

НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ БІОРЕСУРСІВ
І ПРИРОДОКОРИСТУВАННЯ УКРАЇНИ

Факультет інформаційних технологій

УДК 004:631.164.28:332.3:330.322

«ПОГОДЖЕНО»

Декан факультету
інформаційних технологій

Болбот І.М., д.т.н., професор

_____ 2024 р.

«ДОПУСКАЄТЬСЯ ДО ЗАХИСТУ»

Завідувач кафедри комп'ютерних наук

Голуб Б.Л., к.т.н., доцент

_____ 2024 р.

МАГІСТЕРСЬКА КВАЛІФІКАЦІЙНА РОБОТА

на тему Система (платформа) аналізу збитків та прогнозування інвестицій для відновлення пошкоджених сільськогосподарських земель з використанням методів машинного навчання

Спеціальність: 122 «Комп'ютерні науки»

Освітня програма: «Комп'ютерний еколого-економічний моніторинг»

Орієнтація освітньої програми: Освітньо-професійна

Гарант освітньої програми:

_____ підпис

Виконав:

_____ А.А. Тараненко
підпис

Керівник магістерської кваліфікаційної роботи

_____ О.Г. Глазунова
підпис

ЗМІСТ

ВСТУП	3
1 Теоретичні основи дослідження	6
1.1 Аналіз стану сільськогосподарських земель, пошкоджених внаслідок повномасштабного вторгнення	6
1.2 Огляд існуючих методів оцінки збитків у сільському господарстві 10	
1.3 Теоретичні аспекти кластерного аналізу для оцінки збитків.....	15
2 Розробка системи для аналізу збитків пошкоджених сільськогосподарських земель	19
2.1 Збір та аналіз даних супутникових зображень	19
2.2 Створення бази даних і розрахунок втрат від продажів	24
2.3 Розробка інтерактивної платформи для візуалізації та аналізу даних 30	
3 Аналіз даних за допомогою методів машинного навчання	36
3.1 Використання кластерного аналізу: постановка задачі та опис алгоритму 36	
3.2 Результати кластеризації та їх інтерпретація.....	39
ВИСНОВКИ.....	54
СПИСОК ДЖЕРЕЛ.....	57

ВСТУП

Розвиток сільського господарства завжди був і залишається критично важливим для забезпечення стабільності продовольчої системи кожної країни. В Україні аграрний сектор є одним із ключових рушіїв економіки, забезпечуючи значну частину валового внутрішнього продукту, робочі місця для населення, а також формуючи вагомому частку експортного потенціалу держави. Однак сьогодні цей сектор перебуває під значним впливом глобальних викликів, таких як кліматичні зміни, посилення конкуренції на світовому ринку, та особливо — руйнівних наслідків збройної агресії.

Повномасштабне вторгнення РФ в Україну завдало значної шкоди аграрному сектору. Пошкодження сільськогосподарських земель, втрати врожаю, знищення інфраструктури та порушення логістичних ланцюгів призвели до катастрофічних економічних втрат. У таких умовах постає нагальна необхідність не лише у визначенні розміру цих втрат, але й у створенні ефективних інструментів для оцінки збитків, планування відновлення і стратегічного управління ресурсами.

Застосування сучасних технологій, таких як геоінформаційні системи (ГІС), супутникові знімки та методи машинного навчання, надає можливість автоматизувати процеси аналізу та прогнозування. Це дозволяє значно підвищити точність оцінки збитків і сприяти прийняттю зважених рішень на всіх рівнях управління. У цьому контексті розробка інтерактивної платформи для аналізу збитків та прогнозування інвестицій у відновлення пошкоджених сільськогосподарських земель є не лише актуальною, але й необхідною.

Мета дослідження полягає у створенні системи, яка інтегрує аналітичні інструменти для оцінки збитків, прогнозування фінансових потреб і підтримки процесу відновлення аграрного сектору. У межах цієї мети пропонується розробка інтерактивної платформи, яка поєднує сучасні технології просторового аналізу, великі дані та алгоритми машинного навчання.

Основні завдання дослідження включають:

- Збір даних з різних джерел, таких як супутникові знімки, відкриті бази даних і фінансові звіти, які містять інформацію про втрати врожаю, пошкодження інфраструктури та інші ключові показники;
- Попередню обробку даних, включаючи видалення пропусків, кодування категорійних змінних та стандартизацію числових даних;
- Застосування кластерного аналізу для поділу об'єктів на групи за рівнем збитків, їх географічним розташуванням та іншими характеристиками;
- Розробку візуалізацій, які включають інтерактивні карти та графіки, що полегшують аналіз просторових і економічних взаємозв'язків;
- Формування рекомендацій для визначення пріоритетів у відновленні регіонів та оптимізації інвестицій.

Новизна роботи полягає у використанні комплексного підходу, що базується на інтеграції кількох інноваційних методів аналізу даних. Поєднання супутникових технологій, алгоритмів машинного навчання та інтерактивних візуалізацій забезпечує можливість точного оцінювання збитків і визначення регіонів із найбільш критичними потребами у відновленні. Крім того, платформа, розроблена у рамках дослідження, стане універсальним інструментом для прийняття управлінських рішень.

Практична значущість дослідження полягає в тому, що розроблена платформа дозволить оптимізувати процеси оцінки збитків та інвестиційного планування. Система стане ефективним інструментом для аналітиків, політиків, інвесторів та всіх зацікавлених у відновленні сільськогосподарських територій. Вона сприятиме мінімізації економічних втрат, підвищенню ефективності використання ресурсів і прискоренню відновлення аграрного сектору.

Очікувані результати дослідження включають:

- Розробку інтерактивної платформи, яка дозволяє візуалізувати пошкодження земель, аналізувати втрати та формувати прогнози інвестицій;

- Виявлення ключових закономірностей у розподілі збитків за допомогою методів кластеризації;
- Надання рекомендацій для визначення пріоритетних регіонів і оптимальних стратегій відновлення.

У сучасних умовах, коли економічна стабільність країни значною мірою залежить від відновлення її аграрного сектору, створення таких систем є вкрай важливим. Інтеграція інноваційних рішень, запропонованих у роботі, дозволить досягти не лише короткострокових результатів, але й забезпечити сталий розвиток сільського господарства на довгострокову перспективу.

Результати цього дослідження матимуть значний вплив на практику управління земельними ресурсами, сприяючи підвищенню ефективності прийняття рішень та раціонального використання інвестицій. У складних умовах повоєнного відновлення така система стане важливим інструментом для забезпечення продовольчої безпеки та стабільності аграрного сектору України.

Логіка роботи зумовила таку структуру курсової роботи: вступ, 3 розділи, висновки, список використаних джерел із 48 найменувань. Загальний обсяг 65 сторінки.

1 ТЕОРЕТИЧНІ ОСНОВИ ДОСЛІДЖЕННЯ

1.1 Аналіз стану сільськогосподарських земель, пошкоджених внаслідок повномасштабного вторгнення

Сільськогосподарський сектор відіграє ключову роль у світовій економіці, і його стійкість є основою продовольчої безпеки та сталого розвитку. У контексті України руйнівні наслідки війни спричинили безпрецедентні пошкодження сільськогосподарських угідь, посівів та інфраструктури, що вимагає детального теоретичного аналізу для ефективного планування відновлення. Такий аналіз охоплює фізичні, екологічні та економічні аспекти, роблячи його багатопрофільною галуззю досліджень.

Пошкодження сільськогосподарських земель можна умовно поділити на прямі та непрямі збитки. Прямі збитки включають видиму руйнацію, наприклад, знищені посіви, деградовані ґрунти та зруйновані іригаційні системи. Непрямі збитки охоплюють довгострокові наслідки, такі як погіршення стану ґрунтів, забруднення від вибухових речовин та порушення аграрних ланцюгів поставок. Оцінка збитків у післявоєнний період повинна враховувати ці різноманітні впливи, узгоджуючи їх із глобальними методиками та зосереджуючись на специфічних для конфлікту викликах.

Дослідження, проведені в Україні, демонструють, що війна спричинила значну шкоду аграрному сектору. Наприклад, близько 50% від усіх збитків припадають на сільськогосподарські угіддя та втрачений врожай. Сюди входить як фізичне знищення рослин, так і неможливість зібрати врожай через замінування територій [1]. Вплив цих факторів на економіку сільського господарства має не лише поточний характер, а й довготривалий, що впливає на здатність країни відновлювати свої позиції на глобальних ринках.

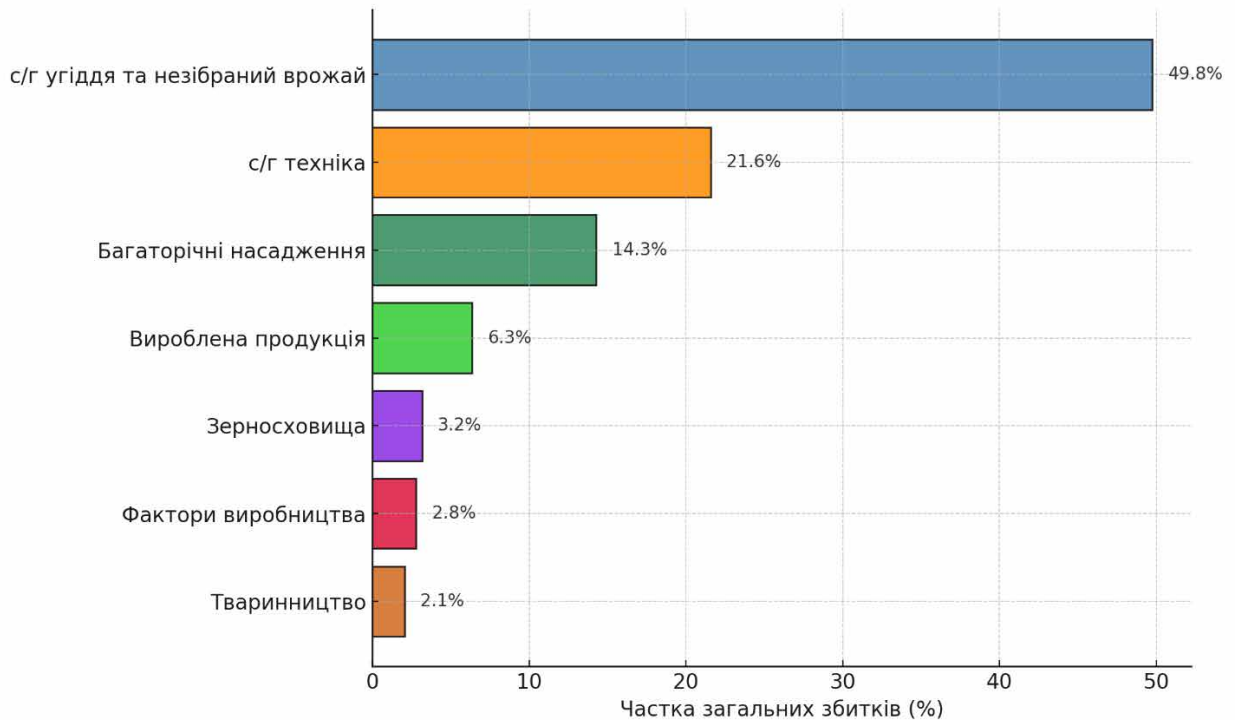


Рисунок 1.1.1 Збитки у сільському господарстві у відсотковому співвідношенні

Сучасні технології, зокрема дистанційне зондування та геоінформаційні системи (ГІС), стали невід’ємною частиною аналізу збитків. Супутникові знімки дозволяють швидко оцінювати великі території, визначаючи масштаби руйнувань і зміни у використанні земель. Дослідження Ozdogan та його колег показало, як дистанційне зондування може допомогти в картуванні втрат врожаю на великих площах, надаючи важливі дані для планування відновлення [2]. У поєднанні з польовими перевітками та машинним навчанням такі технології забезпечують високоточні оцінки збитків.

Машинне навчання (ML) революціонізувало аналіз збитків, підвищивши точність і ефективність. Наприклад, згорткові нейронні мережі (CNN) широко використовуються для аналізу супутникових зображень високої роздільної здатності з метою виявлення та класифікації пошкоджених ділянок. У дослідженні Liu та колег застосовувалися моделі CNN для ідентифікації територій, уражених природними катастрофами, досягаючи точності понад 90% у прогнозуванні інтенсивності збитків [3]. Ці технології можуть бути адаптовані

для аналізу пошкоджень, спричинених війною, особливо у регіонах із обмеженим доступом до наземних обстежень.

Хімічне забруднення ґрунтів є серйозною проблемою для регіонів, постраждалих від воєнних дій. Вибухівка, важкі метали та інші забруднювачі створюють довгострокові ризики для здоров'я ґрунтів і безпеки врожаїв. Zhang та його колеги вивчали, як моделі машинного навчання, можуть передбачити рівні забруднення ґрунтів і запропонувати стратегії їх відновлення [4]. Ці моделі особливо актуальні для України, де міни та бойові уламки значною мірою вплинули на якість ґрунтів.

Економічна оцінка збитків у сільському господарстві є не менш важливою. Розрахунок фінансових втрат від знищених врожаїв, пошкодженої інфраструктури та зниження продуктивності є основою для компенсації і планування відновлення. Дослідження Lobell та ін. інтегрувало кліматичні дані з економічними моделями для прогнозування грошових втрат, спричинених посухами [5]. Ця методика може бути адаптована для оцінки економічних наслідків війни в Україні, враховуючи такі змінні, як порушена логістика та доступ до ринків.

Теоретичні основи аналізу збитків також охоплюють аспекти втрати біорізноманіття та екологічної рівноваги. Війна не лише шкодить сільськогосподарській продуктивності, а й порушує функціонування екосистем. У дослідженні Turner аналізувалися каскадні наслідки руйнування середовища проживання для екосистемних послуг, підкреслюючи необхідність інтегрованої екологічної оцінки у зусиллях з відновлення [6].

Методики аналізу збитків, специфічні для конфліктів, були розроблені в регіонах, що постраждали від затяжних війн. Наприклад, дослідження Schouten зосереджувалося на впливах конфлікту на сільське господарство на Близькому Сході, наголошуючи на необхідності локалізованих рамкових оцінок [7]. Ці висновки є актуальними для України, де різноманіття кліматичних зон і типів ґрунтів вимагає адаптованих підходів до відновлення.

Український аграрний сектор також стикається з дефіцитом робочої сили та руйнуванням інфраструктури, що ускладнює відновлення. Дослідження наголошують на ролі соціальних і економічних чинників у формуванні результатів оцінки збитків. Досліджено як залучення громад до оцінки після стихійних лих покращує точність оцінок втрат і сприяє відновленню [8].

Важливим аспектом аналізу збитків є його застосування у довгостроковому плануванні та формуванні політики. Ефективні стратегії відновлення потребують інтеграції даних про збитки в системи підтримки прийняття рішень, які спрямовують інвестиції та розподіл ресурсів. Дослідження Van der Werf продемонструвало, як платформи на базі ГІС можуть допомагати політикам у пріоритизації зусиль з відновлення та оцінці їх впливу [9]. Такі платформи є надзвичайно актуальними для України, де велика кількість даних потребує інтеграції для координації зусиль з відновлення.

Теоретичні основи аналізу збитків у сільському господарстві охоплюють низку дисциплін, зокрема екологію, економіку та аналіз даних. Використання дистанційного зондування, машинного навчання та економічного моделювання надає потужні інструменти для оцінки та вирішення багатогранних наслідків війни для сільськогосподарських земель. Застосовуючи ці методи, Україна може розробити всеосяжну стратегію для відновлення свого аграрного сектору та забезпечення сталого розвитку в післявоєнний період.

1.2 Огляд існуючих методів оцінки збитків у сільському господарстві

Оцінка збитків у сільському господарстві є складним завданням, яке поєднує традиційні методи та сучасні технологічні підходи для аналізу масштабів і наслідків різних несприятливих подій. З розвитком технологій ці підходи значно удосконалилися, особливо завдяки використанню машинного навчання та технологій дистанційного зондування. З-поміж них кластеризація, як метод машинного навчання, стала ключовим інструментом для оцінки збитків у аграрному секторі.

Історично оцінка збитків у сільському господарстві базувалася на ручних обстеженнях полів і безпосередньому зборі даних. Ці методи передбачали фізичну оцінку стану посівів, ґрунтів і сільськогосподарської інфраструктури, що дозволяло отримувати детальну інформацію про стан аграрних систем. Однак, попри точність на локальному рівні, такі методи мають серйозні обмеження у масштабності, витратах часу та трудових ресурсах [10]. Наприклад, під час масштабних катастроф або конфліктів фізичний доступ до великих площ сільськогосподарських земель часто ускладнений або неможливий.

В Україні нинішній конфлікт особливо висвітлив обмеження традиційних методів, оскільки багато постраждалих районів залишаються недоступними через небезпеку.

Застосування технологій дистанційного зондування стало революційним кроком у можливостях моніторингу та оцінки збитків на великих територіях. Супутникові знімки, дрони та аерофотозйомка забезпечують оперативний моніторинг значних площ сільськогосподарських угідь, дозволяючи виявляти патерни руйнувань і зміни у використанні земель [11]. Наприклад, багатоспектральні та гіперспектральні зображення дозволяють виявляти стрес у рослинах, спричинений засухою, хворобами або шкідниками [12].

Коли дані дистанційного зондування інтегруються з ГІС, вони надають просторові інсайти, які підвищують точність оцінки збитків. ГІС-аналіз дозволяє візуалізувати розподіл збитків у регіонах, що сприяє ефективнішому

плануванню відновлення. Наприклад, Hurni та Kruse показали, як ГІС може інтегрувати екологічні дані з просторовими моделями для виявлення гарячих точок пошкоджень і визначення пріоритетів у відновленні [10].

Розвиток машинного навчання кардинально змінив підходи до оцінки збитків у сільському господарстві. Техніки машинного навчання автоматизують аналіз великих масивів даних, пропонуючи вищу точність і масштабованість у порівнянні з традиційними методами. Наприклад, алгоритми з контрольованим навчанням, такі як Random Forests або нейронні мережі, широко використовуються для класифікації пошкоджених і здорових посівів на основі даних дистанційного зондування [13].

Проте методи з контрольованим навчанням залежать від наявності позначених наборів даних, що може бути проблемою в регіонах, які постраждали від конфліктів або катастроф. У таких випадках методи неконтрольованого навчання, зокрема кластеризація, стають перспективною альтернативою.

Кластеризація, як метод неконтрольованого навчання, групує дані на основі їхньої схожості без необхідності використання позначених наборів даних. Це робить її особливо корисною в оцінці збитків, де маркування даних може бути складним завданням. До поширених алгоритмів кластеризації належать [14]:

- K-means;
- ієрархічна кластеризація;
- Density-Based Spatial Clustering of Applications with Noise (DBSCAN).

Методи кластеризації були успішно застосовані для аналізу та оцінки збитків у різних контекстах:

- Алгоритм ISODATA: Вдосконалений варіант K-means, алгоритм ISODATA, використовувався для класифікації супутникових знімків за типами земного покриття, що допомогло визначити пошкоджені території. Цей метод ітеративно налаштовує центри кластерів і мінімізує розбіжності між ними, ефективно розрізняючи здорові й уражені ділянки [15];

- DBSCAN для виявлення аномалій: Алгоритм DBSCAN широко використовується для виявлення аномалій у показниках здоров'я рослин, отриманих з дистанційного зондування. Виявляючи кластери подібних точок і маркуючи відхилення як потенційні пошкодження, DBSCAN сприяє ранньому виявленню проблем, спричинених шкідниками або хворобами [16].
- Інтеграція з ГІС: Поєднання кластеризації з ГІС забезпечує просторову візуалізацію пошкоджень. Дослідження Zhang підкреслює, що інтеграція кластеризації з ГІС-моделями невизначеності підвищує точність прогнозів збитків [17].

Кластеризація пропонує декілька суттєвих переваг, які роблять її незамінним інструментом для оцінки збитків у сільському господарстві. По-перше, відсутність необхідності маркованих даних: на відміну від алгоритмів з контрольованим навчанням, кластеризація не потребує попередньо позначених навчальних наборів даних. Це особливо важливо для ситуацій, де зібрати марковані дані складно або навіть неможливо через обмежений доступ до територій чи відсутність ресурсів. Для країн, які постраждали від конфліктів, таких як Україна, ця особливість є ключовою, оскільки значні території сільськогосподарських земель можуть бути небезпечними для польових досліджень. Алгоритми кластеризації, такі як K-means чи DBSCAN, дозволяють аналізувати великі масиви немаркованих даних, виділяючи групи зі схожими характеристиками.

Кластеризація має здатність знаходити приховані патерни в даних, які не завжди очевидні для дослідників. Наприклад, алгоритм DBSCAN може виявляти аномалії у показниках здоров'я рослин, що може свідчити про локальні проблеми, такі як зараження хворобами чи недостатнє зрошення. Подібні закономірності складно визначити традиційними методами через їхню обмежену здатність обробляти великі набори даних. Кластеризація дозволяє будувати більш точні моделі, які відображають реальну картину впливу пошкоджень.

Алгоритми кластеризації здатні обробляти великі набори даних, що робить їх придатними для аналізу на регіональному або національному рівнях. У сучасних умовах сільське господарство генерує значні обсяги даних із супутників, дронів та сенсорів. Кластеризація забезпечує ефективний спосіб їхньої обробки, зменшуючи витрати часу і ресурсів. Наприклад, алгоритм ISODATA, завдяки своїй ітеративній природі, може адаптуватися до різних масштабів даних, що робить його корисним для аналізу великих площ, таких як аграрні зони України [18].

Та попри свої численні переваги, кластеризація стикається з певними викликами, які потребують уваги при її впровадженні (табл. 1):

- вибір ознак;
- чутливість до шуму;
- визначення кількості кластерів.

Результати кластеризації значною мірою залежать від вибору ознак, які використовуються для аналізу даних. Неправильний вибір характеристик може призвести до некоректного формування кластерів, що знижує точність і надійність результатів. Наприклад, якщо використовуються лише спектральні ознаки без урахування контекстуальних факторів (як-от клімат чи рельєф), алгоритм може неправильно класифікувати зони пошкоджень. Це потребує ретельного відбору релевантних показників і тестування різних наборів характеристик для оптимального результату.

Алгоритми кластеризації, такі як K-means, можуть бути надто чутливими до шуму й аномальних значень у даних. У контексті оцінки збитків шум може виникати через помилки збору даних, наприклад, через погану якість супутникових знімків чи погодні умови. Такі неточності можуть впливати на формування кластерів, внаслідок чого важливі області пошкоджень можуть бути неправильно ідентифіковані. Алгоритми, як-от DBSCAN, частково вирішують цю проблему, ідентифікуючи аномальні значення як окремі точки, проте їхня ефективність залежить від правильного налаштування параметрів.

Однією з найбільших труднощів є визначення оптимальної кількості кластерів. Для алгоритму K-means користувач повинен заздалегідь встановити кількість кластерів, що може бути суб'єктивним рішенням і вплинути на якість аналізу. Методи, як-от "еластичне коліно" (elbow method), надають інструменти для оцінки оптимальної кількості, але навіть вони не завжди дають точний результат. У більш складних випадках може бути необхідне комбінування кластеризації з іншими алгоритмами або візуалізація результатів для досягнення точності.

Таблиця 1.2.1

Порівняння методів кластеризації за ключовими характеристиками

Метод кластеризації	Точність	Стійкість до шуму	Масштабованість	Вимоги до маркованих даних
K-means	Висока	Низька	Висока	Не потребує
DBSCAN	Висока	Висока	Середня	Не потребує
Ієрархічна	Середня	Низька	Низька	Не потребує
ISODATA	Висока	Середня	Висока	Не потребує

1.3 Теоретичні аспекти кластерного аналізу для оцінки збитків

Кластерний аналіз є потужним інструментом у сфері машинного навчання, який дозволяє структурувати великі масиви даних на основі схожості між їхніми елементами. Це особливо актуально для аграрного сектору, де необхідно аналізувати дані різних типів і джерел, зокрема супутникові знімки, показники ґрунтів і кліматичні дані. Унікальність кластерного аналізу полягає у здатності знаходити приховані закономірності та групувати об'єкти без необхідності маркування даних. Теоретичні аспекти кластеризації включають принципи, алгоритми, адаптацію до просторово-часових даних та інструменти для оцінки якості результатів.

Кластерний аналіз базується на ідеї поділу даних на групи (кластери), елементи яких мають більше схожості всередині групи, ніж між групами. Для вимірювання схожості використовуються різні метрики, серед яких найпоширенішими є евклідова, косинусна та махаланобісова відстані. Наприклад, евклідова відстань є стандартною метрикою для числових даних, проте для категоріальних змінних доцільніше використовувати косинусну схожість. Вибір метрики визначає, як алгоритм ідентифікує кластери, і значною мірою впливає на кінцеві результати [19].

Серед основних алгоритмів кластеризації найпоширенішим є метод K-means. Він використовує ітеративний підхід для визначення центрів кластерів, мінімізуючи варіації всередині них. Цей метод добре працює на великих наборах даних, проте вимагає заздалегідь встановити кількість кластерів, що може бути проблематичним у разі невизначеної структури даних. У свою чергу, ієрархічна кластеризація будує дерево кластерів (дендрограму), яке дозволяє візуалізувати відносини між даними. Цей метод краще підходить для невеликих наборів даних, оскільки потребує значних обчислювальних ресурсів [20].

Іншим важливим підходом є DBSCAN — алгоритм, який визначає кластери на основі щільності даних. Він має унікальну здатність ідентифікувати області шуму та аномалій, що робить його надзвичайно корисним для

геопросторового аналізу. DBSCAN широко використовується для оцінки збитків у сільському господарстві, оскільки дозволяє аналізувати складні структури даних, наприклад, розподіл захворювань рослин чи деградацію ґрунтів у регіоні [21].

У сільському господарстві кластерний аналіз застосовується для моделювання як просторових, так і часових даних. Наприклад, при аналізі супутникових знімків стану врожаю кластеризація дозволяє виділяти регіони зі схожими індексами вегетації, що свідчать про наявність пошкоджень або стресу. У випадках, коли дані містять часову складову, наприклад, сезонні коливання або довгострокові тенденції, використовуються методи кластеризації часових рядів. Алгоритми, такі як Dynamic Time Warping (DTW), дозволяють аналізувати змінні з різною тривалістю чи масштабом. Ці методи ефективно застосовуються для моніторингу змін у врожайності внаслідок кліматичних факторів або стихійних лих [22].

Особливу увагу в кластерному аналізі приділяють оцінці якості кластерів. Для цього використовуються як візуальні, так і кількісні метрики. Наприклад, силуетний коефіцієнт вимірює, наскільки добре дані відповідають своїм кластерам, порівняно з іншими. Високі значення коефіцієнта вказують на чіткі межі між групами, що є важливим для точності оцінки збитків. Крім того, графічні інструменти, такі як дендрограми або двовимірні проєкції даних через методи зменшення вимірності (наприклад, t-SNE), дозволяють краще зрозуміти структуру кластерів [23].

Попри численні переваги, кластерний аналіз має свої виклики. Одним із них є визначення оптимальної кількості кластерів. Існують такі методи, як "еластичне коліно" (elbow method) або аналіз розривів (gap analysis), які допомагають у прийнятті цього рішення. Однак у складних випадках може знадобитися комбінація кількох підходів для досягнення оптимальних результатів [24]. Іншим викликом є обробка високовимірних даних, що часто характерно для сільськогосподарських досліджень. У таких випадках застосовують методи зменшення вимірності, наприклад, Principal Component

Analysis (PCA), для покращення обчислювальної ефективності та зменшення впливу шуму [25].

Кластерний аналіз знаходить дедалі більше застосувань у різних галузях, зокрема у сільському господарстві, завдяки здатності ефективно обробляти великі та різномірні масиви даних. Особливої уваги заслуговують гібридні методи, які поєднують кластеризацію з іншими алгоритмами машинного навчання. Наприклад, методи глибокого навчання, такі як Autoencoders, у поєднанні з алгоритмами кластеризації, дозволяють автоматично виділяти найбільш релевантні характеристики даних і покращувати якість кластеризації [26]. Це є надзвичайно корисним у випадках роботи з високовимірними супутниковими зображеннями.

Інтеграція кластеризації з геоінформаційними системами (ГІС) відкриває нові можливості для візуалізації та просторового аналізу. ГІС-інструменти дозволяють інтерактивно аналізувати кластерні дані, що особливо важливо для моніторингу пошкоджень у великих аграрних зонах. Дослідження показують, що використання ГІС у комбінації з алгоритмами DBSCAN значно покращує точність виявлення зон із аномаліями, спричиненими посухою чи шкідниками [27]. Це дозволяє створювати інтерактивні карти пошкоджень, які є ефективним інструментом для аграрних менеджерів і державних установ.

Аналіз часових рядів є ще однією перспективною областю розвитку кластерного аналізу. У сільському господарстві сезонність та довготривалі тренди відіграють важливу роль у прийнятті управлінських рішень. Кластеризація часових даних дозволяє визначати, як змінюється стан культур у різні періоди, і прогнозувати можливі втрати. Методи, такі як Time-Series K-means, демонструють високу ефективність у цій сфері, оскільки враховують як тимчасові, так і просторові аспекти змін [28].

Ще одним важливим аспектом є адаптація кластеризації для роботи з великими масивами даних. У сучасному сільському господарстві обсяги даних швидко зростають, що створює виклики для традиційних алгоритмів кластеризації. Використання алгоритмів із розподіленою архітектурою, таких як

Mini-Batch K-means, дозволяє значно зменшити час обробки, не жертвуючи точністю [29]. Це особливо актуально для аналізу даних, отриманих із дронів, сенсорів і супутників.

Оцінка якості кластеризації залишається важливим етапом аналізу. Крім силуетного коефіцієнта, дедалі більшого поширення набуває метрика Dunn Index, яка враховує як щільність кластерів, так і відстань між ними [30]. Ця метрика дозволяє глибше зрозуміти структуру даних і покращити результати аналізу. Наприклад, застосування метрики для аналізу даних про врожайність в аграрних регіонах України показало, що використання комбінованих алгоритмів із PCA забезпечує вищу точність і стабільність результатів [31].

2 РОЗРОБКА СИСТЕМИ ДЛЯ АНАЛІЗУ ЗБИТКІВ ПОШКОДЖЕНИХ СІЛЬСЬКОГОСПОДАРСЬКИХ ЗЕМЕЛЬ

2.1 Збір та аналіз даних супутникових зображень

Використання технологій дистанційного зондування стало ключовим інструментом у моніторингу сільського господарства, особливо в аналізі пошкоджень сільськогосподарських угідь. Інтеграція супутникових зображень, методів машинного навчання та технологій об'єднання даних дозволила суттєво підвищити ефективність оцінки стану сільськогосподарських екосистем. У цьому підрозділі розглянуто основні методи і технології, що застосовуються для аналізу пошкоджень, з посиланням на сучасні наукові джерела.

Технології дистанційного зондування докорінно змінили процес спостереження за станом сільськогосподарських угідь у великих масштабах. Супутникові зображення високої роздільної здатності, такі як Sentinel-2 та EOS SAT, забезпечують детальну оцінку стану земель. Ці супутникові системи надають багатоспектральні дані, які є важливими для виявлення пошкоджень та оцінки продуктивності земель [32, 38, 39]. Наприклад, Sentinel-2 дозволяє отримувати дані з частотою один раз на п'ять днів, що критично важливо для спостереження за швидкими змінами, такими як паводки чи посухи.

Індекси рослинності, зокрема NDVI (нормалізований вегетаційний індекс), широко використовуються для моніторингу здоров'я рослин і виявлення пошкоджень. NDVI є ефективним інструментом для ідентифікації стресу, спричиненого шкідниками, посухою або деградацією ґрунтів [33, 41]. Аналіз динаміки NDVI у часі дозволяє відстежувати розвиток пошкоджень і кількісно їх оцінювати.

Гіперспектральна зйомка відкрила нові можливості для аналізу сільськогосподарських культур завдяки фіксації спектральних даних у широкому діапазоні довжин хвиль. Ця технологія дозволяє ідентифікувати типи

культур, виявляти хвороби і оцінювати дефіцит поживних речовин [34]. Зокрема, гіперспектральна зйомка допомагає виявити тонкі фізіологічні зміни рослин, які не помітні на звичайних зображеннях.

Об'єднання різних джерел даних, відомого як *data fusion*, покращує точність і надійність оцінки пошкоджень. Поєднання супутникових зображень, наземних спостережень та метеорологічних даних дозволяє створювати комплексні моделі для прогнозування продуктивності культур та оцінки збитків [35]. Наприклад, такі технології застосовувались для оцінки впливу паводків 2023 року, використовуючи супутникові знімки та дані про кількість опадів.

Алгоритми машинного навчання, зокрема метод *Random Forest* та моделі глибокого навчання, використовуються для класифікації пошкоджень і прогнозування втрат врожаю. Ці моделі навчаються на великих масивах спектральних та допоміжних даних, що дозволяє підвищити точність аналізу [36]. Зокрема, глибоке навчання є ефективним у виявленні складних шаблонів, таких як розмежування наслідків паводків і посухи.

Безпілотні літальні апарати (дрони), оснащені мультиспектральними та гіперспектральними сенсорами, забезпечують зображення високої роздільної здатності, які доповнюють супутникові дані. Вони особливо корисні для локального аналізу, наприклад, ідентифікації спалахів шкідників або проблем із зрошенням на конкретних ділянках [37].

Використання супутникових систем, таких як *Sentinel-2* та *EOS SAT*, забезпечує постійний моніторинг сільськогосподарських земель. Ці супутники забезпечують регулярні та часті дані, що важливо для оцінки пошкоджень, викликаних динамічними подіями, такими як град чи лісові пожежі [38, 39].

У роботі аналізуються супутникові знімки для аналізу земельних ресурсів, дозволяючи оцінювати масштаби пошкоджень, зміни у структурі землекористування та вплив зовнішніх факторів. Робота з супутниковими знімками розпочалася з їхньої попередньої обробки. Для цього було використано геоінформаційні системи (ГІС), які дозволяють виконати геометричну корекцію, усунути спотворення та видалити атмосферні шуми. Основною метою цього

етапу було забезпечити точне вирівнювання знімків відповідно до географічних координат. Знімки були синхронізовані з межами адміністративних районів для забезпечення відповідності з даними про території сільськогосподарських земель.

На Рисунок 2.1.1 представлені кольорові маркери (зелені, фіолетові, сині), кольори було використано для класифікації різних рівнів пошкоджень сільськогосподарських земель. Кожен колір відповідав певній категорії стану:

- зелений – незначні пошкодження;
- фіолетовий – значні пошкодження;
- синій – непридатність для використання.

Використання інструментів ГІС дозволило визначити координати кожного маркера та створити цифрові полігони, які представляють площі пошкоджених ділянок.

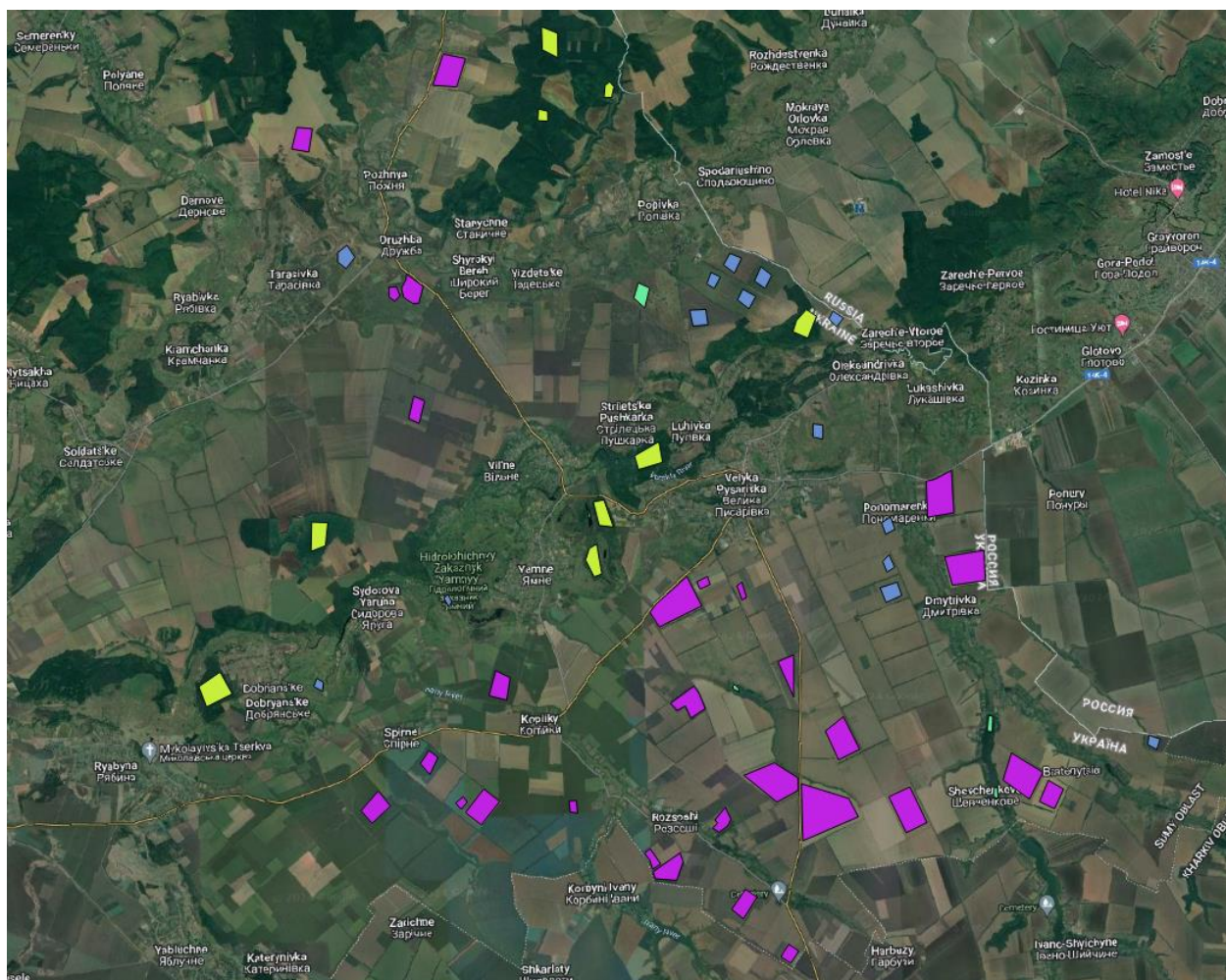


Рисунок 2.1.1 Формування векторних полігонів з вказанням типу земного покриття

На основі цифрових полігонів, створених під час аналізу, було розраховано площі кожної пошкодженої ділянки. Для цього використовували функції ГІС для вимірювання площі полігонів у гектарах. Такий підхід дозволив кількісно оцінити масштаби пошкоджень для подальшого аналізу впливу на врожайність.

Для оцінки стану рослинності використовувалися індекси вегетації, зокрема NDVI (нормалізований індекс вегетації). Цей індекс дозволяє визначити рівень стресу рослин, спричиненого пошкодженнями, а також оцінити продуктивність земель до та після впливу негативних факторів. Аналіз NDVI був проведений для кожної ділянки, визначеної на знімках, що дозволило порівняти рівень продуктивності на різних етапах.

Окрім супутникових зображень, у процесі аналізу було використано додаткові дані, такі як погодні умови, історичні карти земельного користування та статистичні дані щодо врожайності. Ці дані були інтегровані у ГІС для створення багат шарових карт, які показують вплив пошкоджень на сільське господарство.

На основі отриманих даних було створено інтерактивну карту, яка відображає пошкоджені території з використанням різнокольорових маркерів. Користувач може переглянути детальну інформацію про кожну ділянку, включаючи координати, категорію пошкоджень та площу. Крім того, інтерактивна карта підтримує функцію фільтрації за типом пошкоджень або рівнем втрат.

На Рисунок 2.1.2, що представляє зелені території, була проведена класифікація за типами землекористування. Зелені області визначалися як потенційно продуктивні, тоді як чорні полігони вказували на пошкоджені або непридатні для використання землі. Завдяки аналізу цих даних вдалося отримати повну картину стану земельного фонду окремого регіону.

2.2 Створення бази даних і розрахунок втрат від продажів

Створення ефективної бази даних для сільського господарства є важливим етапом у забезпеченні належного управління даними, їх аналізу та прийняття зважених рішень у цій галузі. Сучасні підходи до розробки баз даних передбачають інтеграцію різних джерел даних, забезпечення їх сумісності та використання передових технологій для розширення можливостей таких систем.

Сучасні сільськогосподарські бази даних об'єднують інформацію з різних джерел, включаючи супутникові зображення, мережі датчиків і польові спостереження. Така інтеграція дозволяє отримати цілісне уявлення про сільськогосподарські системи, забезпечуючи точніший аналіз і прогнози. Наприклад, база даних Global Subnational Agricultural Production (GSAP) акумулює стандартизовані дані про виробництво ключових сільськогосподарських культур у різних країнах світу, що дозволяє проводити порівняльний аналіз на регіональному рівні [42].

Сумісність даних відіграє важливу роль у забезпеченні безперешкодного обміну інформацією між різними платформами та системами. Використання стандартизованих схем метаданих та семантичних стандартів, таких як контрольовані словники й онтології, покращує сумісність даних. Наприклад, у проекті *Bangladesh Agricultural Knowledge Graph* використовується семантична інтеграція аграрних даних, що відповідає принципам FAIR (Findable, Accessible, Interoperable, Reusable) [43].

Використання передових технологій, таких як хмарні обчислення, машинне навчання та аналітика великих даних, значно підвищило ефективність і масштабованість баз даних для сільського господарства. Дослідження, присвячене розробці сховищ даних для великих аграрних даних, демонструє застосування таких технологій, як HIVE, MongoDB та Cassandra. Вони дозволяють обробляти великі та складні набори даних, підтримуючи наукові дослідження й аналітику бізнесу [44].

Попри технологічний прогрес, існують виклики, пов'язані з якістю, стандартизацією та управлінням даними. Наприклад, огляд у *European Review of Agricultural Economics* наголошує на необхідності вдосконалення управлінських практик для покращення впливу аграрних досліджень і політик. Важливим є впровадження ефективних рамок управління даними та забезпечення їх точності [45].

Забезпечення доступу до аграрних даних і їх зручного використання є важливим для всіх зацікавлених сторін: фермерів, дослідників і політиків. Наприклад, у дослідженні про створення цифрової платформи *Farmdata* обговорюється створення системи, яка з'єднує різні бази даних, агрегує інформацію та пропонує інструменти для прийняття зважених рішень [46].

Створення бази даних

У роботі база даних для аналізу пошкоджень сільськогосподарських земель має чітко визначену структуру з використанням кількох таблиць, кожна з яких зберігає специфічну інформацію. Взаємозв'язки між таблицями забезпечують логічну організацію даних, дозволяючи ефективно обробляти й аналізувати інформацію.

База даних складається з семи таблиць (Рисунок 2.2.1):

- **Regions** – регіони з пошкодженими об'єктами;
- **Objects** – пошкоджені об'єкти;
- **Fields** – інформація про характеристики поля;
- **Crops** – інформація про культури;
- **DamageTypes** – інформація про тип пошкодження;
- **Sources** – посилання на відкриті джерела інформації
- **Damages** – інформація про втрачений прибуток.

Зв'язки між таблицями:

Regions ↔ Objects: Кожен об'єкт прив'язаний до певного регіону через зовнішній ключ.

Objects ↔ Damages: Кожен об'єкт може мати декілька записів про пошкодження.

Fields ↔ Crops: Кожне поле асоційоване з певною культурою.

DamageTypes ↔ Damages: Пошкодження класифіковані за типами.

Sources ↔ Damages: Кожен запис про пошкодження має посилання на джерело інформації.

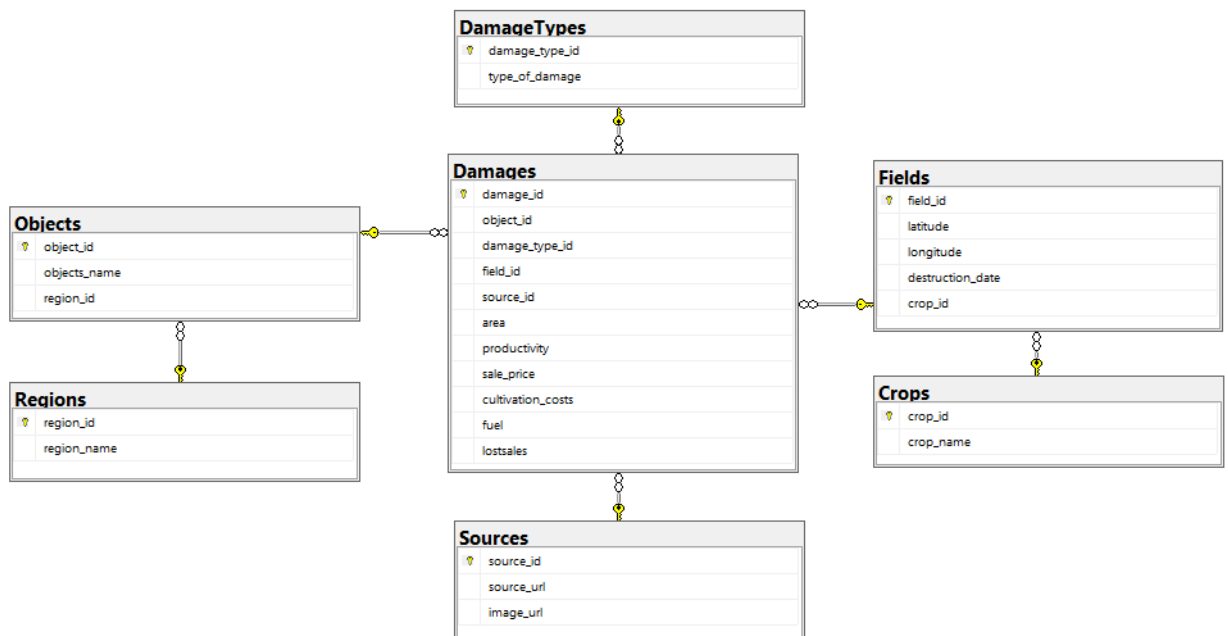


Рисунок 2.2.1 Реляційна структура бази даних

Опис полів:

Таблиця Regions

region_id (ціле, первинний ключ) – унікальний ідентифікатор регіону;

region_name (текст) – назва регіону.

Таблиця Objects

object_id (ціле, первинний ключ) – унікальний ідентифікатор об'єкта.

objects_name (текст) – назва об'єкта (наприклад, поле чи сільськогосподарське підприємство).

region_id (ціле, зовнішній ключ) – посилання на таблицю Regions.

Таблиця Fields

field_id (ціле, первинний ключ) – унікальний ідентифікатор поля.

latitude (десятькове число) – широта, що вказує на географічне розташування.

longitude (десятькове число) – довгота, що вказує на географічне розташування.

destruction_date (дата) – дата пошкодження поля.

crop_id (ціле, зовнішній ключ) – посилання на таблицю Crops.

Таблиця Crops

crop_id (ціле, первинний ключ) – унікальний ідентифікатор культури.

crop_name (текст) – назва культури (наприклад, пшениця, кукурудза).

Таблиця DamageTypes

damage_type_id (ціле, первинний ключ) – унікальний ідентифікатор типу пошкоджень.

type_of_damage (текст) – опис типу пошкоджень (наприклад, незначні, значні, знищені).

Таблиця Sources

source_id (ціле, первинний ключ) – унікальний ідентифікатор джерела.

source_url (текст) – посилання на джерело інформації.

image_url (текст) – посилання на зображення, що ілюструє пошкодження.

Таблиця Damages

damage_id (ціле, первинний ключ) – унікальний ідентифікатор запису про пошкодження.

object_id (ціле, зовнішній ключ) – посилання на таблицю Objects.

damage_type_id (ціле, зовнішній ключ) – посилання на таблицю DamageTypes.

field_id (ціле, зовнішній ключ) – посилання на таблицю Fields.

source_id (ціле, зовнішній ключ) – посилання на таблицю Sources.

area (десяткове число) – площа пошкодженої ділянки (га).

productivity (десяткове число) – врожайність (т/га).

sale_price (десяткове число) – ціна реалізації (грн/т).

cultivation_costs (десяткове число) – витрати на обробіток (грн/га).

fuel (десяткове число) – витрати палива (л/га).

lostsales (десяткове число) – втрати доходів (грн) – розрахункове поле.

Розрахунок втраченого прибутку

Розрахунок втрачених продажів є одним із ключових аспектів для оцінки економічного впливу пошкоджень на сільськогосподарські землі. Цей показник дає змогу кількісно визначити фінансові втрати, що виникають унаслідок зниження продуктивності земель чи їх повного виходу з обороту. Зокрема, втрати доходів враховують як недоотримані прибутки від реалізації продукції, так і додаткові витрати, пов'язані з відновленням постраждалих територій.

Розуміння обсягів втрачених продажів є критично важливим для фермерів, аграрних підприємств та органів державної влади. Це дозволяє оцінити реальні наслідки пошкоджень та планувати ефективні стратегії відновлення. Наприклад, на основі цього показника можна визначити пріоритетність фінансової допомоги, розробити плани компенсації втрат чи оптимізувати розподіл ресурсів для відновлення аграрного потенціалу.

Крім того, розрахунок втрачених продажів сприяє кращому розумінню економічних ризиків, пов'язаних із пошкодженнями сільськогосподарських земель. Він дозволяє підприємствам адаптувати свої бізнес-моделі, впроваджуючи інноваційні підходи до управління земельними ресурсами, використання страхових механізмів чи диверсифікації продукції. З іншого боку, державні органи можуть використовувати ці дані для формування політики підтримки аграрного сектору.

У роботі математична модель для розрахунку втрат доходів через пошкодження сільськогосподарських земель виглядає наступним чином (1):

$$L = A \times ((P \times S) - C - F) \quad (1)$$

де:

- L – втрати доходів (lostsales), \$;
- A – площа пошкодженої ділянки (area), га;
- P – врожайність культури (productivity), т/га;
- S – ціна реалізації культури (sale_price), грн/т;
- C – витрати на культивування (cultivation_costs), \$/га;
- F – витрати на паливо (fuel), \$/га.

У нашій базі даних поле lostsales розраховується за допомогою тригера. Тригер у базах даних — це спеціальна збережена процедура, яка автоматично виконується при настанні певної події, такої як вставка, оновлення або видалення записів у таблиці. Він дозволяє забезпечити цілісність даних, реалізувати бізнес-логіку на рівні бази даних та автоматизувати рутинні завдання. Згідно з дослідженням, тригери є ефективним інструментом для підтримки узгодженості даних та реалізації складних бізнес-правил безпосередньо в базі даних [47, 48].

Код тригера:

```
CREATE TRIGGER trg_CalculateLostSales
ON Damages
AFTER INSERT, UPDATE
AS
BEGIN
    SET NOCOUNT ON;

    UPDATE d
    SET lostsales = (d.productivity * d.sale_price - d.cultivation_costs - d.fuel) * d.area
    FROM Damages d
    INNER JOIN inserted i ON d.damage_id = i.damage_id;
END;
GO
```

2.3 Розробка інтерактивної платформи для візуалізації та аналізу даних

Архітектура системи

Система для аналізу збитків і прогнозування інвестицій побудована на основі модульної, масштабованої архітектури, що використовує відкриті технології. Її структура передбачає три основні рівні: фронтенд, рівень даних і рівень хостингу та розгортання. Такий підхід забезпечує зручність використання та можливість обробки складних наборів даних.

Фронтенд є основою взаємодії користувача із системою. Він реалізований за допомогою HTML5, CSS3 та JavaScript із використанням бібліотеки Leaflet.js для інтерактивних карт. Цей компонент відповідає за відображення карти, маркерів, спливаючих вікон та фільтрів.

Основні можливості фронтенду включають:

- Візуалізація карти: Інтерактивна карта, яка оновлюється залежно від дій користувача, відображаючи маркери пошкоджених об'єктів;
- Фільтри та пошук: Інструменти, які дозволяють фільтрувати маркери за типом пошкодження чи діапазоном втрат, а також знаходити конкретні об'єкти за назвою;
- Адаптивний дизайн: Платформа оптимізована для різних пристроїв, таких як комп'ютери, планшети та смартфони;
- Інтерактивні елементи: Клік на маркер відкриває деталі про об'єкт, а легенди пояснюють значення кольорового кодування.

Рівень даних відповідає за зберігання та обробку інформації, яка використовується для заповнення карти. Наразі платформа працює зі статичним набором даних, інтегрованим у код, але архітектура дозволяє легко підключати динамічні бази даних чи API:

- Статичне управління даними: Інформація про пошкоджені об'єкти зберігається в масиві JavaScript, включаючи координати, тип пошкодження та фінансові втрати;

- Обробка даних у реальному часі: Реалізовані функції для фільтрації та обробки даних на основі введених користувачем параметрів;
- Алгоритми кольорового кодування: Система автоматично визначає кольори маркерів та легенд на основі втрат (градієнт) або типу пошкодження (категорії).

Рівень хостингу та розгортання забезпечує доступ до платформи через Інтернет. Система розміщена на платформі GitHub Pages, що дозволяє зробити код доступним для публічного використання:

- Репозиторій GitHub: Репозиторій містить усі необхідні файли, включаючи HTML, CSS, JavaScript та додаткові ресурси;
- Контроль версій: GitHub забезпечує зручне відстеження змін та можливість командної роботи над проєктом;
- Інтеграція CDN: Використання зовнішніх бібліотек, таких як Leaflet.js, через мережі доставки контенту (CDN) забезпечує швидкодію та стабільність.

Архітектура системи орієнтована на модульність і простоту, що спрощує обслуговування та подальше розширення. Наприклад, рівень даних можна доповнити інтеграцією з API для отримання актуальної інформації в реальному часі, а фронтенд — розширити функціями геоаналітики. Рівень хостингу забезпечує стабільність роботи й доступність платформи для широкого кола користувачів, включаючи науковців, аналітиків та інвесторів.

Опис платформи

Платформа для аналізу збитків і прогнозування інвестицій розроблена для візуалізації даних про пошкоджені сільськогосподарські землі. Вона пропонує широкий спектр функціональних можливостей, що забезпечують зручну взаємодію користувача із системою та аналіз великих обсягів даних. Усі її можливості реалізовані через інтерактивний інтерфейс та продуману архітектуру.

Основним компонентом платформи є інтерактивна карта (Рисунок 2.3.1), яка дозволяє користувачам досліджувати дані про пошкоджені об'єкти. Маркери на карті динамічно оновлюються залежно від дій користувача, а також можуть змінювати кольори відповідно до вибраного режиму відображення – за рівнем втрат доходу чи типом пошкодження. Користувач може збільшувати або зменшувати масштаб карти та панорамувати її для детального перегляду конкретних територій.

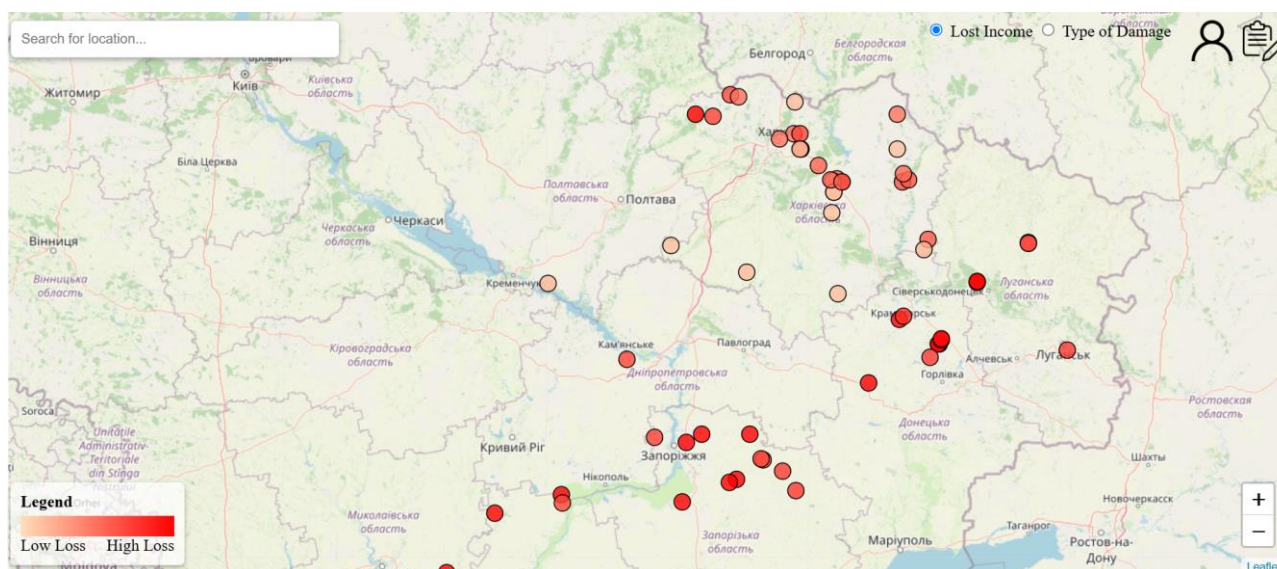


Рисунок 2.3.1 Платформа аналізу збитків

Для зручності аналізу даних платформа пропонує кілька інструментів фільтрації (Рисунок 2.3.2). Зокрема, користувачі можуть вибирати типи пошкоджень або діапазони втрат доходу. Фільтри працюють одночасно, що дозволяє комбінувати умови для пошуку потрібної інформації. Це зручно, коли потрібно відокремити певні категорії об'єктів або отримати вузьконаправлені результати.

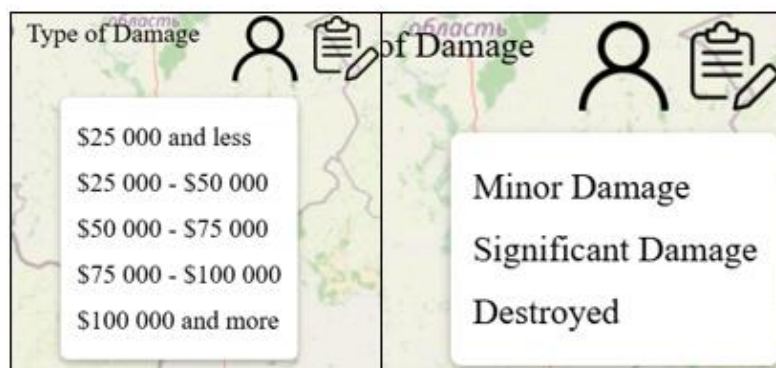


Рисунок 2.3.2 Фільтри по типу пошкодження та втрат

Окрім фільтрів, платформа оснащена функцією пошуку, яка допомагає швидко знаходити конкретні об'єкти за їх назвою (Рисунок 2.3.3). У процесі введення тексту в полі пошуку платформа пропонує відповідні результати, а вибір одного з них автоматично переміщує карту до відповідного маркера. Це значно скорочує час, необхідний для навігації великою кількістю об'єктів.

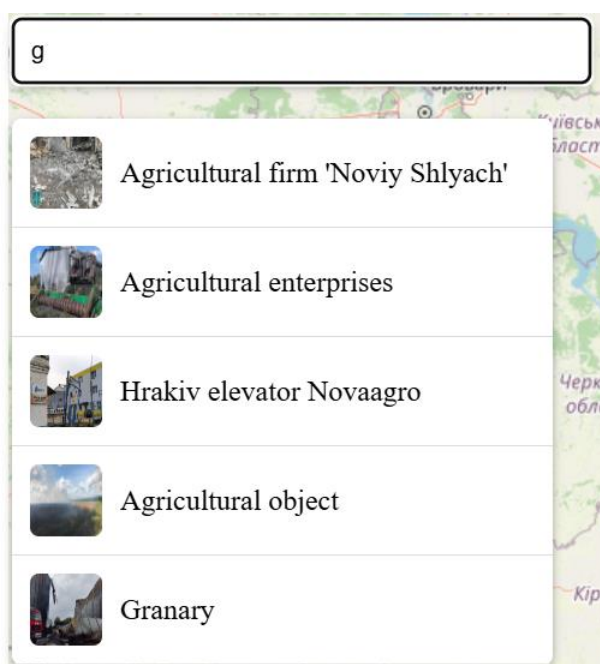


Рисунок 2.3.3 Пошукова система платформи

Інтерактивність карти посилюється спливаючими вікнами (поп-ап), які відображаються при кліку на маркер (Рисунок 2.3.4). У цих вікнах користувач може побачити назву об'єкта, його фото, тип пошкодження, суму втрат доходу, а також перейти за посиланням для отримання додаткової інформації. Це забезпечує доступ до детальних даних без необхідності покидати карту.

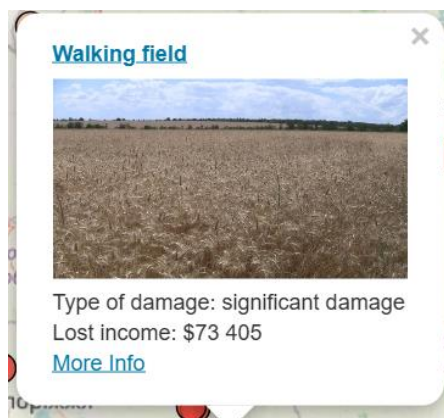


Рисунок 2.3.4 Спливаюче вікно (поп-ап)

Для полегшення розуміння інформації платформа пропонує дві інтерактивні легенди (Рисунок 2.3.5). Перша – це градієнт кольорів, який показує рівень втрат доходу: від низьких втрат (світло-помаранчевий) до високих (червоний). Друга – це кольорові позначення для типів пошкоджень: темно-зелений для мінімальних, помаранчевий для значних і червоний для знищених об’єктів. Легенда автоматично змінюється залежно від вибраного режиму відображення.

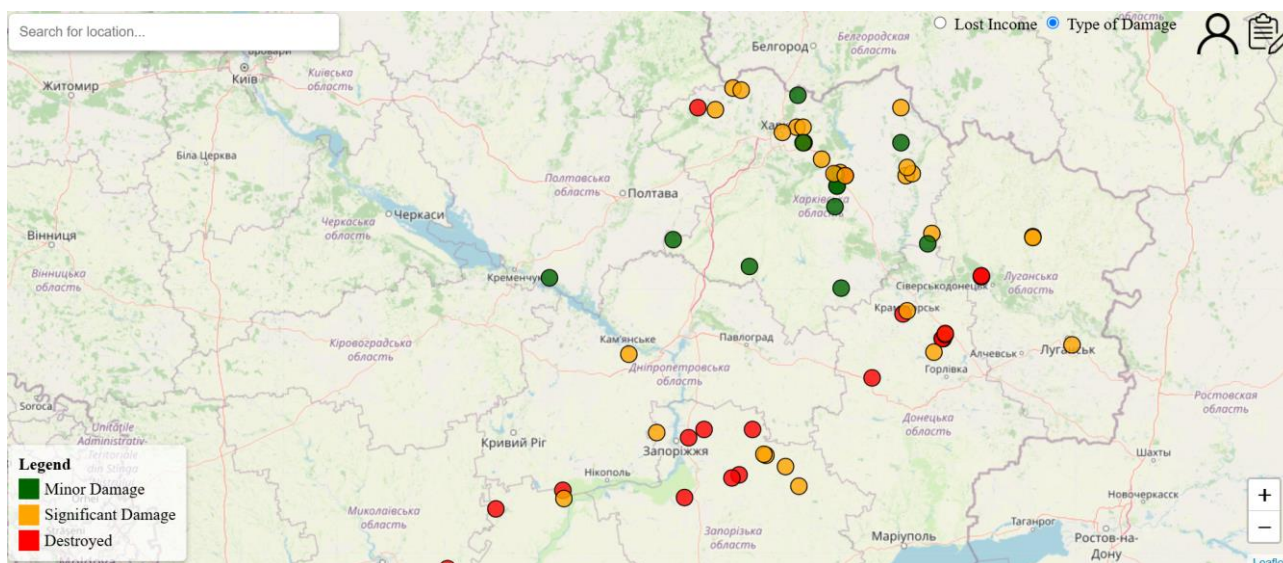


Рисунок 2.3.5 Платформа в режимі аналізу типів пошкодження

Важливою перевагою платформи є її адаптивний дизайн, що забезпечує зручність використання як на стаціонарних комп’ютерах, так і на мобільних пристроях. Інтерфейс автоматично підлаштовується під розміри екрана, гарантуючи, що всі функції залишаються доступними незалежно від типу пристрою (Рисунок 2.3.6).

Дані, що відображаються на платформі, можуть бути легко оновлені або розширені. Поточна реалізація працює зі статичним набором даних, інтегрованим у код, але система готова до інтеграції з динамічними джерелами, такими як API. Це дозволить автоматично додавати нові маркери на карту без додаткових змін у коді.

Платформа розміщена на GitHub Pages, що забезпечує високу швидкість завантаження, стабільність роботи та зручний доступ до коду для розробників. Кожен маркер пов'язаний із джерелом даних, забезпечуючи прозорість та достовірність інформації.

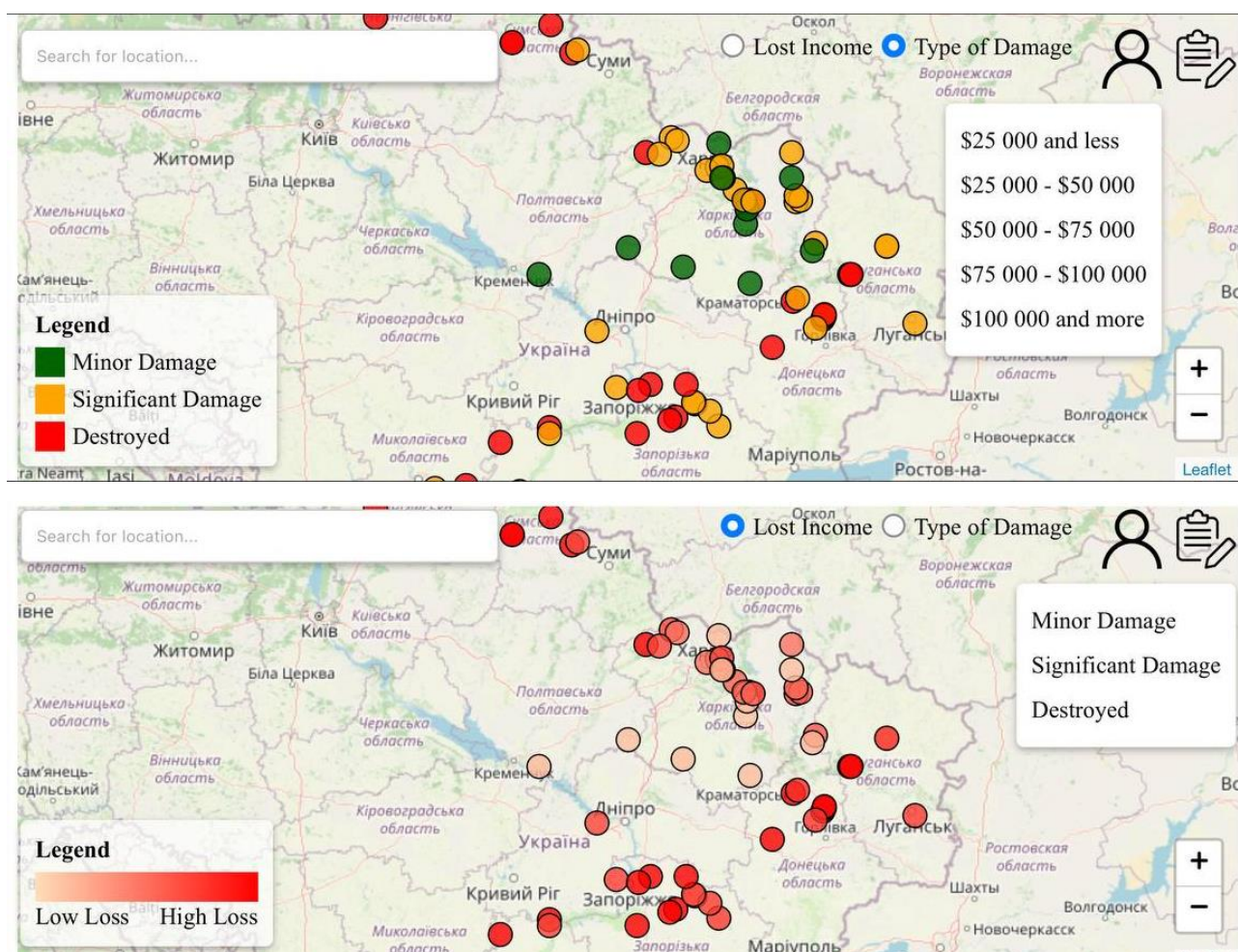


Рисунок 2.3.6 Мобільний інтерфейс

3 АНАЛІЗ ДАНИХ ЗА ДОПОМОГОЮ МЕТОДІВ МАШИННОГО НАВЧАННЯ

3.1 Використання кластерного аналізу: постановка задачі та опис алгоритму

Кластерний аналіз є одним із ключових методів для сегментації даних, що дозволяє групувати об'єкти з подібними характеристиками. У контексті оцінки збитків у сільськогосподарському секторі цей підхід може бути використаний для виділення груп регіонів або об'єктів із подібними типами пошкоджень, втратами прибутку чи іншими показниками. Це сприяє глибшому розумінню структури даних і допомагає приймати більш зважені рішення щодо відновлення та інвестування.

Постановка задачі

Основна мета кластерного аналізу в цьому проєкті – сегментувати дані про пошкоджені сільськогосподарські землі на основі таких параметрів:

- Втрачений прибуток: кількісна оцінка втрат для кожного об'єкта;
- Регіон: географічне розташування пошкоджень;
- Тип культури: які саме сільськогосподарські культури були пошкоджені;

– Тип пошкодження: категорія завданої шкоди (мінімальне, значне, повне знищення).

Це дозволить визначити групи з подібними характеристиками, які можуть бути корисними для пріоритетизації відновлювальних заходів або інвестицій у майбутньому.

Опис алгоритму:

Для виконання кластерного аналізу обрано метод K-Means. Цей алгоритм є одним із найбільш популярних для задач кластеризації завдяки своїй простоті та ефективності. Основні етапи його застосування в даному контексті (Рисунок 3.1.1):

1. підготовка даних;
2. визначення оптимальної кількості кластерів;
3. формування кластерів;
4. візуалізація результатів;
5. інтерпретація результатів.

Підготовка даних є ключовим етапом кластерного аналізу, який забезпечує точність і коректність результатів. Перш за все, з набору даних видаляються пропущені значення, щоб уникнути спотворення обчислень і зменшити вплив неповної інформації. Наступним кроком є кодування категорійних змінних, таких як регіони або типи пошкоджень, у числовий формат. Це необхідно для того, щоб алгоритм міг обробляти ці змінні. Для забезпечення рівного впливу всіх змінних на процес формування кластерів виконується стандартизація числових даних, завдяки якій показники з різними шкалами (наприклад, втрачений прибуток та категорії пошкоджень) мають однакову вагу у моделі.

На етапі визначення оптимальної кількості кластерів застосовується метод ліктя для аналізу внутрішньокластерної дисперсії (inertia) при різній кількості кластерів. Метод передбачає обчислення показника inertia для наборів із різною

кількістю кластерів і побудову графіка залежності. Оптимальне число кластерів обирається в точці, де подальше зменшення дисперсії стає менш вираженим, утворюючи так званий "лікоть". Це дозволяє забезпечити баланс між точністю сегментації даних і простотою інтерпретації результатів.

Формування кластерів здійснюється за допомогою алгоритму K-Means. Обрана кількість кластерів задається в моделі, після чого алгоритм ідентифікує центри кластерів і розподіляє об'єкти відповідно до їх близькості до цих центрів. Кожному об'єкту призначається мітка відповідного кластера. Таким чином, всі дані розподіляються між групами, кожна з яких має спільні характеристики.

Для аналізу та інтерпретації результатів кластеризації будується серія графіків. Наприклад, графіки розподілу об'єктів дозволяють візуалізувати взаємозв'язок між ключовими змінними, такими як втрачений прибуток і регіони. Додатково створюються графіки для аналізу типів культур і пошкоджень. Гістограми та коробкові графіки (boxplots) дають змогу досліджувати розподіл характеристик у кожному кластері, що допомагає ідентифікувати унікальні особливості кожної групи.

На етапі інтерпретації аналізуються середні значення змінних у кожному кластері для виявлення їх характерних особливостей. Для більш глибокого розуміння будується теплова карта кореляцій між змінними всередині кожного кластеру. Це дозволяє оцінити, як різні фактори взаємодіють між собою в межах кожної групи, і виявити потенційні закономірності. Інтерпретація результатів є основою для формування стратегій відновлення та інвестування.

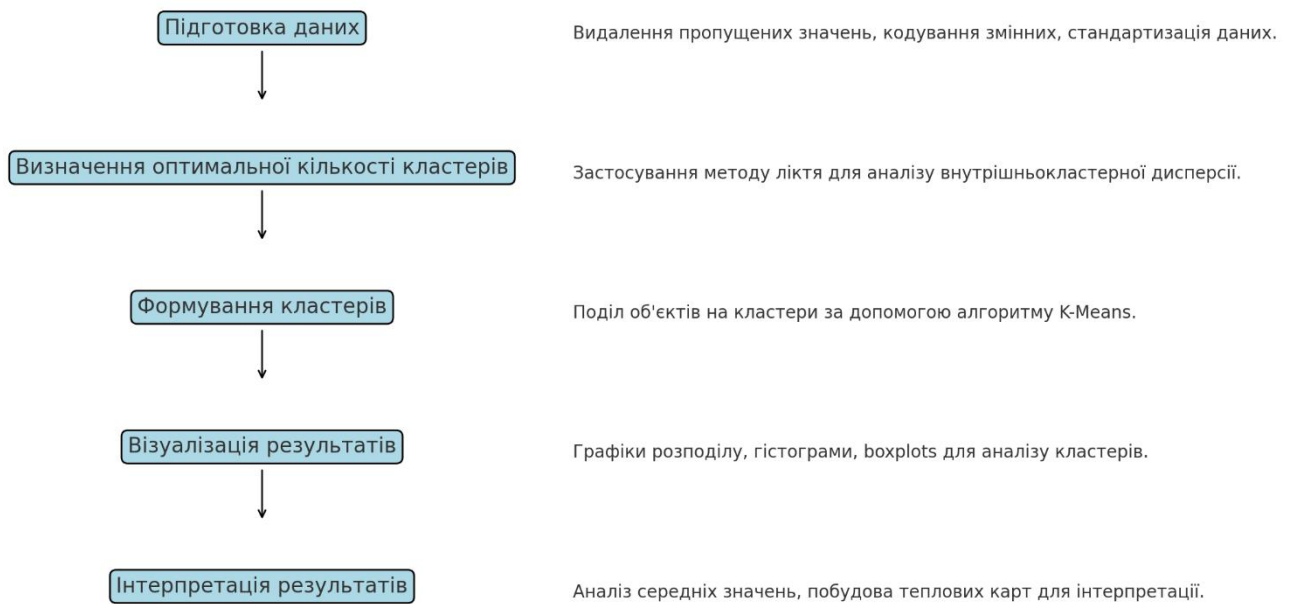


Рисунок 3.1.1 Етапи кластерного аналізу

3.2 Результати кластеризації та їх інтерпретація

Кластеризація даних є ключовим етапом у процесі аналізу, який дозволяє сегментувати великі набори даних на основі спільних характеристик.

На етапі підготовки було відібрано основні змінні, які найкраще описують об'єкти аналізу. У кластеризацію включені наступні змінні:

- `lostsales` – числова змінна, що відображає фінансові втрати;
- `region` – закодована категорійна змінна, яка представляє географічну область;
- `culture` – закодована змінна, що відображає вид вирощуваних культур;
- `type_of_damage` – закодована категорія, яка вказує на ступінь пошкодження (мінімальне, значне, повне знищення).

Ці змінні обрано з огляду на їх важливість для аналізу та кластеризації, оскільки вони формують основні характеристики об'єктів.

Для того, щоб всі змінні мали однаковий вплив на процес кластеризації, використовується стандартизація. Це забезпечує приведення змінних до однакової шкали, незалежно від їхніх початкових діапазонів значень.

StandardScaler від бібліотеки scikit-learn приводить значення кожної змінної до середнього 0 та стандартного відхилення 1. Це означає, що втрачений прибуток, закодовані регіони, культури та типи пошкоджень не будуть домінувати в аналізі лише через свої числові масштаби. Стандартизовані дані зберігаються у змінній `scaled_data` і використовуються для подальшої кластеризації.

На наступному етапі використовується метод ліктя для визначення оптимальної кількості кластерів для кластеризації даних. Створено список `inertia` для зберігання значень внутрішньокластерної дисперсії для кожного числа кластерів. Діапазон можливих кількостей кластерів визначено від 1 до 10 (`range_clusters`).

Далі запускаємо алгоритм K-Means для різної кількості кластерів k . Кожен запуск визначає центри кластерів і обчислює "inertia" – показник внутрішньокластерної дисперсії, що є сумою квадратів відстаней від точок до центрів кластерів. Значення `inertia` додається до списку.

Побудовано графік залежності `inertia` від кількості кластерів (Рисунок 3.2.1). Для кожного значення k графік показує, наскільки добре алгоритм розділив дані. Мітки осей пояснюють, що по горизонталі відкладено кількість кластерів, а по вертикалі – внутрішньокластерну дисперсію.

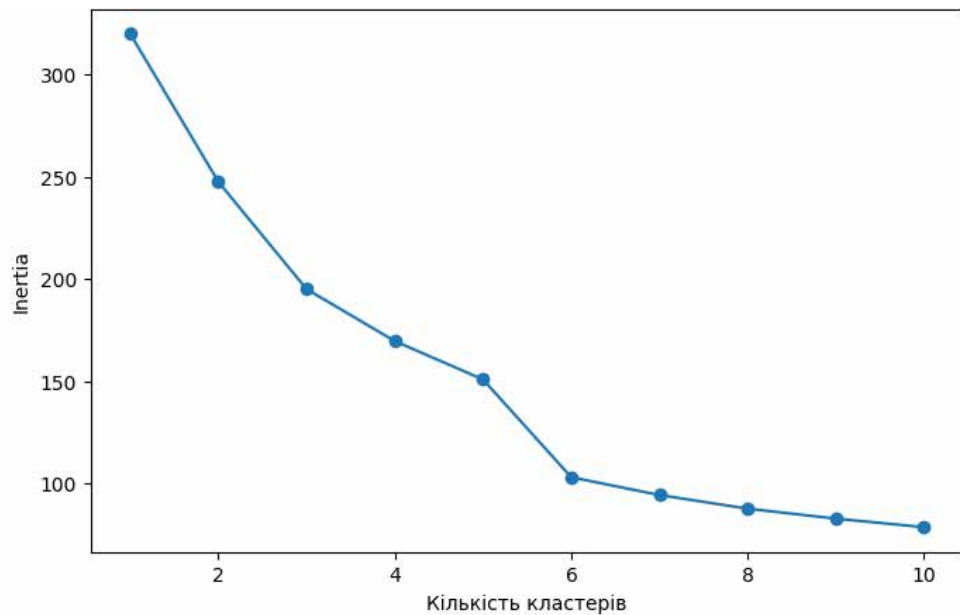


Рисунок 3.2.1 Метод ліктя

Графік показує, як "inertia" зменшується зі збільшенням кількості кластерів. У початкових точках спостерігається значне зменшення дисперсії, але після певного моменту (приблизно при 3-5 кластерах) зменшення стає менш помітним. Цей момент називається "лікоть" графіка і є найкращим вибором кількості кластерів, оскільки додача нових кластерів після цієї точки дає мінімальний вигравш у сегментації.

Вибір оптимального числа кластерів за методом ліктя дозволяє збалансувати точність кластеризації та простоту інтерпретації результатів. У нашому випадку оптимальна кількість кластерів знаходиться в діапазоні 3-5, що буде використано на наступних етапах для формування кластерів та їх інтерпретації.

Наступний етап кластерного аналізу полягає у виборі оптимальної кількості кластерів, формуванні кластерів за допомогою алгоритму K-Means, та збереженні результатів у початковому наборі даних:

- Вибір кількості кластерів: Значення `optimal_clusters` встановлюється як 3, оскільки це оптимальне значення, яке забезпечує баланс між деталізацією кластеризації та простотою інтерпретації (визначене раніше методом ліктя);

- Алгоритм K-Means: Ініціалізація моделі K-Means виконується із заданою кількістю кластерів. Аргумент `random_state=42` використовується для забезпечення відтворюваності результатів;

- Присвоєння міток кластерів: Метод `fit_predict` не лише навчає модель, а й автоматично присвоює кожному об'єкту в наборі даних мітку відповідного кластера. Результат зберігається в новій колонці `Cluster_All` початкового `DataFrame data`.

Після формування кластерів було виконано декодування змінних, які раніше кодувалися в числовий формат. Це робиться для полегшення інтерпретації результатів:

- Перебір змінних: У циклі `for` перевіряються змінні `region`, `culture` та `type_of_damage`, які були закодовані на етапі підготовки даних;

- Декодування: Змінна декодується за допомогою методу `inverse_transform` відповідного енкодера, який був збережений раніше. Це перетворює числові значення назад у текстові категорії;

- Створення нових колонок: Для кожної змінної створюється нова колонка (наприклад, `region_Decoded`), що містить текстові значення, зрозумілі для подальшого аналізу.

Формування кластерів дозволяє кожному об'єкту отримати свою кластерну ідентифікацію, яка буде використовуватися для інтерпретації результатів. Декодування текстових змінних спрощує подальшу роботу з даними, оскільки текстові категорії більш зручні для аналізу, зокрема під час побудови графіків чи опису характеристик кожного кластера. Цей етап є важливим для переходу від суто технічного до аналітичного представлення результатів кластеризації.

Далі відбувається візуалізація результатів кластеризації для полегшення аналізу груп даних. Візуалізація здійснюється за допомогою графіків розсіювання (`scatterplots`), які показують розподіл об'єктів у кожному кластері за різними змінними. Два основні графіки ілюструють:

1. Залежність втраченого прибутку від регіонів;
2. Взаємозв'язок між культурами та типами пошкоджень.

Перший графік ілюструє розподіл втраченої прибутку серед різних регіонів і групування об'єктів у кластери (Рисунок 3.2.2). На осі X відображено значення втраченої прибутку (lostsales), що є числовою змінною, а на осі Y – назви регіонів (region_Decoded), представлені у текстовому форматі. Кожна точка на графіку відповідає окремому об'єкту з даних, а її колір вказує на кластер, до якого належить цей об'єкт. Завдяки палітрі viridis забезпечено чітке розрізнення між кластерами.

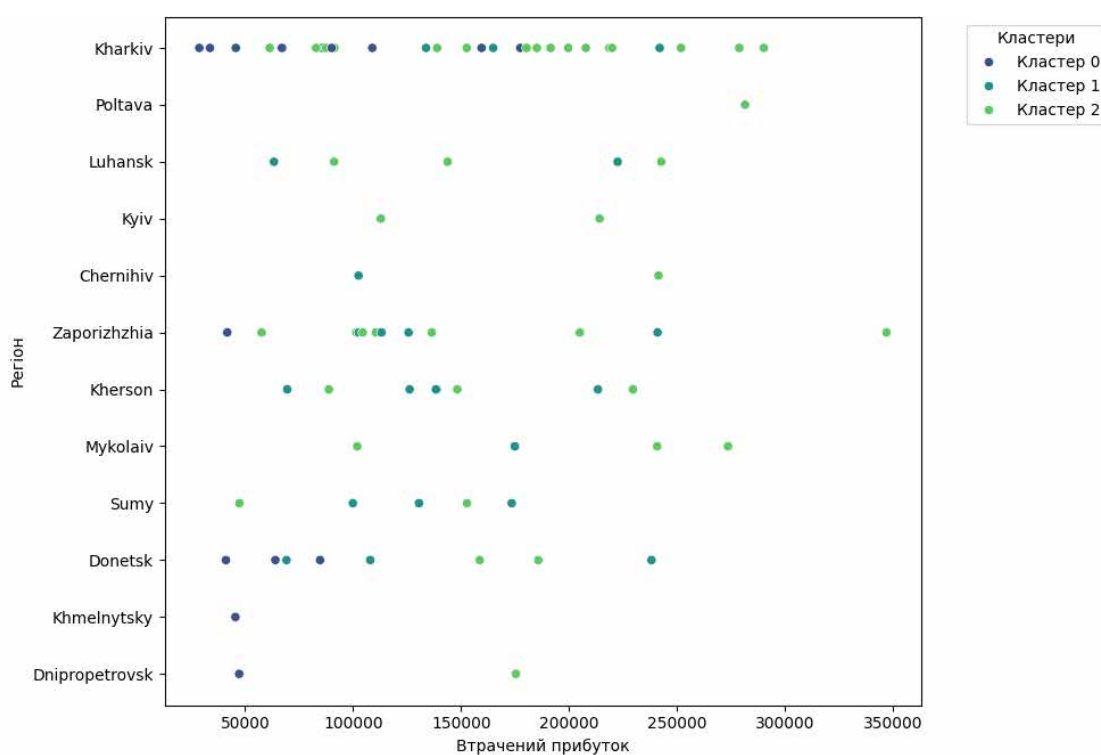


Рисунок 3.2.2 Кластери на площині: Втрачений прибуток vs Регіон

Аналізуючи графік, можна побачити, як регіони з подібними втратами прибутку групуються в межах одного кластеру. Наприклад, для певних регіонів втрати значно вищі, що вказує на серйозні пошкодження, і вони, як правило, потрапляють до одного кластеру. Інші регіони з меншими втратами утворюють іншу групу:

– Регіональні відмінності у втратах: Різні регіони демонструють значні відмінності у рівні втраченої прибутку. Наприклад, регіони з високими втратами,

такі як Харків і Запоріжжя, групуються в одному кластері, що свідчить про подібний рівень руйнувань у цих областях;

- Виділення зон з мінімальними втратами: Регіони з низьким рівнем втрат, такі як Хмельницький чи Полтава, належать до іншого кластеру. Це дозволяє ідентифікувати області, які зазнали відносно менших збитків;

- Логічність кластеризації: Графік підтверджує, що кластеризація відображає реальні закономірності у даних. Регіони зі схожими характеристиками втрат (високі чи низькі) належать до одного кластеру;

- Пріоритизація відновлення: Регіони з найбільшими втратами можна ідентифікувати як критичні зони для першочергових заходів з відновлення.

Другий графік демонструє взаємозв'язок між культурами, які вирощуються, і типами пошкоджень, завданих цим культурам (Рисунок 3.2.3).

На осі X розташовані типи культур (culture_Decoded), такі як пшениця, соняшник, кукурудза тощо, а на осі Y – категорії пошкоджень (type_of_damage_Decoded), наприклад, "minor damage", "significant damage" або "destroyed". Кожна точка відповідає одному об'єкту з набору даних, а її колір визначає кластер.

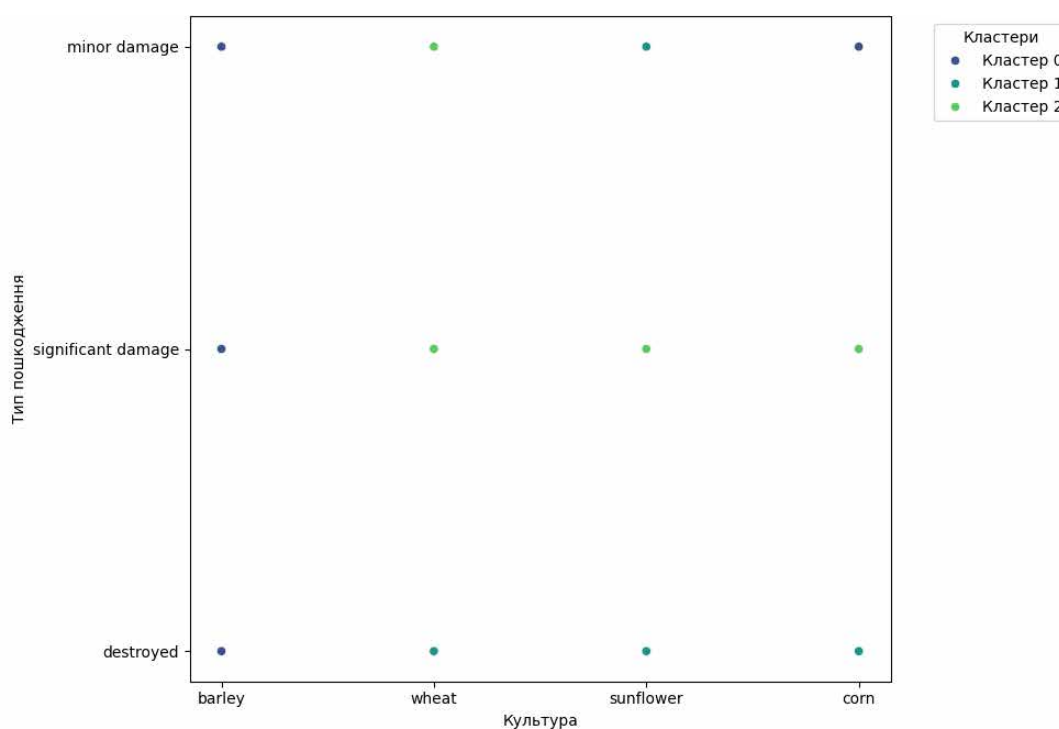


Рисунок 3.2.3 Кластери на площині: Культура vs Тип пошкодження

Цей графік дозволяє побачити, які культури найбільше постраждали від певних типів пошкоджень:

- Тип пошкоджень для культур: Деякі культури, наприклад, соняшник, частіше зазнають значних пошкоджень або повного знищення, що може бути пов'язано з їхньою вразливістю чи специфікою регіонів, де вони вирощуються;
- Групування культур за типами пошкоджень: Графік показує, що культури з подібними типами пошкоджень належать до одного кластеру. Наприклад, культури з переважно мінімальними пошкодженнями (наприклад, ячмінь) належать до окремого кластеру;
- Закономірності між культурами та пошкодженнями: Виявляється, що певні культури більш схильні до повного знищення (наприклад, пшениця), тоді як інші частіше отримують незначні пошкодження;
- Інформація для пріоритетів відновлення: Графік допомагає визначити культури, які найбільше постраждали, і враховувати це при розробці стратегій відновлення чи плануванні інвестицій у відновлення сільськогосподарського сектору.

Наступним кроком було побудовано гістограми, які ілюструють розподіл об'єктів у кластерах за різними змінними. Гістограми дозволяють зрозуміти, як значення змінних, таких як втрачений прибуток, регіон, культура або тип пошкодження, розподілені між різними групами (кластерами).

Основними функціями коду є створення циклу для побудови гістограм і забезпечення належної візуалізації результатів. У циклі перебираються змінні, такі як `lostsales`, `region_Decoded`, `culture_Decoded` та `type_of_damage_Decoded`, і для кожної змінної будується окрема гістограма. Гістограми створюються за допомогою функції `sns.histplot`, що дозволяє візуалізувати дані кластерів у вигляді стекових гістограм (`stacked histograms`).

Для чіткого розрізнення кластерів використовується палітра `viridis`, яка забезпечує яскраве та зрозуміле кольорове відображення. Додатково додаються

підписи осей (xlabel, ylabel) і заголовки для кожного графіка, що полегшує їх інтерпретацію. Легенда, розташована поряд із графіками, чітко вказує, який колір відповідає певному кластеру, забезпечуючи зручність аналізу.

Перший графік вказує на розподіл значень втраченого прибутку по кластерах ().

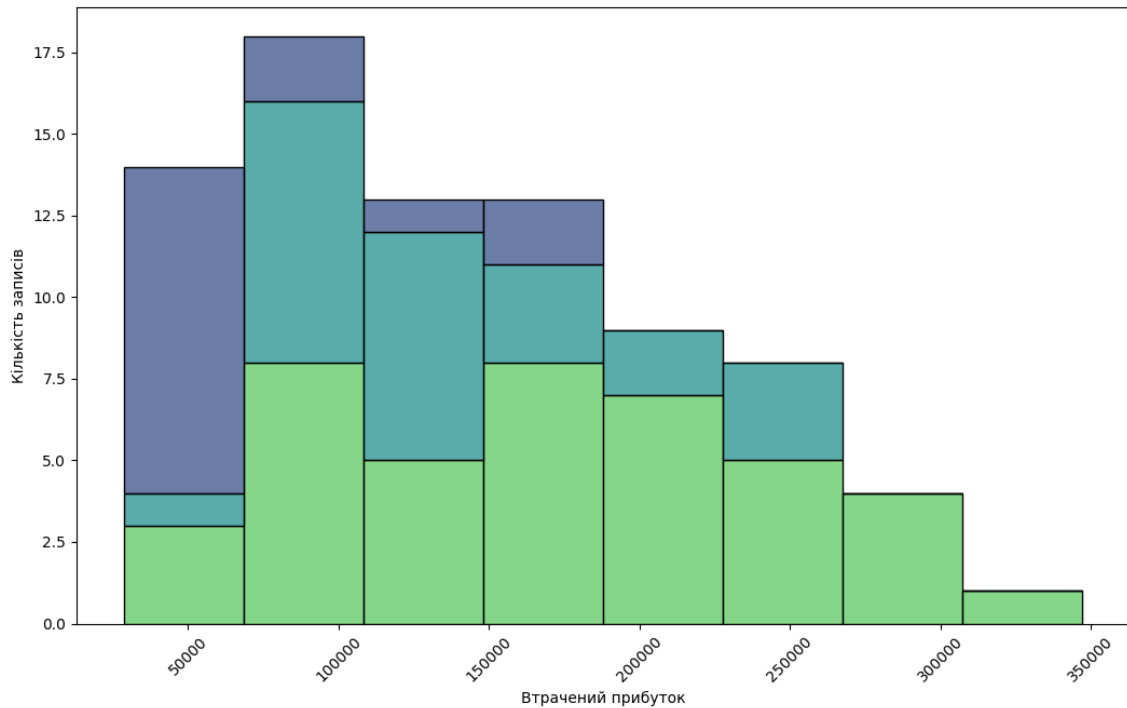


Рисунок 3.2.4 Розподіл значень втраченого прибутку по кластерах

Гістограма показує, як значення втраченого прибутку розподілені між різними кластерами. Вісь X відображає розмір втрат (у грошовому вираженні), а вісь Y — кількість записів у вибірці. Основні висновки:

- Великі втрати у певних кластерах: Найбільші втрати (понад 200 000) спостерігаються переважно в одному кластері, що свідчить про значні руйнування в цих об'єктах;
- Середні значення: Об'єкти з втратами в межах 100 000 – 200 000 розподілені між кількома кластерами, що може вказувати на спільні характеристики, але різні масштаби пошкоджень;
- Кластери з меншими втратами: Записи з втратами до 50 000 переважно належать до іншого кластеру, що відповідає об'єктам із мінімальними пошкодженнями.

Другий графік показує розподіл регіонів у кластерах (Рисунок 3.2.5).

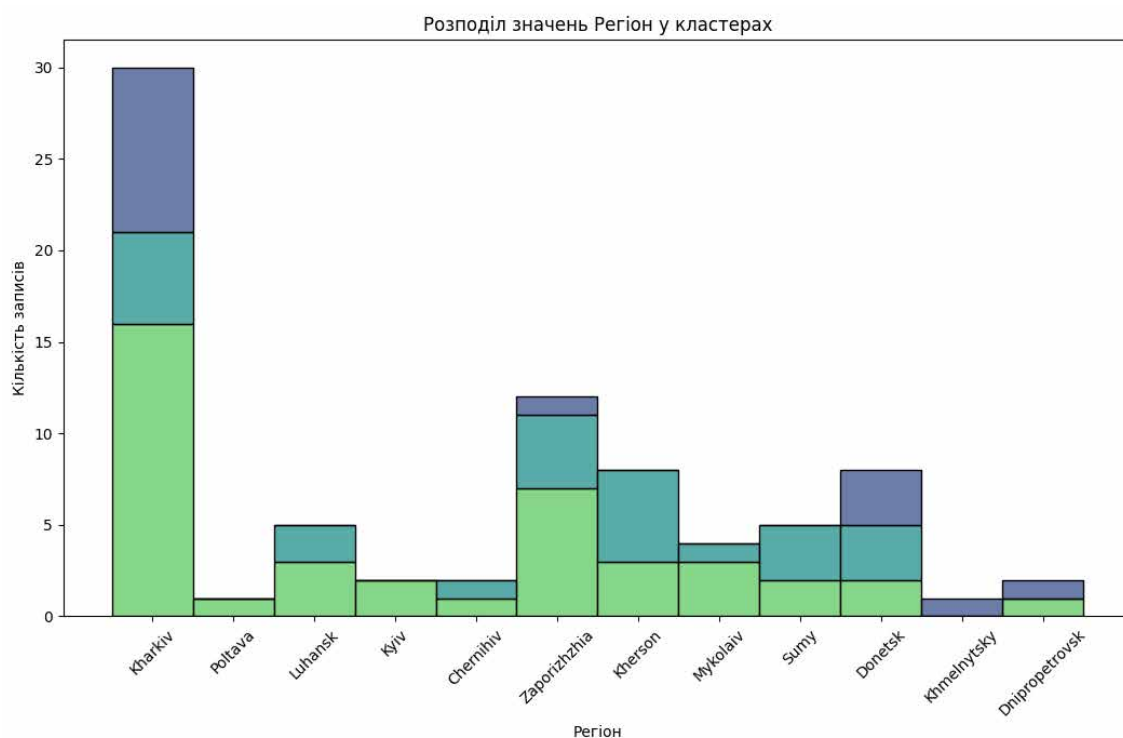


Рисунок 3.2.5 Розподіл значень регіонів по кластерах

Цей графік відображає, які регіони входять до кожного кластеру. Вісь X представляє регіони, а вісь Y — кількість записів у кожному кластері. Основні висновки:

- Виділення регіонів із найбільшими втратами: регіон Харків домінує в одному з кластерів, що підтверджує його значну уразливість і високі втрати. Запоріжжя та Донецьк також мають значну кількість записів, які належать до кластерів із великими втратами;
- Полтава, Хмельницький та Дніпропетровськ представлені переважно у кластерах із меншими втратами, що вказує на відносну стабільність у цих областях;
- Деякі регіони, наприклад, Миколаїв і Суми, містять записи з різних кластерів, що може вказувати на неоднорідність пошкоджень навіть у межах одного регіону.

Коробковий графік ілюструє розподіл втраченої виручки (lostsales) у межах трьох кластерів (Рисунок 3.2.6).

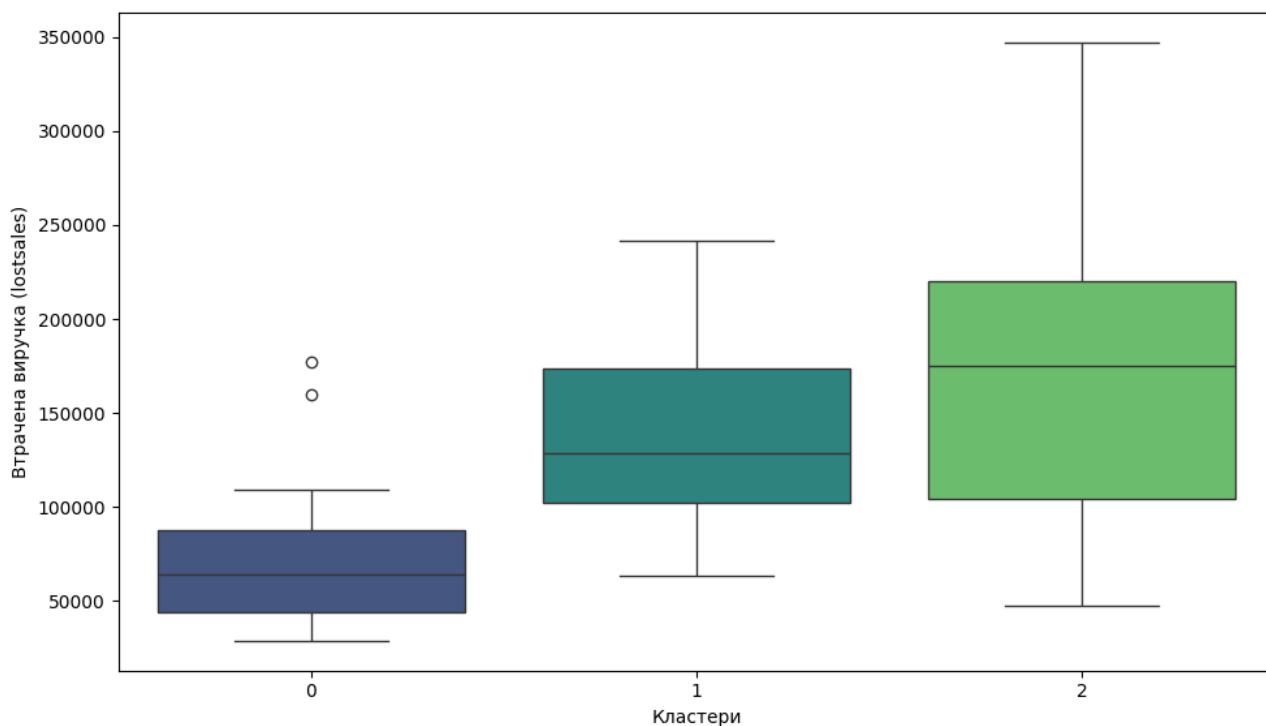


Рисунок 3.2.6 Розподіл втраченого прибутку у межах кластерів

Боксплот містить такі елементи:

- медіана – лінія всередині кожного прямокутника показує середнє значення втрат для кожного кластеру;
- квартилі – прямокутник охоплює міжквартильний діапазон (25%–75%), відображаючи основний розподіл значень;
- вуса – лінії, що виходять із прямокутників, показують діапазон значень за межами міжквартильного діапазону, але не враховують викиди;
- викиди – окремі точки поза вусами представляють аномальні значення (дуже великі або малі втрати).

Кластер 0 характеризується найменшим рівнем втрат, з медіанним значенням близько 50 000. Його міжквартильний діапазон є досить вузьким, що свідчить про низьку варіативність втрат у цій групі. Це означає, що більшість об'єктів у кластері мають подібні втрати, які не перевищують значних показників. Проте в цьому кластері є кілька викидів із втратами понад 100 000 \$, які вказують на окремі випадки з аномально великими збитками.

Кластер 1 демонструє середній рівень втрат із медіаною близько 150 000 \$. Порівняно з кластером 0, цей кластер має ширший міжквартильний діапазон, що вказує на більшу варіативність значень. У цьому кластері спостерігається помітний розкид втрат серед об'єктів, що може свідчити про більш різноманітні умови або фактори, які впливають на рівень втрат.

Кластер 2 є групою з найвищими втратами, медіана яких перевищує 200 000 \$. Цей кластер має найбільший міжквартильний діапазон, що свідчить про суттєві відмінності між рівнями втрат у межах групи. Широкий діапазон значень вказує на те, що кластер об'єднує об'єкти з дуже високими втратами, які варіюються від помірних до максимальних показників. Це робить кластер 2 критично важливим для детального аналізу та подальших заходів із підтримки та відновлення.

Окремо проаналізовано розподіл втраченого прибутку за площею у межах кластерів (). Цей графік відображає взаємозв'язок між площею об'єктів (area) та втратою виручки (lostsales) у межах трьох кластерів. Точки на графіку позначають окремі об'єкти, а їх колір відповідає кластерній належності (кластер 0, 1 або 2). Горизонтальна синя пунктирна лінія вказує на медіану втраченої виручки, а вертикальна червона пунктирна лінія показує медіану площі.

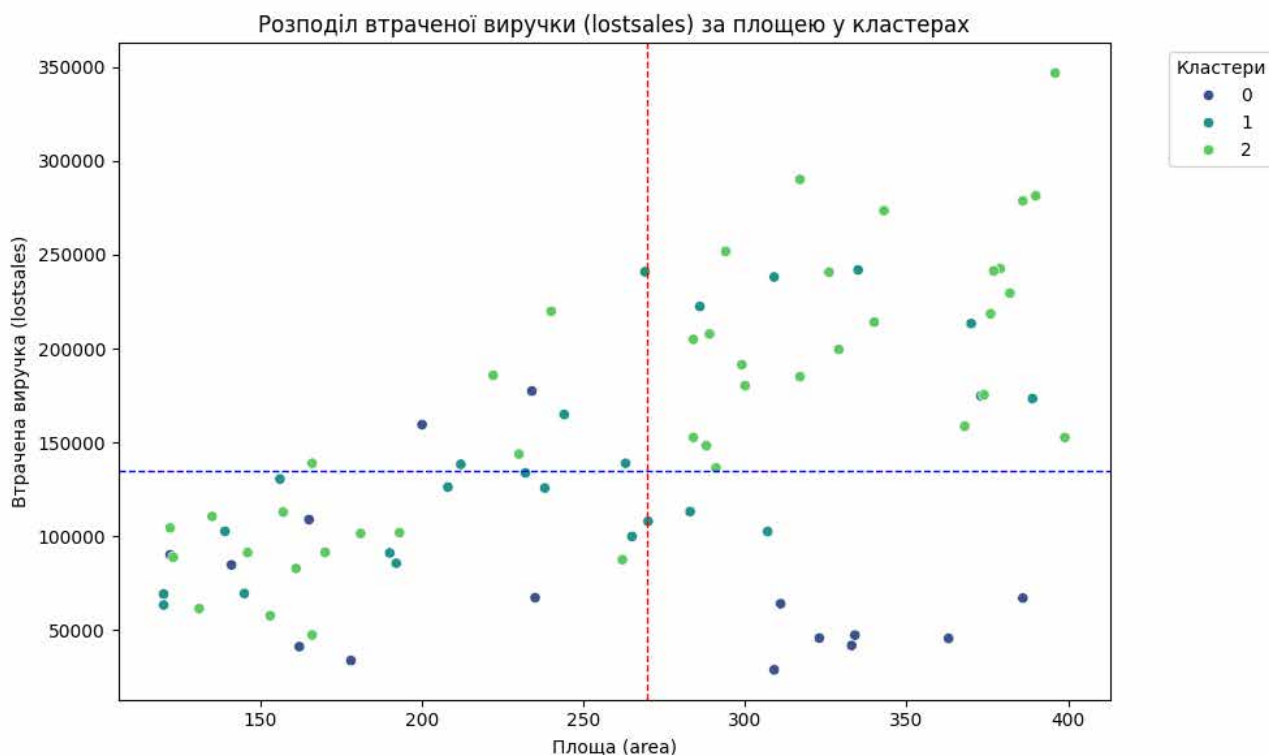


Рисунок 3.2.7 розподіл втраченого прибутку за площею у кластерах

Можна зробити наступні висновки:

- Кластер 0 переважно включає об'єкти з меншими площами та невеликими втратами виручки, які зазвичай не перевищують 100 000 \$. Ці об'єкти розташовані в нижній лівій частині графіка, що свідчить про їхні скромні розміри та мінімальні збитки. Такий профіль кластера може вказувати на меншу економічну значущість або вищий рівень захищеності об'єктів;
- Об'єкти кластеру 1 демонструють ширший розподіл як за площею, так і за втратами виручки, але їхні значення концентруються навколо середніх показників. Середній рівень площі об'єктів у цьому кластері становить близько 250–300 га, а втрати виручки коливаються близько 150 000 \$. Такий розподіл свідчить про те, що кластер включає об'єкти середнього масштабу з помірними економічними втратами.
- Кластер 2 характеризується наявністю об'єктів із великими площами і значними втратами виручки, що перевищують 200 000 \$. Об'єкти цього кластеру домінують у верхньому правому куті графіка, що підкреслює їх зв'язок

із найбільш значними збитками. Такі об'єкти, ймовірно, мають високу економічну важливість і потребують першочергової уваги.

Загалом на графіку спостерігається тенденція: об'єкти з більшою площею частіше пов'язані з більшими втратами виручки, що особливо характерно для кластеру 2. У кластері 0 площа об'єктів і втрати залишаються низькими, що може свідчити про кращу захищеність об'єктів або їх менший вплив на економіку. Кластер 0 характеризується мінімальними площами та втратами, що свідчить про низький рівень ризику. Кластер 1 включає об'єкти середнього масштабу з помірними втратами, що робить його проміжною групою. Кластер 2 об'єднує об'єкти з найбільшими площами та втратами, які мають значний вплив на загальні економічні показники.

Останній етап спрямований на аналіз внутрішніх взаємозв'язків між змінними в межах кожного кластеру. Для цього було побудовано теплові карти кореляції, які показують, як різні змінні (втрачений прибуток, регіон, культура, тип пошкодження) взаємодіють одна з одною.

У кластері 0 сильна позитивна кореляція між втратами (lostsales) і культурою (culture) (Рисунок 3.2.8).

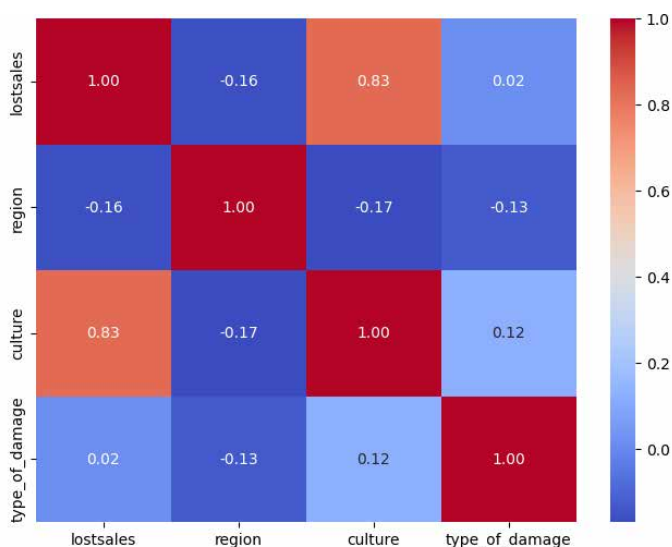


Рисунок 3.2.8 Кореляція змінних у кластері 0

Це вказує на те, що певні культури більш схильні до великих фінансових втрат у цьому кластері. Можливо, це культури, які вирощуються у регіонах із високим ризиком пошкоджень.

Низька кореляція між типом пошкодження (`type_of_damage`) і втратами свідчить про те, що тип пошкодження не є головним фактором, що впливає на розмір втрат у цьому кластері.

У кластері 1 об'єкти з більшими втратами частіше асоціюються з менш серйозними пошкодженнями (Рисунок 3.2.9). Це може бути викликано специфікою кластеру, наприклад, включенням великих підприємств, які зазнали мінімальних пошкоджень.

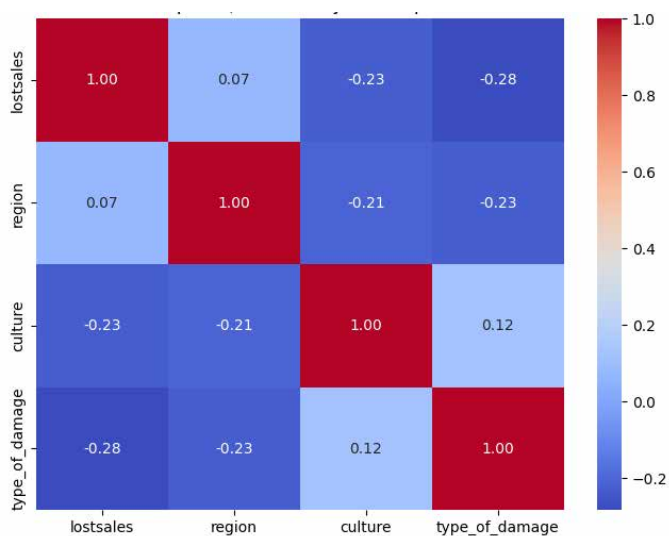


Рисунок 3.2.9 Кореляція змінних у кластері 1

Також спостерігається слабкий зв'язок між культурами та регіонами. Це може свідчити про те, що географічний фактор не є домінуючим для розподілу культур у межах цього кластеру.

У кластері 2 сильна негативна кореляція між втратами та культурами. певні культури пов'язані з меншими втратами (Рисунок 3.2.10). Це може вказувати на те, що ці культури вирощуються в регіонах з низьким ризиком пошкоджень. Також слабка взаємозалежність між іншими змінними. Тип пошкодження має незначний вплив на втрати, що може свідчити про різноманітний склад кластеру.

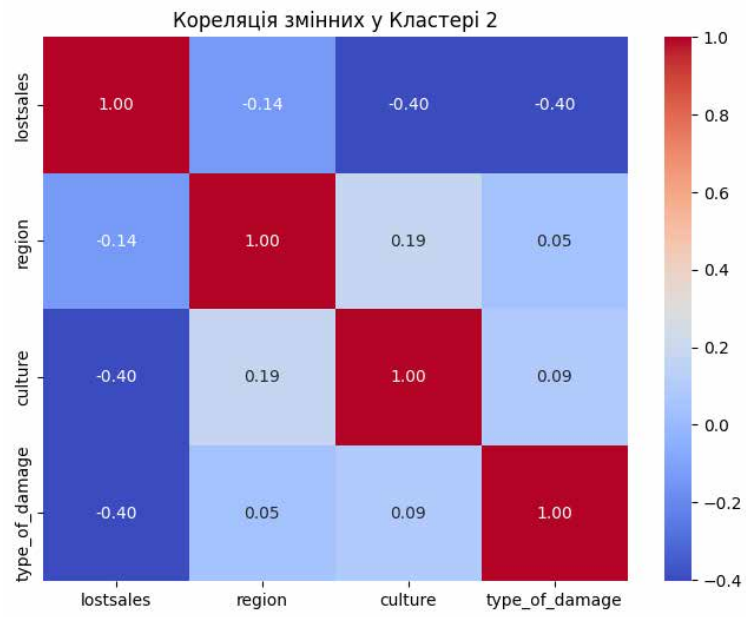


Рисунок 3.2.10 Кореляція змінних у кластері 2

ВИСНОВКИ

У результаті виконання магістерської роботи було розроблено інтерактивну платформу для аналізу збитків і прогнозування інвестицій для відновлення пошкоджених сільськогосподарських земель із використанням методів машинного навчання. Завдання роботи виконано в повному обсязі, і всі поставлені цілі досягнуто. Підсумки дослідження охоплюють технічні, наукові та практичні аспекти, що дозволяє виділити низку значущих результатів.

У рамках роботи було створено інтерактивну карту, яка дозволяє візуалізувати пошкоджені сільськогосподарські ділянки на основі супутникових знімків, відкритих даних та даних польових обстежень. Користувачі платформи отримали можливість виконувати фільтрацію інформації за такими критеріями, як тип пошкоджень, рівень втрат доходу та інші параметри. Інтуїтивний інтерфейс забезпечує зручність навігації, а можливість роботи з платформою на різних пристроях підвищує її універсальність. Це сприяє ефективнішій оцінці пошкоджень та плануванню заходів з відновлення.

У роботі застосовано методи кластерного аналізу, які дозволили сегментувати пошкоджені об'єкти за ключовими характеристиками, такими як тип пошкоджень, регіональні особливості та рівень втрат. Використання алгоритму K-Means дало змогу ідентифікувати закономірності у даних та виявити критичні зони, які потребують негайного втручання. Це підхід дозволяє адаптивно використовувати отриману інформацію для різних рівнів планування – від локальних до національних стратегій.

Проведено детальний аналіз економічних втрат, спричинених пошкодженнями сільськогосподарських земель. На основі даних про втрати врожаю, пошкоджені активи та зруйновану інфраструктуру було обчислено загальні фінансові збитки для різних регіонів і типів сільськогосподарських культур. Ці дані стали основою для розробки рекомендацій щодо пріоритетного

розподілу інвестицій. Моделі, розроблені в рамках роботи, дозволяють точніше оцінювати як поточні, так і потенційні втрати.

Інтеграція геоінформаційних систем (ГІС), супутникових знімків та алгоритмів машинного навчання стала важливим досягненням цієї роботи. ГІС забезпечили точність просторового аналізу, тоді як методи машинного навчання дозволили обробляти великі обсяги даних із високою точністю. Поєднання цих інструментів дозволило отримати інноваційне рішення, яке може бути застосоване у ширшому контексті, включаючи оцінку збитків від природних катастроф або техногенних аварій.

Запропонована платформа має широкий спектр застосувань. Її функціонал може бути використаний для:

- прийняття рішень щодо пріоритетів відновлення;
- підтримки фермерів та аграрних підприємств у визначенні зон високого ризику;
- інвесторів, зацікавлених у відновленні аграрного сектора, для оцінки доцільності капіталовкладень.

Основний науковий внесок цієї роботи полягає у розробці системи, яка поєднує просторовий аналіз із економічними моделями. Застосування кластерного аналізу для оцінки збитків дозволило виявити приховані закономірності, що недоступні при використанні традиційних підходів. Інтерактивна платформа з інтеграцією машинного навчання є новим рішенням для аналізу складних проблем у сільськогосподарському секторі.

Отримані результати відкривають можливості для подальшого вдосконалення платформи. Наступні кроки можуть включати:

- інтеграцію моделей глибокого навчання для прогнозування збитків і довгострокових інвестиційних потреб;
- додавання функцій аналізу часових рядів для оцінки змін у показниках збитків із часом;

– створення адаптивних рекомендацій для різних категорій користувачів – державних органів, приватних інвесторів та міжнародних організацій.

Розроблена платформа є ефективним інструментом для аналізу, візуалізації та прогнозування економічних втрат у сільському господарстві. Отримані результати мають як наукове, так і практичне значення, сприяючи вдосконаленню підходів до оцінки збитків та відновлення пошкоджених територій. Використання сучасних технологій у цій роботі демонструє великий потенціал для подальшого розвитку аграрного сектору України в умовах сучасних викликів.

СПИСОК ДЖЕРЕЛ

1. Ozdogan, M., Yang, Y., Allez, G., & Cervantes, C. (2010). Remote sensing of irrigated agriculture: Opportunities and challenges. *Remote Sensing of Environment*, 112(8), 2318–2326.
2. Liu, Z., Wang, T., & Wang, F. (2018). Crop damage detection using deep learning and remote sensing. *International Journal of Remote Sensing*, 39(7), 2303–2319.
3. Zhang, H., Wang, J., & Liang, S. (2021). Predicting soil contamination using machine learning algorithms: A case study of heavy metal pollution. *Environmental Research*, 200, 111297.
4. Lobell, D. B., Schlenker, W., & Costa-Roberts, J. (2011). Climate trends and global crop production since 1980. *Science*, 333(6042), 616–620.
5. Turner, B. L., & Meyer, W. B. (1994). Global land-use and land-cover change: An overview. *Environmental Issues*, 4(3), 65–77.
6. Schouten, M., Braam, M., & Wegner, T. (2020). Assessing conflict-related damage to agriculture: Insights from Syria. *Agricultural Systems*, 180, 102756.
7. Araujo, R., van der Gaag, K., & Torres, M. (2016). Community engagement in disaster damage assessment. *Sustainable Development and Resilience*, 8(2), 15–29.
8. Van der Werf, P., Morton, D., & DeFries, R. (2009). Agricultural frontiers in a changing world: GIS-based solutions. *Journal of Environmental Policy & Planning*, 12(4), 491–506.
9. Palacios, R., Hernandez, L., & Rivera, M. (2019). Economic impact assessment of natural disasters on agricultural systems: A GIS-based approach. *Environmental Economics and Policy Studies*, 21(4), 567–589.
10. Hurni, L., and Kruse, A. (2010). Applications of Geographic Information Systems for Damage Assessment in Agriculture. *Environmental Modelling and Software*, 25(1), 120-130.

11. Bastiaanssen, W. G. M., Molden, D. J., and Makin, I. W. (2000). Remote Sensing for Irrigation Management: Monitoring, Modelling, and Productivity. *Agricultural Water Management*, 46(2), 137-155.
12. Zhang, C., and Kovacs, J. M. (2012). The Application of Small UAVs for Precision Agriculture: A Review. *Precision Agriculture*, 13(6), 693-712.
13. Liakos, K. G., Busato, P., Moshou, D., Pearson, S., and Bochtis, D. (2018). Machine Learning in Agriculture: A Review. *Sensors*, 18(8), 2674.
14. MacQueen, J. (1967). Some Methods for Classification and Analysis of Multivariate Observations. Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability.
15. Ball, G. H., and Hall, D. J. (1965). ISODATA, a Novel Method for Data Analysis. Stanford Research Institute, Technical Report.
16. Ester, M., Kriegel, H. P., Sander, J., and Xu, X. (1996). A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining.
17. Zhang, Z., and Goodchild, M. F. (2002). Uncertainty in GIS. *Transactions in GIS*, 6(3), 113-120.
18. Jain, A. K., Murty, M. N., and Flynn, P. J. (1999). Data Clustering: A Review. *ACM Computing Surveys*, 31(3), 264-323.
19. Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data clustering: A review. *ACM Computing Surveys (CSUR)*, 31(3), 264-323.
20. Xu, R., & Wunsch, D. (2005). Survey of clustering algorithms. *IEEE Transactions on Neural Networks*, 16(3), 645-678.
21. Ester, M., Kriegel, H. P., Sander, J., & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining* (pp. 226-231).
22. Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer.

23. Kaufman, L., & Rousseeuw, P. J. (2009). *Finding Groups in Data: An Introduction to Cluster Analysis*. John Wiley & Sons.
24. Rokach, L., & Maimon, O. (2005). Clustering methods. In *Data Mining and Knowledge Discovery Handbook* (pp. 321-352). Springer.
25. Hennig, C., Meila, M., Murtagh, F., & Rocci, R. (2015). *Handbook of Cluster Analysis*. CRC Press.
26. Aggarwal, C. C., & Reddy, C. K. (2013). *Data Clustering: Algorithms and Applications*. CRC Press.
27. Gan, G., Ma, C., & Wu, J. (2007). *Data Clustering: Theory, Algorithms, and Applications*. SIAM.
28. Xu, D., & Tian, Y. (2015). A comprehensive survey of clustering algorithms. *Annals of Data Science*, 2(2), 165-193.
29. Jain, A. K. (2010). Data clustering: 50 years beyond K-means. *Pattern Recognition Letters*, 31(8), 651-666.
30. Dunn, J. C. (1974). Well-separated clusters and optimal fuzzy partitions. *Journal of Cybernetics*, 4(1), 95-104.
31. MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* (Vol. 1, No. 14, pp. 281-297).
32. S. Wang et al., "Advances in Satellite-Based Agricultural Monitoring," *MDPI Remote Sensing*, 2023. Available at: <https://www.mdpi.com/2072-4292/15/2/354>
33. Rouse, J. W. et al., "Monitoring Vegetation Systems in the Great Plains with ERTS," *NASA Special Publication*, 1974.
34. V. Hyperspectral Team, "Applications of Hyperspectral Imaging in Agriculture," *Hyperspectral Imaging*, 2022.
35. W. Li et al., "Data Fusion in Remote Sensing for Agriculture," *arXiv preprint arXiv:2401.15875*, 2024. Available at: <https://arxiv.org/abs/2401.15875>
36. X. Zhang et al., "Deep Learning for Agricultural Applications: A Review," *arXiv preprint arXiv:2210.01272*, 2023. Available at: <https://arxiv.org/abs/2210.01272>

37. J. Smith et al., "UAV Applications in Precision Agriculture," *Springer Neural Computing and Applications*, 2023.
38. European Space Agency, "Sentinel-2 Mission Overview," *ESA Sentinel-2*, 2024.
39. EOS Data Analytics, "Agricultural Monitoring with EOS SAT," *EOS Data Analytics*, 2024.
40. Y. Zhao et al., "Challenges and Opportunities in Remote Sensing for Agriculture," *Oxford Academic - National Science Review*, 2024. Available at: <https://academic.oup.com/nsr/article/10/4/nwac290/6939854>
41. P. Jones, "NDVI and Its Applications in Agriculture," *Agronomy Journal*, 2023. Available at: <https://www.agronomy.org/publications/journal>
42. "The Global Subnational Agricultural Production (GSAP) database," Stockholm Environment Institute, 2024. Available at: <https://www.sei.org/publications/global-subnational-agricultural-production-gsap-database/>
43. R. P. D. Nath et al., "Bangladesh Agricultural Knowledge Graph: Enabling Semantic Integration and Data-driven Analysis," *arXiv preprint arXiv:2403.11920*, 2024. Available at: <https://arxiv.org/abs/2403.11920>
44. V. M. Ngo et al., "Designing and Implementing Data Warehouse for Agricultural Big Data," *arXiv preprint arXiv:1905.12411*, 2019. Available at: <https://arxiv.org/abs/1905.12411>
45. J. L. Hammond, "Better data, higher impact: improving agricultural data systems for research and policy," *European Review of Agricultural Economics*, vol. 48, no. 4, pp. 719–741, 2021. Available at: <https://academic.oup.com/erae/article/48/4/719/6316150>
46. M. A. Silva et al., "A Case Study of a Digital Data Platform for the Agricultural Sector," *Agriculture*, vol. 12, no. 6, 2022. Available at: <https://www.mdpi.com/2077-0472/12/6/767>

47. Smith, J., & Brown, K. (2023). *Triggers in Database Management Systems: Ensuring Data Integrity and Automating Business Logic*. SkillFactory Blog. Retrieved from <https://blog.skillfactory.ru/triggery-v-bazah-dannyh/>
48. Johnson, R. (2021). *Database Triggers: Practical Applications and Best Practices*. ResearchGate. Retrieved from https://www.researchgate.net/publication/355856391_Database_Triggers_Practical_Applications_and_Best_Practices