

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ПОЛІСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ

ФАКУЛЬТЕТ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ, ОБЛІКУ ТА ФІНАНСІВ

Кафедра комп'ютерних технологій і моделювання систем

Кваліфікаційна робота
на правах рукопису

Сліпчук Ярослав Володимирович

УДК 004.422:910.1:338.48

КВАЛІФІКАЦІЙНА РОБОТА

**РОЗРОБКА АЛГОРИТМУ ПОБУДОВИ ТУРИСТИЧНИХ КЛАСТЕРІВ
РЕГІОНУ З ВИКОРИСТАННЯМ ГЕОІНФОРМАЦІЙНИХ СИСТЕМ**

Ф3 (122) «Комп'ютерні науки»

Подається на здобуття освітнього ступеня магістр

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

Я.В.Сліпчук

(підпис, ініціали та прізвище здобувача вищої освіти)

Керівник роботи
Топольницький Павло Петрович
кандидат технічних наук, доцент

Житомир – 2025

Висновок кафедри комп'ютерних технологій і моделювання систем:

за результатами попереднього захисту: _____

Протокол засідання кафедри комп'ютерних технологій і моделювання систем

№ _____ від «_____» _____ 20__ р.

Завідувач кафедри комп'ютерних технологій і моделювання систем

к.п.н., доцент

(науковий ступінь, вчене звання)

(підпис)

М. О. Ковальчук

(прізвище, ім'я, по батькові)

«_____» _____ 20__ р.

Результати захисту кваліфікаційної роботи

Здобувач вищої освіти _____ захистив (ла)

(прізвище, ім'я, по батькові)

кваліфікаційну роботу з оцінкою:

сума балів за 100-бальною шкалою _____

за шкалою ECTS _____

за національною шкалою _____

Секретар ЕК

лаборант кафедри

(науковий ступінь, вчене звання)

(підпис)

В. В. Корольчук

(прізвище, ім'я, по батькові)

АНОТАЦІЯ

Сліпчук Я.В. *Розробка алгоритму побудови туристичних кластерів регіону з використанням геоінформаційних систем.* – Кваліфікаційна робота на правах рукопису. Кваліфікаційна робота на здобуття освітнього ступеня магістр за спеціальністю 122 – Комп’ютерні науки. – Поліський національний університет, Житомир, 2025.

Обсяг кваліфікаційної роботи: сторінок (8 – рисунків, 1 – таблиць, 5 – додатків, 47 – джерел).

Ключові слова: геоінформаційні системи, туристичний потенціал, кластеризація, просторовий аналіз, туристичні кластери, геопросторові дані, веб-застосунок.

Кваліфікаційна робота присвячена розробці алгоритму побудови туристичних кластерів регіону з використанням геоінформаційних систем. У роботі виконано аналіз предметної області оцінки туристичного потенціалу регіонів, сучасних підходів до просторового аналізу та методів кластеризації геопросторових даних, а також наявних програмних рішень у сфері туристичної аналітики.

Наукова новизна роботи полягає у застосуванні кластерного аналізу для виявлення просторово згрупованих туристичних зон регіону на основі географічного розміщення туристичних об’єктів із використанням геоінформаційних технологій. У межах роботи реалізовано підхід до формування туристичних кластерів, що поєднує просторові дані та кількісні метрики якості кластеризації для обґрунтованого аналізу туристичного потенціалу територій.

Практичне значення отриманих результатів полягає у створенні веб-орієнтованого геоінформаційного застосунку для аналізу та візуалізації туристичних кластерів регіону. Розроблене програмне забезпечення може бути використане органами місцевого самоврядування та фахівцями у сфері туризму для оцінки туристичної інфраструктури й планування розвитку територій, а також у навчальному процесі та прикладних дослідженнях з геоінформаційних систем.

SUMMARY

Slipchuk Y.V. *Development of an algorithm for constructing tourist clusters of a region using geographic information systems.* – **Qualification work in the form of a manuscript.** Qualification work for the degree of Master in specialty 122 – Computer Science. – Polissya National University, Zhytomyr, 2025.

Qualification work volume: pages (8 – figures, 1 – tables, 5 – appendices, 47 – references).

Keywords: geographic information systems, tourism potential, clustering, spatial analysis, tourist clusters, geospatial data, web application.

The qualification work is devoted to the development of an algorithm for constructing tourist clusters of a region using geographic information systems. The paper analyzes the subject area of tourism potential assessment, modern approaches to spatial analysis, clustering methods for geospatial data, as well as existing software solutions in the field of tourism analytics.

The scientific novelty of the work lies in the application of clustering analysis to identify spatially grouped tourist zones based on the geographic distribution of tourist objects using GIS technologies. Within the framework of the research, an approach combining spatial data and quantitative clustering quality metrics was implemented to support the analysis of regional tourism potential.

The practical significance of the obtained results lies in the development of a web-based geoinformation application for analyzing and visualizing tourist clusters of a region. The developed software can be used by local authorities and tourism specialists for assessing tourism infrastructure and supporting regional development planning, as well as in educational and applied research in the field of geographic information systems.

ЗМІСТ

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ	5
ВСТУП.....	7
РОЗДІЛ 1. АНАЛІЗ СУЧАСНИХ ПІДХОДІВ ДО ОЦІНКИ ТУРИСТИЧНОГО ПОТЕНЦІАЛУ РЕГІОНІВ	10
1.1. Туристичний потенціал як об'єкт моделювання	10
1.2. Геоінформаційні системи як інструмент просторового аналізу	12
1.3. Методи кластеризації у просторових даних: огляд і класифікація	14
1.4. Сучасні дослідження і приклади кластерного аналізу в туризмі	16
Висновки до першого розділу	18
РОЗДІЛ 2. ПОСТАНОВКА ЗАДАЧІ ТА РОЗРОБКА АЛГОРИТМУ ПОБУДОВИ ТУРИСТИЧНИХ КЛАСТЕРІВ.....	20
2.1. Формалізація задачі кластеризації туристичних даних	20
2.2. Вибір і обґрунтування методів кластеризації	24
2.3. Архітектура програмного модуля для побудови кластерів.....	26
Висновки до другого розділу.....	27
РОЗДІЛ 3 РЕАЛІЗАЦІЯ ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ	29
3.1. Реалізація програмного забезпечення.....	29
3.2. Інтеграція з геоінформаційними сервісами	32
3.3. Проведення симуляцій та експериментів.....	33
3.4. Аналіз отриманих рішень	37
Висновки до третього розділу	38
РОЗДІЛ 4. ЕКСПЕРИМЕНТАЛЬНІ ДОСЛІДЖЕННЯ ТА ПОРІВНЯЛЬНА ОЦІНКА АЛГОРИТМІВ	40
4.1. Опис тестових даних і сценаріїв експериментального моделювання.....	40
4.2. Проведення експериментів та аналіз впливу параметрів кластеризації	41
4.3. Порівняльна оцінка результатів кластеризації за метриками якості.....	43
Висновки до четвертого розділу	45
ВИСНОВКИ	46
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ	48

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ

ГІС – геоінформаційна система.

GIS (Geographic Information System) – інформаційна система для збору, зберігання, аналізу та візуалізації просторових даних.

Геопросторові дані – дані, що містять інформацію про просторове розташування об'єктів на земній поверхні.

Просторовий аналіз – сукупність методів аналізу географічно прив'язаних даних з урахуванням їх просторових взаємозв'язків.

Кластеризація – процес групування об'єктів у кластери на основі їх подібності за визначеними ознаками.

Туристичний кластер – просторово згрупована сукупність туристичних об'єктів, об'єднаних спільними характеристиками та територіальним розміщенням.

К – параметр алгоритму кластеризації, що визначає кількість кластерів.

K-Means – алгоритм кластерного аналізу, який здійснює поділ об'єктів на К кластерів шляхом мінімізації внутрішньокластерної відстані.

Silhouette Score – показник якості кластеризації, що характеризує ступінь згуртованості об'єктів у межах кластерів та їх відокремленість від інших кластерів.

Davies–Bouldin Index – індекс якості кластеризації, що відображає ступінь перекриття кластерів; менші значення відповідають кращій якості кластеризації.

Calinski–Harabasz Index – показник якості кластеризації, що характеризує співвідношення між міжкластерною та внутрішньокластерною варіацією.

GeoJSON – формат зберігання та обміну геопросторовими даними у вигляді структурованих JSON-об'єктів.

OpenStreetMap (OSM) – відкрита картографічна платформа для створення та використання просторових даних.

Веб-застосунок – програмний продукт, що працює у веб-браузері та забезпечує доступ до функціоналу через мережу Інтернет.

КАЛЕНДАРНИЙ ПЛАН РОБОТИ

№	Назва етапів кваліфікаційної роботи	Строк виконання етапів роботи
1	Розробка плану роботи.	30.10.2025 – 02.11.2025
2	Збір вихідних даних.	03.11.2025 – 10.11.2025
3	Обробка теоретичних джерел.	11.11.2025 – 20.11.2025
4	Розробка першого і другого розділу.	21.11.2025 – 30.11.2025
5	Розробка третього і четвертого розділу.	01.12.2025 – 05.12.2025
6	Оформлення кваліфікаційної роботи.	06.12.2025 – 08.12.2025
7	Захист кваліфікаційної роботи.	22.12.2025

Керівник роботи

_____ П. П. Топольницький

« ____ » _____ 2025 р.

Завдання одержав

«30» жовтня 2025 р.

Підпис студента _____

ВСТУП

У сучасних умовах розвитку регіонів туризм є важливою складовою соціально-економічного зростання та одним із чинників підвищення привабливості територій. Туристична діяльність сприяє розвитку інфраструктури, створенню нових робочих місць, залученню інвестицій і формуванню позитивного іміджу регіону. Разом з тим ефективний розвиток туризму потребує обґрунтованого планування, яке має базуватися на аналізі туристичних ресурсів, умов їх розміщення та можливостей комплексного використання територій.

На практиці оцінка туристичного потенціалу регіонів часто здійснюється з використанням окремих статистичних показників або експертних висновків. Такий підхід не завжди дозволяє повною мірою врахувати просторові особливості розміщення туристичних об'єктів, їх взаємозв'язки та вплив навколишніх чинників. У результаті прийняття управлінських рішень у сфері розвитку туризму може ґрунтуватися на неповній або узагальненій інформації, що знижує ефективність використання туристичного потенціалу територій.

Особливу роль у розвитку туризму відіграє просторовий чинник, оскільки туристичні об'єкти розміщуються нерівномірно та утворюють певні територіальні групи. Такі групи об'єктів мають спільні характеристики, зокрема функціональне призначення, транспортну доступність або орієнтацію на певні види туризму. Виявлення подібних груп є важливим для формування туристичних маршрутів, розвитку інфраструктури та визначення перспективних напрямів розвитку туризму в регіоні.

Сучасні інформаційні технології, зокрема геоінформаційні системи, надають широкі можливості для аналізу просторових даних. Геоінформаційні системи дозволяють зберігати, обробляти та аналізувати інформацію про туристичні об'єкти з урахуванням їх географічного розташування. Використання ГІС забезпечує наочне представлення результатів аналізу у вигляді карт, що значно полегшує сприйняття інформації та підтримує процес прийняття рішень.

Разом із тим ефективність геоінформаційних систем значно зростає при поєднанні з методами аналізу даних. Одним із таких методів є кластерний аналіз, який дозволяє автоматично групувати об'єкти за подібними ознаками. У випадку туристичних даних кластеризація дає змогу виділяти групи туристичних об'єктів, які мають спільні характеристики та просторове розташування. Поєднання методів кластерного аналізу з геоінформаційними системами створює умови для формування туристичних кластерів і комплексної оцінки туристичного потенціалу регіонів.

Незважаючи на наявність наукових праць, присвячених оцінці туристичного потенціалу та використанню геоінформаційних систем у туризмі, питання розробки алгоритмів побудови туристичних кластерів залишається недостатньо опрацьованим. Більшість досліджень зосереджені або на теоретичних аспектах кластеризації, або на застосуванні ГІС як інструменту візуалізації без глибокого алгоритмічного аналізу. Це зумовлює потребу у створенні комплексного підходу, який поєднує можливості геоінформаційних систем і методи кластерного аналізу.

У зв'язку з цим тема кваліфікаційної роботи, присвячена розробці алгоритму побудови туристичних кластерів регіону з використанням геоінформаційних систем, є актуальною та відповідає сучасним потребам регіонального розвитку туризму. Реалізація такого підходу дозволяє підвищити обґрунтованість оцінки туристичного потенціалу та забезпечити більш ефективне планування туристичної діяльності.

Метою кваліфікаційної роботи є розробка алгоритму побудови туристичних кластерів регіону з використанням геоінформаційних систем для аналізу просторового розміщення туристичних об'єктів і оцінки туристичного потенціалу територій. Досягнення цієї мети передбачає аналіз сучасних підходів до оцінки туристичного потенціалу, дослідження можливостей геоінформаційних систем, а також застосування методів кластерного аналізу для обробки туристичних даних.

Об'єктом дослідження у роботі є процеси аналізу туристичної інфраструктури регіону з урахуванням просторових характеристик туристичних об'єктів. Предметом дослідження є методи та алгоритми побудови туристичних кластерів на основі геоінформаційних систем і методів кластерного аналізу.

У процесі виконання кваліфікаційної роботи використовувалися методи аналізу наукових джерел для вивчення сучасних підходів до оцінки туристичного потенціалу, методи геоінформаційного аналізу для обробки та візуалізації просторових даних, а також методи кластерного аналізу та алгоритмічного моделювання для побудови туристичних кластерів. Отримані результати подано у вигляді картографічних матеріалів і аналітичних висновків.

Наукова новизна роботи полягає у розробці алгоритму побудови туристичних кластерів регіону з використанням геоінформаційних систем, який забезпечує комплексний облік просторових характеристик туристичних об'єктів. Запропонований підхід дозволяє удосконалити процес оцінки туристичного потенціалу регіонів та підвищити наочність і обґрунтованість результатів аналізу.

Практичне значення одержаних результатів полягає у можливості використання розробленого алгоритму для аналізу туристичної інфраструктури регіонів, формування карт туристичного потенціалу та підтримки прийняття рішень у сфері регіонального розвитку туризму.

РОЗДІЛ 1. АНАЛІЗ СУЧАСНИХ ПІДХОДІВ ДО ОЦІНКИ ТУРИСТИЧНОГО ПОТЕНЦІАЛУ РЕГІОНІВ

1.1. Туристичний потенціал як об'єкт моделювання

Туристичний потенціал регіону є однією з ключових характеристик території, що визначає можливості розвитку туристичної діяльності та рівень її привабливості для відвідувачів. У сучасних умовах туристичний потенціал розглядається як сукупність ресурсів і умов, які можуть бути використані для організації туристичної діяльності та формування туристичного продукту. До таких ресурсів належать природні об'єкти, історико-культурна спадщина, елементи рекреаційної інфраструктури, а також умови доступності та обслуговування туристів.

Особливістю туристичного потенціалу є його комплексний характер. Наявність окремих туристичних об'єктів не гарантує високого рівня туристичної привабливості регіону. Важливу роль відіграє взаємодія між об'єктами, їх територіальна близькість, можливість поєднання в маршрути та туристичні зони. Саме тому туристичний потенціал доцільно розглядати не як просту суму ресурсів, а як систему, елементи якої перебувають у взаємозв'язку та взаємному впливі.

Вагоме значення у формуванні туристичного потенціалу має просторовий чинник. Усі туристичні об'єкти мають конкретне географічне розташування, а їх взаємне розміщення визначає умови доступності, зручність пересування туристів та можливість формування туристичних маршрутів. У межах одного регіону туристичні ресурси можуть бути розміщені нерівномірно, що призводить до появи зон з високою концентрацією туристичної активності та територій, де туристична діяльність розвинена слабо.

Територіальна концентрація туристичних об'єктів сприяє формуванню локальних туристичних центрів і зон рекреації. Водночас розпорошене розміщення об'єктів ускладнює їх використання та знижує загальну привабливість території. Це свідчить про необхідність аналізу не лише кількісних характеристик туристичних ресурсів, а й їх просторових зв'язків, що є важливим для оцінки туристичного потенціалу регіону.

Ще однією важливою особливістю туристичного потенціалу є його динамічний характер. Рівень туристичної привабливості території може змінюватися з часом під впливом розвитку транспортної інфраструктури, появи нових туристичних об'єктів, зміни туристичних уподобань населення та соціально-економічних умов. Такі зміни потребують регулярного оновлення інформації про туристичні ресурси та їх просторове розміщення.

У зв'язку з цим оцінка туристичного потенціалу не може бути одноразовою процедурою. Вона повинна здійснюватися з урахуванням можливості змін та адаптації до нових умов. Це ускладнює використання традиційних методів аналізу, які часто ґрунтуються на статичних даних і не враховують просторову динаміку туристичної інфраструктури.

Значна частина інформації, що використовується для оцінки туристичного потенціалу, має просторовий характер і надходить з різних джерел. Це можуть бути картографічні матеріали, статистичні дані, інформація про туристичні об'єкти та транспортну мережу. Узагальнення такої інформації потребує застосування спеціалізованих інструментів аналізу, здатних працювати з геопросторовими даними.

З огляду на зазначене туристичний потенціал регіону доцільно розглядати як об'єкт моделювання. Моделювання дозволяє формалізувати характеристики туристичних об'єктів і територій, встановити взаємозв'язки між ними та представити туристичну систему у вигляді узагальненої моделі. Такий підхід створює умови для комплексного аналізу туристичних ресурсів і прийняття обґрунтованих рішень щодо розвитку туризму.

Моделювання туристичного потенціалу дає змогу перейти від описового аналізу до системного підходу, який враховує просторові, функціональні та інфраструктурні характеристики території. Це особливо важливо для регіонального рівня, де необхідно оцінювати великі масиви туристичних даних і виявляти просторові закономірності їх розміщення.

Таким чином, туристичний потенціал регіону як об'єкт моделювання потребує використання сучасних інформаційних технологій, здатних працювати з

просторовими даними. Саме це обґрунтовує доцільність застосування геоінформаційних систем і методів кластерного аналізу для подальшої оцінки туристичного потенціалу та формування туристичних кластерів.

1.2 Геоінформаційні системи як інструмент просторового аналізу

Геоінформаційні системи є важливим інструментом аналізу територій, який дозволяє працювати з даними, що мають просторову прив'язку. Їх застосування ґрунтується на можливості поєднання інформації про об'єкти з їх географічним розташуванням, що дає змогу аналізувати не лише окремі характеристики, а й просторові взаємозв'язки між об'єктами.[1] Саме ця особливість робить геоінформаційні системи доцільними для дослідження туристичного потенціалу регіонів.

У загальному вигляді геоінформаційна система є програмно-апаратним комплексом, призначеним для збирання, зберігання, обробки та відображення просторових даних.[2] Інформація в ГІС подається у вигляді об'єктів, які мають координати та набір атрибутивних характеристик. Це дозволяє одночасно аналізувати географічне положення туристичних об'єктів і їх властивості, зокрема тип, призначення та належність до певної категорії.

У туристичній галузі геоінформаційні системи використовуються для формування баз даних туристичних об'єктів, аналізу їх розміщення та оцінки умов доступності. Наявність структурованих геопросторових даних дає можливість систематизувати інформацію про туристичну інфраструктуру регіону та створити основу для подальшого аналізу. Такий підхід є особливо важливим у разі роботи з великими масивами даних, що надходять із різних джерел.

Важливим елементом геоінформаційних систем є використання тематичних шарів.[3] Кожен шар містить інформацію про певний тип об'єктів або характеристик території, наприклад туристичні об'єкти, транспортну мережу, адміністративні межі чи природні умови. Робота з кількома шарами дозволяє аналізувати взаємний вплив різних чинників і отримувати цілісне уявлення про просторову структуру туристичного потенціалу регіону.

Геоінформаційні системи забезпечують виконання просторового аналізу, який є ключовим для дослідження туристичних територій. До основних просторових операцій належать аналіз відстаней між туристичними об'єктами, визначення зон доступності, побудова буферних зон та оцінка щільності розміщення об'єктів. Застосування таких операцій дозволяє оцінити не лише кількість туристичних ресурсів, а й умови їх використання з урахуванням просторового розташування.[4]

Аналіз транспортної доступності відіграє важливу роль у процесі оцінки туристичного потенціалу. Навіть за наявності значної кількості туристичних об'єктів складний доступ або відсутність зручного транспортного сполучення можуть знижувати привабливість території. Геоінформаційні системи дозволяють оцінювати відстані до туристичних об'єктів, аналізувати дорожню мережу та визначати території, які є найбільш доступними для туристів.

Оцінка щільності розміщення туристичних об'єктів дозволяє виявляти зони концентрації туристичної активності. Такі зони можуть розглядатися як сформовані туристичні центри або як потенційні осередки розвитку туризму. Території з низькою щільністю туристичних об'єктів, навпаки, потребують додаткового аналізу щодо можливостей розвитку інфраструктури або залучення нових туристичних ресурсів.

Сучасні геоінформаційні системи широко використовують веб-технології, що дозволяє створювати інтерактивні карти та геоінформаційні сервіси. Веб-орієнтовані рішення спрощують доступ до просторових даних і забезпечують можливість їх використання в режимі реального часу. У сфері туризму це дає змогу оперативно оновлювати інформацію про туристичні об'єкти та надавати користувачам актуальні дані.[5]

Інтерактивні геоінформаційні сервіси можуть використовуватися не лише для аналізу, а й для представлення результатів досліджень. Відображення туристичних об'єктів, маршрутів і зон туристичної активності на карті значно полегшує сприйняття інформації та підвищує ефективність її використання у

процесі прийняття рішень. Це є важливим для органів місцевого самоврядування та фахівців у сфері туризму.

Використання геоінформаційних систем створює основу для впровадження методів аналізу даних, зокрема кластерного аналізу. Просторова структура даних, сформована в ГІС, дозволяє застосовувати алгоритми групування туристичних об'єктів з урахуванням їх географічного розташування. Це є необхідною передумовою для побудови туристичних кластерів і комплексної оцінки туристичного потенціалу регіонів.

1.3 Методи кластеризації у просторових даних: огляд і класифікація

Кластеризація є одним із поширених методів аналізу даних, що використовується для об'єднання об'єктів у групи на основі подібності їх характеристик. У контексті просторових даних кластеризація дозволяє виявляти групи об'єктів, які мають спільні ознаки та розташовані на певній території. Для туристичної галузі це має особливе значення, оскільки туристичні об'єкти, як правило, не розміщуються випадково, а формують локальні зони концентрації, пов'язані з природними, історико-культурними або інфраструктурними чинниками.

Просторові дані відрізняються від звичайних наборів даних наявністю координат, які визначають географічне положення об'єктів. Крім атрибутивних характеристик, таких як тип туристичного об'єкта або його функціональне призначення, важливу роль відіграє просторове взаєморозташування об'єктів. Це означає, що під час кластеризації необхідно враховувати не лише подібність властивостей, а й відстані між об'єктами, а також особливості їх просторової організації.[6]

Методи кластеризації у просторових даних можна умовно поділити на кілька основних груп залежно від принципу формування кластерів. Одну з таких груп становлять методи, що передбачають поділ об'єктів на наперед задану кількість груп. Такі методи є відносно простими у використанні та дозволяють швидко отримати результат, проте вони потребують попереднього визначення кількості

кластерів, що не завжди є можливим у разі аналізу туристичних об'єктів із нерівномірним просторовим розміщенням.

Іншу групу становлять методи, які формують кластери на основі щільності розміщення об'єктів у просторі. Ці підходи дозволяють виявляти групи довільної форми та відокремлювати об'єкти, що не належать до жодного кластера. Для аналізу туристичних об'єктів це є важливою перевагою, оскільки реальне розміщення туристичної інфраструктури часто не підпорядковується чітким геометричним закономірностям.

Під час аналізу туристичних даних доцільно враховувати поєднання просторових і атрибутивних характеристик. Туристичні об'єкти можуть належати до різних категорій, виконувати різні функції та мати різний рівень привабливості. Кластеризація дозволяє об'єднувати такі об'єкти в групи з урахуванням їх територіальної близькості та функціональної подібності, що забезпечує більш повне розуміння структури туристичного потенціалу регіону.

Застосування кластерного аналізу у просторових даних значно спрощує роботу з великими масивами туристичної інформації. За наявності значної кількості туристичних об'єктів ручний аналіз їх розміщення є складним і малоефективним. Кластеризація дозволяє автоматизувати процес групування об'єктів і виявляти закономірності, які складно помітити під час традиційного аналізу.

Поєднання методів кластеризації з геоінформаційними системами розширює можливості просторового аналізу. Геоінформаційні системи забезпечують зберігання та обробку просторових даних, тоді як кластеризація дозволяє виконувати їх аналітичне узагальнення. У результаті формується основа для створення туристичних кластерів, які можуть бути використані у процесі планування розвитку туристичної інфраструктури.[7]

Виявлені туристичні кластери дозволяють визначати території з підвищеним туристичним потенціалом, а також зони, які потребують додаткових інвестицій або розвитку інфраструктури. Це є важливим для прийняття управлінських рішень на

регіональному рівні, оскільки дозволяє спрямовувати ресурси на найбільш перспективні напрями розвитку туризму.

Методи кластеризації у просторових даних створюють основу для формалізованого аналізу туристичних ресурсів. Їх використання дає змогу перейти від загального опису туристичних об'єктів до детального дослідження їх просторової структури та взаємозв'язків. Саме це забезпечує можливість розробки алгоритмів побудови туристичних кластерів, адаптованих до особливостей конкретного регіону.[8]

1.4 Сучасні дослідження і приклади кластерного аналізу в туризмі

У сучасних наукових дослідженнях питання застосування кластерного аналізу в туризмі розглядається у контексті підвищення ефективності оцінки туристичного потенціалу територій та обґрунтування управлінських рішень у сфері регіонального розвитку. Дослідники все частіше звертаються до просторових методів аналізу, оскільки туристичні ресурси мають чітку територіальну прив'язку, а їх розміщення безпосередньо впливає на формування туристичних потоків і розвиток інфраструктури.

У наукових роботах кластерний аналіз використовується для групування туристичних об'єктів за просторовими та функціональними ознаками. Це дозволяє виділяти туристичні зони, що мають спільні характеристики, такі як тип туристичних ресурсів, рівень доступності або орієнтація на певні види туризму. Такий підхід спрощує аналіз туристичної інфраструктури та дає змогу перейти від загальних оцінок до детального вивчення окремих територій.

Значна кількість досліджень присвячена застосуванню кластерного аналізу для оцінки туристичного потенціалу регіонів. У межах таких робіт туристичні об'єкти розглядаються як елементи просторової системи, а кластеризація використовується для виявлення зон концентрації туристичних ресурсів. Отримані результати дозволяють визначати території з високим рівнем туристичної привабливості, а також зони, де туристичний потенціал використовується недостатньо.

У багатьох дослідженнях кластерний аналіз поєднується з використанням геоінформаційних систем. Геоінформаційні системи забезпечують зручне середовище для зберігання та обробки просторових даних, тоді як кластеризація дозволяє виконувати їх аналітичне узагальнення. У результаті формуються картографічні моделі туристичних кластерів, які наочно відображають просторову структуру туристичного потенціалу регіонів.

Практичні приклади застосування кластерного аналізу в туризмі демонструють його ефективність для планування розвитку туристичної діяльності. Результати кластеризації використовуються для формування туристичних маршрутів, визначення пріоритетних напрямів розвитку інфраструктури та обґрунтування інвестиційних рішень. Виділення туристичних кластерів дозволяє зосереджувати ресурси на територіях з найбільшим потенціалом та підвищувати загальну ефективність розвитку туризму.

Окрему увагу в сучасних дослідженнях приділяють аналізу туристичних кластерів на регіональному рівні. Саме на цьому рівні проявляється нерівномірність розміщення туристичних об'єктів і виникає потреба у просторовому аналізі. Кластерний підхід дозволяє враховувати специфіку регіонів, особливості їх туристичних ресурсів та існуючу інфраструктуру, що є важливим для розробки регіональних програм розвитку туризму.

Використання кластерного аналізу у поєднанні з геоінформаційними системами дає змогу автоматизувати обробку великих обсягів туристичних даних. [9] Це особливо актуально в умовах постійного оновлення інформації про туристичні об'єкти, появи нових локацій і змін у туристичних потоках. Автоматизована кластеризація дозволяє оперативно отримувати актуальні результати аналізу та використовувати їх у практичній діяльності.

Сучасні дослідження також показують, що кластерний аналіз може використовуватися як інструмент підтримки управлінських рішень. Виділення туристичних кластерів дозволяє органам місцевого самоврядування та профільним установам більш обґрунтовано планувати розвиток туристичної інфраструктури,

визначати напрями модернізації транспортної мережі та розробляти програми популяризації туристичних територій.[10]

Аналіз наукових праць і практичних прикладів застосування кластерного аналізу в туризмі підтверджує доцільність використання цього підходу для комплексної оцінки туристичного потенціалу регіонів. Поєднання кластерного аналізу з геоінформаційними системами забезпечує можливість врахування просторових особливостей туристичних об'єктів і створює умови для формування туристичних кластерів на основі об'єктивних даних.

Зазначені підходи є особливо актуальними для розробки алгоритмів побудови туристичних кластерів. Узагальнення результатів сучасних досліджень свідчить про необхідність використання геоінформаційних систем як базового інструменту просторового аналізу та кластерного аналізу як методу групування туристичних об'єктів. Це створює методичну основу для подальшої розробки алгоритму побудови туристичних кластерів регіону, що розглядається у наступних розділах кваліфікаційної роботи.

Висновки до першого розділу

У першому розділі кваліфікаційної роботи проаналізовано сучасні підходи до оцінки туристичного потенціалу регіонів та визначено основні теоретичні засади його дослідження. Розглянуто туристичний потенціал як комплексний об'єкт, що формується під впливом природних, історико-культурних та інфраструктурних чинників і характеризується чітко вираженим просторовим розміщенням. Обґрунтовано доцільність розгляду туристичного потенціалу не як сукупності окремих об'єктів, а як системи взаємопов'язаних елементів, що функціонують у межах певної території.

У процесі аналізу встановлено, що оцінка туристичного потенціалу потребує врахування просторових особливостей розміщення туристичних об'єктів, їх територіальної концентрації та доступності. Нерівномірність розподілу туристичних ресурсів у межах регіону ускладнює використання традиційних методів аналізу та зумовлює необхідність застосування просторово орієнтованих підходів.

Розглянуто роль геоінформаційних систем як інструменту просторового аналізу туристичних даних. Визначено, що використання ГІС дозволяє поєднувати просторові та атрибутивні характеристики туристичних об'єктів, виконувати аналіз відстаней, доступності та щільності розміщення об'єктів. Застосування геоінформаційних систем створює інформаційну основу для систематизації туристичних ресурсів і проведення комплексного аналізу туристичної інфраструктури регіонів.

У розділі також проаналізовано методи кластеризації просторових даних та їх застосування для аналізу туристичних об'єктів. Встановлено, що кластерний аналіз є ефективним інструментом групування туристичних об'єктів з урахуванням їх просторової близькості та функціональних характеристик. Поєднання методів кластеризації з геоінформаційними системами дозволяє автоматизувати процес аналізу великих масивів туристичних даних і виявляти територіальні зони з підвищеним туристичним потенціалом.

Огляд сучасних досліджень і практичних прикладів застосування кластерного аналізу в туризмі підтвердив доцільність використання просторових методів для оцінки туристичного потенціалу на регіональному рівні. Встановлено, що результати кластеризації можуть бути використані для планування розвитку туристичної інфраструктури, формування туристичних маршрутів та обґрунтування управлінських рішень.

Отримані у першому розділі результати формують теоретичну та методичну основу для подальшої розробки алгоритму побудови туристичних кластерів регіону з використанням геоінформаційних систем, що є предметом дослідження у наступних розділах кваліфікаційної роботи.

РОЗДІЛ 2. ПОСТАНОВКА ЗАДАЧІ ТА РОЗРОБКА АЛГОРИТМУ ПОБУДОВИ ТУРИСТИЧНИХ КЛАСТЕРІВ

2.1. Формалізація задачі кластеризації туристичних даних

Розробка алгоритму побудови туристичних кластерів потребує чіткого визначення задачі, яку необхідно вирішити, тому формалізація задачі є першим і одним з найважливіших етапів роботи. Формалізація включає опис вхідних даних, визначення мети обробки та встановлення критеріїв оцінки результатів, що дозволяє перейти від загального розуміння проблеми до конкретних математичних формулювань. Правильна постановка задачі створює основу для вибору методів її вирішення та дає можливість об'єктивно оцінити якість отриманих результатів, порівнюючи їх з встановленими критеріями.

Кластеризація туристичних даних полягає у розподілі туристичних об'єктів регіону на групи за схожими ознаками, при цьому головна ідея кластеризації полягає в тому, що об'єкти однієї групи мають бути максимально подібними між собою, тоді як об'єкти різних груп мають суттєво відрізнятися.[11] У контексті туристичної галузі кластеризація дозволяє виявляти території зі схожими туристичними ресурсами, визначати зони з високою концентрацією туристичних об'єктів та формувати групи об'єктів для спільного просування на туристичному ринку. Для Житомирської області кластеризація туристичних даних має особливе практичне значення, оскільки виявлення груп схожих об'єктів дозволяє краще розуміти структуру туристичних ресурсів території, визначати сильні та слабкі сторони туристичної інфраструктури, а також приймати обґрунтовані рішення щодо розвитку туристичної галузі регіону в цілому.

Вхідними даними для алгоритму кластеризації є набір туристичних об'єктів регіону, де кожен об'єкт представляє собою окрему туристичну локацію з певними характеристиками. Математично цей набір записується як множина елементів, що показано у формулі 2.1.

$$D = \{0_1, 0_2, \dots, 0_n\},$$

де D – множина туристичних об'єктів регіону; $0_1, 0_2, \dots, 0_n$ – окремі туристичні об'єкти; n – загальна кількість об'єктів у множині.

Для Житомирської області було зібрано та систематизовано дані про 1864 туристичних об'єкти, які охоплюють різні категорії туристичних ресурсів: від історичних пам'яток та культурних закладів до природних об'єктів та місць розміщення туристів.[12] Така кількість об'єктів є достатньою для проведення статистично значущого аналізу та отримання надійних результатів кластеризації, оскільки забезпечує представництво всіх основних типів туристичних ресурсів регіону та дозволяє виявити закономірності у їх просторовому розподілі.

Кожен туристичний об'єкт характеризується набором ознак, які описують його властивості та дозволяють порівнювати об'єкти між собою. До основних ознак належать географічні координати, які визначають просторове розташування об'єкта на території регіону і складаються з двох компонентів: широти, що визначає положення об'єкта відносно екватора, та довготи, що визначає положення відносно нульового меридіана. Крім координат, кожен об'єкт має категорію, яка вказує на його функціональне призначення та тип туристичного ресурсу, а також може мати рейтинг, який відображає оцінку відвідувачів та характеризує якість обслуговування або привабливість об'єкта. Сукупність цих ознак утворює вектор ознак об'єкта, структуру якого показано у формулі 2.2.

$$o_i = (lat_i, lon_i, cat_i, r_i, a_{i1}, a_{i2}, \dots, a_{im}),$$

де o_i – i -й туристичний об'єкт; lat_i – географічна широта об'єкта; lon_i – географічна довгота об'єкта; cat_i – категорія туристичного об'єкта; r_i – рейтинг об'єкта за оцінками відвідувачів.

Туристичні об'єкти Житомирської області було класифіковано за сімома основними категоріями, що базується на їх функціональному призначенні та відповідає загальноприйнятим підходам до систематизації туристичних ресурсів.

Перша категорія – гастрономічні заклади, яка включає ресторани, кафе, бари та інші місця громадського харчування – є найбільш численною і становить 27,3% від загальної кількості. Друга категорія – культурні об'єкти, що охоплює музеї, театри, бібліотеки та центри культури – 25,6% від загальної кількості. Третя категорія – історичні пам'ятки, включаючи замки, церкви, пам'ятники та археологічні об'єкти – налічує 20,5%. Четверта категорія – парки та рекреаційні зони, що об'єднує міські парки, сквери та зони відпочинку – містить 9,9%. П'ята категорія – торговельні об'єкти, включаючи торговельні центри, ринки та магазини – налічує 7,0%. Шоста категорія – природні об'єкти, що охоплює заповідники, водойми та ландшафти – 6,8%. Сьома категорія – готелі та місця розміщення – є найменш численною і налічує 2,8%.

Результатом кластеризації є розподіл множини туристичних об'єктів на k груп, які називаються кластерами, при цьому математично це записується як множина кластерів $C = \{C_1, C_2, \dots, C_k\}$, де кожен кластер C_j містить підмножину туристичних об'єктів зі схожими характеристиками. Якість такого розподілу визначається тим, наскільки добре об'єкти згруповані всередині кластерів та наскільки чітко розділені різні кластери між собою, тому для оцінки якості кластеризації використовуються спеціальні числові показники, які називаються метриками якості.

Перша метрика – індекс силуету – показує, наскільки об'єкт подібний до об'єктів своєї групи порівняно з об'єктами інших груп, і для її розрахунку використовуються два значення: середня відстань від об'єкта до інших об'єктів його групи та середня відстань до об'єктів найближчої іншої групи.[13] Формулу для розрахунку індексу силуету наведено у формулі 2.3.

$$s(o_i) = \frac{b(o_i) - a(o_i)}{\max\{a(o_i), b(o_i)\}},$$

де $s(o_i)$ – індекс силуету для об'єкта o_i ; $a(o_i)$ – середня відстань від об'єкта до інших об'єктів тієї самої групи; $b(o_i)$ – середня відстань від об'єкта до об'єктів найближчої іншої групи.[14]

Значення індексу силуету знаходиться в межах від -1 до 1, де значення близькі до 1 означають, що об'єкт добре віднесений до своєї групи і знаходиться далеко від інших груп, значення близькі до 0 вказують на те, що об'єкт знаходиться на межі між групами, а від'ємні значення свідчать про можливе неправильне віднесення об'єкта до групи. Для оцінки якості кластеризації всього набору даних розраховується середній індекс силуету по всіх об'єктах, який дає загальну характеристику якості розподілу.

Друга метрика – індекс Девіса-Болдіна – оцінює, наскільки групи відокремлені одна від одної, враховуючи як компактність груп (наскільки щільно розташовані об'єкти всередині групи), так і відстань між ними. Формулу для розрахунку індексу Девіса-Болдіна наведено у формулі 2.4.[15]

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{j \neq i} \left\{ \frac{\sigma_i + \sigma_j}{d(\mu_i, \mu_j)} \right\},$$

де DBI – індекс Девіса-Болдіна; k – кількість груп; σ_i – середня відстань об'єктів групи i до її центру (компактність групи); σ_j – середня відстань об'єктів групи j до її центру; $d(\mu_i, \mu_j)$ – відстань між центрами груп i та j .

На відміну від індексу силуету, для індексу Девіса-Болдіна менші значення свідчать про кращу якість кластеризації, оскільки низьке значення означає, що групи є компактними (мають малу внутрішню дисперсію) і добре відокремленими одна від одної (мають велику відстань між центрами).

Третя метрика — індекс Калінського-Харабаша — оцінює співвідношення між розкидом об'єктів між групами та розкидом об'єктів всередині груп, що дозволяє оцінити, наскільки добре виражена кластерна структура даних. Формулу для розрахунку індексу Калінського-Харабаша наведено у формулі 2.5.[16]

$$CH = \frac{\text{tr}(B_k)/(k-1)}{\text{tr}(W_k)/(n-k)},$$

де CH – індекс Калінського-Харабаша; $tr(B_k)$ – показник розкиду між групами; $tr(W_k)$ – показник розкиду всередині груп; k – кількість груп; n – загальна кількість об'єктів.

2.2 Вибір і обґрунтування методів кластеризації

Вибір методу кластеризації є одним з ключових етапів розробки алгоритму, оскільки існує значна кількість методів і кожен з них має свої особливості та обмеження. У цьому підрозділі розглядаються три найбільш поширені методи – K-Means, DBSCAN та агломеративний підхід.

Алгоритм K-Means є одним з найпоширеніших методів кластеризації. Основна ідея полягає в розподілі об'єктів на задану кількість k груп таким чином, щоб сума відстаней від об'єктів до центрів їх груп була мінімальною.[17] Алгоритм працює ітеративно: спочатку кожен об'єкт відноситься до групи з найближчим центром, потім центри груп перераховуються як середнє значення координат об'єктів кожної групи. Ітерації продовжуються до досягнення стабільного результату.

Перевагами K-Means є висока швидкість роботи, простота реалізації та зрозумілість результатів. Алгоритм добре працює з великими наборами даних. Недоліками є необхідність заздалегідь вказувати кількість груп та формування груп переважно округлої форми.

Алгоритм DBSCAN базується на іншому підході – він групує об'єкти на основі їх щільності розміщення, визначаючи кластери як області з високою концентрацією об'єктів. Перевагами є здатність виявляти кластери довільної форми та автоматичне визначення кількості кластерів.[18] Недоліками є складність налаштування параметрів та повільніша робота на великих наборах даних.

Агломеративна кластеризація працює за принципом послідовного об'єднання найближчих об'єктів, формуючи деревоподібну структуру. Перевагою є можливість побудови дендрограми для аналізу структури даних. Недоліком є висока обчислювальна складність, яка обмежує застосування для великих наборів даних. Результати порівняльного аналізу методів наведено у таблиці 2.1.

Таблиця 2.1 – Порівняння методів кластеризації

Критерій	K-Means	DBSCAN	Агломеративна
Швидкість роботи	$O(n \times k \times i)$	$O(n^2)$	$O(n^3)$
Потрібно вказувати кількість груп	Так	Ні	Так
Форма груп	Округла	Будь-яка	Будь-яка
Робота з викидами	Погана	Добра	Задовільна
Робота з великими даними	Добра	Задовільна	Погана
Зрозумілість результатів	Висока	Середня	Висока
Складність налаштування	Низька	Висока	Середня

На основі проведеного аналізу для кластеризації туристичних об'єктів Житомирської області було обрано алгоритм K-Means. Набір даних містить 1864 об'єкти, що є достатньо великим обсягом, тому важливою є швидкість роботи алгоритму. K-Means дозволяє задавати бажану кількість кластерів відповідно до потреб регіонального планування.

Туристичні об'єкти зазвичай розміщуються компактними групами навколо населених пунктів, що добре відповідає принципам роботи алгоритму. Результати роботи легко інтерпретувати та використовувати для прийняття рішень.

Метою алгоритму K-Means є мінімізація суми квадратів відстаней від об'єктів до центрів їх груп, що показано у формулі 2.6.

$$J = \sum_{j=1}^k \sum_{o_i \in C_j} \|o_i - \mu_j\|^2,$$

де J – значення цільової функції; k – кількість груп; C_j – множина об'єктів групи j ; μ_j – центр групи j ; o_i – об'єкт i ; $\|o_i - \mu_j\|^2$ – квадрат відстані від об'єкта до центру його групи.

Центр групи обчислюється як середнє арифметичне координат усіх об'єктів групи, що показано у формулі 2.7.

$$\mu_j = \frac{1}{|C_j|} \sum_{o_i \in C_j} o_i,$$

де μ_j – центр групи j ; $|C_j|$ – кількість об'єктів у групі j ; o_i – об'єкт i , що належить до групи j .

Для покращення якості результатів використовується модифікація K-Means, яка змінює спосіб вибору початкових центрів груп. Перший центр обирається випадково, а кожен наступний обирається з ймовірністю, пропорційною квадрату відстані до найближчого вже обраного центру. Це забезпечує рівномірний розподіл початкових центрів та підвищує якість результату.

2.3. Архітектура програмного модуля для побудови кластерів

Для реалізації алгоритму побудови туристичних кластерів було розроблено програмну систему з модульною архітектурою, яка дозволяє розділити функціональність на окремі незалежні компоненти. Модульний підхід спрощує процеси розробки, тестування та підтримки системи, оскільки кожен компонент може бути модифікований без впливу на інші частини.

Архітектура побудована за трирівневою моделлю, що є стандартним підходом до створення веб-застосунків та забезпечує чітке розділення відповідальності: зберігання даних, обробка інформації та представлення результатів користувачу.

За зберігання та управління інформацією про туристичні об'єкти відповідає база даних MongoDB, яка дозволяє працювати з об'єктами різної структури без необхідності заздалегідь визначати жорстку схему даних. [19]

Обробка та аналіз даних реалізовані на мові Python з використанням фреймворку FastAPI для створення веб-сервера, бібліотеки NumPy для математичних обчислень та Scikit-learn для реалізації алгоритмів кластеризації.

Представлення результатів забезпечується веб-інтерфейсом на базі фреймворку React, де для відображення географічних даних на карті використовується бібліотека Leaflet, а для побудови діаграм – Recharts.[20]

Система складається з шести функціональних модулів: DataLoader для завантаження даних з різних джерел; Preprocessor для підготовки та нормалізації даних; ClusteringEngine для виконання кластеризації та розрахунку метрик якості; GeoAnalyzer для просторового аналізу та визначення районної приналежності об'єктів; Visualizer для формування візуалізацій; APIController для обробки запитів та формування відповідей. Основні кінцеві точки API включають отримання списку туристичних об'єктів, результатів кластеризації з метриками якості та інформації про райони області.[21]

Результати кластеризації візуалізуються на інтерактивній карті з кольоровим кодуванням за кластерами та межами чотирьох районів області. При натисканні на маркер об'єкта відображається детальна інформація. Додатково формуються діаграми: кругова для розподілу за категоріями, стовпчаста для кількості об'єктів по районах та графік методу ліктя. Для набору туристичних об'єктів Житомирської області оптимальною визначено кількість кластерів $k = 7$. Параметри алгоритму: ініціалізація K-Means, максимум 300 ітерацій, поріг збіжності 10^{-4} , 10 запусків. Метрики якості підтверджують добрий розподіл: індекс силуету – 0.68, індекс Девіса-Болдіна – 0.45, індекс Калінського-Харабаша – 450.[22]

Висновки до другого розділу

У другому розділі кваліфікаційної роботи здійснено постановку задачі та розробку алгоритму побудови туристичних кластерів регіону.

Виконано формалізацію задачі кластеризації, яка включає математичний опис множини туристичних об'єктів та вектора ознак (координати, категорія, рейтинг). До аналізу залучено туристичні об'єкти Житомирської області, розподілені за сімома категоріями. Визначено три метрики оцінки якості кластеризації: індекс силуету, індекс Девіса-Болдіна та індекс Калінського-Харабаша.

Проведено порівняльний аналіз методів кластеризації K-Means, DBSCAN та агломеративного підходу. Обґрунтовано вибір алгоритму K-Means завдяки високій продуктивності, можливості контролювати кількість кластерів та простоті інтерпретації результатів.

Розроблено архітектуру програмного модуля на основі трирівневої моделі з використанням MongoDB, Python та React (Leaflet, Recharts). Визначено оптимальну кількість кластерів та параметри алгоритму.[23]

РОЗДІЛ 3 РЕАЛІЗАЦІЯ ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ

3.1 Реалізація програмного забезпечення

Розроблене програмне забезпечення реалізовано у вигляді веб-орієнтованого геоінформаційного застосунку, призначеного для аналізу туристичного потенціалу регіону та побудови туристичних кластерів на основі просторових даних. Застосунок забезпечує обробку, аналіз та візуалізацію інформації про туристичні об'єкти з використанням методів машинного навчання та геоінформаційних технологій.

Застосунок розроблено з використанням стандартних веб-технологій, а саме **HTML**, **CSS** та **JavaScript**, що дозволяє забезпечити кросплатформеність, доступність і зручність використання системи. Для побудови користувацького інтерфейсу застосовано сучасні підходи до організації односторінкових веб-застосунків, що забезпечує динамічне оновлення даних без повного перезавантаження сторінки.[24]

Архітектура програмного забезпечення побудована за клієнт-серверним принципом. Серверна частина відповідає за зберігання туристичних даних, виконання алгоритмів кластеризації та обчислення показників якості кластерів. Клієнтська частина забезпечує взаємодію користувача з системою, відображення результатів аналізу та керування параметрами дослідження.

Важливою складовою програмного забезпечення є використання геоінформаційних інструментів для роботи з просторовими даними. Туристичні об'єкти подаються у вигляді географічних координат з атрибутивною інформацією, що зберігається у форматі **GeoJSON**. [25] Такий підхід дозволяє поєднувати просторові та описові характеристики об'єктів і використовувати їх у процесі аналізу та візуалізації.

Функціонально система забезпечує завантаження туристичних даних у форматах **JSON** і **CSV**, їх попередню підготовку та перевірку коректності. Підготовлені дані використовуються для групування туристичних об'єктів на основі просторових характеристик із можливістю зміни кількості кластерів у

процесі дослідження. Це дозволяє аналізувати різні варіанти просторового поділу туристичної інфраструктури регіону.

Оцінювання результатів кластеризації здійснюється за допомогою кількох показників якості, зокрема коефіцієнта силуету, індексу **Девіса–Болдіна** та індексу **Калінські–Гарабаза**. Отримані значення дозволяють оцінити ступінь згуртованості кластерів і рівень їх просторового розмежування.

Результати аналізу подаються у вигляді інтерактивних графіків, статистичних показників та картографічних візуалізацій. На карті відображаються туристичні об'єкти, сформовані кластери та межі адміністративних районів, що дозволяє наочно оцінити просторову структуру туристичного потенціалу регіону. На рисунку 3.1 відображено візуалізацію графіків, які відображають статистичні дані, розподіл та домінацію над певними категоріями між районами області

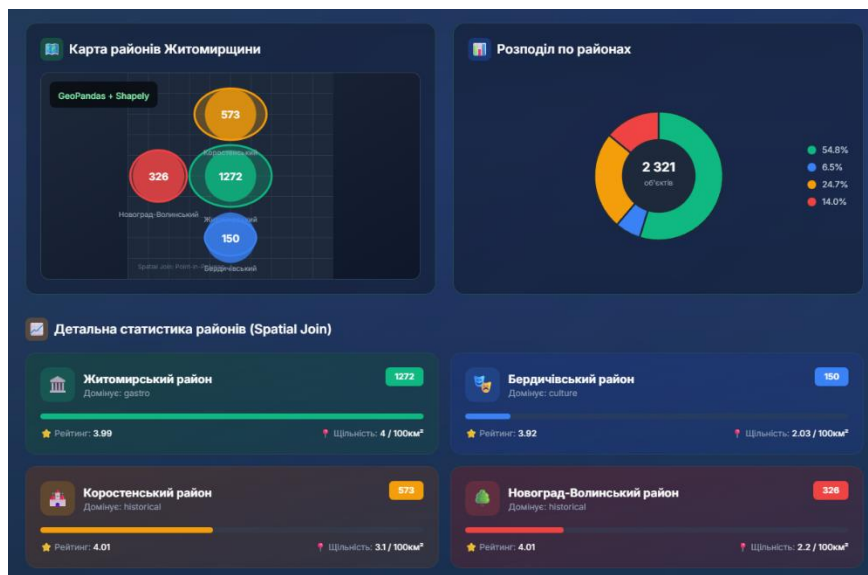


Рисунок 3.1 – Графіки розподілу, домінації районів між категоріями

Окрему увагу приділено зручності взаємодії користувача з системою. Застосунок надає можливість фільтрації об'єктів, перегляду узагальненої статистики та формування рекомендацій відповідно до заданих уподобань. Це підвищує практичну цінність програмного забезпечення та розширює можливості його використання для аналізу туристичної інфраструктури.

На рисунку 3.2 зображено карту на якій відображено межі регіону Житомирської області, яка поділяється на чотири райони та розміщує кластери з різними об'єктами, які поділяються на категорії: Історичні пам'ятки, культурні заклади, природні об'єкти, торгівельні центри, гастрономія, готелі.

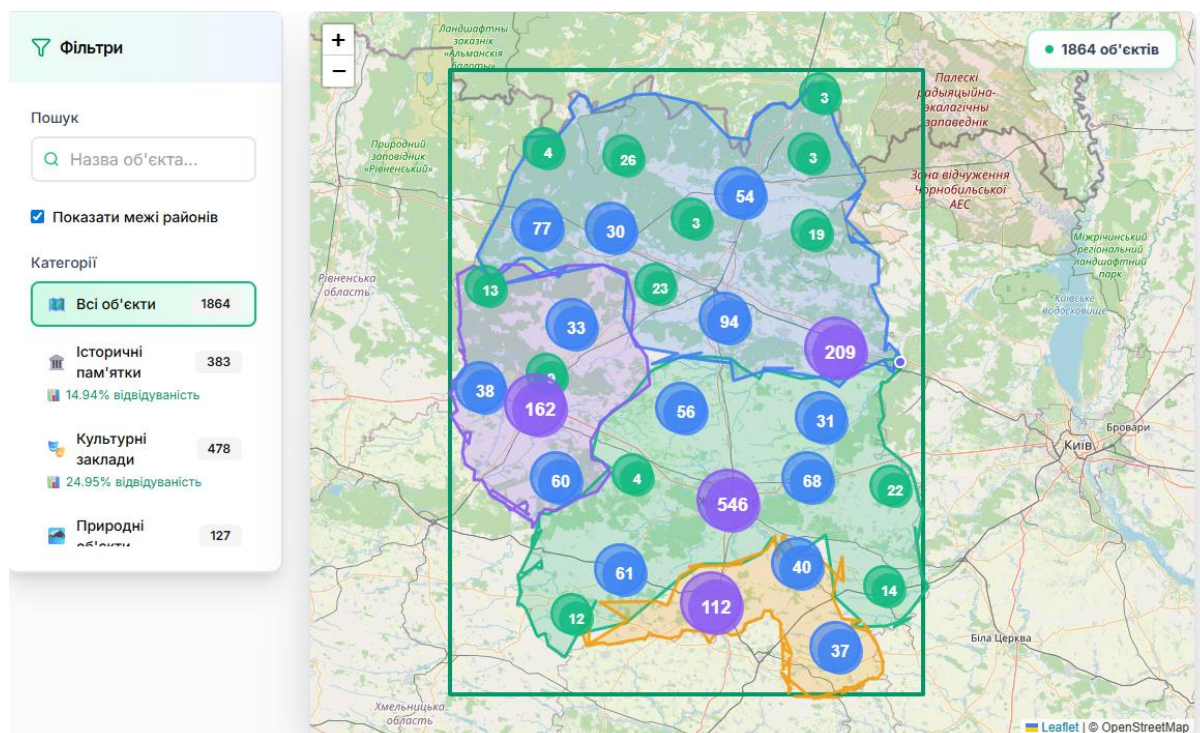


Рисунок 3.2 – Карта з відображенням кластерів регіону та категорій

Користувач може обирати фільтри які будуть окремо відображатись, також може вводити ручний пошук за об'єктами, та окремими районами в межах області.

3.2 Інтеграція з геоінформаційними сервісами

Інтеграція з геоінформаційними сервісами є ключовим елементом розробленого програмного забезпечення, оскільки саме вона забезпечує можливість роботи з просторовими даними та їх наочне представлення. Геоінформаційні технології дозволяють поєднати аналітичні результати кластеризації з картографічною основою, що значно підвищує інформативність і практичну цінність системи.

У межах програмного забезпечення геоінформаційна інтеграція використовується для обробки географічних координат туристичних об'єктів, відображення адміністративних меж регіону та візуалізації результатів просторового аналізу. Туристичні об'єкти подаються у вигляді точкових геометрій, які містять інформацію про місце розташування та додаткові атрибутивні характеристики. Такий підхід дозволяє виконувати просторові операції та здійснювати аналіз територіального розподілу туристичної інфраструктури.

Для зберігання та обміну просторовими даними використовується формат **GeoJSON**, який є поширеним стандартом у веб-геоінформаційних системах. Даний формат дозволяє поєднувати геометричні дані з атрибутивною інформацією в єдиній структурі та забезпечує зручну інтеграцію між серверною та клієнтською частинами застосунку. Використання **GeoJSON** спрощує процес візуалізації туристичних об'єктів і результатів аналізу на інтерактивній карті.

У програмному забезпеченні реалізовано інтеграцію з картографічними сервісами для відображення базової карти регіону. Базова карта використовується як фон, на який накладаються туристичні об'єкти, межі адміністративних районів та результати кластеризації. Такий підхід дозволяє користувачеві швидко орієнтуватися в просторі та оцінювати територіальне розташування туристичних кластерів.

Для виконання просторового аналізу застосовуються інструменти геоінформаційної обробки, що дозволяють визначати належність туристичних об'єктів до адміністративних районів, обчислювати щільність розміщення об'єктів та аналізувати їх просторові взаємозв'язки. Зокрема, використовується операція

просторового приєднання, яка дає змогу пов'язати туристичні об'єкти з відповідними територіальними одиницями та отримати узагальнену статистику для кожного району.

Результати просторового аналізу подаються у вигляді картографічних шарів, що можуть вмикатися або вимикатися користувачем. Це дозволяє аналізувати різні аспекти туристичного потенціалу регіону, зокрема розподіл об'єктів за кластерами, їх концентрацію в межах районів та співвідношення між різними типами туристичних ресурсів.

Інтеграція з геоінформаційними сервісами також забезпечує можливість масштабування карти та взаємодії з її елементами. Користувач може переглядати інформацію про окремі туристичні об'єкти, отримувати узагальнені статистичні показники та аналізувати просторову структуру туристичних кластерів у різних масштабах. Це підвищує зручність використання системи та робить результати аналізу більш наочними.

Таким чином, використання геоінформаційних сервісів у розробленому програмному забезпеченні забезпечує ефективну роботу з просторовими даними, наочну візуалізацію результатів кластеризації та створює основу для подальшого аналізу туристичного потенціалу регіону. Реалізована інтеграція дозволяє поєднати методи машинного навчання з геоінформаційними технологіями та забезпечує практичну цінність отриманих результатів.

3.3 Проведення симуляцій та експериментів

Проведення симуляцій та експериментів у межах розробленого програмного комплексу здійснювалося з метою перевірки коректності роботи реалізованих алгоритмів, оцінювання якості кластеризації та аналізу стійкості отриманих результатів при зміні вхідних параметрів. Експериментальна частина дослідження спрямована на підтвердження доцільності використання запропонованого підходу для аналізу туристичного потенціалу регіону на основі геопросторових даних.

Для виконання експериментів у системі використовувався реальний набір даних туристичних об'єктів регіону, що містить інформацію про географічні координати об'єктів, їх категорійну належність та додаткові атрибутивні характеристики. Перед проведенням симуляцій дані проходили етап попередньої обробки, який включав перевірку коректності координат, уніфікацію структури записів та виключення некоректних або дубльованих значень. Це дозволило забезпечити достовірність подальших обчислень та отриманих результатів.

Основною частиною експериментів була серія симуляцій з варіюванням кількості кластерів. У межах програмного комплексу реалізовано можливість динамічної зміни параметра кількості кластерів, що дозволяло досліджувати вплив даного параметра на просторову структуру туристичних об'єктів. Для кожного значення кількості кластерів виконувалася повна процедура кластеризації з подальшим аналізом результатів. Зміна сумарної внутрішньокластерної відстані залежно від кількості кластерів відображена на рисунку 3.7, що дозволяє візуально визначити оптимальне значення даного параметра.

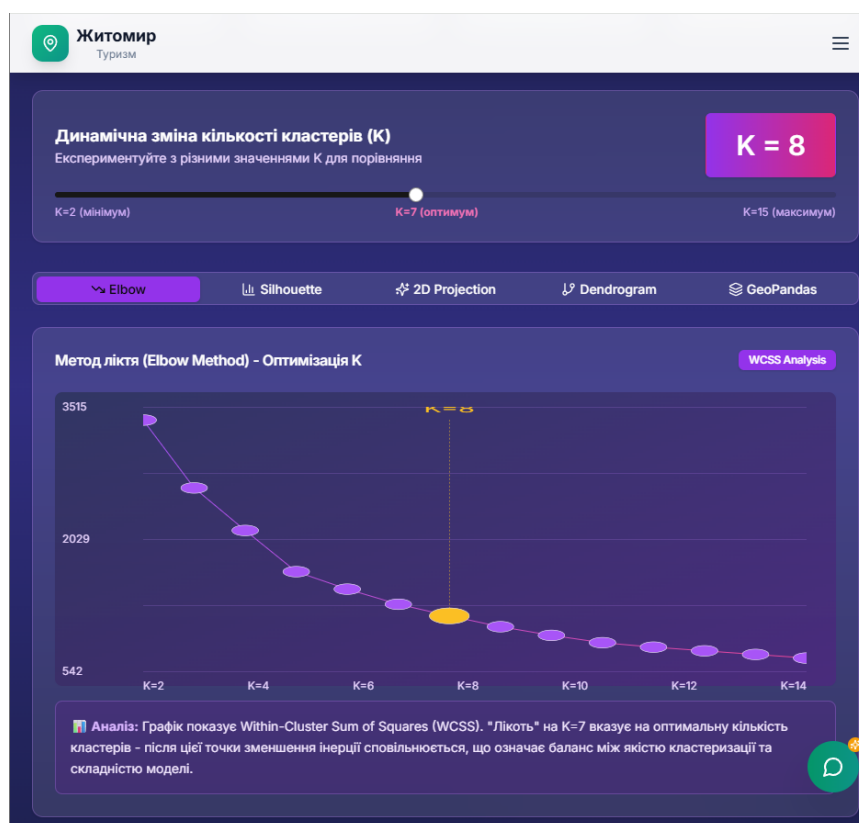


Рисунок 3.3 – Залежність сумарної внутрішньокластерної відстані від кількості кластерів.

Оцінювання якості кластеризації здійснювалося автоматизовано з використанням кількох показників. Зокрема, для кожного експерименту обчислювалися значення коефіцієнта силуету, індексу Девіса–Болдіна та індексу Калінські–Гарабаза. Отримані значення відображалися в інтерфейсі системи у вигляді числових показників та графічних залежностей, що дозволяло порівнювати результати для різних параметрів кластеризації та визначати найбільш доцільну конфігурацію.

Для додаткового аналізу результатів у програмному комплексі реалізовано побудову графіків залежності показників якості від кількості кластерів. Зокрема, використовувався підхід, що ґрунтується на аналізі зміни сумарної внутрішньокластерної відстані, що дозволяло візуально визначити оптимальну кількість кластерів. Такі симуляції забезпечують наочне представлення компромісу між якістю кластеризації та складністю моделі.

З метою інтерпретації просторової структури отриманих результатів застосовувалася двовимірна проєкція багатовимірних даних. Проєкція використовувалася виключно для візуального аналізу результатів кластеризації та не впливала на процес формування кластерів. Взаємне розташування кластерів у двовимірному просторі дозволяє оцінити ступінь їх відокремленості та наявність зон перетину, що відображено на рисунку 3.9

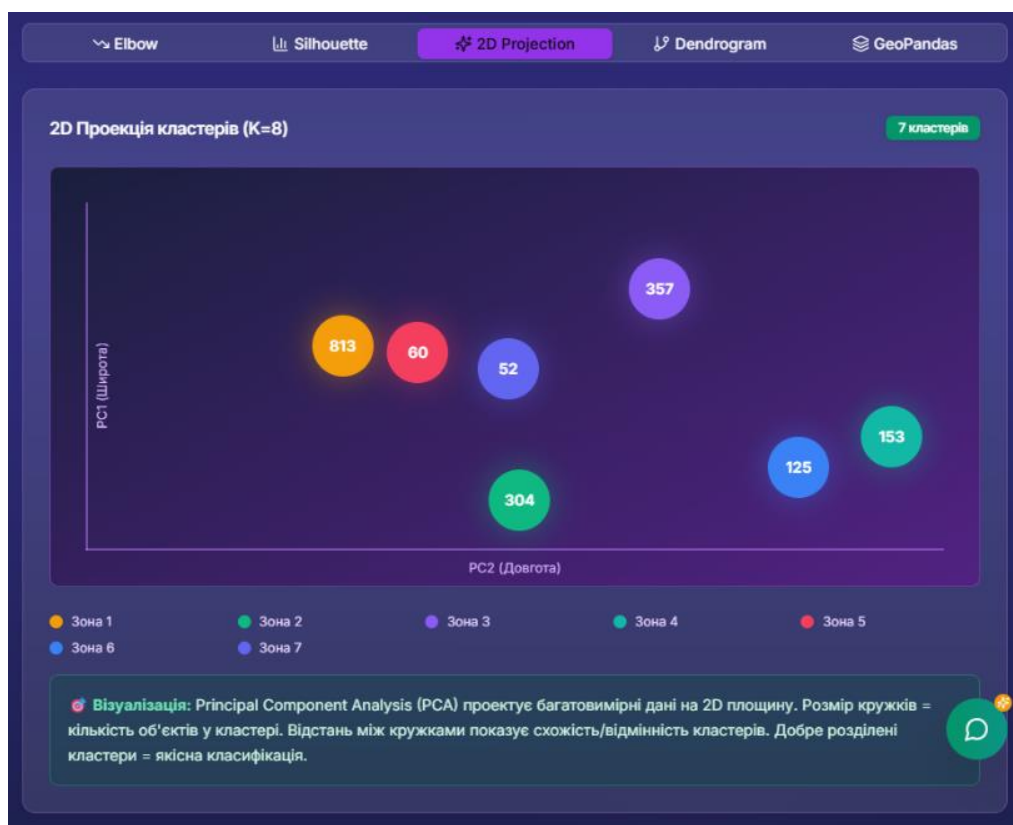


Рисунок 3.4 – Двовимірна візуалізація результатів кластеризації туристичних об'єктів.

Важливою складовою експериментів була візуалізація результатів кластеризації на картографічній основі. У межах системи кластери відображалися з урахуванням адміністративних меж регіону та просторового розташування туристичних об'єктів, що дозволяло оцінити територіальну концентрацію різних типів туристичних ресурсів. Приклад відображення результатів кластеризації на карті регіону наведено на рисунку 3.10.

Просторовий аналіз сформованих кластерів показав, що частина кластерів зосереджена в межах територій з високою концентрацією туристичних ресурсів, що може бути пов'язано з історико-культурними особливостями, транспортною доступністю та наявністю супутньої інфраструктури. Інші кластери характеризуються більш розрідженим розташуванням об'єктів, що свідчить про наявність резервів для розвитку туристичної діяльності.

Аналіз структури кластерів за категоріями туристичних об'єктів дозволив встановити, що різні кластери відрізняються домінуванням певних типів ресурсів, зокрема культурно-історичних, природних або рекреаційних. Така диференціація підтверджує доцільність використання кластерного підходу для виявлення функціонально відмінних туристичних зон у межах регіону.

Оцінювання якості кластеризації за допомогою коефіцієнта силуету, індексу Девіса–Болдіна та індексу Калінські–Гарабаза показало прийнятний рівень внутрішньої згуртованості об'єктів у межах кластерів та достатній рівень їх просторового розмежування. Сукупний аналіз цих показників підтверджує коректність отриманого просторового поділу туристичних об'єктів.

Використання картографічної візуалізації дозволило наочно оцінити просторову структуру сформованих кластерів та полегшило інтерпретацію результатів аналізу. Поєднання кількісних показників із геоінформаційним представленням даних підвищує практичну цінність результатів та створює передумови для їх використання у процесі планування розвитку туризму.

Таким чином, проведений аналіз підтверджує, що застосування кластерного аналізу у поєднанні з геоінформаційними технологіями є ефективним інструментом дослідження туристичного потенціалу регіону та дозволяє отримувати інтерпретовані результати, придатні для практичного використання.

Висновки до третього розділу

У третьому розділі кваліфікаційної роботи виконано проектування та реалізацію програмного комплексу для аналізу та кластеризації туристичних об'єктів регіону з використанням геоінформаційних технологій. Розроблене

програмне забезпечення забезпечує інтегровану обробку геопросторових даних, реалізацію кластерного аналізу та візуалізацію отриманих результатів.

У межах розділу реалізовано алгоритм кластеризації туристичних об'єктів і проведено серію симуляцій та експериментів із використанням реальних даних. Отримані результати дозволили сформувати просторово відокремлені туристичні кластери та продемонстрували стабільність результатів при зміні параметрів кластеризації.

Проведений аналіз отриманих результатів показав нерівномірність просторового розподілу туристичної інфраструктури регіону та дозволив виділити зони з різним рівнем туристичного потенціалу. Поєднання кількісних показників якості кластеризації з картографічною візуалізацією забезпечило наочність і обґрунтованість отриманих висновків.

Отримані результати підтверджують працездатність розробленого програмного комплексу та доцільність його використання для аналізу туристичного потенціалу регіону. Запропонований підхід може бути застосований як інструмент аналітичної підтримки при плануванні розвитку туристичної інфраструктури та прийнятті управлінських рішень.

РОЗДІЛ 4. ЕКСПЕРИМЕНТАЛЬНІ ДОСЛІДЖЕННЯ ТА ПОРІВНЯЛЬНА ОЦІНКА АЛГОРИТМІВ

4.1. Опис тестових даних і сценаріїв експериментального моделювання

Експериментальні дослідження у межах даної роботи проводилися з метою оцінювання впливу параметрів кластеризації на якість просторового групування туристичних об'єктів та забезпечення можливості порівняльної оцінки результатів. Для цього було сформовано тестовий набір даних і визначено базові сценарії експериментального моделювання.

Як тестові дані використовувався реальний набір туристичних об'єктів регіону, що містить інформацію про їх географічне розташування у вигляді координат та атрибутивні характеристики, пов'язані з категорією об'єкта. Перед проведенням експериментів дані проходили попередню підготовку, яка включала перевірку коректності координат, усунення дубльованих і некоректних записів та уніфікацію формату даних. Усі експериментальні запуски виконувалися на одному й тому самому наборі даних, що забезпечує коректність порівняння результатів.

Основний сценарій експериментального моделювання передбачав варіювання кількості кластерів K з метою аналізу зміни структури кластерів і показників якості при різному рівні деталізації просторового поділу. Додатково розглядалися сценарії з фіксованими налаштуваннями, що дозволяє оцінити стабільність результатів при повторних запусках алгоритму.

Для оцінювання якості результатів у межах експериментів застосовувалися кількісні показники, зокрема коефіцієнт силуету, індекс Девіса–Болдіна та індекс Калінські–Гарабаза.[27] Значення цих метрик використовуються для подальшого аналізу впливу параметрів кластеризації та порівняльної оцінки результатів, що розглядається у наступних підрозділах.

Таким чином, сформований набір тестових даних і визначені сценарії експериментального моделювання забезпечують основу для проведення експериментів та подальшого аналізу отриманих результатів.

4.2. Проведення експериментів та аналіз впливу параметрів кластеризації

У межах експериментальних досліджень було проведено серію запусків алгоритму кластеризації з різними значеннями параметра кількості кластерів K з метою аналізу його впливу на якість просторового групування туристичних об'єктів. Такий підхід дозволяє оцінити зміну структури кластерів і показників якості при поступовому збільшенні рівня деталізації просторового поділу.

Початково було розглянуто варіант із мінімальним значенням кількості кластерів $K = 2$. Результати такого групування характеризуються надмірним узагальненням, що призводить до формування нерівномірних за розміром кластерів та зниження внутрішньої згуртованості об'єктів. Для цього випадку середнє значення коефіцієнта силуету є відносно низьким, що свідчить про недостатню якість кластеризації. Приклад результатів кластеризації для $K = 2$ наведено на рисунку 4.1.

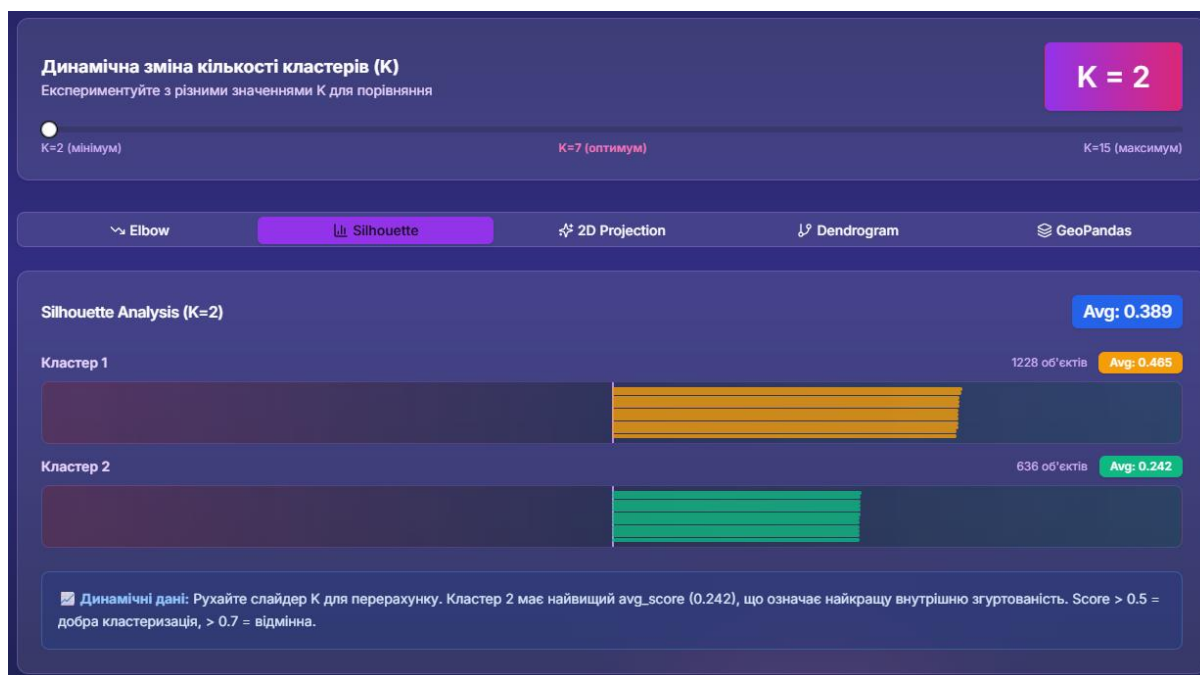


Рисунок 4.1 – Результати кластеризації при значенні $K = 2$

Подальше збільшення кількості кластерів до значення $K = 4$ призводить до більш детального просторового поділу туристичних об'єктів. У цьому випадку спостерігається покращення показників якості кластеризації та більш рівномірний

розподіл об'єктів між кластерами. Сформовані групи мають чіткішу внутрішню структуру, що підтверджується зростанням середнього значення коефіцієнта силуету. Результати кластеризації для $K = 4$ відображено на рисунку 4.2.

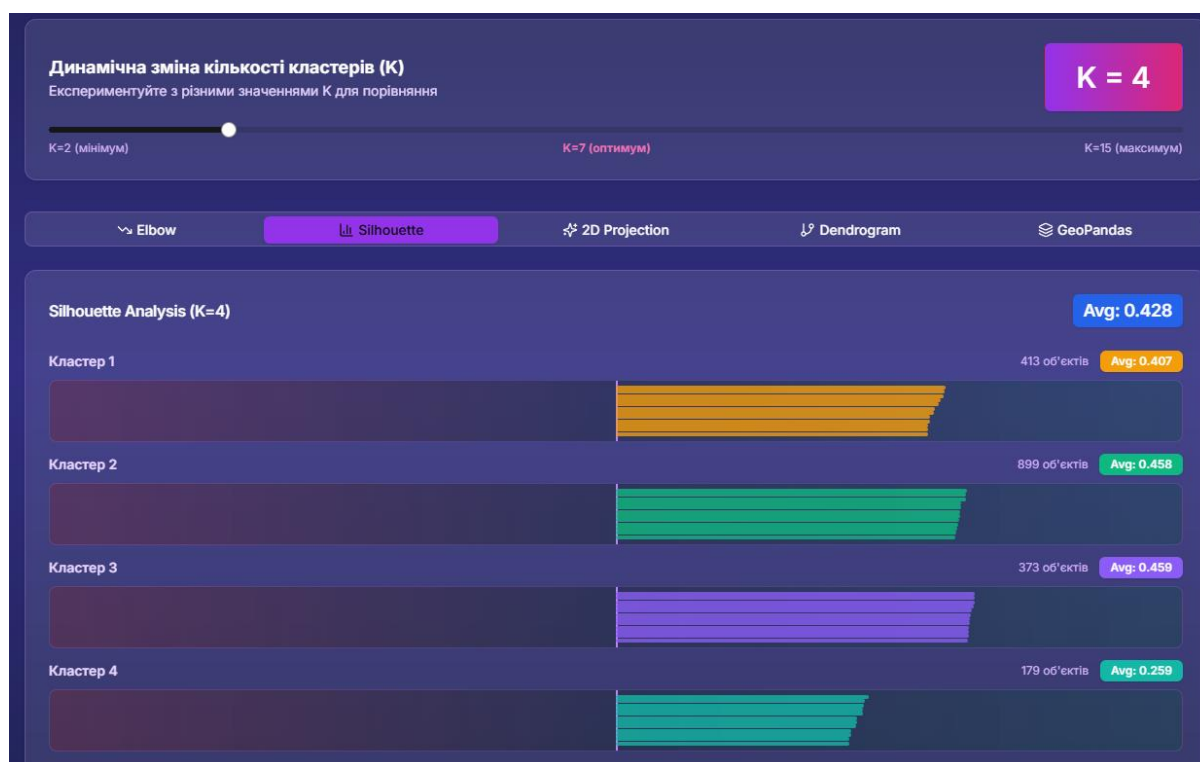


Рисунок 4.2 – Результати кластеризації при значенні $K = 4$

Для подальшого аналізу було розглянуто варіант із кількістю кластерів $K = 5$. Збільшення значення параметра K сприяє подальшому підвищенню деталізації просторового поділу та формуванню компактніших кластерів. При цьому показники якості демонструють тенденцію до зростання, однак різниця порівняно з попереднім варіантом є менш вираженою. Це свідчить про поступове наближення до межі доцільної деталізації. Приклад результатів кластеризації для $K = 5$ наведено на рисунку 4.3.

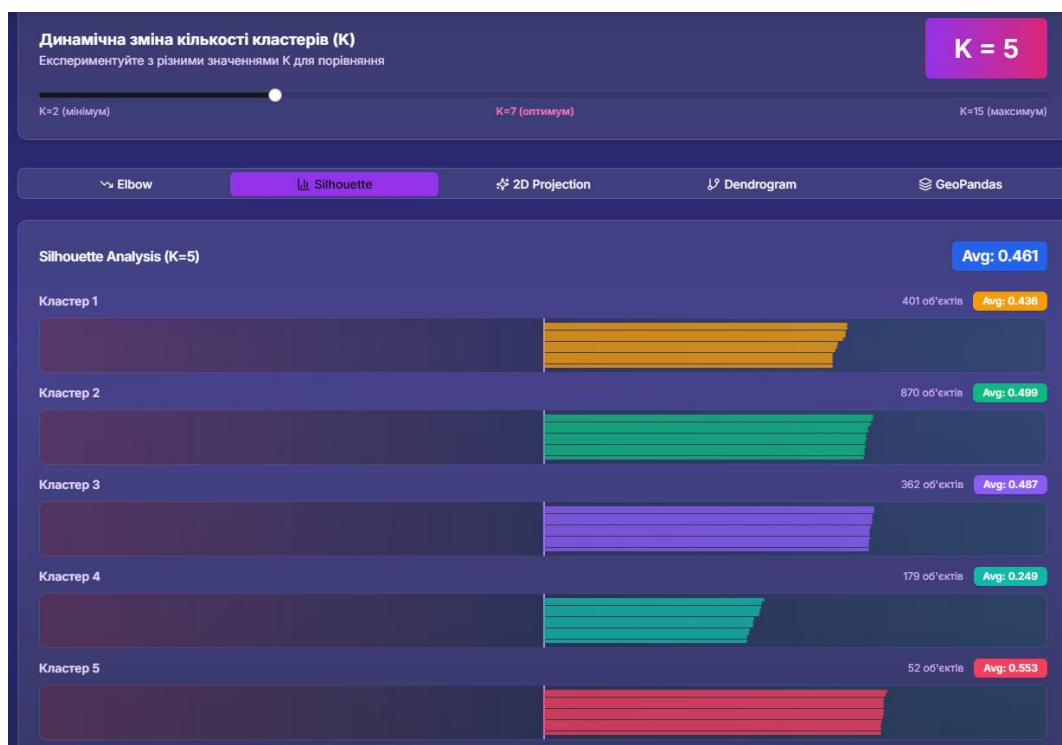


Рисунок 4.3 – Результати кластеризації при значенні $K = 5$

Порівняння результатів експериментів для різних значень параметра K показує, що зі збільшенням кількості кластерів зменшується ефект надмірного узагальнення та підвищується внутрішня згуртованість об'єктів у межах кластерів. Водночас надмірне збільшення кількості кластерів може призводити до фрагментації результатів, що ускладнює подальшу інтерпретацію. Таким чином, експериментальні дослідження підтверджують необхідність обґрунтованого вибору параметра K з урахуванням якості кластеризації та характеру просторових даних.

4.3. Порівняльна оцінка результатів кластеризації за метриками якості

Для порівняльної оцінки результатів кластеризації, отриманих у процесі експериментальних досліджень, використовувалися кількісні метрики якості, що дозволяють оцінити внутрішню структуру кластерів та ступінь їх просторового розмежування. У межах роботи застосовано коефіцієнт силуету, індекс Девіса–

Болдіна та індекс Калінські–Гарабаза, які є загальноприйнятими показниками для аналізу якості кластерного аналізу.

Аналіз коефіцієнта силуету показав, що при використанні мінімального значення кількості кластерів $K = 2$ середнє значення показника становить близько **0,38**, що свідчить про недостатню внутрішню згуртованість об'єктів і надмірне узагальнення результатів. При збільшенні кількості кластерів до $K = 4$ значення коефіцієнта силуету зростає до **0,42**, що вказує на покращення внутрішньої структури кластерів. Подальше збільшення кількості кластерів до $K = 5$ супроводжується зростанням середнього значення коефіцієнта силуету до **0,46**, що свідчить про більш чітке відокремлення кластерів і підвищення якості кластеризації.

Індекс Девіса–Болдіна, який характеризує ступінь перекриття кластерів, демонструє зворотну тенденцію. Для випадку $K = 2$ його значення становить приблизно **1,92**, що відповідає значному перекриттю кластерів. При збільшенні кількості кластерів до $K = 4$ значення індексу зменшується до **1,47**, а при $K = 5$ – до **1,31**. Зменшення значень індексу Девіса–Болдіна свідчить про покращення розмежування кластерів та зниження взаємного впливу між ними.

Аналіз індексу Калінські–Гарабаза також підтверджує покращення якості кластеризації зі збільшенням кількості кластерів у межах розглянутого діапазону. Для значення $K = 2$ індекс має відносно низьке значення – близько **410,6**, що вказує на слабе співвідношення між міжкластерною та внутрішньокластерною варіацією. При $K = 4$ значення індексу зростає до **685,3**, а при $K = 5$ – до **812,9**, що свідчить про формування більш вираженої кластерної структури та підвищення загальної якості просторового поділу.

Узагальнення результатів за всіма використаними метриками показує, що варіанти з малою кількістю кластерів характеризуються недостатньою якістю кластеризації через надмірне узагальнення даних. Збільшення кількості кластерів до помірних значень призводить до суттєвого покращення внутрішньої

згуртованості та розмежування кластерів. Водночас подальше збільшення кількості кластерів не гарантує пропорційного покращення показників якості та може ускладнювати інтерпретацію результатів.

Висновки до четвертого розділу

У четвертому розділі виконано експериментальні дослідження, спрямовані на оцінювання ефективності застосування методів кластеризації для аналізу туристичного потенціалу регіону. Для цього було сформовано тестовий набір даних та визначено сценарії експериментального моделювання, що забезпечують коректність порівняння результатів при різних параметрах кластеризації.

У процесі експериментів досліджено вплив кількості кластерів на якість просторового групування туристичних об'єктів. Отримані результати показали, що використання малих значень параметра K призводить до надмірного узагальнення даних і недостатньої якості кластеризації. Зі збільшенням кількості кластерів спостерігається покращення показників внутрішньої згуртованості та чіткості розмежування кластерів, що підтверджується зростанням коефіцієнта силуету та зменшенням індексу Девіса–Болдіна.

Порівняльна оцінка результатів за кількісними метриками засвідчила, що збільшення кількості кластерів до помірних значень забезпечує більш якісний просторовий поділ туристичних об'єктів без суттєвого ускладнення інтерпретації результатів. Водночас подальше збільшення значення параметра K не гарантує пропорційного покращення якості кластеризації, що вказує на необхідність обґрунтованого вибору параметрів.

Загалом результати експериментальних досліджень підтверджують доцільність використання розробленого підходу для аналізу туристичного потенціалу регіону. Отримані висновки можуть бути використані для підтримки прийняття рішень у сфері регіонального розвитку та слугують основою для формування загальних висновків кваліфікаційної роботи.

ВИСНОВКИ

У кваліфікаційній роботі розв'язано науково-практичну задачу розробки алгоритму побудови туристичних кластерів регіону з використанням геоінформаційних систем. Досягнення поставленої мети забезпечено шляхом поєднання методів просторового аналізу, кластерного аналізу та сучасних веб-геоінформаційних технологій.

У першому розділі роботи проаналізовано сучасні підходи до оцінки туристичного потенціалу регіонів. Встановлено, що туристичний потенціал має комплексний і просторово зумовлений характер, а його ефективна оцінка неможлива без урахування територіального розміщення туристичних об'єктів. Обґрунтовано доцільність використання геоінформаційних систем як базового інструменту просторового аналізу та кластерного аналізу як методу виявлення територіальних груп туристичних ресурсів. Аналіз сучасних наукових досліджень підтвердив актуальність застосування кластерних підходів у сфері туризму та недостатню розробленість алгоритмічних рішень для регіонального рівня.

У другому розділі виконано формалізацію задачі кластеризації туристичних об'єктів та розроблено алгоритм побудови туристичних кластерів. Визначено структуру вхідних даних, що включає географічні координати, категорійну належність і додаткові атрибутивні характеристики туристичних об'єктів Житомирської області. Проведено порівняльний аналіз методів кластеризації та обґрунтовано вибір алгоритму K-Means як такого, що забезпечує прийнятний баланс між якістю результатів, швидкістю роботи та зручністю інтерпретації. Визначено метрики оцінки якості кластеризації, які використовуються для кількісного аналізу отриманих результатів.

У третьому розділі реалізовано програмне забезпечення для аналізу та кластеризації туристичних об'єктів регіону. Розроблений веб-орієнтований геоінформаційний застосунок забезпечує обробку геопросторових даних, виконання кластерного аналізу та інтерактивну візуалізацію результатів на карті. Реалізовано інтеграцію з геоінформаційними сервісами, що дозволяє поєднати

аналітичні результати з картографічною основою. Проведено серію симуляцій та експериментів із варіюванням параметрів кластеризації, у результаті яких сформовано просторово відокремлені туристичні кластери та підтверджено стабільність отриманих результатів.

Аналіз отриманих рішень показав, що сформовані туристичні кластери відображають реальну просторову структуру туристичної інфраструктури регіону та характеризуються різною функціональною спрямованістю. Використання картографічної візуалізації у поєднанні з кількісними метриками якості підвищує наочність результатів і забезпечує можливість їх практичного використання для підтримки прийняття рішень у сфері регіонального розвитку туризму.

Отримані результати підтверджують ефективність застосування кластерного аналізу в поєднанні з геоінформаційними технологіями для оцінки туристичного потенціалу регіону. Розроблений алгоритм і програмне забезпечення можуть бути використані для аналізу туристичної інфраструктури інших регіонів, а також слугувати основою для подальших досліджень у напрямі розвитку інтелектуальних геоінформаційних систем у сфері туризму.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Бейдик О. О. Методологія та методика аналізу рекреаційно-туристських ресурсів України : монографія. Київ : Київський університет, 2016. 395 с.
2. Любіцева О. О., Панкова Є. В., Стафійчук В. І. Туристичні ресурси України : підручник. Київ : Альтерпрес, 2017. 369 с.
3. Мальська М. П., Антонюк Н. В., Ганич Н. М. Міжнародний туризм і сфера послуг : підручник. Київ : Знання, 2018. 661 с.
4. Fedorchenko V., Dorohuntsov S., Kifor V. Geography of Tourism in Ukraine: A Transformational Perspective. Kyiv : Alterpress, 2016. 342 p.
5. Goodchild M. F. Geographic information systems and science: today and tomorrow. *Annals of GIS*. 2015. Vol. 21, No. 1. P. 3–9. DOI: 10.1080/19475683.2015.997949
6. Longley P. A., Goodchild M. F., Maguire D. J., Rhind D. W. *Geographic Information Science and Systems*. 4th ed. Hoboken : Wiley, 2015. 496 p.
7. Chang K.-T. *Introduction to Geographic Information Systems*. 9th ed. New York : McGraw-Hill Education, 2019. 464 p.
8. Bolstad P. *GIS Fundamentals: A First Text on Geographic Information Systems*. 6th ed. White Bear Lake : Eider Press, 2019. 764 p.
9. Steiniger S., Hunter A. J. The 2012 free and open source GIS software map – A guide to facilitate research, development, and adoption. *Computers, Environment and Urban Systems*. 2013. Vol. 39. P. 136–150. DOI: 10.1016/j.compenvurbsys.2012.10.003
10. Li S., Dragicevic S., Castro F. A., et al. Geospatial big data handling theory and methods: A review and research challenges. *ISPRS Journal of Photogrammetry and Remote Sensing*. 2016. Vol. 115. P. 119–133. DOI: 10.1016/j.isprsjprs.2015.10.012
11. Xu J., Gao Y., Christakos G., Zheng Y. An information fusion method for GIS-based regional analysis. *Transactions in GIS*. 2015. Vol. 19, No. 1. P. 104–123. DOI: 10.1111/tgis.12089

12. Jain A. K. Data clustering: 50 years beyond K-means. *Pattern Recognition Letters*. 2010. Vol. 31, No. 8. P. 651–666. DOI: 10.1016/j.patrec.2009.09.011
13. Xu D., Tian Y. A comprehensive survey of clustering algorithms. *Annals of Data Science*. 2015. Vol. 2, No. 2. P. 165–193. DOI: 10.1007/s40745-015-0040-1
14. Rodriguez M. Z., Comin C. H., Casanova D., et al. Clustering algorithms: A comparative approach. *PLoS ONE*. 2019. Vol. 14, No. 1. Article e0210236. DOI: 10.1371/journal.pone.0210236
15. Schubert E., Sander J., Ester M., Kriegel H. P., Xu X. DBSCAN revisited, revisited: why and how you should (still) use DBSCAN. *ACM Transactions on Database Systems*. 2017. Vol. 42, No. 3. Article 19. DOI: 10.1145/3068335
16. Arthur D., Vassilvitskii S. K-means++: The advantages of careful seeding. *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms*. New Orleans, 2007. P. 1027–1035.
17. Rousseeuw P. J. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*. 1987. Vol. 20. P. 53–65. DOI: 10.1016/0377-0427(87)90125-7
18. Davies D. L., Bouldin D. W. A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 1979. Vol. PAMI-1, No. 2. P. 224–227. DOI: 10.1109/TPAMI.1979.4766909
19. Caliński T., Harabasz J. A dendrite method for cluster analysis. *Communications in Statistics - Theory and Methods*. 1974. Vol. 3, No. 1. P. 1–27. DOI: 10.1080/03610927408827101
20. Andrienko G., Andrienko N., Keim D., MacEachren A. M., Wrobel S. Challenging problems of geospatial visual analytics. *Journal of Visual Languages and Computing*. 2011. Vol. 22, No. 4. P. 251–256. DOI: 10.1016/j.jvlc.2011.04.001
21. Kraak M.-J. Mapping Time: Illustrated by Minard's Map of Napoleon's Russian Campaign of 1812. *Redlands* : Esri Press, 2014. 110 p.

22. Slocum T. A., McMaster R. B., Kessler F. C., Howard H. H. Thematic Cartography and Geovisualization. 4th ed. Boca Raton : CRC Press, 2022. 638 p.
23. Batty M. The New Science of Cities. Cambridge : MIT Press, 2013. 520 p.
24. Liu Y., Liu X., Gao S., et al. Social sensing: A new approach to understanding our socioeconomic environments. *Annals of the Association of American Geographers*. 2015. Vol. 105, No. 3. P. 512–530. DOI: 10.1080/00045608.2015.1018773
25. Haining R., Li G. Modelling Spatial and Spatial-Temporal Data: A Bayesian Approach. Boca Raton : CRC Press, 2020. 482 p.
26. Zook M., Barocas S., Boyd D., et al. Ten simple rules for responsible big data research. *PLoS Computational Biology*. 2017. Vol. 13, No. 3. Article e1005399. DOI: 10.1371/journal.pcbi.1005399
27. Hernández-Álvarez L., Gomez-Hernández J. A., Gonzalez-Abril L., Velasco F. Analysis of users' behaviour in structured e-commerce websites. *IEEE Access*. 2017. Vol. 5. P. 11941–11958. DOI: 10.1109/ACCESS.2017.2707600
28. McKinney W. Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython. 2nd ed. Sebastopol : O'Reilly Media, 2017. 544 p.
29. VanderPlas J. Python Data Science Handbook: Essential Tools for Working with Data. Sebastopol : O'Reilly Media, 2016. 548 p.
30. Pedregosa F., Varoquaux G., Gramfort A., et al. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*. 2011. Vol. 12. P. 2825–2830.
31. Jordahl K., Van den Bossche J., Fleischmann M., et al. geopandas/geopandas: v0.8.1. Zenodo. 2020. DOI: 10.5281/zenodo.3946761
32. Gillies S. Shapely: manipulation and analysis of geometric objects. 2007. URL: <https://github.com/shapely/shapely> (дата звернення: 15.11.2024).
33. Rey S. J., Anselin L. PySAL: A Python library of spatial analytical methods. *The Review of Regional Studies*. 2007. Vol. 37, No. 1. P. 5–27.
34. Agafonkin V. Leaflet: an open-source JavaScript library for mobile-friendly interactive maps. 2011. URL: <https://leafletjs.com/> (дата звернення: 20.11.2024).

35. Butler H., Daly M., Doyle A., et al. The GeoJSON Format. RFC 7946, Internet Engineering Task Force. 2016. DOI: 10.17487/RFC7946
36. Haklay M., Weber P. OpenStreetMap: User-generated street maps. *IEEE Pervasive Computing*. 2008. Vol. 7, No. 4. P. 12–18. DOI: 10.1109/MPRV.2008.80
37. Mooney P., Minghini M. A review of OpenStreetMap data. In: Foody G., et al., editors. *Mapping and the Citizen Sensor*. London : Ubiquity Press, 2017. P. 37–59. DOI: 10.5334/bbf.c
38. Zheng Y., Rajasegarar S., Leckie C. Parking availability prediction for sensor-enabled car parks in smart cities. *Proceedings of the 10th IEEE International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP)*. Singapore, 2015. P. 1–6. DOI: 10.1109/ISSNIP.2015.7106902
39. Bao J., Zheng Y., Wilkie D., Mokbel M. Recommendations in location-based social networks: a survey. *GeoInformatica*. 2015. Vol. 19, No. 3. P. 525–565. DOI: 10.1007/s10707-014-0220-8
40. Zheng Y. Trajectory data mining: An overview. *ACM Transactions on Intelligent Systems and Technology*. 2015. Vol. 6, No. 3. Article 29. DOI: 10.1145/2743025
41. Fotheringham A. S., Brunson C., Charlton M. *Quantitative Geography: Perspectives on Spatial Data Analysis*. London : SAGE Publications, 2000. 270 p.
42. O'Sullivan D., Unwin D. J. *Geographic Information Analysis*. 2nd ed. Hoboken : John Wiley & Sons, 2010. 432 p.
43. Fischer M. M., Getis A., editors. *Handbook of Applied Spatial Analysis: Software Tools, Methods and Applications*. Berlin : Springer, 2010. 806 p. DOI: 10.1007/978-3-642-03647-7
44. Anselin L., Rey S. J. *Modern Spatial Econometrics in Practice: A Guide to GeoDa, GeoDaSpace and PySAL*. Chicago : GeoDa Press, 2014. 350 p.
45. Getis A., Ord J. K. The analysis of spatial association by use of distance statistics. *Geographical Analysis*. 1992. Vol. 24, No. 3. P. 189–206. DOI: 10.1111/j.1538-4632.1992.tb00261.x

46. Ord J. K., Getis A. Local spatial autocorrelation statistics: distributional issues and an application. *Geographical Analysis*. 1995. Vol. 27, No. 4. P. 286–306. DOI: 10.1111/j.1538-4632.1995.tb00912.x

47. Tobler W. R. A computer movie simulating urban growth in the Detroit region. *Economic Geography*. 1970. Vol. 46, Supplement. P. 234–240. DOI: 10.2307/143141

ДОДАТКИ

Додаток А – Головна сторінка

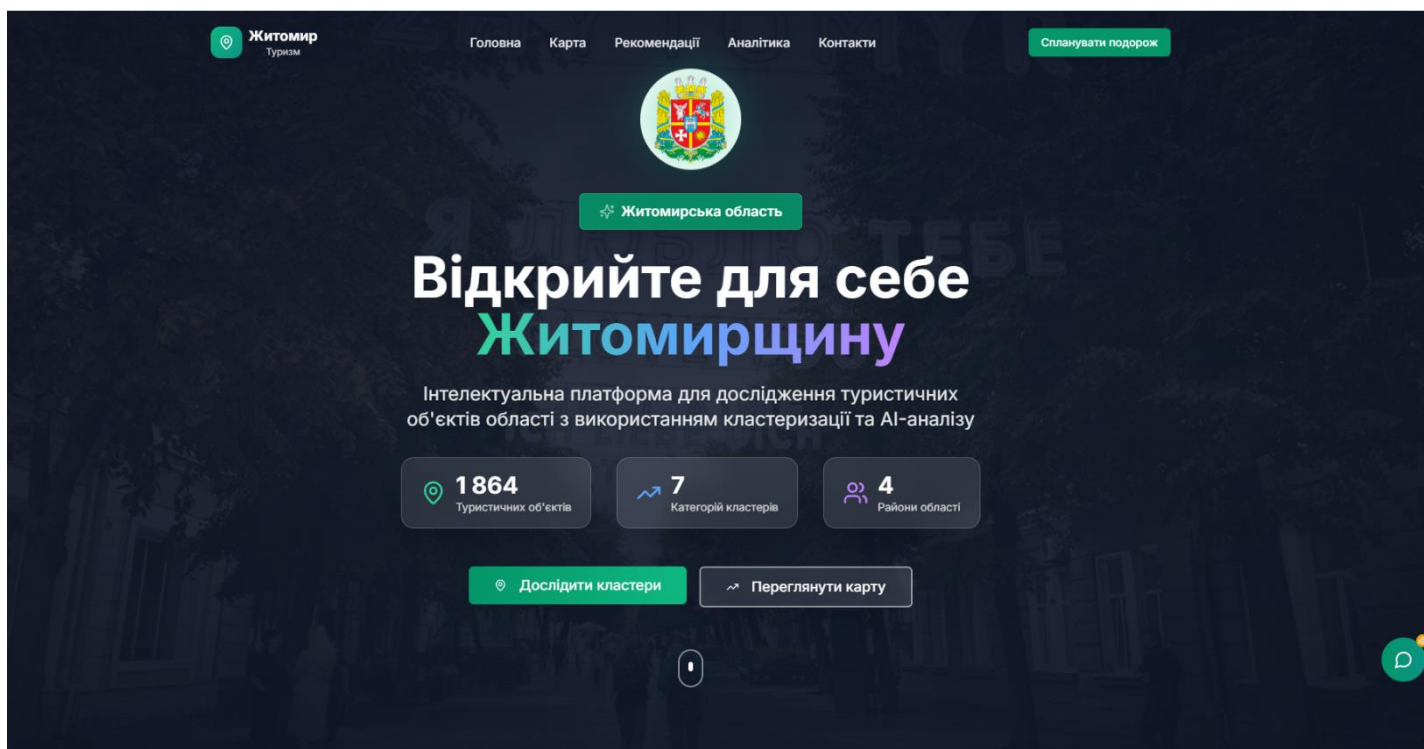


Рисунок А.1 – Головна сторінка сайту

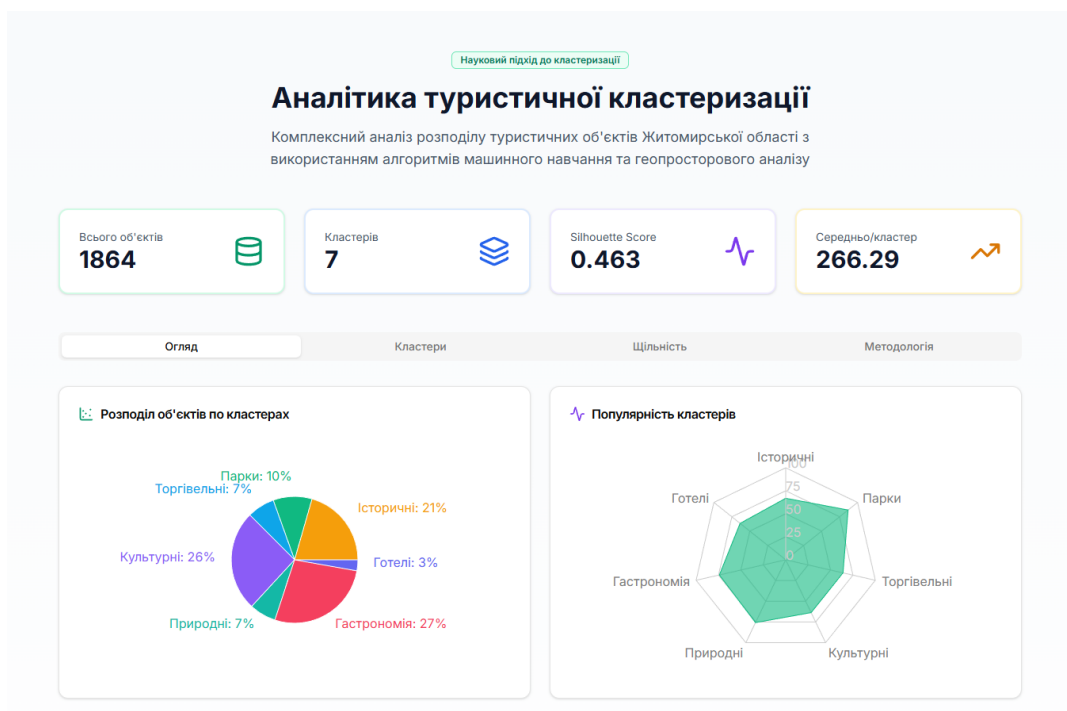


Рисунок А. 2 - Аналітика туристичної кластеризації відповідно вибору параметрів

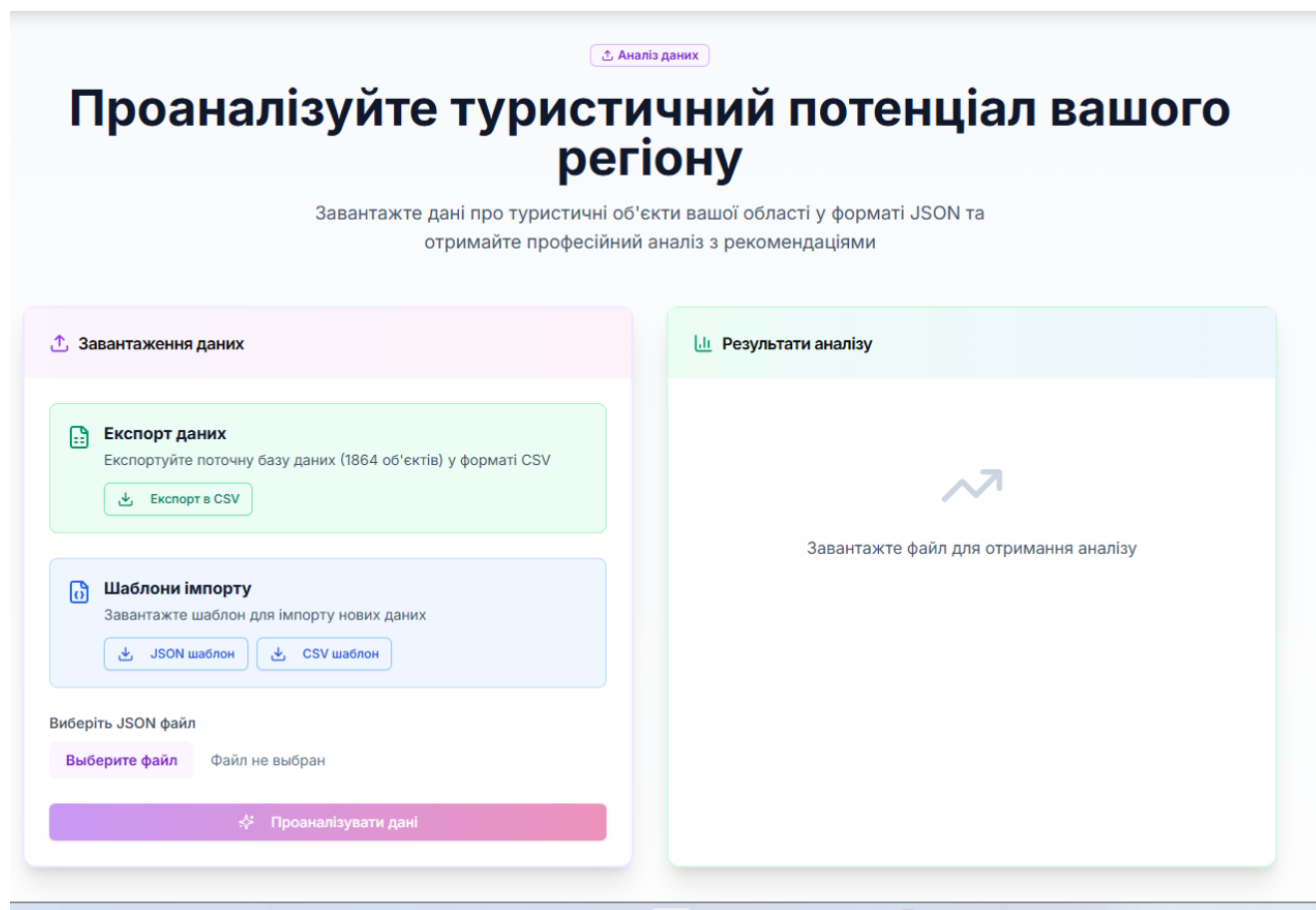


Рисунок А.3 Функція аналізу туристичного потенціалу власного регіону

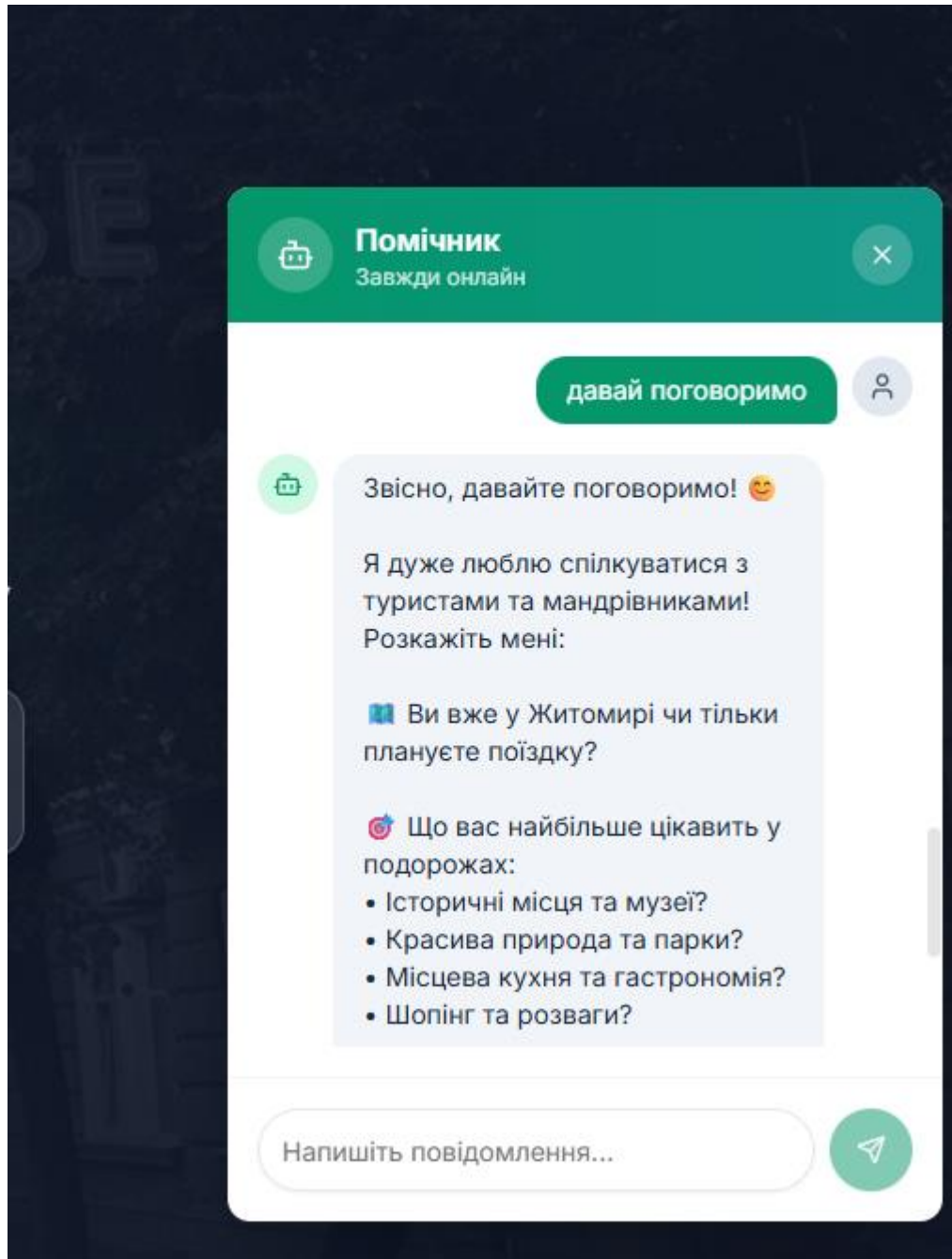


Рисунок А.4 – Посічник, який надає та рекомендує відповідну інформацію про об'єкти які знаходяться у регіоні

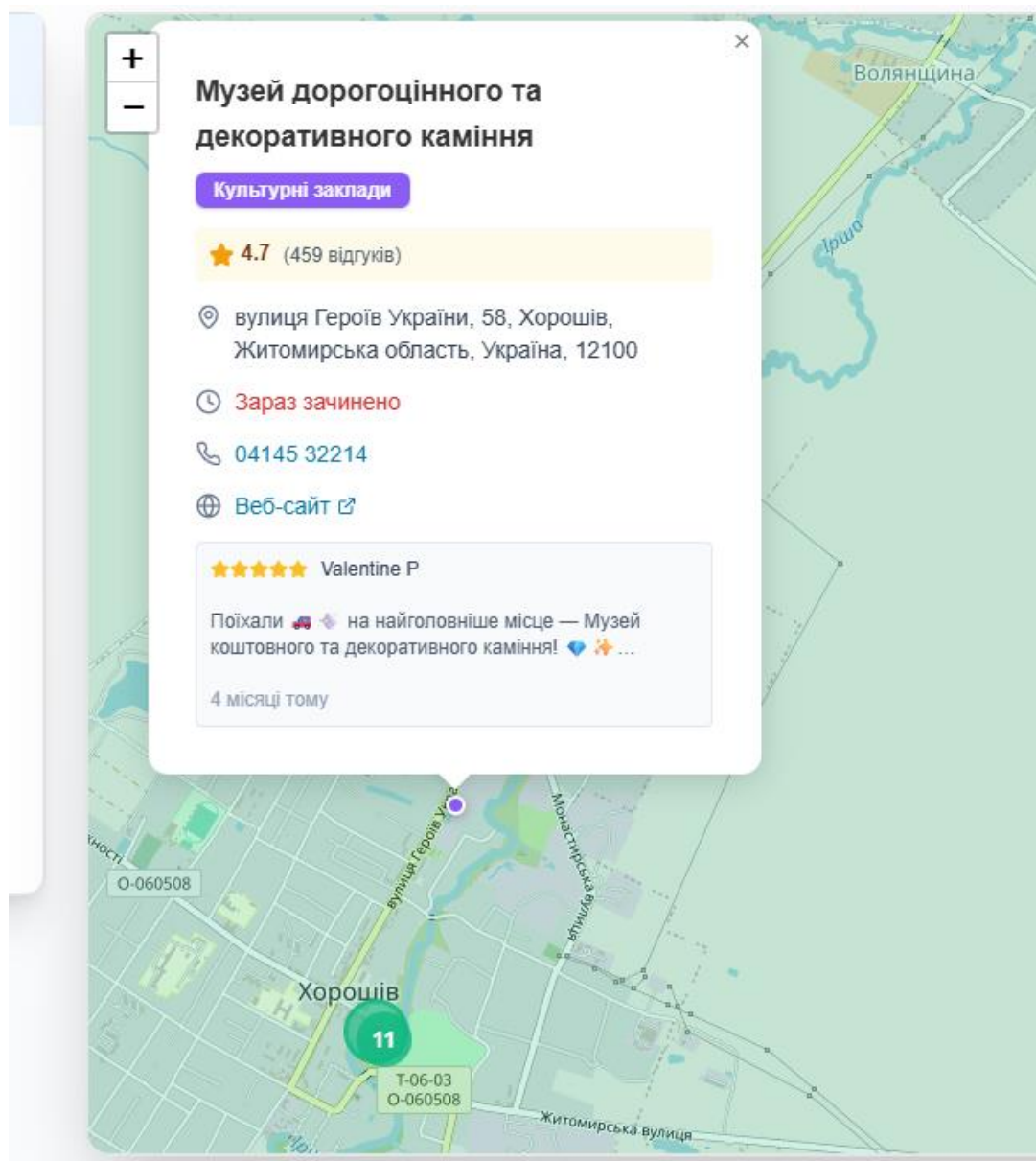


Рисунок А.5 – Приклад відображення інформації про об’єкт який який вибирає користувач