_{FE}} * \frac{nT - n - d}{n - 1} \sim F(n - 1, nT - n - d), \quad (2.21)$$

де R_{FE}^2 – коефіцієнт множинної кореляції моделі з фіксованими ефектами;

R_{pool}^2 – коефіцієнт множинної кореляції об'єднаної моделі;

Q_{pool} – сума квадратів залишків об'єднаної моделі;

Q_{FE} – сума квадратів залишків моделі з фіксованими ефектами;

n – кількість об’єктів дослідження;

T – кількість періодів;

d – кількість незалежних змінних моделі.

Якщо справедлива гіпотеза H_0 і виконується передумова про нормальний розподіл помилок, тестова статистика має F-розподіл з (n-1) і (nT-n-k) ступенями свободи [15].

Перевірку на значимість випадкових ефектів моделі здійснюють за допомогою тесту множників Лагранжа, запропонованого Бреушем і Паганом, що базується на відповідній статистиці:

$$LM = \frac{nT}{2(T-1)} * \left(\frac{\sum_{i=1}^{n} (\sum_{t=1}^{T} e_{it})^2}{\sum_{i=1}^{n} \sum_{t=1}^{T} {e_{it}}^2} - 1 \right)^2, \quad (2.22)$$

де e_{it} – залишки об'єднаної моделі регресії.

З цього тесту висувають нульову гіпотезу H_0, яка полягає у тому, що об'єднана модель регресії є частковим випадком моделі з випадковими ефектами, де відсутні помилки u_i або $\sigma_u^2 = 0$. Відповідно в гіпотезі H_1 приймається $\sigma_u^2 > 0$. Якщо гіпотеза H_0 є вірною і виконується передумова щодо нормального розподілу помилок, статистика LM має асимптотичний χ^2 розподіл з одним ступенем свободи [16].

Оскільки найважливішою відмінністю у підходах до моделювання гетерогенності об'єктів спостереження є співвідношення ефектів, що включаються, з регресорами: випадкові ефекти мають не корелювати з регресорами, у той час як фіксовані ефекти можуть з ними корелювати, вибір моделі з фіксованими або випадковими ефектами залежить від того, корелюють ефекти з регресорами чи ні.

При справедливості гіпотези H_1 оцінки моделі з фіксованими ефектами спроможні, а оцінки моделі з випадковими ефектами неспроможні. У цьому випадку можна очікувати суттєві відмінності між оцінками даних двох моделей. Виявлення такої відмінності досліджують за допомогою відповідного тесту, що базується на статистиці Хаусмана:

$$H = \left(\widehat{\beta_{FE}} - \widehat{\beta_{RE}}\right)` \widehat{\Phi}^{-1} \left(\widehat{\beta_{FE}} - \widehat{\beta_{RE}}\right), \quad (2.23)$$

де $\widehat{\Phi}$ – оцінка матриці коваріацій $\left(\widehat{\beta_{FE}} - \widehat{\beta_{RE}}\right)$, що має асимптотичний розподіл χ^2 з d ступенями свободи;

$\widehat{\beta_{FE}}$ – вектор оцінок моделі з фіксованими ефектами;

$\widehat{\beta_{RE}}$ – вектор оцінок моделі з випадковими ефектами.

З цього тесту на підставі отриманих оцінок моделей з фіксованими і випадковими ефектами перевіряють гіпотезу H_0 – таку, що оцінки моделі з випадковими ефектами є обґрунтованими і не мають відрізнятися від оцінок моделі з фіксованими ефектами. Відповідно, якщо справедлива альтернативна гіпотеза H_1, різниця між оцінками моделі з випадковими і фіксованими ефектами є суттєвою, але оцінки моделі з фіксованими ефектами є обґрунтованими [17-18].

Результати проведення тестів дозволили обрати специфікацію регресії на панельних даних, що найкращим чином описує залежність між результуючою та факторною змінною.

Для побудови моделі класифікації, було обрано ансамблеву модель, що базується на ансамблі дерев прийняття рішень. Це модель де одночасно працюють багато дерев. Саме по собі дерево, це не зовсім надійний алгоритм, він підходить далеко не до всіх задач, тому що якщо трохи змінити свої дані, створене дерево рішень може бути дуже різним.) У такому випадку може бути створена надійна модель (зі зменшеним значенням дисперсії) за допомогою пакетування, коли створюються різні моделі, шляхом перекомпонування даних, щоб зробити отриману модель більш надійною – тобто утворюється ансамбль з дерев [25].

Існує два типи моделей на основі ансамблів дерев – це беггінг (Bagging) та бустинг (Boosting).

На основі беггінгу працює RandomForest, що є метаоцінювачем, який підходить до ряду класифікаторів дерев рішень на різних під-зразках набору даних і використовує усереднення для підвищення точності прогнозування та контролю над приміркою. Розмір підпроби завжди такий самий, як вихідний розмір вхідного зразка, але зразки витягуються із заміною[lib site]. Тобто він будує багато дерев рішень, та розділяє всідні дані на невеликі шматки, на яких будує прогноз, а остаточний результат визначається методом середнього арифметичного.

На основі бустингу працює XGB алгоритм (метод градієнтного спуску), на відмінну від беггінгу, він навчається на своїх помилках. Алгоритми схожі, але якщо в RandomForest всі дерева працюють паралельно, то в бустингу помилки першого дерева направляються до наступного, щоб алгоритм краще засвоїв свої помилки і так

він повторює до останнього дерева, а при обчисленні кінцевого результату він бере до уваги ті місця де він помилявся.

Бустинг зменшує дисперсію та упередженість моделі, оскільки ми використовуємо декілька моделей (ансамбль). Також в параметрах, можна обрати метод бустингу, одним з яких є Adaboost. Це оригінальний алгоритм; ви говорите наступним моделям, щоб покарати більш суворі спостереження, помилкові твердження попередньої моделі. Підвищення градієнта: тренуємо кожну наступну модель за допомогою залишків (різниця між передбачуваними та справжніми значеннями).У цих ансамблях базовий учень повинен бути слабким. Якщо це перевищує дані, не буде жодних залишків або помилок для наступних моделей, на яких можна розвиватись.

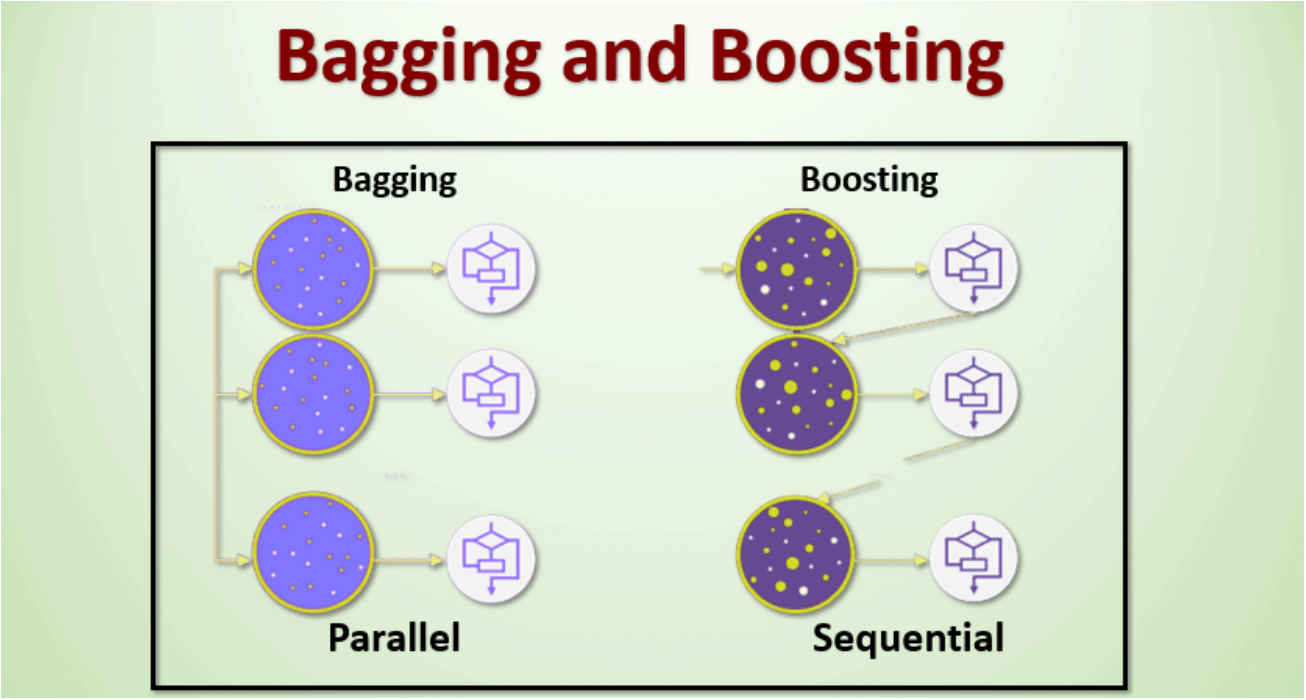

Рис. 2.3. Приклад роботи бустингу та беггінгу

Збільшення градієнта, зокрема, також є досить гарним методом моделювання, що забезпечує простий спосіб використання різних функцій втрат, навіть коли похідна не опукла. Наприклад, використовуючи ймовірнісний прогноз, можна використовувати функцію пінболу і функцію втрати, що складніше, порівняно з нейронними мережами (тому що похідна завжди постійна), рис.2.3 [19].

Отже, XGBoost – це оптимізована бібліотека для збільшення градієнтів, розроблена як високоефективна, гнучка та портативна. Вона реалізує алгоритми машинного навчання в рамках Gradient Boosting. XGBoost забезпечує паралельне збільшення дерев (також відоме як GBDT, GBM), яке швидко та точно вирішує багато проблем із інформацією про дані. Цей же код працює у великих розподілених середовищах (Hadoop, SGE, MPI) і може вирішувати проблеми за мільярдами прикладів. Оскільки ми працюємо на історичних даних, ми будемо використовувати сама бустингову модель.

2.3 Висновки за розділом 2

У другому розділі даної роботи була побудована концептуальна схема моделювання рівня соціально-економічного розвитку регіонів України, що включає всі проведені етапи дослідження. Також був визначений математичний інструментарій моделювання рівня життя населення – методи багатовимірного аналізу рівня розвитку регіонів, методи побудови таксономічного показника для оцінки соціально-економічного розвитку регіонів.

Для визначення наявності взаємозв’язку між регіональними показниками соціального та економічного аспектів розвитку країни серед розглянутих методів в другому розділі даної роботи був обраний канонічний аналіз кореляцій.

Вибір методу кластерного аналізу зумовлений класифікацією регіонів за рівнем розвитку, що стало основою для проведення класифікації регіонів та побудови моделі панельних даних.

Вихідними даними для побудови моделей багатовимірної статистики виступають регіональні статистичні показники – середня заробітна плата, економічно активне населення, безробітне населення, валовий регіональний продукт у розрахунку на одну особу, індекси промислової продукції, сальдо (експорт-імпорт), капітальні інвестиції, обсяги викидів забруднюючих речовин у період з 2012 по 2019 роки.

3 АНАЛІЗ ТА ЕКСПЕРЕМЕНТАЛЬНА ОЦІНКА РІВНЯ СОЦІАЛЬНО-ЕКОНОМІЧНОГО РОЗВИТКУ РЕГІОНІВ

3.1 Визначення послідовності використання розроблених програмних застосувань для експериментальної оцінки рівня соціально-економічного розвитку регіонів

Структура послідовності експериментального дослідження з використання програмних застосувань подана на рис. 3.1.

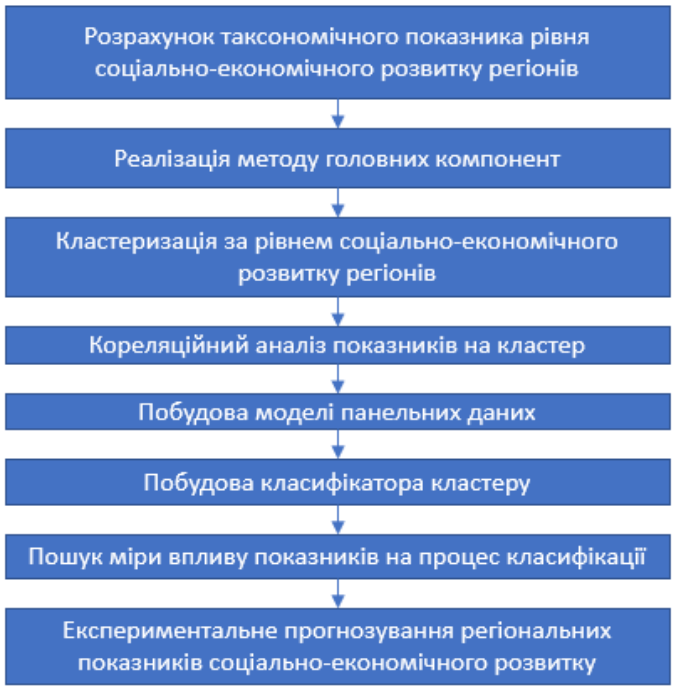

Рис. 3.1. Структура послідовності експериментального дослідження з використання програмних застосувань

Для проведення експериментальної оцінки рівня соціально-економічного розвитку регіонів використовується таке програмне забезпечення:

– для розрахунку таксономічного показника рівня соціально-економічного розвитку регіонів методом Хельвіга створений шаблон з використанням функцій пакету Microsoft Office Excel;

– для реалізації методу головних компонент таксономічного показника створений шаблон з використанням функцій пакету Microsoft Office Excel;

– для кластеризації регіонів за рівнем соціально-економічного розвитку з використанням алгоритму DBSCAN розроблено програмне застосування мовою Python (код програми поданий у додатку Б);

– для кореляційного аналізу показників на кластер отриманого алгоритмом DBSCAN з використанням алгоритму кореляції Пірсона розроблено програмне застосування мовою Python (код програми поданий у додатку Б);

– для побудови моделі панельних даних кластеру отриманого алгоритмом DBSCAN розроблено програмне застосування мовою Python, що використовує модель PanelOLS (код прогами поданий у додатку Б);

– для побудову класифікатора кластеру створеного алгоритмом DBSCAN розроблено програмне застосування мовою Python, що використовує алгоритм XGBoost (код програми поданий у додатку Б);

– для пошуку міри впливу показників на процес класифікації кластерів отриманих алгоритмом DBSCAN розроблено програмне застосування мовою Python, з використанням вбудованої функції «feature_importances_» (код програми поданий у додатку Б);

– для експериментального прогнозування регіональних показників соціально-економічного розвитку створений шаблон з використанням функцій пакету Microsoft Office Excel.

3.2 Експериментальна реалізація методу таксономії та моделі кластерного аналізу рівня соціально-економічного розвитку регіонів України

Для дослідження соціально-економічного розвитку регіонів України, обраний такий перелік факторів, які є найбільш вагомими (див. пп. 2.1). Однак залишилося відкритим питанням, який показник має бути обраним для оцінки рівня розвитку. В пп. 2.2 була запропонована процедура кластеризації регіонів на основі середнього значення таксономічного показника (у річному розрізі) для регіонів, що ввійшли в той чи інший кластер. Визначені дані, за якими буде проводитися експериментальне моделювання, подані в табл.3.1.

Таблиця 3.1 – Легенда даних, на яких буде відбуватися моделювання

Показник	Одиниці вимірювання
Середня заробітна плата	грн
Економічно активне населення	тис. осіб
Безробітне населення (за методологією МОП)	тис. осіб
Валовий регіональний продукт у розрахунку на одну особу	грн
Індекси промислової продукції	відсотків до відповідного періоду
Сальдо (експорт-імпорт)	млн. дол США
Капітальні інвестиції	млн. грн
Обсяги викидів забруднюючих речовин	тис. тон

За вхідними даними проводилися розрахунки таксономічного показника. Для розрахунків використовувався метод Хельвіга (обраний в пп.2.2), для кожного регіону за кожним роком [22-23]. Результати подані в табл. 3.2 та на рис. 3.2.

Отже, за аналізом таблиці 3.2, можна зробити висновок, що таксономічний показник знаходиться на однаково низькому рівні (менш «0,5») майже для всіх регіонів. Середнє значення для Харківського регіону складає 54,17%. Найвищий результат отримав Дніпропетровський регіон, його середнє значення за 8 років сягнуло 87,5%.

Таблиця 3.2 – Результати розрахунку таксономічного показника регіонів України

Регіони	Таксономічний показник								Ср.знач. за 8 років
	2012	2013	2014	2015	2016	2017	2018	2019	
Вінницька	0,3464	0,3426	0,3581	0,3806	0,3948	0,4358	0,4276	0,4638	0,3937
Волинська	0,2971	0,3001	0,2997	0,3198	0,3289	0,3577	0,3525	0,3805	0,3296
Дніпропетровська	0,8323	0,8022	0,8067	0,8256	0,8993	0,9477	0,9325	0,9528	0,8749
Донецька	0,9142	0,9445	0,9321	0,9375	0,8050	0,6302	0,6465	0,6634	0,8092
Житомирська	0,3153	0,3228	0,3311	0,3368	0,3422	0,3730	0,3619	0,3892	0,3466
Закарпатська	0,3448	0,3214	0,3312	0,3378	0,3568	0,3579	0,3712	0,3912	0,3516
Запорізька	0,4878	0,4673	0,4674	0,4748	0,5201	0,5671	0,5401	0,5758	0,5126
Івано-Франківська	0,3352	0,3599	0,3556	0,3482	0,3793	0,3994	0,3783	0,4135	0,3712
Київська	0,5207	0,5369	0,5575	0,5830	0,6171	0,6366	0,6568	0,6299	0,5923
Кіровоградська	0,2990	0,3071	0,3185	0,3288	0,3380	0,3428	0,3618	0,3639	0,3325
Луганська	0,4911	0,5093	0,4885	0,5056	0,4091	0,2488	0,3322	0,2718	0,4070
Львівська	0,4654	0,4762	0,4782	0,4801	0,5016	0,5576	0,5561	0,5815	0,5121
Миколаївська	0,3830	0,3611	0,3704	0,3865	0,3954	0,4278	0,4427	0,4448	0,4014
Одеська	0,5127	0,4451	0,4939	0,4994	0,4962	0,5253	0,5396	0,5593	0,5090
Полтавська	0,4686	0,4495	0,4698	0,4670	0,4968	0,5326	0,5394	0,5511	0,4968
Рівненська	0,3273	0,3084	0,3120	0,3156	0,3388	0,3603	0,3368	0,3651	0,3331
Сумська	0,3141	0,3294	0,3265	0,3409	0,3353	0,3751	0,3501	0,3747	0,3433
Тернопільська	0,2735	0,2814	0,2855	0,2901	0,3107	0,3180	0,3151	0,3441	0,3023
Харківська	0,5166	0,5167	0,5281	0,5193	0,5334	0,5713	0,5759	0,5719	0,5417
Херсонська	0,2855	0,2809	0,2915	0,2926	0,2999	0,3306	0,3303	0,3551	0,3083
Хмельницька	0,3122	0,3167	0,3226	0,3258	0,3405	0,3714	0,3609	0,3808	0,3414
Черкаська	0,3458	0,3345	0,3452	0,3520	0,3584	0,3817	0,3848	0,4005	0,3629
Чернівецька	0,2618	0,2450	0,2526	0,2707	0,2609	0,2821	0,2642	0,2955	0,2666
Чернігівська	0,3003	0,3010	0,3148	0,3156	0,3338	0,3560	0,3543	0,3695	0,3307

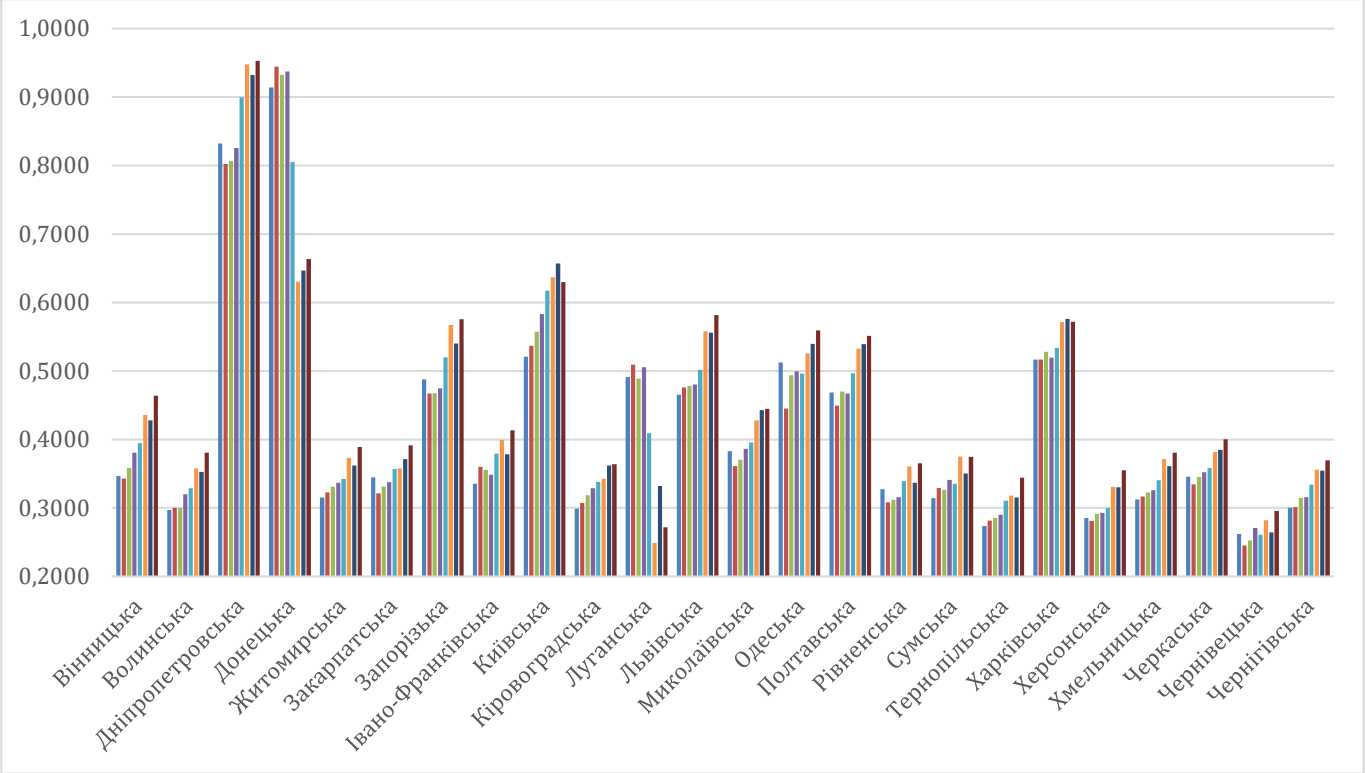

Рис. 3.2. Результати таксономічного показника за регіонами

До регіонів з найнижчим рівнем увійшли: Чернівецький, Херсонський, Волинський, Кіровоградський та Чернігівський регіони, значення їхнього показника в середньому знаходиться в межах менш 30%.

Можемо спостерігати, що регіони України мають досить високий розкид, що вказує на відсутність спільної динаміки розвитку, також можемо спостерігати процес дивергенції регіонів.

Оскільки визначені 8 факторів, то перед проведенням кластеризації, необхідно зробити звуження простору ознак методом головних компонент, та кластеризувати регіони по рокам за першими двома компонентами.

Метод головних компонент, має наступні значення варіативності кожної компоненти (табл 3.3).

Таблиця 3.3 – Варіативність кожної компоненти методу головних компонент

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC1+PC2, %
2012	70,19	12,92	9,94	3,82	1,54	1,04	0,35	0,21	83,10
2013	71,25	13,10	8,23	4,71	1,29	0,92	0,36	0,13	84,35
2014	70,05	12,73	9,54	4,98	1,23	0,85	0,46	0,16	82,78
2015	70,22	11,69	9,85	4,39	2,18	1,15	0,38	0,14	81,91
2016	65,58	18,44	6,92	4,68	2,13	1,63	0,45	0,17	84,02
2017	57,31	21,04	8,50	5,51	3,42	2,48	1,50	0,25	78,35
2018	55,27	19,00	11,98	5,33	4,62	2,26	1,26	0,28	74,27
2019	56,02	22,18	7,85	5,78	4,75	1,96	1,01	0,44	78,19

З аналізу табл. 3.1 випливає, що в середньому перші дві головні компоненти обумовлюють 80,8% варіативності всієї вибірки. Найменше значення 74,27% у 2016 році, найбільше – 84,35% у 2011 році. Оскільки, перші дві головні компоненти приймають значення від 74 до 84%, кластерний аналіз проводився на двомірному звуженому просторі ознак, першої та другої головної компоненти.

Для кластеризації було обрано метод DBSCAN (див. пп. 2.2).

DBSCAN – це алгоритм кластеризації даних, який запропонували Мартін Естер, Ганс-Петер Крігель, Йорґ Сандер та Сяовей Су у 1996 році. Він є алгоритмом кластеризації заснованим на щільності: для заданої множини точок у деякому просторі він відносить в одну групу точки, які розташовані найбільш щільно (точки з багатьма сусідами) та розмічає точки, які лежать в областях з невеликою щільністю (чиї сусіди розташовані занадто далеко) як викиди. DBSCAN є одним з найпоширеніших алгоритмів кластеризації, а також найбільш цитованим у науковій літературі [24].

Для проведення кластеризації по регіонах за роками, використовувалися таки параметри алгоритму:

– eps = 0.4 (відстань, за якою будуть групуватися точки в просторі),

– min_samples = 3 (мінімальна кількість спостережень в групі).

Алгоритм видавав від 2-до 4 кластерів, у кожному році, тому було використане середнє значення таксономічного показника для групування регіонів вже на два (2) кластери, показники яких є залежною змінною у класифікації.

Загалом, отримані такі середні значення таксономічного показника, табл. 3.4.

Таблиця 3.4 – Результати розподілу та середнє значення таксономічного показника

	кластер 0	ср. Значення показника	кластер 1	ср. Значення показника
2012	9	0,3085	15	0,4782
2013	8	0,3276	16	0,4524
2014	11	0,3237	13	0,4981
2015	6	0,34	18	0,4552
2016	17	0,3769	7	0,5691
2017	13	0,3596	11	0,5464
2018	13	0,3478	11	0,5626
2019	13	0,3751	11	0,5647

Загалом, розглядалися 102 регіони з високим рівнем соціально-економічного розвитку (середнє значення таксономічного показника – 0,5158), та 90 регіонів з низьким рівнем розвитку (0,3449). Проте слід зазначити, що в нашому випадку експериментального дослідження, регіони розділялися скоріше на клас з низьким та середнім рівнем розвитку, оскільки середня значення першого кластеру складає 0,51, а це не високий результат. Як зазначалося вище, тільки Дніпропетровська та Донецька області мають середнє значення показника таксономії за 8 років вище ніж 0,75.

Слід зазначити, що за результатами отриманий майже рівний баланс класу (53/47). Це означає, що результати класифікації будуть більш надійними, а в подальшому дослідженні не потрібно вирішувати проблему дисбалансу класу, тому що для відмінних від DBSCAN over-sampling та under-sampling замало

спостережень, а синтетична генерація даних за методикою SMOTE може спричинити до «перенавчання» моделі.

Динаміка потрапляння регіону до класу з високим чи низьким рівнем розвиту подана в табл. 3.5 та табл. 3.6.

Таблиця 3.5 – Матриці динаміки регіонів за роками (частина 1)

year	Vin	Vol	Dnipro	Don	Zhut	Zak	Zapor	Ivan	Kyiv	Kirov	Lug	Lviv	Mukol	Odes
2012	0	1	1	1	0	1	1	1	1	0	1	1	0	1
2013	0	1	1	1	1	0	1	1	1	0	1	1	0	1
2014	0	1	1	1	1	0	1	0	1	0	1	1	1	1
2015	1	1	1	1	1	0	1	0	1	1	1	1	0	1
2016	0	0	1	1	0	0	0	0	1	0	1	0	0	1
2017	1	0	1	1	0	1	1	0	1	0	1	1	0	1
2018	1	0	1	1	0	0	1	0	1	0	1	1	1	1
2019	1	0	1	1	0	1	1	0	1	0	1	1	0	1
клас 0	4	4	0	0	5	5	1	6	0	7	0	1	6	0
клас 1	4	4	8	8	3	3	7	2	8	1	8	7	2	8

За даними табл. 3.5, спостерігається, що у період з 2012 по 2019 рр. Дніпропетровський, Донецький, Київський, Луганський, Одеський та Полтавський регіони не змінювали свій клас, та усі 8 років були у кластері з високим рівнем розвитку.

Таблиця 3.6 – Матриці динаміки регіонів за роками (частина 2)

year	Polt	Rivne	Sums	Tern	Khakr	Kherson	Khmel	Cherk	Cherniv	Chernihiv
2012	1	1	1	0	1	0	0	1	0	0
2013	1	0	0	1	1	1	0	0	1	1
2014	1	0	0	0	1	0	0	0	1	0
2015	1	1	1	0	1	1	0	0	1	1
2016	1	0	0	0	0	0	0	0	1	0
2017	1	0	0	0	1	0	0	0	0	0
2018	1	0	0	0	1	0	0	0	0	0
2019	1	0	0	0	1	0	0	0	0	0
клас 0	0	6	6	7	1	6	8	7	4	6
клас 1	8	2	2	1	7	2	0	1	4	2

За даними табл. 3.6 слід відмітити, що Запорізький, Львівський та Харківський регіони 7 років були у кластері з високим рівнем розвитку і лише в 2016 році потрапили до кластеру з низьким рівнем.

Отже, метод таксономії показав, що між регіонами України відсутня спільна динаміка розвитку, про що свідчить великий розкид показника таксономії. Лідерами за значенням показника стали Дніпропетровський та Донецький регіон, найнижче значення у Чернівецького, Херсонського та Волинського регіону.

Слід також зазначити, що у розрізі 8 років, майже у всіх регіонів спостерігається позитивна динаміка таксономічного показника, проте на досить низькому рівні, на це вказує низьке значення стандартного відхилення. Найбільший розкид показника за роками мають Донецький та Луганський регіон, за останні 5 років показник таксономії почав падати, що пов’язано з політичними подіями в країні. Метод головних компонент транслює експериментальну вибірку на двомірний простір, що в свою чергу сумісно працює з методом кластеризації.

Отже за отриманими даними визначено, що в Україні немає чітко сформованого середнього класу і спостерігається незбалансованість класів з високим та низьким рівнем соціально-економічного розвитку регіонів України.

На наступному етапі проводився кореляційний аналіз показників на кластер. Оскільки існують тільки показники типу «numeric» (залежна змінна має бінарний тип) то використовується кореляція Пірсона (див. рис.3.3).

Найменше на кластер впливають: середня заробітна плата (СЗП) (-1%), індекс промисловості (–2%) та валовий регіональний продукт (ВРП) (5%). Найбільше – викиди в атмосферу (30%), сальдо (22%) та кількість безробіття (17 відсотків).

Слід також зазначити що між показниками спостерігається мультиколеніарність. Так, наприклад, між економічно активним населенням та кількістю безробіття зв’язок 89%, між сальдо та викидами в атмосферу – 84%, між ВРП та СЗП – 78%, між сальдо та кількістю безробіття – 70%.

Проте на першому етапі моделювання використовувалася операція позбування від показників, що корелюють між собою. Під час навчання моделі робився аналіз

важливості факторів, а у разі незадовільних результатів – вирішувалася задача позбавлення від мультиколеніарності методом покрокового виключення показників.

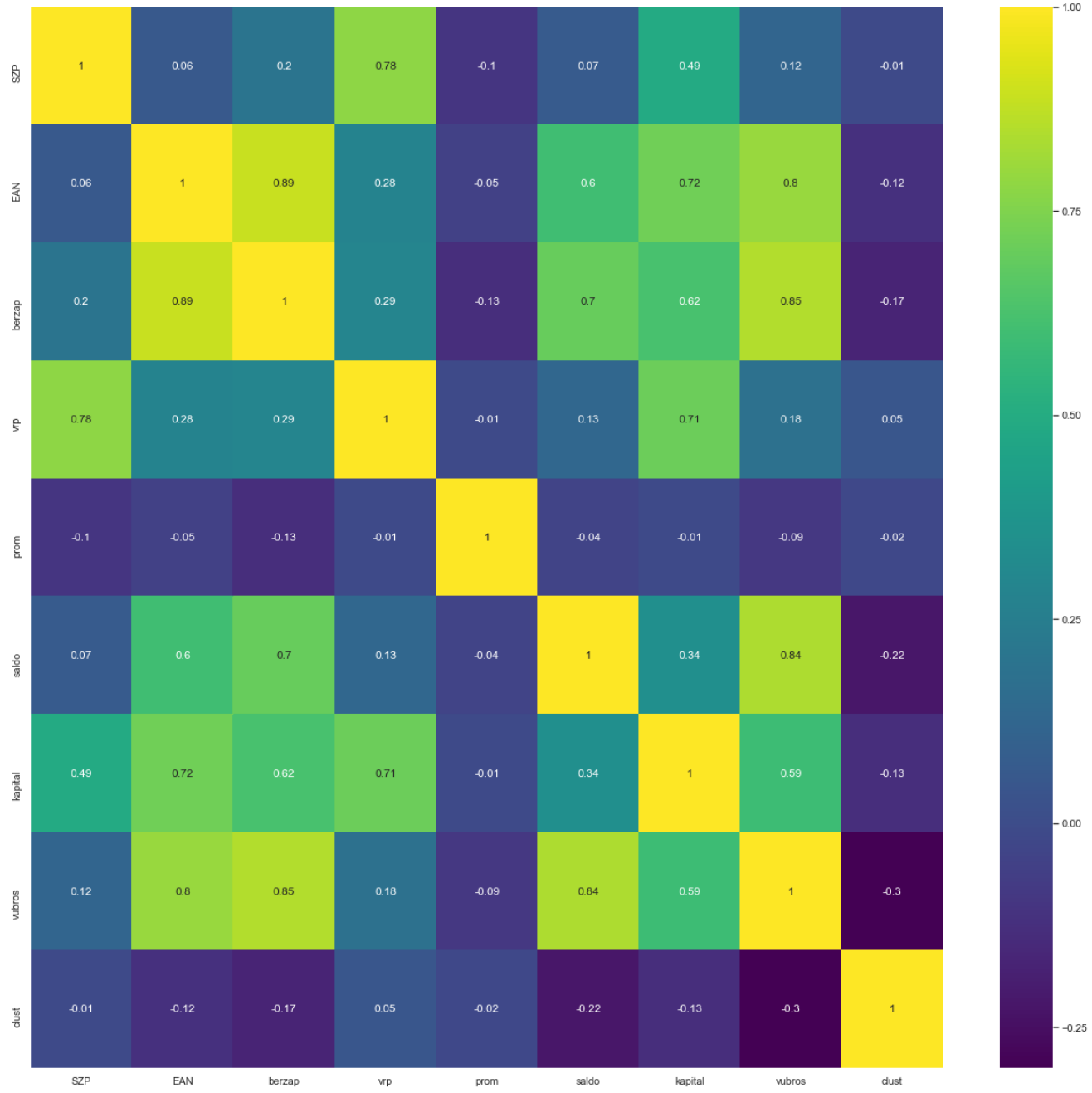

Рис. 3.3 Кореляційний аналіз незалежних показників

Для дослідження впливу показників на таксономічний показник, будувалася модель панельних даних з фіксованим ефектом. Для побудови моделі на панельних даних, був обраний тип моделі PanelOLS з бібліотеки linearmodels. Побудова моделі відбувалася мовою програмування Python.

Перед проведенням моделювання дані стандартизувалися, за допомогою функції preprocessing.scale(). Після чого дані розділялися на дві вибірки (навчальну та тестову), оскільки використовуються панельні дані, що містять в собі часову компоненту. В якості навчальної вибірки використовувалися дані за 2012-2018 роки. Для валідації моделі використовувалися дані за 2019 рокі, що в розробленій моделі є тестовою вибіркою. Якість моделювання оцінювалося за допомогою коефіцієнта кореляції, з бібліотеки scikit-learn. Код рівняння та результати моделювання подані на рис 3.4 та рис.3.5 відповідно.

```
from linearmodels import PanelOLS
mod = PanelOLS(data_train['index'], data_train[features], time_effects=False,
entity_effects=True,
        other_effects=None, singletons=False, drop_absorbed=True)
res = mod.fit(cov_type='clustered', cluster_entity=True)
```

Рис.3.4 Код моделі на панельних даних

Для моделювання було обрано наступні гіперпараметри моделі: time_effects (часова компонента – в розглядаємому випадку дані історичні, тому підтверджувалася наявність часової компоненти), entity_effects (просторова компонента, оскільки використовуються панельні дані цьому параметру присвоювалося значення «0»), singletons (тип даних, лінійний чи просторовий, в нашому випадку спостерігалося, що данні нелінійні), drop_absorbed (видалення стороннього шуму). За отриманими даними спостерігається, шо модель має високу якість, на що вказує коефіцієнт кореляції, 88,57%, та F-статистика.

Також не всі показники виявилися значимі. Середня заробітна плата, безробіття та викиди в атмосферу за результатами моделі на панельних даних виявилися незначимі. Слід зазначити, що розроблена модель показала такі результати на навчальній вибірці.

Далі за допомогою функції predict(), модель перевірялася на тестовій вибірці. Результати подані в табл. 3.6. Коефіцієнт кореляції складає 64,7%. Рівняння моделі має вигляд:

$$y = -SZP * 0{,}0099 + EAN * 0{,}0632 - bezrob * 0.0008 + VRP * 0.0550 + PROM * 0{,}0052 - SALDO * 0{,}0109 + KAPITAL * 0{,}0049 - VUBROS * 0{,}0083 \quad (3.1)$$

Отже можна зробити висновок, що модель не високої якості, а результати моделювання показують, що між регіонами України відсутня спільна динаміка розвитку, спостерігається процес дивергенції, як було зазначено вище.

Для підвищення якості моделі, було побудовано модель панельних даних окремо для регіонів з високим та низьким рівнем соціально-економічного розвитку. Результати моделювання подані у табл. 3.8–3.9. Коефіцієнт кореляції моделі для кластеру з високим рівнем соціально-економічного розвитку складає 98,76% на навчальній вибірці, а на тестовій – 71,62% (модель мала такі ж самі параметри, як і модель одразу для всіх регіонів). Також в моделі панельних даних, з високим рівнем соціально-економічного розвитку, із всіх незалежних факторів, що були обрані методом аналізу, статистично незначимий тільки один – це «сальдо». Модель має наступний вигляд:

$$y = -SZP * 0{,}07731 + EAN * 0{,}0507 - bezrob * 0.003 + VRP * 0.0622 + PROM * 0{,}009 - SALDO * 0{,}0109 + KAPITAL * 0{,}0290 - VUBROS * 0{,}0534 \quad (3.2)$$

Модель для кластеру регіонів, що мають низький рівень розвитку, показала наступні результати: навчальна вибірка, коефіцієнт кореляції склав 93%, тестова – 68%. Отже, отримані результати моделювання вказують на високу якість кластеризації, оскільки результати моделей панельних даних вищі для кожного кластеру окремо, ніж регіони з високим рівнем розвитку. Це вказує на присутність ефекту конвергенції. Проте низька якість моделі для кластеру з низьким рівнем свідчить про нечітку картину спільної динаміки розвитку регіонів. В моделі для кластеру з низьким рівнем розвитку статистично незначимі два незалежні показника

– кількість безробітних та сальдо (як і в попередній моделі). Рівняння моделі має вигляд:

$$y = -SZP * 0{,}1012 + EAN * 0{,}0579 - bezrob * 0.0042 + VRP * 0.0299 + PROM * 0{,}009 - SALDO * 0{,}0002 + KAPITAL * 0{,}0373 - VUBROS * 0{,}0072 \quad (3.3)$$

PanelOLS Estimation Summary

Dep. Variable:	index	**R-squared:**	0.8857
Estimator:	PanelOLS	**R-squared (Between):**	0.0555
No. Observations:	168	**R-squared (Within):**	0.8857
Date:	Mon, Dec 02 2019	**R-squared (Overall):**	0.0626
Time:	06:11:28	**Log-likelihood**	475.91
Cov. Estimator:	Clustered		
		F-statistic:	131.67
Entities:	24	**P-value**	0.0000
Avg Obs:	7.0000	**Distribution:**	F(8,136)
Min Obs:	7.0000		
Max Obs:	7.0000	**F-statistic (robust):**	453.50
		P-value	0.0000
Time periods:	7	**Distribution:**	F(8,136)
Avg Obs:	24.000		
Min Obs:	24.000		
Max Obs:	24.000		

Parameter Estimates

	Parameter	Std. Err.	T-stat	P-value	Lower CI	Upper CI
SZP	-0.0099	0.0090	-1.0950	0.2754	-0.0278	0.0080
EAN	0.0632	0.0159	3.9639	0.0001	0.0317	0.0947
berzap	-0.0008	0.0078	-0.1038	0.9175	-0.0163	0.0146
vrp	0.0550	0.0099	5.5514	0.0000	0.0354	0.0746
prom	0.0052	0.0017	3.0490	0.0028	0.0018	0.0085
saldo	0.0109	0.0061	1.7791	0.0775	-0.0012	0.0230
kapital	0.0049	0.0037	1.3007	0.1956	-0.0025	0.0122
vubros	-0.0083	0.0287	-0.2906	0.7718	-0.0651	0.0485

Рис.3.5 Результати моделі на панельних даних

Отже, можемо спостерігати те, що якість моделі на тестовій вибірці – нижча, ніж на навчальній (табл. 3.7).

Таблиця 3.7 – Результати моделі панельних даних на тестовій вибірці

region	year	index	predictions
Vin	2019	0.463841	0.217015
Vol	2019	0.380534	0.059859
Dnipro	2019	0.952757	1.101085
Don	2019	0.663445	0.002193
Zhut	2019	0.389165	0.044515
Zak	2019	0.391232	0.165504
Zap	2019	0.575804	0.451398
Ivan	2019	0.413493	0.011586
Kyiv	2019	0.629947	0.596478
Kirov	2019	0.363901	0.019774
Lug	2019	0.271782	0.550860
Lviv	2019	0.581468	0.442639
Mukol	2019	0.444755	0.140852
Odes	2019	0.559309	0.479893
Polt	2019	0.551059	0.694707
Pivne	2019	0.365118	0.078233
Sums	2019	0.374686	0.022654
Tern	2019	0.344096	0.140695
Khakr	2019	0.571935	0.674756
Kherson	2019	0.355121	0.053817
Khmel	2019	0.380805	0.033210
Cherk	2019	0.400509	0.134201
Cherniv	2019	0.295518	0.250384
Chernihiv	2019	0.369481	0.036274

Результати моделей по кластерам (табл. 3.8), мають вищу якість на тестових вибірках, ніж модель на всіх регіонів одразу.

Таблиця 3.8 – Результати моделі панельних даних на тестовій вибірці для кластеру з високим рівнем розвитку

region	year	index	predictions
Vin	2019	0.463841	0.372864
Dnipro	2019	0.952757	1.092556
Don	2019	0.663445	0.596036
Zak	2019	0.391232	0.151444
Zap	2019	0.575804	0.560510
Kyiv	2019	0.629947	0.765874
Lug	2019	0.271782	0.061036
Lviv	2019	0.581468	0.543266
Odes	2019	0.559309	0.531714
Polt	2019	0.551059	0.637966
Khakr	2019	0.571935	0.583259

Оскільки серед регіонів, що належать до одного кластеру, наявний тісний зв’язок, то якість моделей є більш точна ніж для моделі всіх регіонів одразу.

Таблиця 3.9 – Результати моделі панельних даних на тестовій вибірці для кластеру з низьким рівнем розвитку

region	year	index	predictions
Vol	2019	0.380534	0.339502
Zhut	2019	0.389165	0.407123
Ivan	2019	0.413493	0.479771
Kirov	2019	0.363901	0.358359
Mukol	2019	0.444755	0.607540
Pivne	2019	0.365118	0.369636
Sums	2019	0.374686	0.391569
Tern	2019	0.344096	0.266171
Kherson	2019	0.355121	0.348608
Khmel	2019	0.380805	0.436749
Cherk	2019	0.400509	0.469227
Cherniv	2019	0.295518	0.180078
Chernihiv	2019	0.369481	0.336488

Для побудови класифікатора використовувався алгоритм XGBoost (xgboost). Алгоритм був імплементований мовою Python, версія 3.7.

Спочатку проведений розподіл вибірки на тестову та навальну. Пропорція дорівнює 10%, тобто з 192 спостережень до тестової вибірки потрапило 20, а до навчальної – 172.

Навчання було проведено з використання методу крос-валідації. Для цього вибірка випадковим чином розділялася 5 разів на навчальну та тестову (пропорція = 30%). Результати навчання – це середнє значення моделі на 5 фолдах (рис. 3.6).

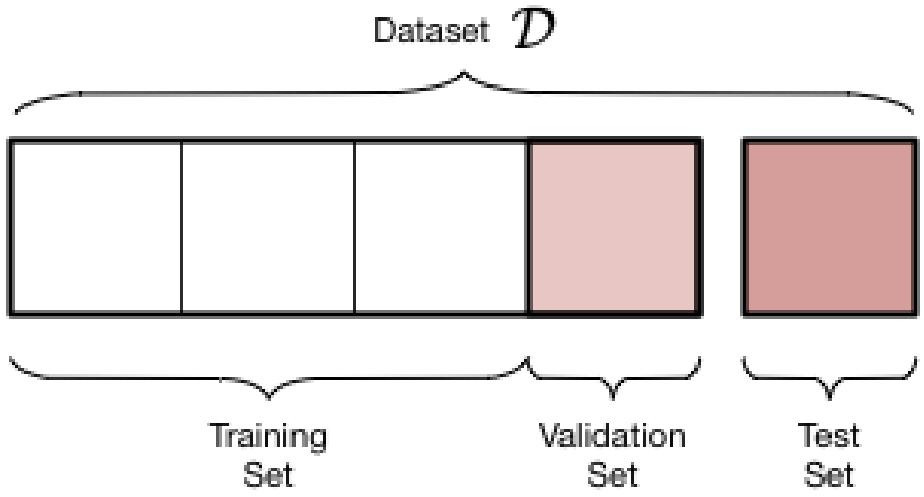

Рис. 3.6 Схема крос-валідації

Також, для перевірки отриманих результатів було обрано 4 метрики:

– accuracy – точність підмножини: набір міток, передбачених для вибірки, повинен точно відповідати відповідному набору міток у y_true.

– f1_score – середньозважене значення точності та відкликання, коли оцінка F1 досягає найкращого значення у 1, а найгірший – у 0. Відносний внесок точності та виклику в рахунок F1 рівний. Формула для оцінки F1 така: F1 = 2 * (precision * recall) / (precision + recall)

– recall – відношення tp / (tp + fn), де tp – кількість справжніх позитивних результатів, а fn – кількість помилкових негативів. Це інтуїтивно зрозуміла здатність

класифікатора знаходити всі позитивні вибірки.

– precision – відношення tp / (tp + fp), де tp – кількість справжніх позитивних результатів, а fp – кількість помилкових позитивних результатів. Це інтуїтивно зрозуміла здатність класифікатора не маркувати як позитивний зразок, який є негативним.

Для класифікатора обиралися 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

Бустинговий класифікатор має в своєму арсеналі більше гіперпараметрів. Оскільки розмір експериментальної вибірки не великий, то обиралися тільки чотире параметри.

Результати крос-валідації показали високі результати (подані в табл. 3.10).

Таблиця 3.10. – Результати крос-валідації класифікатора на навчальній виборці

	accuracy	f1_score	precision	recall
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
MEAN	0.826923	0.826795	0.830590	0.826923

За допомогою вбудованої функції feature_importances_ в програмному застосуванні Python, знаходилася числова міра впливу показників на процес класифікації (рис. 3.7).

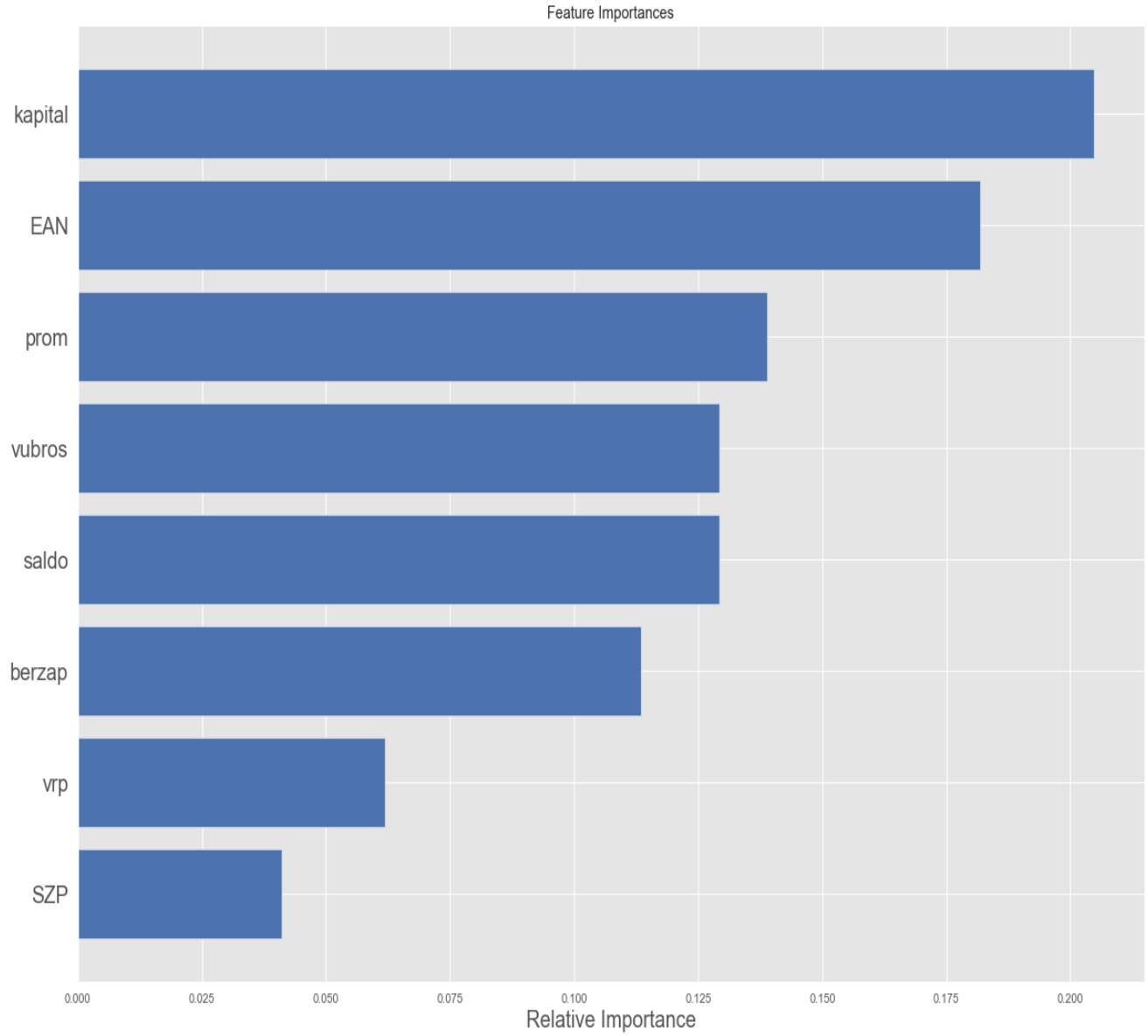

Рис. 3.7. Графік впливу факторів

Отже, капітальні інвестиції впливають: на клас – на 20%, економічно активне населення – на 18%, індекс промисловості – на 13%, викиди в повітря та сальдо на 12,5%, кількість безробіття – на 11,5%, а валовий регіональний продукт та середня заробітна плата не несуть значний вплив на процес класифікації.

Перед навчанням вибірка була розділена і для аналізу залишились 20 спостережень, які не приймали участь в навчанні моделі. Навчена модель використовувалася для перевірки на тестовій вибірці, що подані в табл. 3.11 та 3.12.

Таблиця 3.11 – Результати валідування моделі на тестовій вибірці

	Probability	True	Predicted
Polt_2016	0.852365	1	1
Dnipro_2015	0.980530	1	1
Kherson_2018	0.013524	0	0
Vol_2016	0.282792	0	0
Zap_2017	0.727431	1	1
Chernihiv_2014	0.273373	0	0
Khakr_2012	0.695317	1	1
Odes_2018	0.927397	1	1
Vol_2018	0.178018	0	0
Ivan_2012	0.334574	1	0
Zak_2012	0.558368	1	1
Kherson_2017	0.075045	0	0
Polt_2018	0.879024	1	1
Dnipro_2019	0.984242	1	1
Sums_2018	0.140028	0	0
Chernihiv_2018	0.111619	0	0
Odes_2013	0.788238	1	1
Khakr_2014	0.955387	1	1
Khakr_2017	0.917986	1	1
Dnipro_2017	0.964841	1	1

Аналіз табл. 3.11 показав, що під час валідації навчена модель помилилася тільки на Івано-Франківському регіоні (2012 р.), а 19 з 20 показників було класифіковано правильно. Але під час використання даних табл. 3.11 слід враховувати, що це не до кінця збалансована вибірка (маємо 13 регіонів кластеру з високим рівнем соціально-економічного розвитку та 8 регіонів з низьким рівнем розвитку).

Проте якість класифікації на тестовій вибірці є якісною, модель є статистично значима, на що вказує велике значення метрик (> 92%), що були обрані для даної моделі. Слід зазначити, що однією з метрик, ми досягли 100% точності.

Таблиця 3.12 – Результати основних метрик моделі класифікації при валідуванні

accuracy	0,95
f1_score	0,95
precision	1
recall	0.92

Для більш чіткого трактування результатів, обиралася перехресна матриця валідації (рис. 3.8).

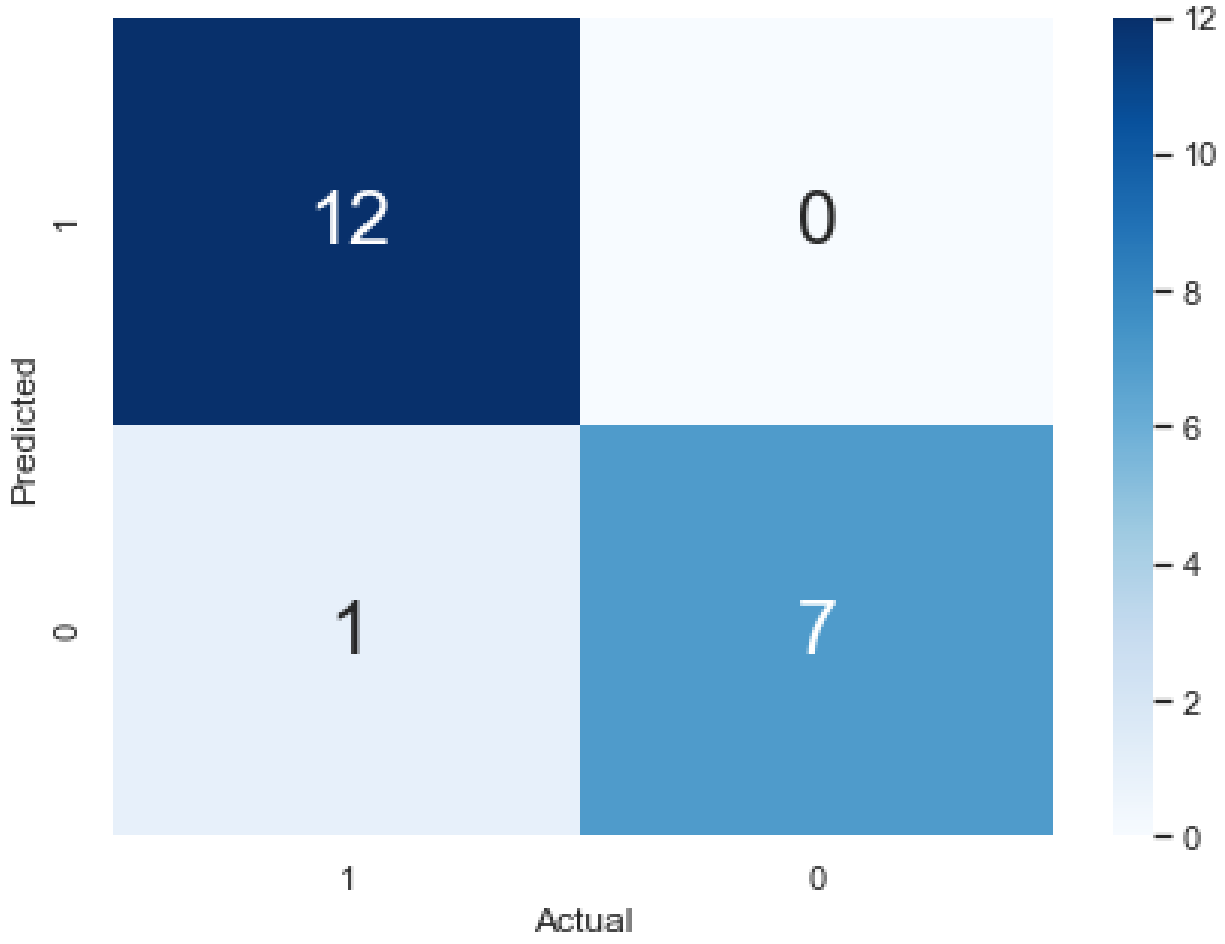

Рис. 3.8. Перехресна матриця класифікації на тестовій вибірці

3.3 Експериментальна побудова прогнозів регіональних показників соціально-економічного розвитку

Для прогнозування таксономічного показника Харківського регіону на 2020-2021 рр. та для визначення його класу були побудовані прогнозні моделі для показників, які ще не розраховані для цих років.

Щоб зробити прогноз значення таксономічного показника та прогноз ймовірності були потрібні значення незалежних факторів за 2020 та 2021 роки або їх прогнозування для тих показників, котрі ще не розраховані Держкомстатом.

Отже, на 2020 рік вимагалося спрогнозувати валовий регіональний продукт та масу викидів в атмосферу, на 2021 рік – валовий регіональний продукт, індекс промисловості, сальдо, капітальні інвестиції та масу шкідливих викидів в атмосферу.

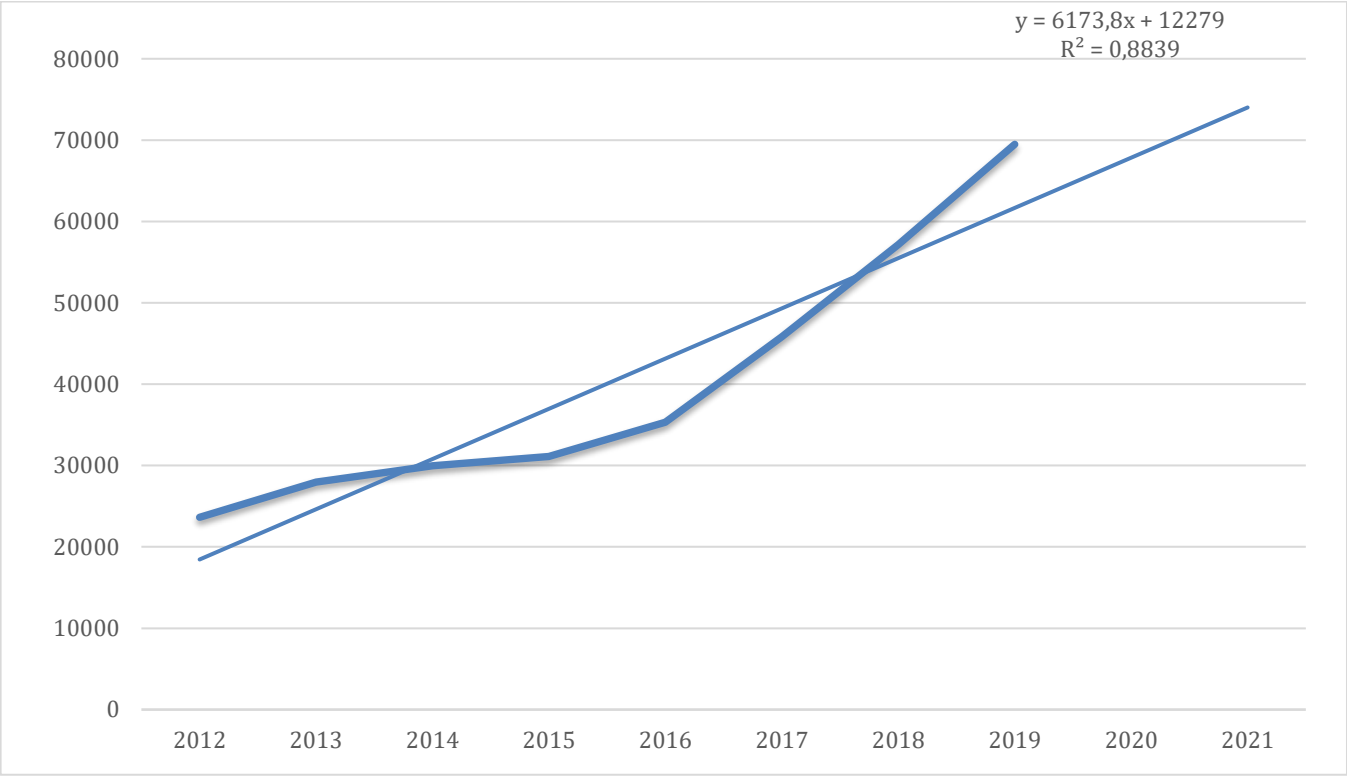

Рис. 3.9. Прогноз валового регіонального продукту на 2020-2021 роки

Валовий регіональний продукт прогнозувався за допомогою лінійного тренду в MS Excel. Результати подані на рис. 3.9. За даними рис. 3.9 можемо спостерігати, що точність прогнозу висока, на що вказує коефіцієнт кореляції (складає 88,39%). Також

в найближчі два роки буде зберігатися тенденція зростання показника. За наведеним рівнянням (рис. 3.9), у 2020 році показник складе 67843,2 грн, а в 2021 – 74017 грн.

Індекс промисловості прогнозувався за допомогою ковзкого середнього за 2 періоди. На рис.3.10 поданий графік реальних та розрахункових значень. Оскільки за останні роки індекс немає чіткої динаміки росту чи занепаду, якість прогнозу не дуже висока. У 2021 році, за прогнозом, значення показника складе 104,5% до попереднього року.

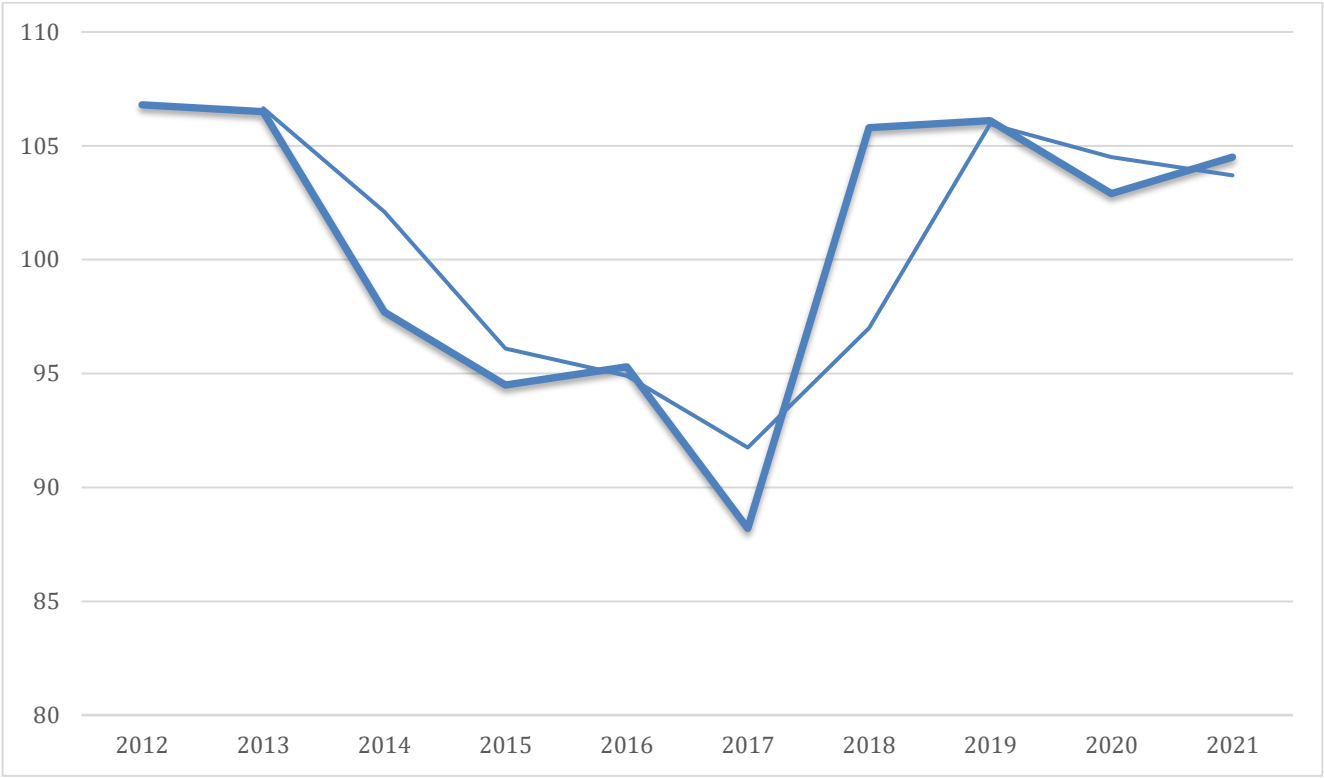

Рис. 3.10. Прогноз індексу промисловості на 2021 рік

Сальдо прогнозувалося за аналогією індексу промисловості, ковзким середнім за двома регіонами. Результати прогнозу подані на рис. 3.11. За прогнозом, у 2021 році, це показник складе -4438,688 млн. дол США.

Незалежний показник – капітальні інвестиції прогнозувалися за допомогою поліноміального тренду другого ступеня. Результати прогнозу та рівняння моделі подані на рис. 3.11. У 2021 році цей показник складе 27948 млн. грн. Модель має високу точність, на що вказує високе значення коефіцієнту кореляції (коефіцієнт складає 74,2%).

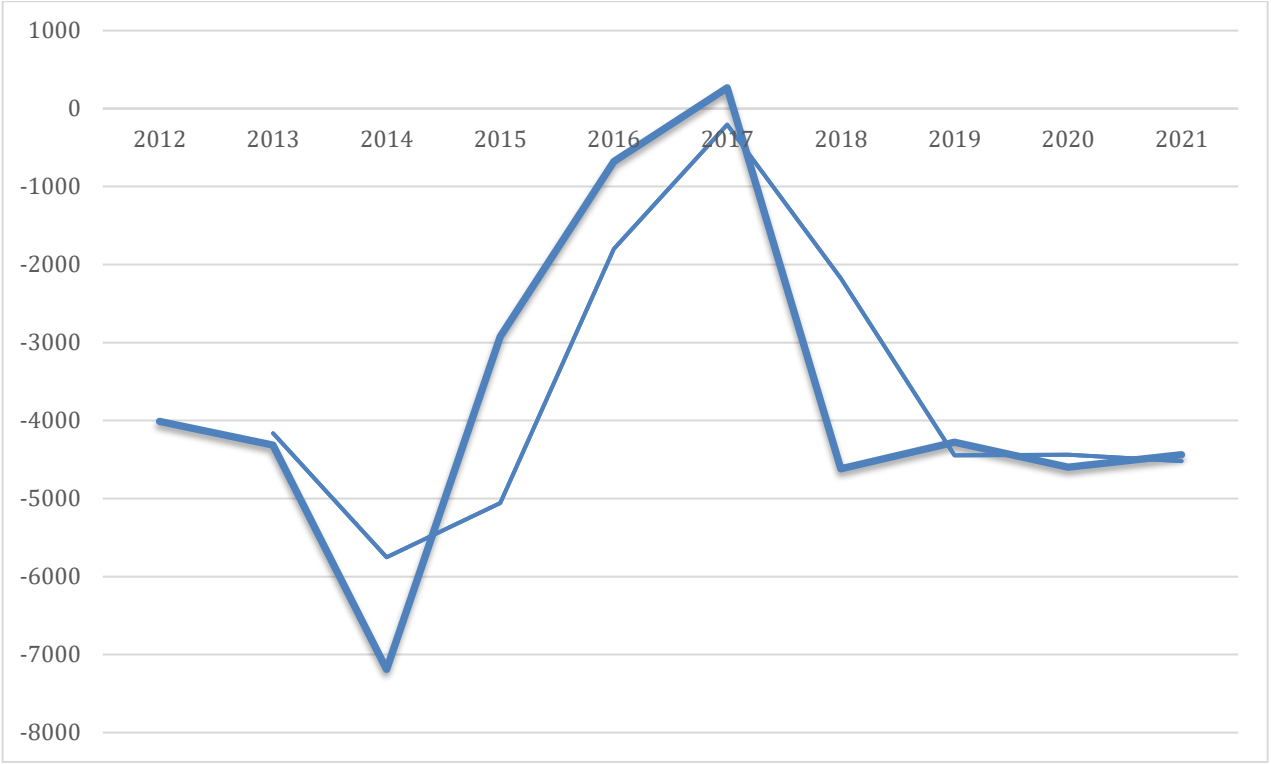

Рис. 3.11. Прогноз сальдо на 2021 рік

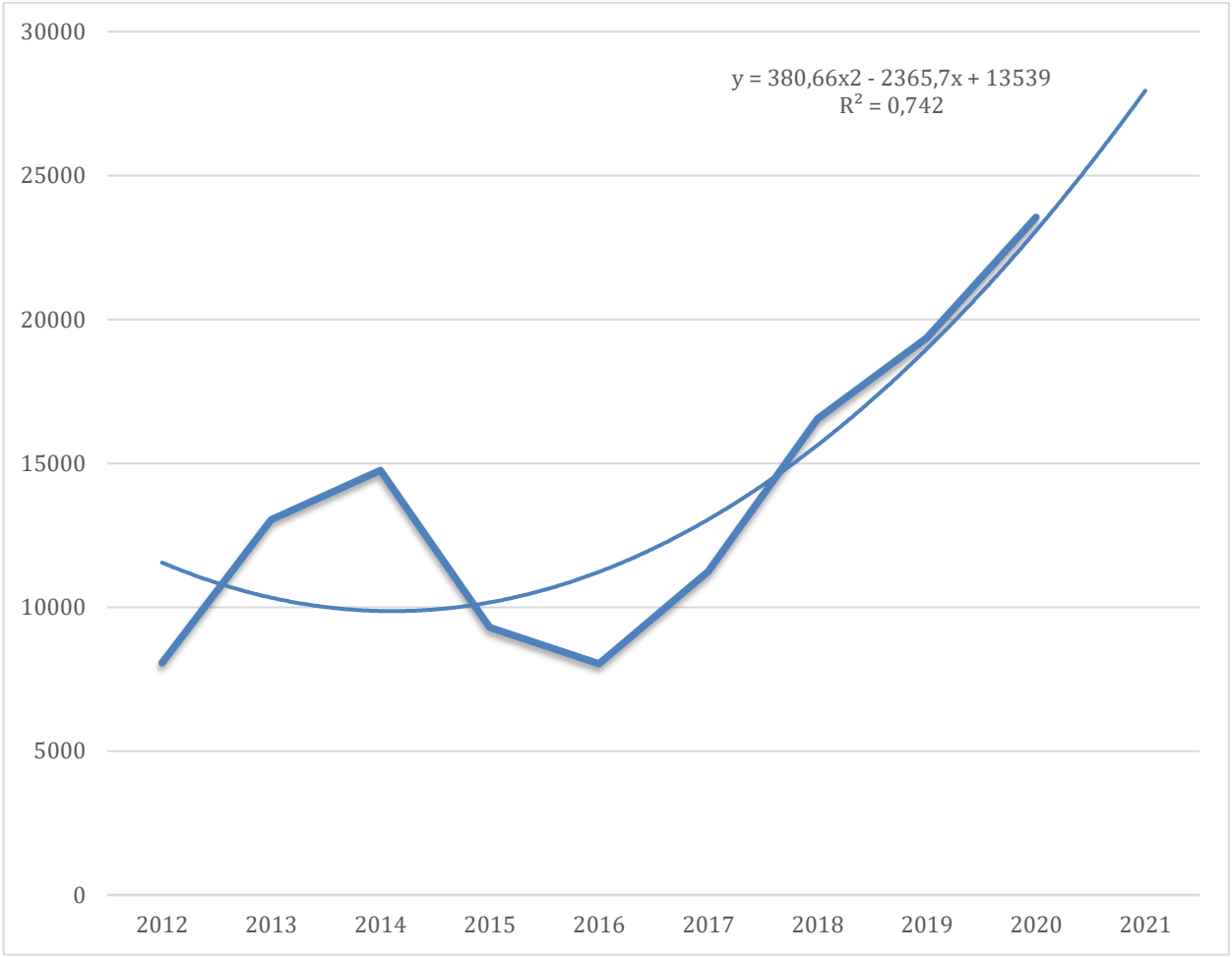

Рис. 3.12. Прогноз капітальних інвестицій на 2021 рік

Кількість шкідливих викидів атмосферу на 2020-2021 роки прогнозувалося за допомогою експоненціального тренду. Результати прогнозу та рівняння моделі подані на рис. 3.13. Модель має не високу точність, на що вказує коефіцієнт кореляції, (сягає лише 58,5%), але в моделі панельних даних цей показник є статистично незначимим і точність його прогнозу в бустинговому ансамблевому класифікаторі за коефіцієнтом Джині для його фактора є не високим. Цей факт не впливає на точність прогнозу коефіцієнта танксономії та ймовірності потрапляння Харківського регіону у 2020-2021 роках до кластеру регіонів з високим рівнем соціально-економічного розвитку. За рівням моделі, було розраховане прогнозне значення показника на наступні два періоди. Так у 2020 році прогнозне значення шкідливих викидів в атмосферу складе 51,7 тис. тон, а в 2021 – 42,96 тис. тон

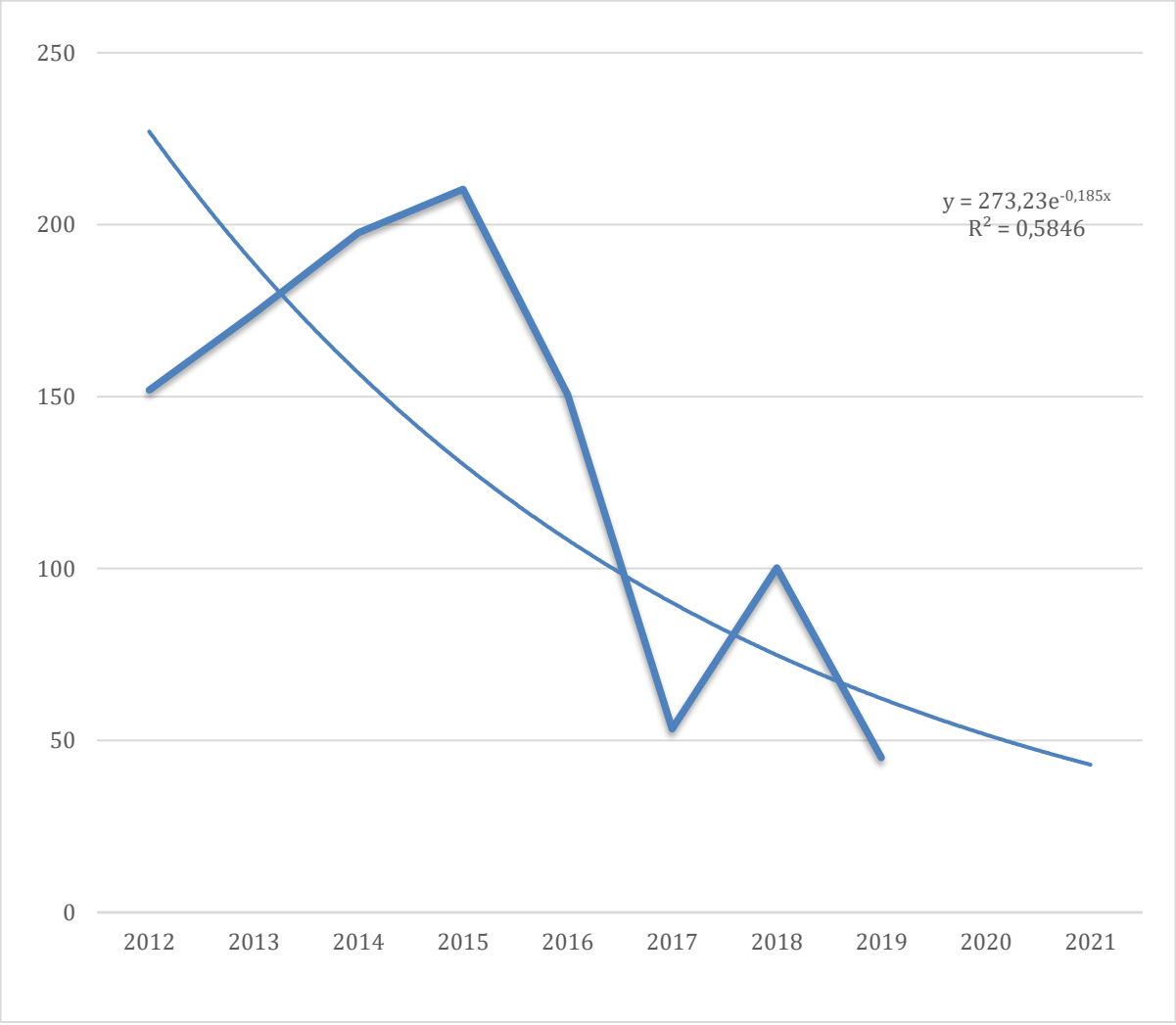

Рис. 3.13. Прогноз шкідливих викидів на 2021 рік

Пісяля отримання прогнозних значень незалежних факторів на 2020-2021 роки, за допомогою моделі панельних даних прогнозувався таксономічний показник для Харківського регіону, тобто для кластеру регіонів з високим рівнем соціально-економічного розвитку. Сама точність цієї моделі була найвищою, що вказує на присутність спільної траєкторії розвитку регіонів України з високим рівнем соціально-економічного розвитку.

Результати наступні: у 2020 році таксономічний показник складе 0.586059, а в 2019 – 0.649359. Отже Харківський регіон зберігає свої позиції у кластері регіонів з високим рівнем розвитку. На рис. 3.14 подана динаміка розрахункового показника таксономії у період з 2012-2019 років та прогнозні значення на 2020-2021 роки.

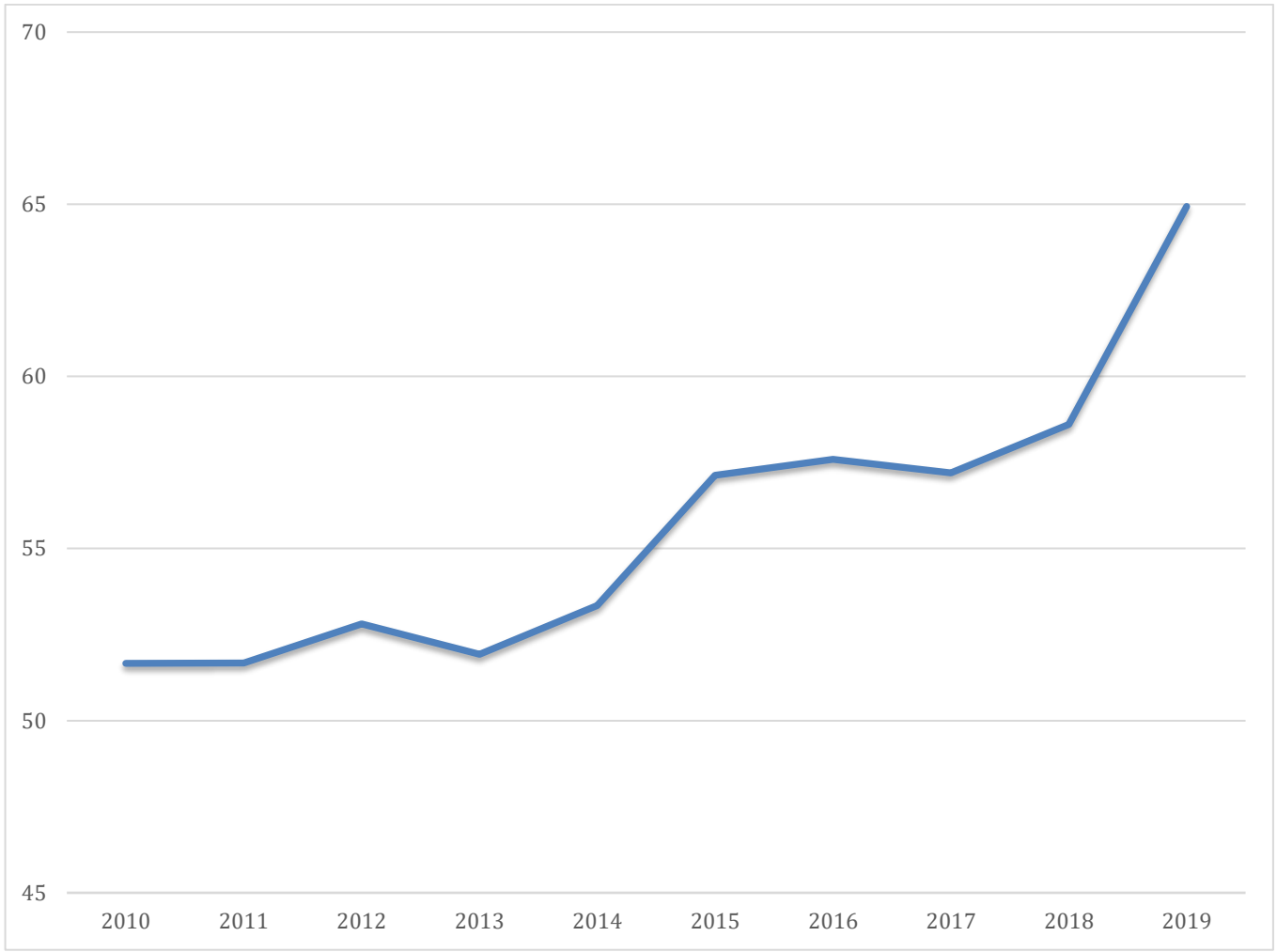

Рис. 3.14. Динаміка індексу таксономії Харківського регіону з прогнозними на 2020-2021 роки значеннями

Також слід зазначити, що зберігається й позитивна динаміка зростання показника. Передумовами до цього стали незалежні фактори, що використовувалися в моделі, адже більшість факторів також зберегли позитивну динаміку на майбутні періоди.

На наступному кроці прогнозувалася ймовірність потрапляння Харківського регіону до кластеру з високим рівнем соціально-економічного розвитку регіонів України. Результати подані на табл.3.13.

Таблиця 3.13 – Результати прогнозування потрапляння Харківського регіону до кластеру регіонів з високим рівнем розвитку

Рік	Ймовірність потрапляння до 1 класу	Прогнозований клас
2020	0.868770	1
2021	0.873845	1

Отже, в пункті 3.2 даної роботи прогнозовані незалежні показники, що впливають на рівень соціально-економічного розвитку регіонів України, та проведений прогноз показника таксономії для Харківського регіону на 2020-2021 роки. Прогноз таксономії підтвердив результати ансамблевої моделі, яка показала, що у 2020-2021 роках з високою ймовірністю Харківський регіон потрапить до класу регіонів з високим рівнем соціально-економічного розвитку.

3.4 Висновки за розділом 3

Отже, проведене експериментальне застосування обраних методів дозволили:

– зробити висновок, що таксономічний показник знаходиться на однаково низькому рівні (менш «0,5») майже для всіх регіонів. Середнє значення для Харківського регіону складає 54,17%. Найвищий результат показав Дніпропетровський регіон, його середнє значення за 8 років сягнуло 87,5%.

– до регіонів з най нищим рівнем увійшли: Чернівецький, Херсонський, Волинський, Кіровоградський та Чернігівський регіони, значення їхнього показника в середньому знаходиться в межах менш 30%.

– прогнозовані незалежні показники, що впливають на рівень соціально-економічного розвитку регіонів України;

– проведений прогноз показника таксономії для Харківського регіону на 2020-2021 роки. Прогноз таксономії підтвердив результати ансамблевої моделі, яка показала, що у 2020-2021 роках з високою ймовірністю Харківський регіон потрапить до класу регіонів з високим рівнем соціально-економічного розвитку.

ВИСНОВКИ

За проведеним аналізом факторів, що відображують рівень соціально-економічного розвитку в якості вихідних даних визначені такі регіональні статистичні показники: середня заробітна плата, економічно активне населення, безробітне населення, валовий регіональний продукт у розрахунку на одну особу, індекси промислової продукції, сальдо (експорт-імпорт), капітальні інвестиції, обсяги викидів забруднюючих речовин у період з 2012 по 2019 роки.

Для оцінки рівня соціально-економічного розвитку регіонів України побудована концептуальна схема моделювання з визначеними етапами. У роботі виконано дослідження та вибір:

– методу Хельвіга – для розрахунку таксономічного показника рівня соціально-економічного розвитку регіонів;

– метод головних компонент – для трансляції вибірок на двовимірний простір;

– алгоритм класифікації DBSCAN – для кластеризації регіонів за рівнем соціально-економічного розвитку;

– алгоритм розрахунку кореляції Пірсона – для кореляційного аналізу лінійної взаємозалежності показників кластеру;

– модель PanelOLS – для побудови моделі панельних даних регіонів України;

– алгоритм XGBoost – для побудови класифікатора кластеру.

Проведений аналіз за концептуальною схемою моделювання дозволив визначити рівень розвитку регіонів України у 2018–2019 рр., а також здійснити прогнозування їхнього розвитку на 2021 рік.

Практичне використання розробленої концептуальної схеми моделювання з визначеними методами багатовимірного аналізу даних й розробленими програмними застосуваннями дозволить здійснювати оцінку розвитку регіонів і країни у цілому, включаючи оцінку рівня життя населення. Галузь застосування – департаменти та відділи державної влади, відділи підприємств, організацій та компаній, що займаються оцінкою і прогнозом розвитку різних галузей економіки України.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАНЬ

1. Міністерство та Комітет цифрової трансформації України. Перший у 2020 році OpenData Campus відбувся у Харкові [Електронний ресурс] / Міністерство та Комітет цифрової трансформації України. – 2020. – Режим доступу до ресурсу: https://thedigital.gov.ua/news/pershiy-u-2020-rotsi-opendata-campus-vidbuvsya-u-kharkovi.

2. TAPAS. Open Data Campus. Харків [Електронний ресурс] / TAPAS. – 2020. – Режим доступу: http://tapas.org.ua/media/open-data-campus-kharkiv/.

3. Ісакій К.Г. Дослідження методів багатовимірного аналізу даних для оцінки рівня соціально-економічного розвитку регіонів. // World science: problems, prospects and innovations. Abstracts of the 3rd International scientific and practical conference. Perfect Publishing. Toronto, Canada. 2020. Pp. 21-27. URL: https://sci-conf.com.ua/iii-mezhdunarodnaya-nauchno-prakticheskaya-konferentsiya-world-science-problems-prospects-and-innovations-25-27-noyabrya-2020-goda-toronto-kanada-arhiv/.

4. Sitnikov D., Ryabov O., Mishcheriakov I., Kovalenko A. A Rough Set based algebraic approach to modelling complex systems.//The International Journal of Design&Nature and Ecodynamics. Vol.13, No.3(2018), pp.324–329. ISSN: 1755-7437, 1755-7445 (online), DOI:10.2495/DNE-V13-N3-324-329.

5. Офіційний сайт The Social Progress Imperative – [Електронний ресурс]. – Режим доступу до ресурсу: http://www.socialprogressindex.com.

6. Мандибура В. О. Рівень життя населення та механізми його регулювання : автореф. Дис. На здобуття наук. Ступеня канд. Наук / Інститут економіки НАН України. – К., 1999. – 40 с., с. 79.

7. Global Innovation Index – 2016 [Electronic resource].– Access mode : http://www.globalinnovationindex.org, с. XVШ-XIX.

8. Садова У. Подолання бідності в контексті локалізації цілей розвитку тисячоліття в Україні / У. Садова, Н. Андрусишин // Регіон. Економіка. – 2006. – №3. – 450 с.

9. Макарова О. В. Соціальна політика в Україні: Монографія/ О.В. Макарова// Ін–т демографії та соціальних досліджень ім. М.В. Птухи НАН України. – К., 2015. – 244 с.

10. Федулова Л. І. Україна в міжнародних рейтингових оцінках: чинник інноваційно-технологічного розвитку / Л. І. Федулова / / Актуальні проблеми економіки. – 2009. - № 5. – С. 39-53

11. Соціальна політика в Україні: Монографія / О.В. Макарова ; Ін-т демографії та соціальних досліджень ім. М.В. Птухи НАН України. – К., 2015. – 244 с.

12. Савчук Т. О. Використання кластерного аналізу для вирішення задач цільового маркетингу / Т. О. Савчук. // Вимірювальна та обчислювальна техніка в технологічних процесах. – 2011. – №2. – С. 144–148.

13. Чураков Е. П. Математические методы обработки экспериментальных данных в экономике: Учеб.пособие / Е. П. Чураков, 2004. – 240 с.

14. Соціальна політика в Україні: Монографія / О.В. Макарова ; Ін-т демографії та соціальних досліджень ім. М.В. Птухи НАН України. – К., 2015. – 244 с.

15. Hotelling H. Relations between two sets of variates – Biometrika − 1936 − № 28. – 321-377 p.p. – [Electronic resource].– Access mode : http://www.csulb.edu/~jchang9/OnlinePapers/relations_betwen_two_sets_of_variates.pdf.

16. Klecka W. Discriminant Analysis / William Klecka. – Newbury Park, CA : SAGE publications, Inc., 1980. – P. 7–8.

17. Krus D.J., et al. (1976) Rotation in canonical analysis. Educational and Psychological Measurement, 36, pp. 725-730. – [Електронний ресурс] – Режим доступу: http://www.visualstatistics.net/Statistics/Rotation in CA/Rotation in CA.htm.

18. Liang, K.H ., Krus, D.J., & Webb, J.M. (1995) K-fold crossvalidation in canonical analysis. Multivariate Behavioral Research, 30, pp. 539-545. – [Ел. Ресурс] – Режим доступу: www.visualstatistics.net/Statistics/K-fold CA/K-fold CA.doc.

19. XGBoost Documentation [Електронний ресурс] – Режим доступу до ресурсу: https://xgboost.readthedocs.io/en/latest/.

20. Zhernova, P.Y. Data Stream Online Clustering Based on Fuzzy Expectation-Maximization Approach / Deineko, A.O., Zhernova, P.Y., Gordon, B., Zayika, O.O., Pliss,

I., Pabyrivska, N. // Proceedings of the 2018 IEEE 2nd International Conference on Data Stream Mining and Processing, DSMP 2018. – 2018. – Vol. 8478517. – P. 171–176.

21. Zhernova, P.Y. Adaptive Kernel Data Streams Clustering Based on Neural Networks Ensembles in Conditions of Uncertainty about Amount and Shapes of Clusters / Zhernova, P.Y., Deineko, A.O., Bodyanskiy, Y.V., Riepin, V.O. // Proceedings of the 2018 IEEE 2nd International Conference on Data Stream Mining and Processing, DSMP 2018. – 2018. – Vol. 8478616. – P. 7–12.

22. Stevens J. Applied multivariate statistics for the social sciences. – L. Erlbaum Associates Inc. Hillsdale, NJ, USA ©1986. – [Electronic resource].– Access mode : http://books.google.com/books/about/Applied_multivariate_statistics_for_the.html?id=mK0MtyWa7-QC.

23. Krus D.J., et al. (1976) Rotation in canonical analysis. Educational and Psychological Measurement, 36, pp. 725-730. – [Електронний ресурс] – Режим доступу: http://www.visualstatistics.net/Statistics/Rotation in CA/Rotation in CA.htm;

24. 2.3.7. DBSCAN [Електронний ресурс] – Режим доступу до ресурсу: https://scikit-learn.org/stable/modules/clustering.html#dbscan.

25. Марченко В. М. Эконометрика и экономико-математические методы и модели. В 2 ч. Ч. 1. Эконометрика : учеб. Пособие для студентов учреждений высшего образования по экономическим специальностям / В. М. Марченко, Н. П. Можей//. – Минск : БГТУ, 2011. – 157 с.

26. Варналий З.С. Регіони України: проблеми та пріоритети соціально-економичного розвитку: Монографія. — К.: Знання України, 2005. — 498 с.

27. Василенко В.Н. Архитектура регионального экономического пространства: Монография / НАН Украины. Ин-т экономико-правовых исследований. — Донецк: ООО «Юго-Восток, ЛТД», 2006. — 311 с.

28. Максимова Т.С. Формування механізму діагностування та прогнозування економічного і соціального розвитку регіонів: Дис. Д. екон. Наук. — Донецьк, 2004. — 447 с.

29. Кузьменко Л.М. Управление функционированием и развитием экономики региона: Монография / НАН Украины, Ин-т экономики промышленности. — Донецк, 2004. — 272 с.

30. Самуэльсон П. Экономика: В 2 т.: Пер. С англ. — Т. II. — М.: НПО «Алгон», ВНИИСИ, 1992. — 352 с.