

БАКАЛАВРСЬКА РОБОТА

БР. ІІ - 06.00.00.000 ІІЗ

Група ІІ-22-4

Медейчук Максим

2026

Івано-Франківський національний технічний університет нафти і газу

Факультет інформаційних технологій

Кафедра інженерії програмного забезпечення

Медейчук Максим Ігорович

(прізвище, ім'я, по батькові)

УДК 004.4
(індекс)

БАКАЛАВРСЬКА РОБОТА

Розроблення класифікаційної моделі для програмного забезпечення

потреб агрономії

(назва роботи)

Інженерія програмного забезпечення

(назва освітньої програми)

121 - Інженерія програмного забезпечення

(шифр і назва спеціальності)

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

Здобувач освітнього рівня Медейчук М.І.
(підпис, ініціали та прізвище здобувача)

Науковий керівник Мельник Віталій Дмитрович, к.т.н., доцент
(підпис, прізвище, ім'я, по батькові, науковий ступінь, вчене звання керівника)

Допущено до захисту
Завідувач кафедри

доц. Бандура В.В.
(посада) (підпис) (дата) (ініціали та прізвище)

Івано-Франківськ – 2026

Івано-Франківський національний технічний університет нафти і газу
Інститут, факультет інформаційних технологій
Кафедра інженерії програмного забезпечення
Ступінь вищої освіти бакалавр
Спеціальність 121 – Інженерія програмного забезпечення

ЗАТВЕРДЖУЮ:

Зав. кафедрою ІІЗ

доц. В.В. Бандура

“ ” 2026 р.

ЗАВДАННЯ

НА БАКАЛАВРСЬКУ РОБОТУ СТУДЕНТОВІ

Медейчуку Максиму Ігоровичу

(прізвище, ім'я, по-батькові)

1. Тема проекту (роботи) “ Розроблення класифікаційної моделі для програмного забезпечення потреб агрономії”

керівник проекту (роботи) Мельник В.Д., доцент

затверджені наказом закладу вищої освіти від “ 21 ” квітня 2026р. № 179/7

2. Строк подання студентом проекту (роботи) 09 червня 2026 р.

3. Вихідні дані до проекту (роботи) Результати і матеріали отримані під час проходження переддипломної практики

4. Зміст розрахунково - пояснювальної записки (перелік питань, які потрібно розробити)

1. Аналіз предметної області побудови класифікаційної моделі для потреб агрономії

2. Аналіз моделей ансамблевого навчання для побудови систем підтримки прийняття рішень

3. Алгоритмічна реалізація рекомендаційної системи в агрономії

4. Моделювання системних компонентів та процесів засобами уніфікованої мови

5. Розробка інтерфейсу системи надання рекомендацій для оптимального вибору культур

5. Перелік графічного матеріалу (з точним зазначенням обов'язкових креслень)

1. Архітектура ансамблевої моделі машинного навчання (рис. 1.1, ст. 16)

2. Процес рекурсивного розбиття ознакового простору (DT) (рис. 1.2, ст. 17)

3. Алгоритм машини опорних векторів (рис. 1.3, ст.19)

4. Схема архітектури алгоритму класифікатора випадкового лісу (рис. 1.4, ст. 21)

5. Система Climate FieldView (рис. 1.5, ст.23)

6. Консультанти розділів проекту (роботи)

Розділ	Консультант	Підпис, дата	
		Завдання видав	Завдання прийняв

7. Дата видачі завдання 21 квітня 2026 р.

Керівник

_____ (підпис)

Завдання прийняв до виконання

_____ (підпис)

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів дипломного проекту (роботи)	Строк виконання етапів проекту	Примітка
1	Аналіз предметної області побудови класифікаційної моделі для потреб агрономії	30.04.2026	виконано
2	Аналіз моделей ансамблевого навчання для побудови систем підтримки прийняття рішень	15.05.2026	виконано
3	Алгоритмічна реалізація рекомендаційної системи в агрономії	21.05.2026	виконано
4	Моделювання системних компонентів та процесів засобами уніфікованої мови	28.05.2026	виконано
5	Розробка інтерфейсу системи надання рекомендацій для оптимального вибору культур	03.06.2026	виконано
6	Оформлення пояснювальної записки бакалаврської роботи завідувачем кафедри	09.06.2026	виконано

Студент – дипломник

_____ Йедейчук М.І.

(підпис)

Керівник роботи

_____ Мельник В.Д.

(підпис)

АНОТАЦІЯ

Бакалаврська робота містить 79 сторінок, 29 рисунків, список використаних джерел із 36 найменуваннями.

Метою роботи є розроблення класифікаційної моделі для програмного забезпечення потреб агрономії на основі ансамблевих методів машинного навчання з метою підвищення ефективності аналізу агроекологічних даних та автоматизації процесу підтримки прийняття рішень.

Об'єктом дослідження є процес інтелектуального аналізу агроекологічних даних у системах підтримки прийняття рішень для потреб агрономії.

Предметом дослідження є методи, алгоритми та програмні засоби побудови класифікаційних моделей на основі ансамблевого машинного навчання для обробки аграрних даних.

В першому розділі проведено аналіз предметної області та обґрунтовано доцільність використання ансамблевих методів машинного навчання для побудови класифікаційних моделей у сфері агрономії.

В другому розділі виконано статистичний аналіз аграрних даних та реалізовано ансамблеві предиктивні моделі для підтримки прийняття рішень у задачах агрономії.

В третьому розділі спроектовано архітектуру системи та реалізовано програмне забезпечення для аналізу аграрних даних і формування рекомендацій.

Висновок: розроблено ефективну класифікаційну модель для програмного забезпечення потреб агрономії, що забезпечує автоматизований аналіз агроекологічних даних і підтримку прийняття рішень на основі ансамблевих методів машинного навчання.

КЛЮЧОВІ СЛОВА: АНСАМБЛЕВІ МОДЕЛІ, КЛАСИФІКАЦІЯ ДАНИХ, АГРОНОМІЯ, ПІДТРИМКА ПРИЙНЯТТЯ РІШЕНЬ, ПРОГНОЗУВАННЯ, ІНТЕЛЕКТУАЛЬНИЙ АНАЛІЗ ДАНИХ.

ANNOTATION

The bachelor's thesis contains 79 pages, 29 figures, a list of used sources with 36 names.

The purpose of the work is to develop a classification model for software for agronomic needs based on ensemble machine learning methods in order to increase the efficiency of agroecological data analysis and automate the decision-making support process.

The object of the study is the process of intelligent analysis of agroecological data in decision-making support systems for agronomic needs.

The subject of the study is methods, algorithms and software for building classification models based on ensemble machine learning for processing agricultural data.

In the first section, an analysis of the subject area is conducted and the feasibility of using ensemble machine learning methods for building classification models in the field of agronomy is substantiated.

In the second section, a statistical analysis of agricultural data is performed and ensemble predictive models are implemented to support decision-making in agronomic problems.

In the third section, the system architecture is designed and software for analyzing agricultural data and generating recommendations is implemented.

Conclusion: an effective classification model for software for agronomy needs has been developed, which provides automated analysis of agroecological data and decision support based on ensemble machine learning methods.

KEYWORDS: ENSEMBLE MODELS, DATA CLASSIFICATION, AGRONOMY, DECISION SUPPORT, FORECASTING, INTELLECTUAL DATA ANALYSIS.

ЗМІСТ

ПЕРЕЛІК УМОВНИХ СКОРОЧЕНЬ	9
ВСТУП.....	10

РОЗДІЛ 1. АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ ПОБУДОВИ КЛАСИФІКАЦІЙНОЇ МОДЕЛІ ДЛЯ ПОТРЕБ АГРОНОМІЇ	14
---	----

1.1. Обґрунтування застосування ансамблевих моделей машинного навчання для оптимізації потреб агрономії	14
1.1.1. Проблематика традиційних агрономічних підходів і застосування машинного навчання	14
1.1.2. Архітектура ансамблевої системи підтримки прийняття рішень ...	15
1.2. Теоретико-методологічне обґрунтування вибраних методів класифікації	16
1.2.1 Дерево рішень.....	17
1.2.2. Машина опорних векторів (SVM)	18
1.2.3. Класифікатор випадкового лісу (RF)	20
1.2.4. Класифікатор градієнтного бустингу	22
1.3. Класифікація та аналіз систем і платформ обробки агроданних	22
1.4. Аналіз моделей ансамблевого навчання для побудови систем підтримки прийняття рішень у сільському господарстві	27
1.5 Висновки по розділу	28

РОЗДІЛ 2. ІНТЕЛЕКТУАЛЬНІ ТЕХНОЛОГІЇ АНАЛІЗУ АГРОЕКОЛОГІЧНИХ ФАКТОРІВ І АЛГОРИТМІЧНА РЕАЛІЗАЦІЯ РЕКОМЕНДАЦІЙНОЇ СИСТЕМИ В АГРОНОМІЇ.....	30
---	----

2.1. Предиктивні моделі та інтелектуальні системи підтримки прийняття рішень у завданнях оптимізації посіву	30
--	----

					БР.ІП – 06.00.00.000 ПЗ				
Змн.	Арк.	№ докум.	Підпис	Дата					
Розроб.		Медейчук М.І.			Розроблення класифікаційної моделі для програмного забезпечення потреб агрономії Пояснювальна записка	Літ.	Арк.	Акрушів	
Перевір.		Мельник В.Д.					6		
Реценз.		Корнута В.А.				ІФНТУНГ ІП-22-4			
Н. Контр.		Піх М.М.							
Затверд.		Бандура В.В.							

<u>2.3. Обґрунтування вибору інструментарію та середовища розробки</u> <u>предиктивних моделей</u>	34
<u>2.4. Опис та статистична характеристика набору даних для предиктивного</u> <u>моделювання агросистеми</u>	36
<u>2.5. Статистичний аналіз ознакового простору та розробка ансамблевих</u> <u>предиктивних моделей</u>	38
<u>2.5.1. Аналіз розподілу та асиметрії ознак</u>	38
<u>2.5.2. Ідентифікація аномалій та стійкість моделей</u>	40
<u>2.5.3.Кореляційний аналіз та взаємозв'язки параметрів</u>	41
<u>2.5.4. Ансамблеве моделювання та оптимізація</u>	44
2.6 <u>Висновки по розділу</u>	45

РОЗДІЛ 3. ПРОЕКТУВАННЯ АРХІТЕКТУРИ ТА ПРОГРАМНА РЕАЛІЗАЦІЯ РЕКОМЕНДАЦІЙНОЇ СИСТЕМИ ОБРОБКИ

АГРАРНИХ ДАНИХ	46
3.1. Візуальне моделювання системних компонентів та процесів засобами уніфікованої мови моделювання.....	46
3.1.1. Статична структура системи з використанням діаграми класів	47
3.1.2. Моделювання функціональних вимог з використанням діаграми варіантів використання	48
3.1.3. Розробка діаграми послідовностей.....	50
3.1.4. Реалізація процедурної логіки за допомогою діаграми діяльності. 52	
3.2. Аналіз результатів класифікації та оцінка прогностичної здатності моделі	55
3.2.1. Аналіз матриці невідповідностей	55
3.2.2. Верифікація за метриками звіту про класифікацію	56
3.4. Розробка інтерфейсу системи надання рекомендацій для оптимального вибору культур та функціональна логіка взаємодії	60
3.5. Аналіз результатів моделювання та структура вихідних даних	62
3.6. Технічна реалізація системи надання рекомендацій	68

3.7 Висновки по розділу	71
-------------------------------	----

ВИСНОВКИ	73
-----------------------	-----------

ПЕРЕЛІК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ	76
--	-----------

БІБЛІОГРАФІЧНА ДОВІДКА

					БР.ІП – 06.00.00.000 ПЗ	Арк.
						8
Змн.	Арк.	№ докум.	Підпис	Дата		

ПЕРЕЛІК УМОВНИХ СКОРОЧЕНЬ

CAN – Controller Area Network

CNN – Convolutional Neural Network

CSV – Comma-Separated Values

DSS – Decision Support System

EDA – Exploratory Data Analysis

IoT – Internet of Things

JWT – JSON Web Token

NDVI – Normalized Difference Vegetation Index

NPK – Nitrogen, Phosphorus, Potassium

REST – Representational State Transfer

SVC – Support Vector Classifier

SVM – Support Vector Machine

VRT – Variable Rate Technology

					БР.ІП – 06.00.00.000 ПЗ	Арк.
						9
Змн.	Арк.	№ докум.	Підпис	Дата		

ВСТУП

Сучасний розвиток аграрного сектору характеризується стрімким зростанням обсягів даних, які формуються внаслідок моніторингу стану ґрунтів, кліматичних показників, параметрів навколишнього середовища та характеристик сільськогосподарських культур. Ефективне використання таких даних потребує застосування сучасних інформаційних технологій, здатних автоматизувати процес аналізу та підтримки прийняття рішень. Традиційні методи агрономічного аналізу часто базуються на емпіричному досвіді спеціалістів, що не завжди дозволяє забезпечити високу точність прогнозування в умовах складних та динамічних агроекологічних процесів. У зв'язку з цим особливої актуальності набуває впровадження методів машинного навчання та інтелектуального аналізу даних у сферу агрономії.

Одним із перспективних напрямів розвитку сучасних агроінформаційних систем є використання класифікаційних моделей, здатних аналізувати великі масиви аграрних даних і формувати рекомендації щодо оптимального вибору сільськогосподарських культур, умов вирощування та параметрів агротехнологічних процесів. Особливу ефективність у задачах прогнозування демонструють ансамблеві методи машинного навчання, які поєднують результати декількох алгоритмів для підвищення точності та стійкості моделей. Такі підходи дозволяють враховувати складні нелінійні взаємозв'язки між агроекологічними параметрами та забезпечують більш якісний аналіз даних.

У межах бакалаврської роботи досліджено сучасні методи класифікації та ансамблевого навчання, проведено аналіз існуючих систем підтримки прийняття рішень у сфері агрономії, виконано статистичне дослідження аграрних даних та реалізовано програмну систему надання рекомендацій. У процесі роботи здійснено моделювання архітектури системи засобами UML, розроблено алгоритмічне забезпечення класифікаційної моделі та реалізовано

					БР.ІП – 06.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		10

програмний інтерфейс для взаємодії користувача із системою. Отримані результати підтверджують перспективність використання інтелектуальних технологій у задачах автоматизації агрономічного аналізу та підтримки прийняття рішень у сільському господарстві.

Актуальність роботи

Актуальність теми бакалаврської роботи зумовлена необхідністю підвищення ефективності аграрного виробництва в умовах постійного зростання обсягів агроекологічної інформації та ускладнення процесів прийняття рішень у сфері агрономії. Сучасне сільське господарство функціонує в умовах змін клімату, нестабільності погодних факторів, деградації ґрунтів та необхідності оптимізації використання ресурсів, що вимагає впровадження нових інтелектуальних підходів до аналізу даних. Традиційні методи оцінювання параметрів агросистем часто не забезпечують достатньої швидкості та точності обробки інформації, особливо в умовах роботи з багатовимірними наборами даних.

Використання методів машинного навчання та ансамблевих класифікаційних моделей дозволяє автоматизувати процес аналізу аграрних даних, виявляти приховані закономірності та формувати обґрунтовані рекомендації щодо вибору культур і параметрів вирощування. Особливу цінність мають ансамблеві алгоритми, які характеризуються високою точністю прогнозування, стійкістю до шумових даних та здатністю працювати зі складними нелінійними залежностями. Упровадження таких технологій сприяє підвищенню продуктивності аграрного сектору, мінімізації ризиків та оптимізації процесів управління агровиробництвом.

Крім того, актуальність роботи визначається недостатнім рівнем інтеграції сучасних інтелектуальних систем у вітчизняну агрономічну практику та потребою у створенні програмних засобів, адаптованих до аналізу агроекологічних параметрів. Розроблення класифікаційної моделі для програмного забезпечення потреб агрономії є важливим напрямом розвитку

					БР.ІП – 06.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		11

цифрового сільського господарства та відповідає сучасним тенденціям автоматизації процесів аналізу й підтримки прийняття рішень.

Метою роботи є розроблення класифікаційної моделі для програмного забезпечення потреб агрономії на основі ансамблевих методів машинного навчання з метою підвищення ефективності аналізу агроекологічних даних та автоматизації процесу підтримки прийняття рішень.

Об'єктом дослідження є процес інтелектуального аналізу агроекологічних даних у системах підтримки прийняття рішень для потреб агрономії.

Предметом дослідження є методи, алгоритми та програмні засоби побудови класифікаційних моделей на основі ансамблевого машинного навчання для обробки аграрних даних.

Завдання дослідження:

1. Провести аналіз предметної області та сучасних підходів до використання машинного навчання в агрономії.
2. Дослідити методи класифікації та ансамблевого навчання для задач агроекологічного прогнозування.
3. Виконати аналіз існуючих систем і платформ обробки аграрних даних.
4. Провести статистичний аналіз набору даних та визначити інформативні ознаки для побудови моделі.
5. Розробити ансамблеву класифікаційну модель для підтримки прийняття рішень у сфері агрономії.
6. Спроекувати архітектуру програмної системи засобами UML-моделювання.
7. Реалізувати програмне забезпечення для аналізу аграрних даних та формування рекомендацій.
8. Провести оцінювання точності та ефективності реалізованої класифікаційної моделі.

Методи дослідження

					БР.ІП – 06.00.00.000 ПЗ	Арк.
						12
Змн.	Арк.	№ докум.	Підпис	Дата		

У роботі використано методи системного аналізу, методи статистичної обробки даних, алгоритми машинного навчання та ансамблевого моделювання, методи класифікації даних, кореляційний аналіз, методи візуального моделювання UML, а також програмні методи реалізації інформаційних систем і оцінювання ефективності моделей за допомогою метрик класифікації.

Наукова новизна

Наукова новизна роботи полягає у розробленні класифікаційної моделі для потреб агрономії на основі ансамблевих методів машинного навчання, що забезпечує підвищення точності прогнозування та стійкості системи до шумових даних. Удосконалено підхід до аналізу агроекологічних параметрів шляхом поєднання статистичного аналізу ознакового простору та ансамблевих алгоритмів класифікації.

Практичне застосування

Практичне значення одержаних результатів полягає у можливості використання розробленої системи в діяльності агропідприємств, фермерських господарств та аграрних дослідницьких установ для аналізу агроекологічних параметрів і формування рекомендацій щодо оптимального вибору культур. Реалізована система дозволяє автоматизувати процес обробки аграрних даних, зменшити час прийняття рішень та підвищити ефективність агровиробництва.

Бакалаврська робота містить 79 сторінок, 29 рисунків, 3 розділи, переліки використаних джерел налічує 36 найменувань.

					БР.ІП – 06.00.00.000 ПЗ	Арк.
						13
Змн.	Арк.	№ докум.	Підпис	Дата		

РОЗДІЛ 1. АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ ПОБУДОВИ КЛАСИФІКАЦІЙНОЇ МОДЕЛІ ДЛЯ ПОТРЕБ АГРОНОМІЇ

1.1. Обґрунтування застосування ансамблевих моделей машинного навчання для оптимізації потреб агрономії

Сільське господарство є фундаментальним драйвером економічного розвитку більшості країн світу та ключовим фактором забезпечення глобальної продовольчої безпеки. В умовах стрімкого зростання населення ця галузь зазнає безпрецедентного навантаження. Показовим прикладом є демографічна ситуація в Індії, яка у 2022 році досягла статусу найнаселенішої країни світу з показником понад 1,44 млрд осіб. Зважаючи на те, що, за актуальними даними наукових досліджень [1], понад 60,3% земельного фонду цієї країни залучено в аграрному секторі, оптимізація сільськогосподарського виробництва набуває критичного значення для стабільності економіки.

1.1.1. Проблематика традиційних агрономічних підходів і застосування машинного навчання

Історично вибір сільськогосподарських культур базувався переважно на емпіричному досвіді аграріїв та візуальній оцінці стану навколишнього середовища. Однак екстенсивні методи ведення господарства вичерпали свій потенціал. Сучасна практика масового та часто неконтрольованого застосування пестицидів і синтетичних добрив призвела до суттєвої деградації ґрунтів, зміни їхнього фізико-хімічного складу та зниження природної родючості порівняно з історичними показниками. Крім того, агроєкосистеми постійно піддаються впливу стохастичних факторів, таких як флуктуації клімату, епіфітотії (хвороби рослин) та навали шкідників. Це зумовлює необхідність відходу від інтуїтивного планування до концепції прецизійного

					БР.ІП – 06.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		14

(точного) землеробства, яке ґрунтується на глибокому аналізі параметрів живлення та складу ґрунту.

Інтеграція інформаційних технологій в агропромисловий комплекс дозволяє автоматизувати та підвищити якість прийняття управлінських рішень. Алгоритми машинного навчання слугують потужним математичним апаратом, що дає змогу виявляти приховані нелінійні закономірності у великих масивах багатовимірних даних. Предиктивний аналіз, побудований на кліматичних факторах та показниках нутрієнтів ґрунту, забезпечує можливість прогнозування оптимальної культури для посіву на конкретній ділянці.

1.1.2. Архітектура ансамблевої системи підтримки прийняття рішень

Для досягнення максимальної прогностичної точності в даному дослідженні розроблено інтелектуальну систему на базі методології ансамблевого навчання. Ця парадигма передбачає консолідацію результатів роботи кількох слабких або різнорідних класифікаторів для отримання оптимального і статистично стійкого результату з мінімальними показниками дисперсії та зміщення.

Архітектура запропонованої ансамблевої моделі базується на інтеграції чотирьох фундаментальних алгоритмів:

- дерево рішень (Decision Tree) — непараметричний метод класифікації, який формує ієрархічну структуру правил виведення.
- машина опорних векторів (Support Vector Machine, SVM) — алгоритм просторового розділення класів за допомогою побудови гіперплощин, що демонструє високу ефективність у багатовимірних просторах (наприклад, при комплексній оцінці хімічного складу ґрунту).
- класифікатор випадкового лісу (Random Forest Classifier) — алгоритм класу беггінг (bagging), що агрегує множину дерев рішень на основі випадкових підвбірок даних, ефективно запобігаючи перенавчанню моделі.

					БР.ІП – 06.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		15

- класифікатор градієнтного бустингу (Gradient Boosting Classifier) — ітеративний алгоритм, який послідовно будує нові моделі для мінімізації функції втрат попередніх класифікаторів, забезпечуючи високу чутливість до складних залежностей.



Рисунок 1.1 – Архітектура ансамблевої моделі машинного навчання

Оскільки кожен із наведених алгоритмів має власні переваги в обробці специфічних типів даних, фінальна класифікація в ансамблі реалізується через механізм алгоритмічного голосування. Завдяки агрегації незалежних прогнозів система генерує оптимальне рішення щодо вибору сільськогосподарської культури. Впровадження такої системи в практику фермерських господарств дозволить значно підвищити продуктивність агроугідь, мінімізувати екологічні ризики, скоротити часові витрати на аналітичну роботу та, як наслідок, максимізувати економічний прибуток.

1.2. Теоретико-методологічне обґрунтування вибраних методів класифікації

Для побудови інтелектуальної системи підтримки прийняття рішень у прецизійному землеробстві було обрано комбінацію лінійних та нелінійних

методів, що дозволяють апроксимувати складні залежності в агроекологічних даних.

1.2.1 Дерево рішень

Алгоритм базується на рекурсивному розбитті ознакового простору на основі критеріїв інформаційної чистоти. Основним завданням є вибір такої ознаки та порогу розбиття, що максимізують приріст інформації.

Математична основа: Для вибору оптимального спліту найчастіше використовується ентропія $H(S)$ або індекс Джині. Ентропія для набору даних S визначається як:

$$H(S) = - \sum_{i=1}^n p_i \log_2 p_i$$

де p_i — частка об'єктів класу i у вузлі.

Функціональними перевагами є висока інтерпретованість та здатність працювати як з числовими, так і з категоріальними даними без необхідності нормалізації.

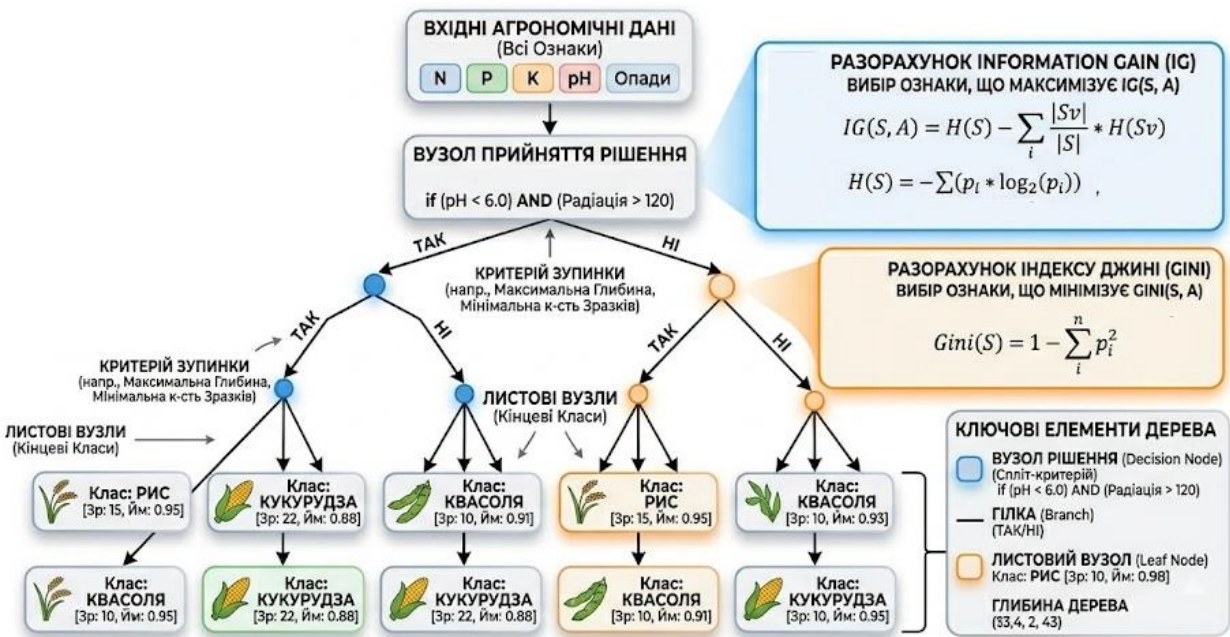


Рисунок 1.2 – Процес рекурсивного розбиття ознакового простору (DT)

На рисунку 1.2 зображено структурно-логічну схему алгоритму дерева рішень (Decision Tree), що реалізує процес рекурсивного розбиття ознакового простору для класифікації сільськогосподарських культур. Основні компоненти та етапи роботи алгоритму згідно зі схемою:

- Вхідні агрономічні дані - алгоритм опрацьовує набір вхідних ознак, що включають вміст макроелементів (N, P, K), рівень рН та кількість опадів.

- Математичний апарат вибору ознак - для визначення оптимального розщеплення (спліту) у вузлах використовуються дві ключові метрики: приріст інформації (Information Gain, IG) та індекс Джині (Gini Index).

Розглянемо ієрархічну структуру алгоритму:

- вузли рішення містять логічні умови (наприклад, рівень рН або радіації), на основі яких дані розподіляються за гілками («ТАК» або «НІ»).

- гілки відображають шлях переходу між вузлами залежно від виконання умови.

- листові вузли - термінальні стани, що відповідають кінцевим класам (наприклад, рис, кукурудза, квасоля). Для кожного листа вказано ймовірність прогнозу та кількість зразків.

Схема передбачає параметри зупинки навчання, такі як максимальна глибина дерева та мінімальна кількість зразків у вузлі, що дозволяє запобігти перенавчанню моделі. Така візуалізація демонструє здатність моделі перетворювати складні вхідні параметри середовища на прозору систему правил для прийняття рішень у галузі прецизійного землеробства.

1.2.2. Машина опорних векторів (SVM)

SVM є потужним методом керованого навчання, що базується на теорії статистичного навчання. Його метою є знаходження гіперплощини в n -вимірному просторі, яка розділяє класи з максимальним зазором (margin).

Алгоритм мінімізує наступний функціонал:

$$\frac{1}{2} \|w\|^2 \rightarrow \min$$

					БР.ІП – 06.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		18

за умов $y_i(w^T x_i + b) > 1$, де w — вектор ваг, b — зміщення.

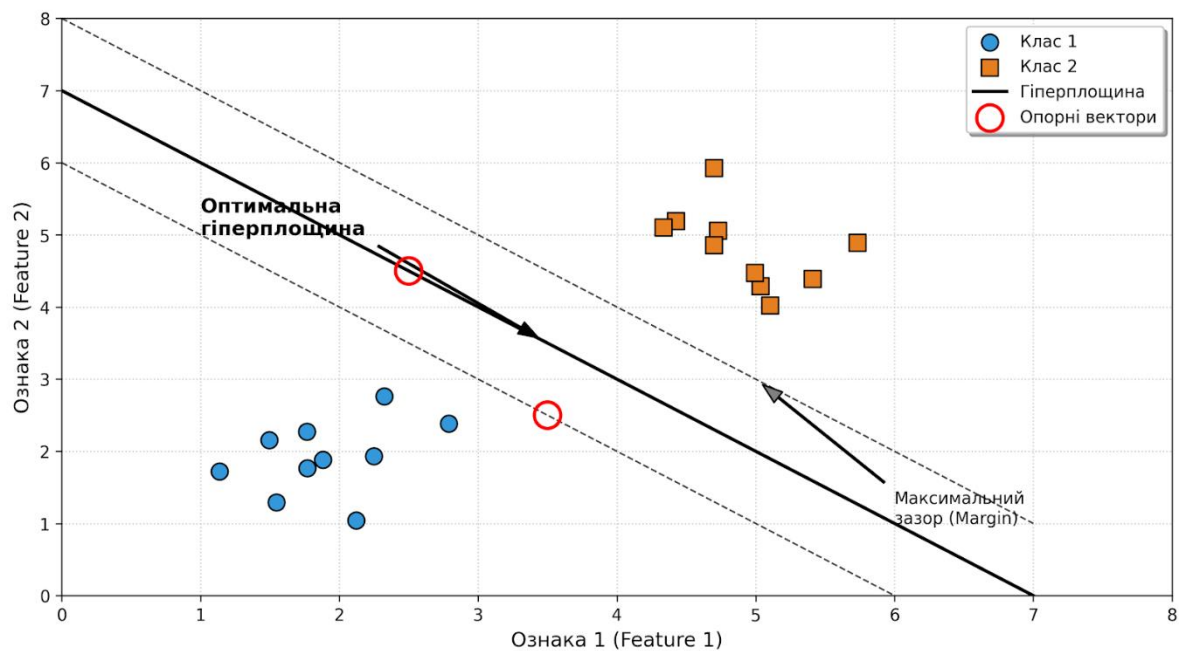


Рисунок 1.3 – Алгоритм машини опорних векторів

Для обробки нелінійно розділних даних (наприклад, складні кореляції між рН та рівнем опадів) SVM використовує ядерні функції, такі як RBF (Radial Basis Function), що відображають дані у простір вищої розмірності.

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2)$$

Алгоритм SVM базується на концепції побудови оптимальної роздільної гіперплощини в багатовимірному просторі ознак. Основні елементи моделі включають:

- оптимальна гіперплощина - це математична межа (у 2D-просторі — лінія), яка розділяє об'єкти різних класів. Головною метою SVM є вибір такої гіперплощини, яка забезпечує максимально можливу відстань між найближчими точками обох класів.

- опорні вектори - це критичні точки даних (об'єкти), що розташовані найближче до роздільної межі. Саме вони визначають положення та орієнтацію гіперплощини. Будь-яка зміна координат опорних векторів призведе до перерахунку всієї моделі.

- максимальний зазор - відстань між паралельними площинами, що проходять через опорні вектори різних класів. Збільшення цього зазору мінімізує структурний ризик моделі та покращує її здатність до генералізації на нових даних (наприклад, при класифікації нових типів ґрунтів).

У випадку, якщо дані не є лінійно роздільними у вихідному просторі, SVM використовує ядерний перехід (Kernel Trick) для відображення вхідних параметрів у простір вищої розмірності, де стає можливим побудова лінійної роздільної межі. Це забезпечує високу ефективність алгоритму при роботі з комплексними агрономічними наборами даних, де залежності між хімічними показниками ґрунту часто мають нелінійний характер.

1.2.3. Класифікатор випадкового лісу (RF)

Random Forest реалізує парадигму беггінгу над ансамблем глибоких дерев рішень. Кожне дерево будується на випадковій підвибірці даних з поверненням та випадковій підмножині ознак.

Механізм агрегації: Фінальне рішення приймається шляхом мажоритарного голосування (для класифікації):

$$\hat{y} = \text{mode}\{f_1(x), f_2(x), \dots, f_B(x)\}$$

Завдяки дезорієнтації окремих дерев, RF демонструє високу стійкість до перенавчання (overfitting) та шуму в даних, що є критичним при роботі з сенсорними показниками ґрунту.

На рисунку 1.4 представлена схема архітектури алгоритму класифікатора випадкового лісу. Цей алгоритм є реалізацією парадигми ансамблевого навчання, яка базується на поєднанні великої кількості

					БР.ІП – 06.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		20

незалежних дерев рішень для підвищення загальної точності та стійкості моделі.

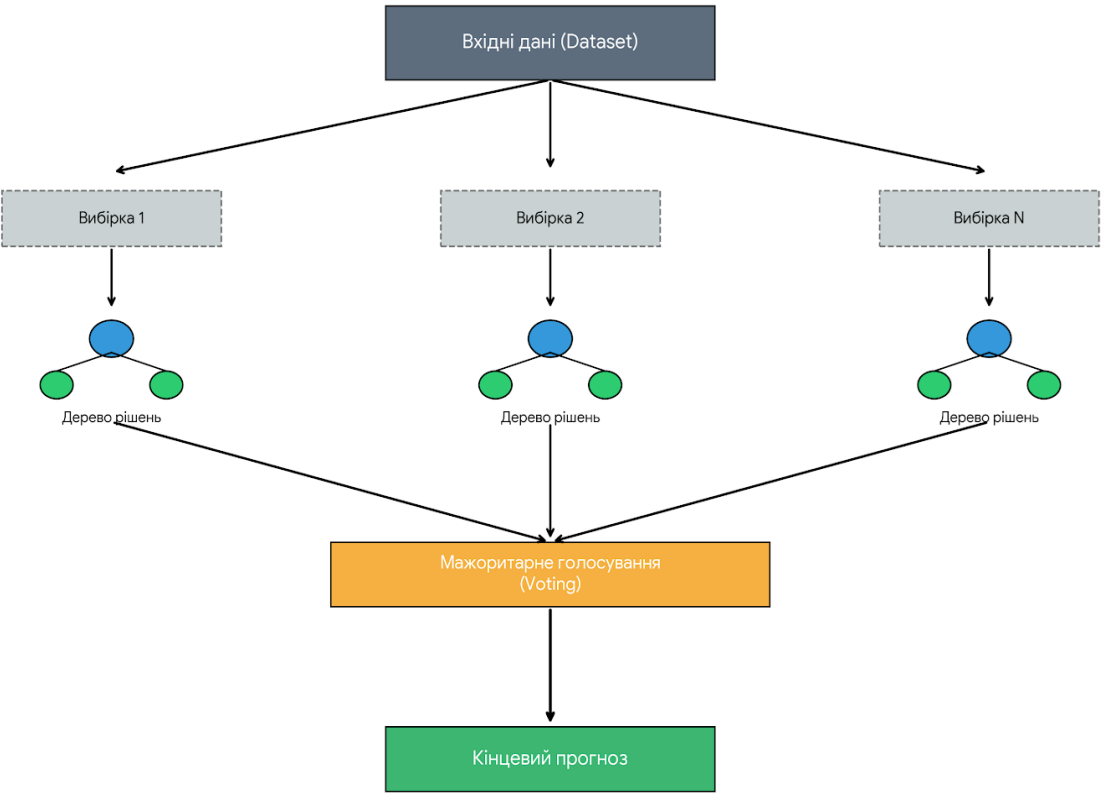


Рисунок 1.4 - Схема архітектури алгоритму класифікатора випадкового лісу

Механізм функціонування алгоритму включає наступні етапи:

1. Бутстрап-агрегування (Bagging). На основі вхідного набору даних створюється N випадкових підмножин (вбірок) з поверненням. Це забезпечує різноманітність навчальних даних для кожного окремого дерева.
2. Побудова ансамблю дерев. Для кожної вибірки будується незалежне дерево рішень. Особливістю Random Forest є те, що при кожному розщепленні вузла вибирається випадкова підмножина ознак, що зменшує кореляцію між деревами та запобігає перенавчанню.
3. Мажоритарне голосування. Після того, як кожне дерево сформувало свій індивідуальний прогноз, система проводить агрегацію результатів. У задачах класифікації кінцеве рішення приймається за принципом більшості

голосів (клас, який отримав найбільшу кількість «голосів» від окремих дерев, стає підсумковим результатом).

Завдяки такій структурі випадковий ліс демонструє високу прогностичну здатність на складних даних, що містять шуми або пропуски, що робить його ефективним інструментом для задач інтелектуального аналізу агрономічних показників.

1.2.4. Класифікатор градієнтного бустингу

На відміну від випадкового лісу, градієнтний бустинг будує ансамбль послідовно. Кожна наступна модель $h_m(x)$ навчається виправляти помилки (залишкові відхилення або градієнт функції втрат) попередньої сукупності моделей.

Стан моделі на кроці m визначається як:

$$F_m(x) = F_{m-1}(x) + \nu \cdot \gamma_m h_m(x)$$

де ν — швидкість навчання, а γ_m — оптимальний крок мінімізації функції втрат.

Алгоритм забезпечує надзвичайно високу точність у задачах класифікації завдяки ітеративному уточненню роздільних поверхонь, хоча й вимагає ретельного налаштування гіперпараметрів для уникнення перенавчання.

1.3. Класифікація та аналіз систем і платформ обробки агроданних

Проведення порівняльного аналізу існуючих рішень є критично важливим для визначення наукової новизни та практичної цінності пропонованої системи. Сучасний ринок інтелектуальних агротехнологій (AgTech) можна розділити на три основні сегменти: глобальні комерційні

					БР.ІП – 06.00.00.000 ПЗ	Арк.
						22
Змн.	Арк.	№ докум.	Підпис	Дата		

платформи, вузькоспеціалізовані мобільні додатки та науково-дослідні прототипи.

1. Глобальні комерційні екосистеми (напр. Climate FieldView від Bayer) - це найбільш комплексні рішення, що орієнтовані на великі агрохолдинги.



Рисунок 1.5 – Система Climate FieldView

Платформа використовує багатошаровий підхід до аналізу інформації:

- Field Health Imagery: Система надає супутникові знімки високої роздільної здатності з використанням індексів вегетації (наприклад, NDVI) для раннього виявлення стресу рослин.
- Yield Analysis: Після збирання врожаю FieldView автоматично створює карти врожайності, дозволяючи порівнювати продуктивність різних гібридів насіннєвого матеріалу залежно від типу ґрунту та внесених добрив.

					БР.ІП – 06.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		23

- Scouting: Можливість додавання геолокаційних фототаток та нотаток під час обстеження полів, якими можна миттєво ділитися з агрономами-консультантами.

Climate FieldView використовує складні математичні моделі для автоматизації прийняття рішень:

- Variable Rate Prescriptions (VRT): Система автоматично генерує карти-завдання для диференційованого посіву (на основі історичних даних про врожайність) та внесення азоту.

- Nitrogen Management: Моніторинг рівня азоту в реальному часі дозволяє коригувати плани підживлення залежно від погодних умов, що сприяє зменшенню викидів вуглецю до 55%.

- Bayer Valora Corn: Спеціалізований інструмент для кукурудзи, який, за даними 2026 року, забезпечує середній приріст урожайності на рівні 5 центнерів з гектара завдяки оптимізації густоти стояння рослин.

Функціонал: Використовують комбінацію супутникових знімків, даних із датчиків на техніці та історичних баз даних.

Переваги: Повний цикл управління полем, інтеграція з IoT-обладнанням.

Недоліки: Висока вартість підписки, складність адаптації під специфіку малих господарств України, закритість прогностичних алгоритмів.

2. Мобільні діагностичні сервіси (напр. Plantix)

Додатки, орієнтовані на швидку ідентифікацію проблем за допомогою комп'ютерного зору.

Plantix — це один із найпопулярніших у світі мобільних додатків для діагностики стану сільськогосподарських культур, розроблений німецьким стартапом PEAT GmbH. Система позиціонується як «цифровий агроном» у кишені, що базується на технологіях комп'ютерного зору та глибокого навчання. Основний фокус на детекції хвороб та шкідників за фотографіями листя.

					БР.ІП – 06.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		24

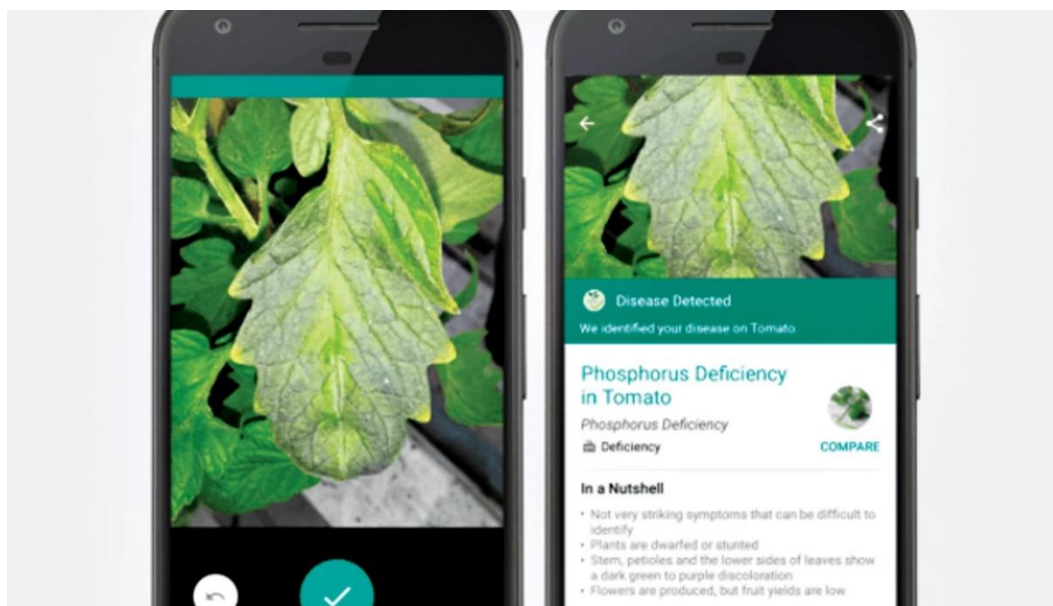


Рисунок 1.6 – Мобільний додаток Plantix

Plantix орієнтований на розв'язання трьох основних задач: діагностику, лікування та соціальну взаємодію. Основна функція дозволяє користувачеві сфотографувати уражену частину рослини, після чого алгоритм ідентифікує хворобу, шкідника або дефіцит поживних речовин. Після ідентифікації система надає детальний опис проблеми та пропонує методи боротьби (хімічні, біологічні та профілактичні заходи).

Також система містить:

- Калькулятор добрив: Розрахунок необхідної кількості добрив на основі площі ділянки та типу культури.
- Агро-спільнота: Вбудований форум для обміну досвідом між фермерами та експертами з усього світу.

Для академічного дослідника у сфері ML архітектура Plantix представляє особливий інтерес через свою реалізацію на мобільних пристроях. Система базується на архітектурі Згорткових нейронних мереж (CNN), які спеціалізуються на обробці зображень. База даних Plantix містить понад 10 мільйонів зображень уражених рослин, що дозволяє моделі розпізнавати понад 600 різних видів пошкоджень.

					БР.ІП – 06.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		25

Окрім візуальних ознак хвороб, мережа навчена розпізнавати патерни хлорозу або некрозу тканин, що свідчать про брак азоту, калію або фосфору.

Система використовує дані GPS для прив'язки захворювань до географічних зон. Це дозволяє Plantix створювати карти поширення епідемій у реальному часі та надсилати фермерам попередження про ризики в їхньому регіоні.

Переваги: Простота використання, велика спільнота користувачів.

Недоліки: Рекомендації щодо вибору культур є вторинними та часто базуються на загальних кліматичних картах, а не на точному аналізі складу ґрунту.

3. Наукові та Open-source прототипи (напр. рішення на базі Kaggle Agricultural Datasets)

Системи, що розробляються дослідницькими групами для тестування нових архітектур машинного навчання.

Функціонал: Використання алгоритмів Random Forest, XGBoost або нейронних мереж для класифікації.

Переваги: Висока прозора точність на верифікованих даних.

Недоліки: Відсутність зручного інтерфейсу для кінцевого користувача та низька масштабованість.

У таблиці нижче наведено параметри оглянутих систем ключовими технічними характеристиками.

Таблиця 1.1. – Параметри порівняння систем

Параметр порівняння	Комерційні платформи	Мобільні додатки (AI)
Архітектура моделі	Глибокі Нейронні мережі	CNN (для зображень)
Вхідні дані	Супутник + IoT	Фото + GPS
Точність (Accuracy)	85–92% (варіюється)	80–90%
Кількість культур	Обмежена (основні)	Широка (діагностика)
Інтерфейс	Складні Dashboard	Спрощений мобільний

1.4. Аналіз моделей ансамблевого навчання для побудови систем підтримки прийняття рішень у сільському господарстві

У контексті глобальних викликів продовольчої безпеки та необхідності інтенсифікації агропромислового виробництва, цифровізація сільського господарства стає стратегічним пріоритетом. Ключовим аспектом підвищення продуктивності агроєкосистем є впровадження інтелектуальних систем підтримки прийняття рішень, які дозволяють оптимізувати процес вибору сільськогосподарських культур на основі мультимодальних даних про стан навколишнього середовища. Ефективне управління валовим збором врожаю безпосередньо корелює з точністю аналізу поживних речовин у ґрунті та моніторингу гідрометеорологічних показників.

Метою даної роботи є розробка та верифікація моделі машинного навчання для прогнозування оптимального типу сільськогосподарських культур. Для навчання та тестування було використано верифікований набір даних із репозиторію Kaggle, що містить 2200 спостережень. Структура вхідних даних представлена набором незалежних змінних у форматі CSV, що включають:

- концентрацію макроелементів у ґрунті (азот N, фосфор P, калій K);
- фізико-хімічні параметри (рівень кислотності pH);
- кліматичні індикатори (температура, відносна вологість повітря, рівень атмосферних опадів).

У роботі проведено порівняльний аналіз двох фундаментальних парадигм ансамблевого навчання: беггінгу (bagging) та бустингу (boosting).

Метод беггінгу: для побудови стійких прогнозів було застосовано базові класифікатори, такі як дерево рішень (Decision Tree) та метод опорних векторів (SVC). Особливу увагу приділено алгоритму випадкового лісу, який базується на агрегації результатів множини незалежних дерев рішень для зниження дисперсії моделі.

					БР.ІП – 06.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		27

Метод бустингу: реалізовано за допомогою класифікатора градієнтного бустингу (Gradient Boosting Classifier), де кожна наступна модель мінімізує функцію втрат попередньої, що дозволяє ефективно працювати зі складними нелінійними залежностями в агрономічних даних.

Експериментальні дослідження показали високу роздільну здатність запропонованих моделей. Після етапу гіперпараметричної оптимізації було встановлено:

- найвищу прогностичну точність (Accuracy) на рівні 99% продемонстрував класифікатор Random Forest. Це обумовлено здатністю алгоритму нівелювати проблему перенавчання (overfitting) та ефективно обробляти багатовимірні ознакові простори.

- алгоритм Gradient Boosting показав стабільний результат із точністю 98%, що підтверджує доцільність використання ітеративних підходів для уточнення класифікаційних границь.

Отримані результати підкреслюють високу ефективність ансамблевих методів у задачах прецизійного землеробства. Застосування моделі Random Forest як ядра інтелектуальної системи рекомендацій дозволяє мінімізувати ризики при плануванні посівних кампаній, забезпечуючи раціональне використання агресурсів. У подальшому дані моделі можуть бути інтегровані в комплексні системи моніторингу на базі IoT-протоколів (наприклад, MQTT) для забезпечення обробки даних у режимі реального часу безпосередньо з польових сенсорів.

1.5 Висновки по розділу

У першому розділі дипломної роботи проведено комплексний аналіз предметної області побудови класифікаційних моделей для потреб агрономії та досліджено сучасні підходи до інтелектуальної обробки аграрних даних. Встановлено, що традиційні методи оцінювання агроекологічних параметрів

					БР.ІП – 06.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		28

мають обмежену ефективність через значну залежність від людського фактора та складність аналізу великих обсягів інформації. У процесі дослідження розглянуто основні алгоритми класифікації, серед яких дерева рішень, машина опорних векторів, випадковий ліс і градієнтний бустинг, а також визначено їхні переваги й недоліки у задачах агрономічного прогнозування. Проведений аналіз існуючих систем і платформ обробки агроданих підтвердив актуальність використання ансамблевих моделей машинного навчання для підтримки прийняття рішень у сільському господарстві. Отримані результати дослідження стали теоретичною основою для подальшого проектування та реалізації класифікаційної системи підтримки агрономічних рішень.

					БР.ІП – 06.00.00.000 ПЗ	Арк.
						29
Змн.	Арк.	№ докум.	Підпис	Дата		

РОЗДІЛ 2. ІНТЕЛЕКТУАЛЬНІ ТЕХНОЛОГІЇ АНАЛІЗУ АГРОЕКОЛОГІЧНИХ ФАКТОРІВ І АЛГОРИТМІЧНА РЕАЛІЗАЦІЯ РЕКОМЕНДАЦІЙНОЇ СИСТЕМИ В АГРОНОМІЇ

2.1. Предиктивні моделі та інтелектуальні системи підтримки прийняття рішень у завданнях оптимізації посіву

Інтелектуальні системи рекомендації сільськогосподарських культур (Crop Recommendation Systems, CRS) на сучасному етапі розвитку агропромислового комплексу виступають ключовим інструментом оптимізації виробничих процесів. Основною функцією таких систем є надання аграріям верифікованої інформації щодо вибору оптимальної культури для посіву, що базується на багатофакторному аналізі нутрієнтного складу ґрунту та динамічних показників кліматичних умов. Моделювання здійснюється шляхом врахування специфічних параметрів локації, зокрема концентрації макроелементів (N, P, K) та гідрометеорологічних індикаторів. Для побудови високоточних рекомендацій застосовується широкий стек методів машинного навчання, серед яких: дерево рішень, випадковий ліс, метод k-найближчих сусідів та наївний байєсівський класифікатор. Даний підхід забезпечує не лише зростання продуктивності врожаю, а й підвищення його якісних характеристик за рахунок адаптивного управління агроресурсами.

Окремим напрямом досліджень є застосування теорії нечіткої логіки. Зокрема, було розроблено систему нечіткого виведення (Fuzzy Inference System, FIS), яка дозволяє фермерам приймати рішення в умовах невизначеності кліматичних чинників. Експериментальна апробація даної моделі на базі ретроспективного набору даних (зібраних у штаті Раджастан протягом десятирічного періоду) продемонструвала прогностичну точність на рівні 91%. Впровадження FIS-моделей сприяє підвищенню врожайності та

					БР.ІП – 06.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		30

фінансової стабільності аграрних підприємств шляхом нівелювання ризиків, пов'язаних із волатильністю метеорологічних умов.

Розвиток концепції Інтернету речей (IoT) дозволив інтегрувати засоби безперервного моніторингу в архітектуру CRS. Сенсорні мережі IoT забезпечують стрімінговий збір даних про параметри ґрунту та стан навколишнього середовища, що в подальшому обробляються алгоритмами багатокласової класифікації. Порівняльний аналіз різноманітних технік машинного навчання показав, що класифікатор випадкового лісу (Random Forest) демонструє найвищу ефективність при роботі з відкритими агрономічними наборами даних. Це підкреслює значний потенціал систем на базі IoT для створення масштабованих та автономних рішень у прецизійному землеробстві.

Для вирішення завдань імовірнісного міркування та оцінки взаємовпливу параметрів доцільним є використання байєсових моделей. Інтегрована байєсова мережа переконань (Bayesian Belief Network, BBN) дозволяє агрегувати кліматичні дані, морфологічні характеристики ґрунту та історичні показники врожайності в єдину структуру.

Основними перевагами BBN у контексті CRS є:

- Врахування невизначеності: здатність моделі оперувати неповними або зашумленими даними без втрати якості прогнозу.
- Динамічна адаптація: корегування рекомендацій у режимі реального часу відповідно до змін зовнішнього середовища та специфічних обмежень господарства.
- Висока інтегративність: можливість поєднання з існуючими алгоритмами навчання та програмними архітектурами.

Таким чином, застосування байєсових мереж у синергії з методами навчання на великих даних забезпечує якісно новий рівень стійкості та індивідуалізації аграрних рішень.

					БР.ІП – 06.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		31

2.2. Предиктивна аналітика врожайності на основі гетерогенних ансамблів та методів стекінгу

Роль аграрного сектору в системі глобальної економіки та забезпеченні сталого розвитку територій залишається визначальною. Зокрема, на прикладі економіки Індії встановлено, що сільське господарство забезпечує зайнятість приблизно 50% працездатного населення, виступаючи фундаментальним драйвером економічного зростання. Однак ефективність галузі нівелюється проблемою ірраціонального вибору сільськогосподарських культур. Основним дестабілізуючим фактором є дефіцит верифікованої інформації щодо фізико-хімічного складу ґрунтів та стохастичних екологічних змінних, що призводить до суттєвого зниження показників врожайності.

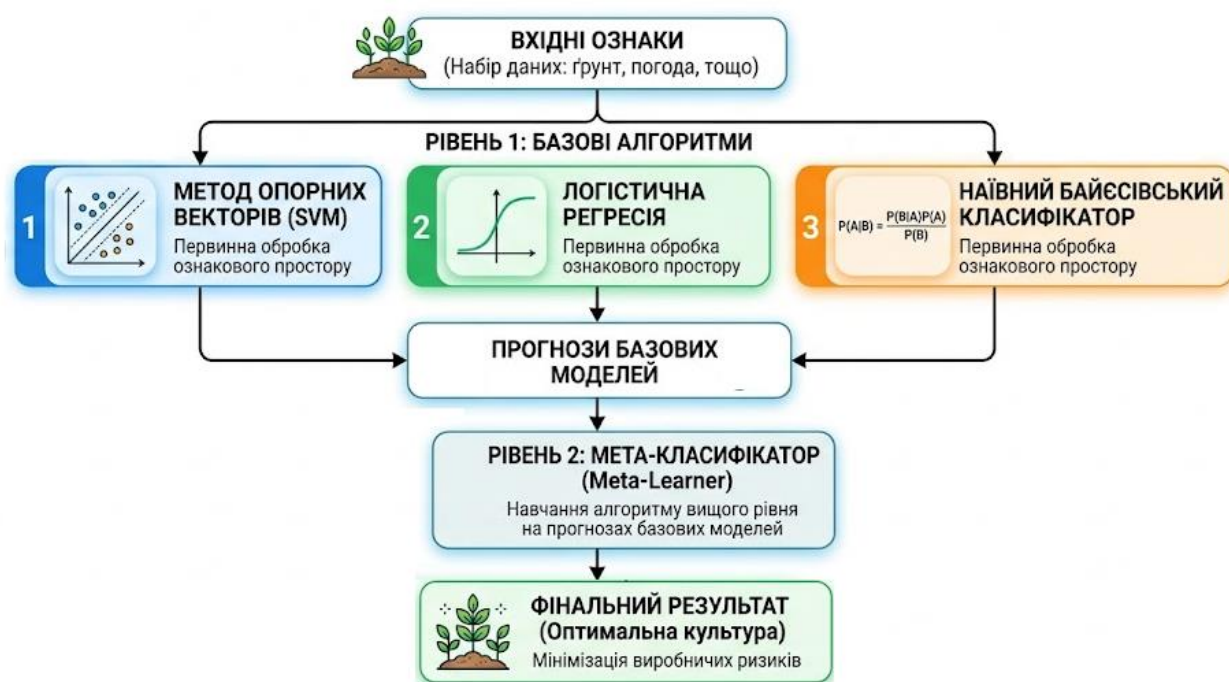


Рисунок 2.1 – Архітектура гетерогенного ансамблю за стратегією стекінгу

Для розв’язання задачі оптимізації вибору культур та мінімізації виробничих ризиків доцільним є впровадження методів предиктивного моделювання. У роботі пропонується застосування гетерогенного ансамблю

за стратегією стекінгу (рис. 2.1). Суть даного підходу полягає у дворівневій архітектурі:

- базові алгоритми (Base Learners): використання комбінації методу опорних векторів (SVM), логістичної регресії та наївного байєсового класифікатора для первинної обробки ознакового простору.

- мета-класифікатор (Meta-Learner): навчання алгоритму вищого рівня на прогнозах базових моделей для отримання фінального результату.

Така стратегія дозволяє досягти високої прогностичної точності — 99%, забезпечуючи надійне підґрунтя для формування продовольчої безпеки та переходу до концепції сталого сільського господарства.

Процес ідентифікації оптимальних культур базується на аналізі багатовимірного вектора ознак x , що включає:

- Хімічні показники: концентрація азоту (N), фосфору (P) та калію (K);
- Фізико-хімічні властивості: показник кислотності ґрунту (pH);
- Метеорологічні параметри: температура навколишнього середовища, відносна вологість повітря та середньорічний об'єм атмосферних опадів.

Для навчання моделі використовуються ретроспективні дані попередніх вегетаційних періодів, що дає змогу адаптувати рекомендації до специфіки конкретної локації та мінімізувати фінансові втрати аграріїв.

Результати експериментальних досліджень підтверджують статистичну перевагу нелінійних ансамблевих методів над традиційними статистичними підходами. Зокрема, порівняння моделі випадкового лісу із класичною моделлю множинної лінійної регресії продемонструвало перевагу RF за всіма ключовими метриками ефективності (R^2 , MSE, MAE).

Висока роздільна здатність алгоритму Random Forest в аналізі гетерогенних даних пояснюється його здатністю до агрегації множини незалежних дерев рішень, що нівелює проблему високої дисперсії та забезпечує стійкість системи в умовах шуму вхідних сенсорних даних.

					БР.ІП – 06.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		33

Впровадження інтелектуальних систем підтримки прийняття рішень на основі ансамблевих методів навчання дозволяє трансформувати емпіричний підхід до землеробства у науково обґрунтовану технологію. Це створює умови для прецизійного управління агропромисловими ресурсами, підвищення якості продукції та зміцнення фінансової стабільності виробників.

2.3. Обґрунтування вибору інструментарію та середовища розробки предиктивних моделей

Для реалізації архітектури запропонованої системи підтримки прийняття рішень було обрано високорівневу мову програмування Python. Вибір обумовлений наявністю розвиненої екосистеми спеціалізованих бібліотек для наукових обчислень, високою швидкістю прототипування та підтримкою сучасних парадигм об'єктно-орієнтованого програмування. Розробка та налагодження програмних модулів здійснювалися в інтегрованому середовищі PyCharm IDE, що забезпечує розширені можливості статичного аналізу коду, рефакторингу та інтеграції з системами контролю версій.

Центральним компонентом підсистеми управління даними є бібліотека Pandas (версія 2.0.3). Вона використовується для реалізації ETL-процесів (Extract, Transform, Load), зокрема для імпорту структурованих даних із форматів CSV, Excel та реляційних баз даних (SQL). За допомогою структур даних DataFrame забезпечується:

- фільтрація та очищення вхідних вибірок від аномалій;
- агрегація та групування агрономічних показників за територіальною ознакою;
- обробка пропущених значень, що є критичним для забезпечення цілісності навчальної вибірки.

					БР.ІП – 06.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		34

Низькорівнева обробка багатовимірних масивів та виконання складних матричних операцій реалізовано на базі бібліотеки NumPy. Завдяки використанню оптимізованих алгоритмів на мові C, NumPy дозволяє мінімізувати споживання оперативної пам'яті та підвищити обчислювальну потужність системи при виконанні лінійно-алгебраїчних перетворень, необхідних для функціонування предиктивних моделей.

Для розробки інтелектуального ядра системи використано фреймворк Scikit-learn. Дана бібліотека надає стандартизований інтерфейс для реалізації повного циклу машинного навчання:

- масштабування ознак: нормалізація даних про вміст нутрієнтів та метеорологічні умови.
- навчання моделей: імплементація алгоритмів класифікації (Random Forest, SVM, Gradient Boosting тощо).
- оцінка якості: використання метрик точності (Accuracy), повноти (Recall) та F1-score для валідації прогнозів.

Безшовна інтеграція Scikit-learn з NumPy та Pandas дозволяє автоматизувати конвеєри (pipelines) обробки даних, забезпечуючи високу відтворюваність експериментів.

Аналіз патернів у розподілі сільськогосподарських культур та ідентифікація кореляційних зв'язків між екологічними факторами реалізовані засобами бібліотек Matplotlib та Seaborn. Matplotlib застосовується для побудови базових 2D та 3D графічних представлень, що дозволяють візуалізувати динаміку врожайності. Seaborn, базуючись на структурах Pandas, використовується для генерації статистичних візуалізацій (теплових карт кореляції, діаграм розсіювання), що сприяє глибшому розумінню впливу окремих змінних (рН ґрунту, рівень опадів) на результативність прогнозу.

Комплексне застосування зазначеного програмного стеку дозволило сформувати стабільну та масштабовану платформу для розгортання системи

					БР.ІП – 06.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		35

рекомендацій, здатної функціонувати в умовах реальних агровиробничих сценаріїв.

2.4. Опис та статистична характеристика набору даних для предиктивного моделювання агросистеми

Для розробки та верифікації моделей інтелектуальної підтримки прийняття рішень у галузі прецизійного землеробства було використано верифікований набір даних із репозиторію Kaggle. Вибірка складається з 2200 спостережень, кожне з яких представлено вектором ознак у восьмивимірному просторі. Структура даних охоплює ключові агроекологічні фактори, що безпосередньо корелюють із продуктивністю сільськогосподарських культур.

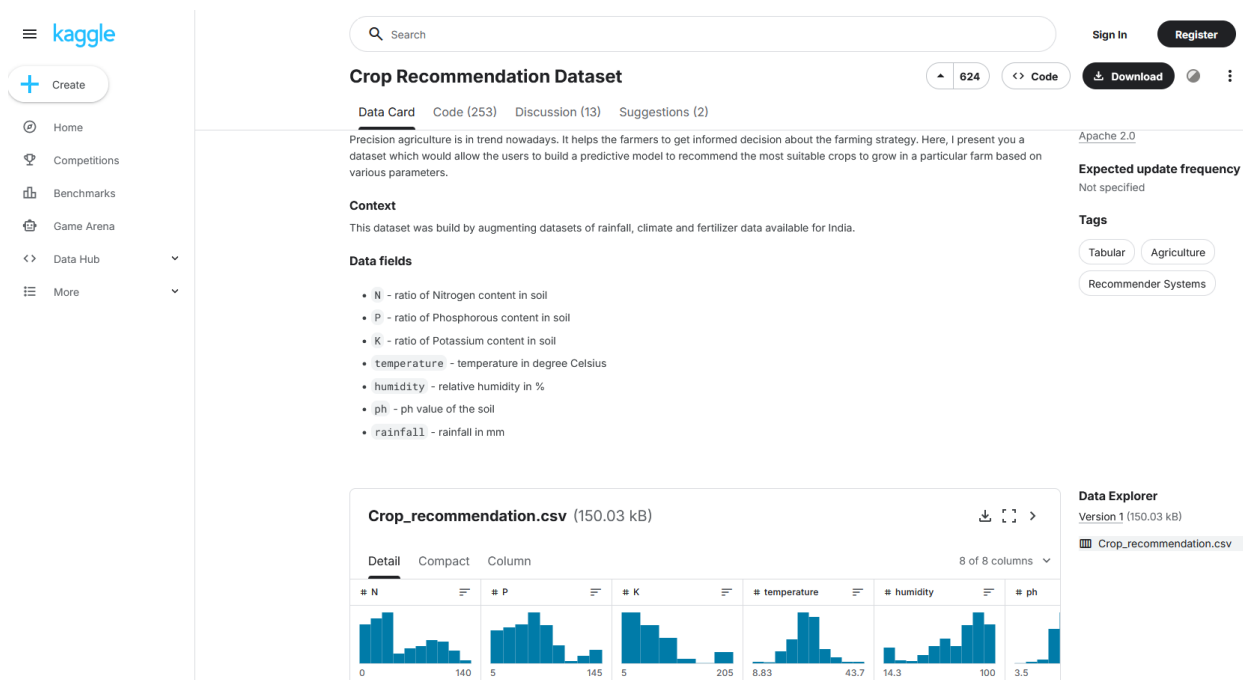


Рисунок 2.2 – Набір даних Crop Recommendation Dataset (kaggle)

Опишемо ознаковий простір моделі. Вхідні параметри системи включають як ґрунтові характеристики, так і метеорологічні показники:

- Макроелементи (NPK): концентрація азоту (N), фосфору (P) та калію (K) у ґрунті.

					БР.ІП – 06.00.00.000 ПЗ	Арк.
						36
Змн.	Арк.	№ докум.	Підпис	Дата		

- Фізико-хімічні показники: рівень кислотності (pH).
- Кліматичні чинники: температура навколишнього середовища, відносна вологість повітря та середньодобова кількість опадів.

Кожен запис у масиві даних відображає унікальну комбінацію ендегенних (грунтових) та екзогенних (атмосферних) умов, що слугує базисом для розробки багатокласної моделі класифікації.

Цільова змінна представлена 22 типами культур, які для системного аналізу згруповано за функціональним призначенням:

- Зернові та зернобобові (продовольчі): рис, кукурудза, нут та інші.
- Плодово-ягідні (фруктові): гранат, банан, манго тощо.
- Технічні культури: бавовна, джут, кава.

Для забезпечення збалансованості навчальної вибірки кожна категорія представлена рівною кількістю спостережень — по 200 записів на кожен тип культури. Це мінімізує ризик зміщення моделі на користь мажоритарних класів під час навчання алгоритмів машинного навчання.

У таблиці 2.1 наведено результати первинної статистичної обробки даних, що демонструють діапазони коливань основних атрибутів.

Таблиця 2.1 - Статистичні характеристики вхідних агроекологічних параметрів

Атрибут (Ознака)	Мінімальне значення	Максимальне значення	Середнє значення
Концентрація азоту (N), мг/кг	0	140	37
Концентрація калію (K), мг/кг	5	145	51
Концентрація фосфору (P), мг/кг	5	205	32
Температура повітря, °C	8	44	25
Відносна вологість, %	14	99	80
Рівень кислотності (pH)	3	10	6
Рівень опадів, мм	20	298	94

Аналіз даних свідчить про значну варіативність ознак, зокрема рівня опадів та концентрації фосфору, що підкреслює складність задачі класифікації та необхідність застосування потужних нелінійних апроксиматорів, таких як ансамблеві методи навчання або глибокі нейронні мережі. Отримані статистичні межі стають основою для етапу нормалізації даних перед подачею на вхід інтелектуальної системи.

2.5. Статистичний аналіз ознакового простору та розробка ансамблевих предиктивних моделей

Етап розвідувального аналізу даних (EDA) є критично важливою складовою розробки інтелектуальних систем підтримки прийняття рішень, оскільки він дозволяє ідентифікувати приховані закономірності, аномалії та специфічні патерни в структурі вхідного набору даних. Детальна візуалізація та статистична інтерпретація параметрів ознакового простору забезпечують глибше розуміння нелінійних взаємозв'язків між агроекологічними факторами та цільовими показниками врожайності на масиві з 2200 спостережень.

2.5.1. Аналіз розподілу та асиметрії ознак

Для оцінки статистичної природи вхідних змінних було побудовано гістограми розподілу (рис. 2.3). Аналіз візуалізацій дозволяє зробити висновки щодо наявності викидів, які, завдяки специфіці обраних алгоритмів, не справляють дестабілізуючого впливу на прогностичну здатність системи.

					БР.ІП – 06.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		38

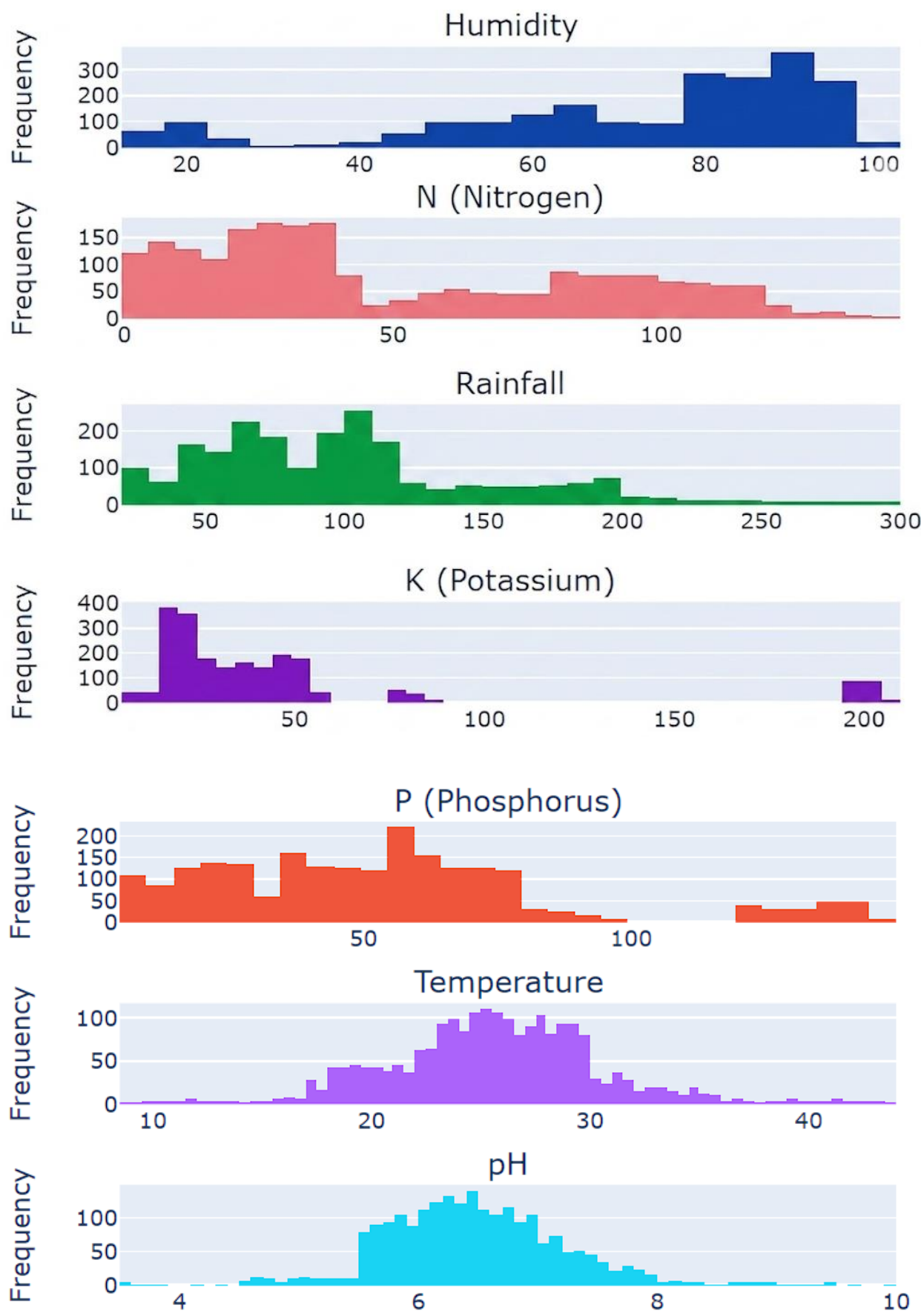


Рисунок 2.3 – Гістограми розподілу ознак

Особлива увага була приділена дослідженню асиметрії розподілів таких факторів, як температура, вологість, концентрація азоту та об'єм опадів:

- Концентрація азоту (N) та рівень опадів: демонструють виражену правосторонню (позитивну) асиметрію, що свідчить про домінування низьких та середніх значень у вибірці.

- Відносна вологість: характеризується лівосторонньою (негативною) асиметрією, що вказує на зміщення даних у бік вищих значень.

- Температура: підпорядковується закону нормального розподілу, що свідчить про рівномірне охоплення варіативного діапазону температурних показників у досліджуваних агроєкосистемах.

2.5.2. Ідентифікація аномалій та стійкість моделей

На рисунку 2.4 представлено діаграми розмаху для таких макроелементів, як фосфор (P) та калій (K). Даний метод візуалізації застосовано для детекції статистичних викидів та аномалій. Аналіз показав, що в даних по калію спостерігаються специфічні скупчення викидів, які потребують особливої уваги при виборі архітектури класифікатора.

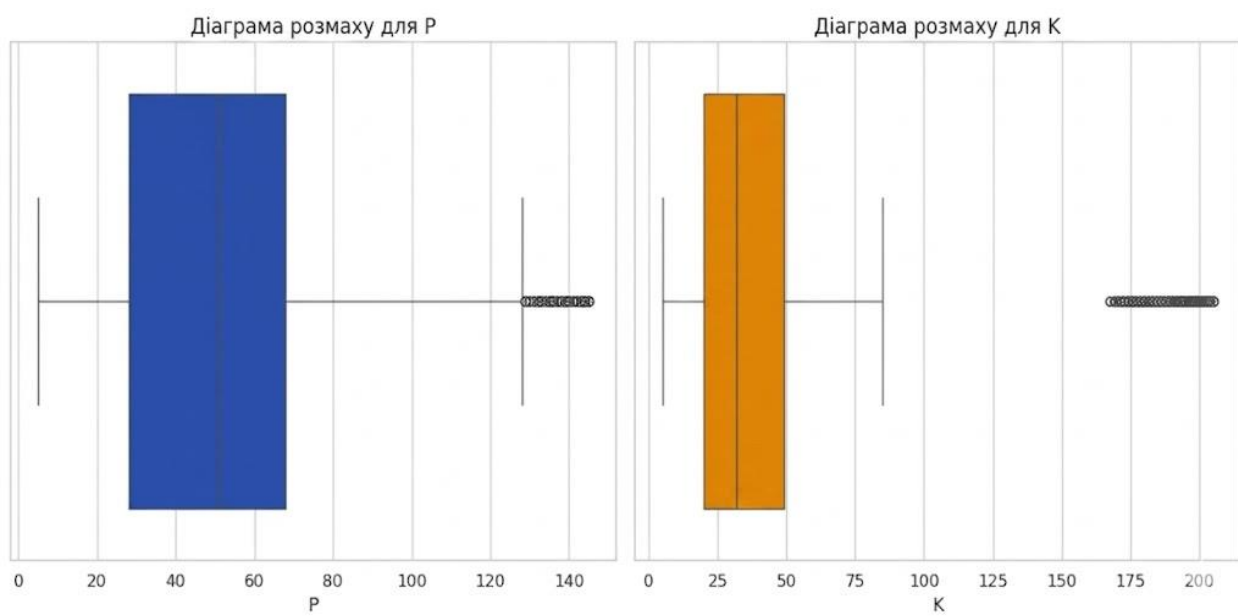


Рисунок 2.4 – Діаграми розмаху для макроелементів

Для нівелювання впливу аномальних значень у роботі було використано стійкі (робастні) техніки машинного навчання, зокрема випадковий ліс та дерева рішень. На відміну від лінійних моделей, зазначені алгоритми є менш чутливими до екстремальних значень завдяки механізмам ієрархічного розбиття простору ознак.

2.5.3. Кореляційний аналіз та взаємозв'язки параметрів

На рисунку 2.5 зображено теплову карту кореляційної матриці, яка відображає ступінь лінійної залежності між нутрієнтами ґрунту та кліматичними факторами. Градація кольорів на карті варіюється від темних відтінків (коефіцієнт кореляції 0.2, слабкий зв'язок) до яскравих (коефіцієнт 1.0, повна кореляція).

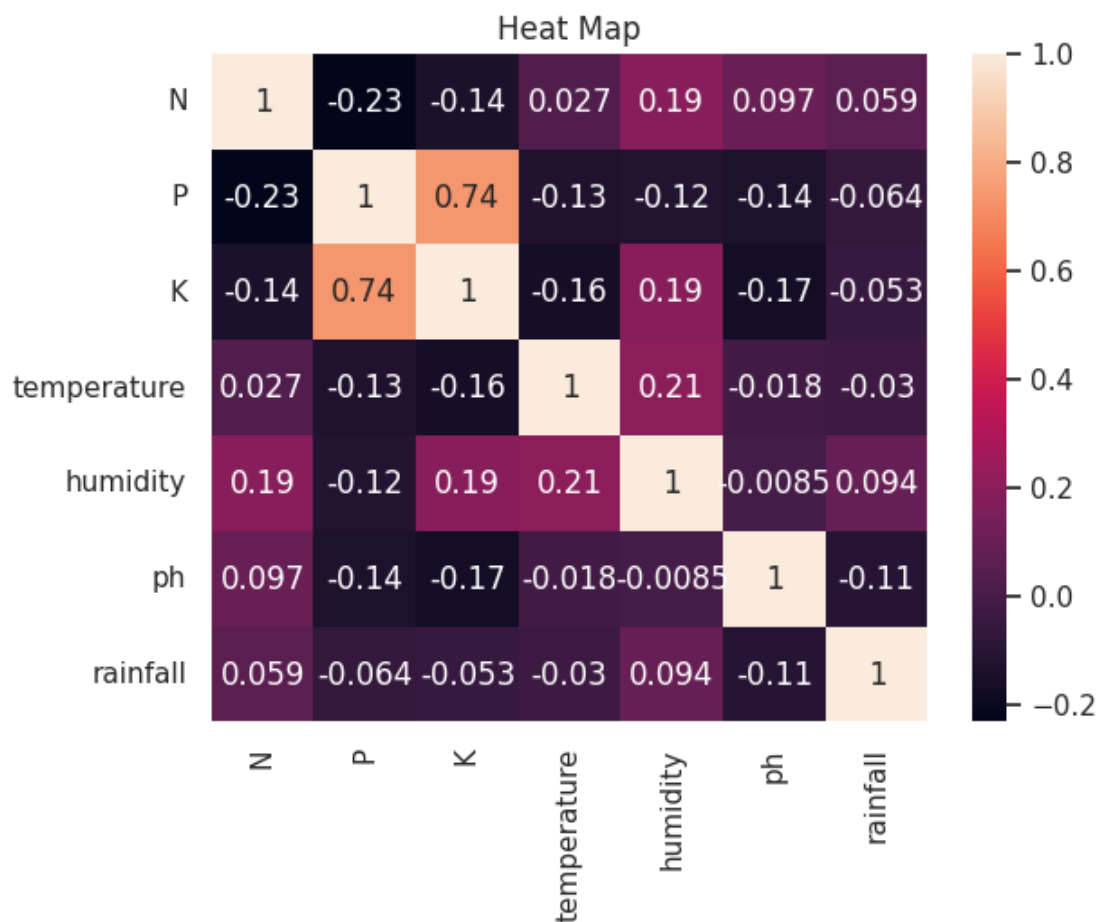


Рисунок 2.5 - Теплова карта кореляційної матриці

Представлена матриця є результатом статистичного аналізу ознакового простору набору даних, що використовується для навчання предиктивних моделей системи AgroSense. Вона візуалізує силу та напрямок лінійних взаємозв'язків між агрохімічними показниками ґрунту та метеорологічними чинниками за допомогою коефіцієнта кореляції Пірсона (r).

1. Інтерпретація кольорової шкали та діагональних значень

Матриця використовує градієнтну колірну схему, де світлі відтінки (кремовий) відповідають позитивній кореляції, а темні (чорний/пурпуровий) — слабкій або негативній.

Діагональні елементи: мають значення 1.0, що підтверджує ідеальну автокореляцію кожної змінної самої з собою.

Діапазон значень: коефіцієнти в матриці варіюються від приблизно -0.23 до +0.74, що вказує на наявність як слабких зворотних, так і сильних прямих зв'язків між окремими предикторами.

2. Аналіз ключових взаємозв'язків (Кореляційні пари)

Сильна позитивна кореляція: Р (Фосфор) та К (Калій)

Найбільш статистично значущим взаємозв'язком у наборі даних є кореляція між концентрацією фосфору та калію, що становить 0.74.

Агрономічне обґрунтування: такий високий показник часто обумовлений комплексним внесенням мінеральних добрив, де ці два елементи присутні одночасно.

Вплив на ML-моделі: наявність мультиколінеарності між цими ознаками може впливати на стабільність деяких лінійних алгоритмів, що підтверджує доцільність використання ансамблевих методів (зокрема, дерев рішень), які є менш чутливими до таких залежностей.

Слабкі та помірні кореляції:

- Температура та вологість (Humidity): коефіцієнт становить 0.21, що відображає помірну кліматичну залежність у досліджуваних умовах.

					БР.ІП – 06.00.00.000 ПЗ	Арк.
						42
Змн.	Арк.	№ докум.	Підпис	Дата		

- Азот (N) та Вологість: зв'язок на рівні 0.19 вказує на слабку позитивну кореляцію, що може бути пов'язано з процесами вимивання азоту або його розчинності в умовах підвищеної вологості.

3. Аналіз незалежних ознак (Слабка кореляція)

Ряд показників демонструє майже повну лінійну незалежність від інших факторів, що робить їх унікальними джерелами інформації для класифікатора:

- Показник кислотності (pH): демонструє дуже слабкі від'ємні або близькі до нуля зв'язки з більшістю параметрів (наприклад, -0.0085 з вологістю та -0.018 з температурою). Це вказує на те, що рівень pH є автономною характеристикою ділянки.

- Опади (Rainfall): показники кореляції з основними нутрієнтами (N, P, K) є мінімальними (від -0.064 до 0.059), що підтверджує статус опадів як незалежної зовнішньої змінної.

Аналіз матриці дозволяє зробити висновок, що ознаковий простір системи не є надлишковим (за винятком пари P–K). Більшість атрибутів мають низьку взаємну кореляцію, що свідчить про їх високу інформативність для моделі.

Відсутність сильної лінійної залежності між більшістю факторів (коефіцієнти в діапазоні -0.2...0.2) підтверджує необхідність використання нелінійних апроксиматорів, таких як ансамблева модель, обрана для системи AgroSense, оскільки вона здатна ефективно виявляти складні патерни, які не описуються простою лінійною кореляцією.

За результатами аналізу встановлено:

- Сильна позитивна кореляція (0.74) між концентрацією фосфору (P) та калію (K), що пояснюється їхньою спільною присутністю в комплексних мінеральних добривах.

- Помірна кореляція (0.21) між температурою та вологістю повітря.

- Автономність параметрів: азот (N) демонструє слабку кореляцію з іншими елементами, тоді як показник кислотності (pH) та кількість опадів є

					БР.ІП – 06.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		43

практично незалежними змінними, що підвищує їхню інформативність як предикторів.

2.5.4. Ансамблеве моделювання та оптимізація

Парадигма ансамблевого навчання полягає в агрегації прогнозів кількох базових моделей для підвищення стабільності та точності фінального результату. У дослідженні було реалізовано та порівняно дві фундаментальні стратегії: беггінг (bagging) та бустинг (boosting).

1. Модель беггінгу на основі Random Forest

Для побудови моделі беггінгу використано ансамбль базових оцінювачів: дерева рішень, випадковий ліс та класифікатор опорних векторів. Оптимізація гіперпараметрів (зокрема `base_estimators` та `max_samples`) здійснювалася за допомогою методу `GridSearchCV`, що забезпечує вичерпний пошук найкращої конфігурації на заданій сітці параметрів.

Процедура верифікації включала перехресну перевірку (cross-validation) за 5 фолдами. За результатами порівняння точність класифікатора випадкового лісу досягла показника 99% на тестовій вибірці.

2. Модель градієнтного бустингу

Модель бустингу базується на алгоритмі `Gradient Boosting Classifier`. В ході експериментів проводилося налаштування параметрів кількості ітерацій (`n_estimators`), швидкості навчання (`learning_rate`) та максимальної глибини дерев (`max_depth`). Аналогічно до попередньої моделі, для пошуку оптимальних значень використовувався `GridSearchCV` з 5-кратною крос-валідацією. Точність даної моделі склала 98%.

На основі отриманих метрик ефективності для подальшої інтеграції в програмний інтерфейс користувача та розгортання в межах інтелектуальної системи було обрано модель на основі беггінгу (Random Forest) як таку, що демонструє найвищу прогностичну здатність та стійкість до викидів.

					БР.ІП – 06.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		44

2.6 Висновки по розділу

У другому розділі роботи досліджено інтелектуальні технології аналізу агроекологічних факторів та реалізовано підхід до побудови предиктивної системи підтримки прийняття рішень у сфері агрономії. Обґрунтовано вибір інструментальних засобів і програмного середовища, які забезпечують ефективну реалізацію алгоритмів машинного навчання та обробки аграрних даних. Проведений статистичний аналіз набору даних дозволив виявити закономірності розподілу ознак, оцінити кореляційні взаємозв'язки між параметрами та визначити найбільш інформативні характеристики для побудови класифікаційної моделі. У процесі дослідження також здійснено виявлення аномалій та оцінено стійкість моделей до впливу шумових даних, що дало можливість підвищити якість прогнозування. Результати ансамблевого моделювання підтвердили високу ефективність використання комбінованих алгоритмів класифікації для задач оптимізації агрономічних рішень.

					БР.ІП – 06.00.00.000 ПЗ	Арк.
						45
Змн.	Арк.	№ докум.	Підпис	Дата		

РОЗДІЛ 3. ПРОЕКТУВАННЯ АРХІТЕКТУРИ ТА ПРОГРАМНА РЕАЛІЗАЦІЯ РЕКОМЕНДАЦІЙНОЇ СИСТЕМИ ОБРОБКИ АГРАРНИХ ДАНИХ

3.1. Візуальне моделювання системних компонентів та процесів засобами уніфікованої мови моделювання

У процесі розробки складних програмних рішень, зокрема інтелектуальних систем підтримки прийняття рішень у сільському господарстві, критично важливим етапом є архітектурне моделювання. Використання діаграм уніфікованої мови моделювання UML забезпечує детальну візуалізацію структурних компонентів системи та їхніх ієрархічних зв'язків. Це дозволяє зацікавленим сторонам (розробникам, дослідникам та кінцевим користувачам) сформулювати цілісне розуміння функціональних можливостей та внутрішньої логіки програмного продукту.

Методологічна значущість UML-моделювання в проєктах машинного навчання полягає в координації етапів життєвого циклу моделі: від препроцесингу та структурування вхідних масивів даних до конфігурації розгортання та операційного управління ансамблевими моделями.

Інтеграція UML у процес проєктування системи рекомендацій сільськогосподарських культур спрямована на досягнення таких цілей:

- Стандартизація візуалізації: надання розробникам експресивної мови графічного моделювання, що мінімізує неоднозначність при інтерпретації архітектурних рішень.
- Концептуальна верифікація: сприяння формалізації нових концепцій на етапі проєктування системи.
- Оптимізація кодогенерації: створення підґрунтя для розробки об'єктно-орієнтованого коду на основі графічних моделей.

					БР.ІП – 06.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		46

- Забезпечення модуляризації: впровадження принципів розподілу відповідальності через використання патернів проєктування, фреймворків та автономних компонентів.

- Інтеграція найкращих практик: акумуляція передового досвіду інженерії програмного забезпечення для створення стійких та масштабованих моделей.

3.1.1. Статична структура системи з використанням діаграми класів

Діаграма класів, представлена на рисунку 3.1, відображає статичну структуру системи, визначаючи типи об'єктів та їхні взаємозв'язки. Архітектура базується на двох фундаментальних класах: «Фермер» (Farmer) та «Ансамблеві моделі» (Ensemble Models).

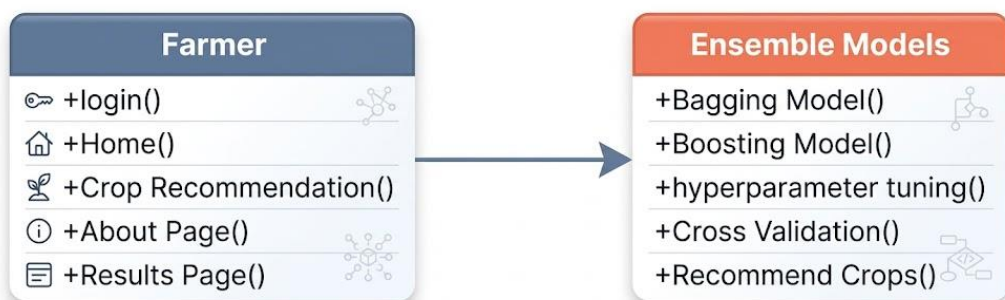


Рисунок 3.1 – Діаграма класів

Клас «Фермер» інкапсулює поведінку кінцевого користувача. Доступ до функціоналу здійснюється через методи навігації між інтерфейсними формами: автентифікації, головної панелі, параметризації запиту та візуалізації результатів.

Клас «Ансамблеві моделі» відповідає за обчислювальну логіку системи. Він агрегує методи обробки вхідних векторів ознак (N, P, K, кліматичні чинники) та генерації предиктивних значень, що згодом передаються на сторінку результатів для інтерпретації.

3.1.2. Моделювання функціональних вимог з використанням діаграми варіантів використання

Функціональні межі системи та взаємодія зовнішніх суб'єктів із її сервісами деталізовані на діаграмі варіантів використання (рисунок 3.2). У системі виділено три ключові актори:

- Фермер ініціює процеси введення агрономічних параметрів, отримує результати предиктивного аналізу та переглядає згенеровані рекомендації.
- Дослідник/Розробник здійснює адміністрування системи, включаючи попередню обробку даних (data cleaning) та навчання (training) ансамблевих моделей.
- Система рекомендацій виступає як автоматизований агент, що забезпечує виконання алгоритму класифікації на основі навчених моделей та оновлення вихідних даних відповідно до преференцій користувача.

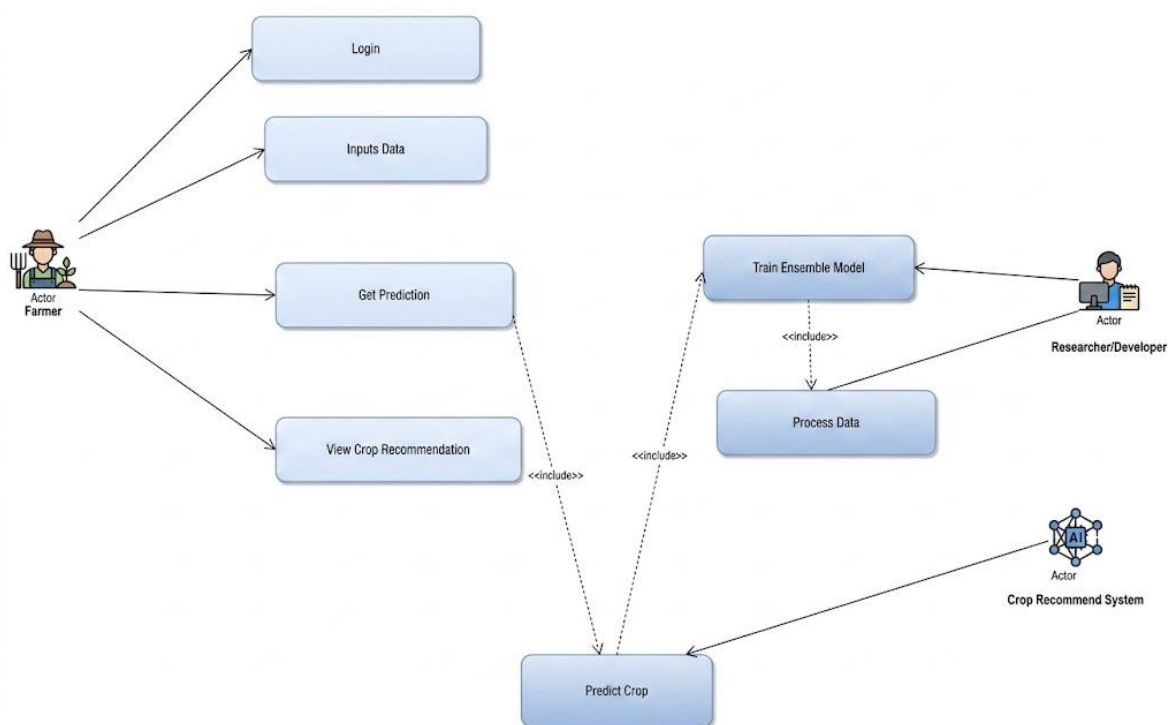


Рисунок 3.2 - Діаграма варіантів використання

Архітектура взаємодії базується на трьох основних акторах, кожен з яких має специфічний набір привілеїв та завдань:

- Актор Фермер виступає як кінцевий користувач системи. Його функціональний профіль охоплює операції з автентифікації («Login»), ініціалізації запиту шляхом введення агрономічних параметрів («Inputs Data»), формування запиту на отримання предиктивного висновку («Get Prediction») та подальшої візуалізації результатів у формі конкретних порад щодо посіву («View Crop Recommendation»).

- Актор Дослідник/Розробник відповідає за технічне забезпечення та високу точність інтелектуального ядра. Його діяльність спрямована на реалізацію складних процесів підготовки даних («Process Data») та безпосереднє навчання предиктивних алгоритмів («Train Ensemble Model»).

- Актор Система рекомендацій функціонує як автоматизований внутрішній агент, що реалізує безпосередній обчислювальний процес класифікації та генерації прогнозу («Predict Crop»).

Діаграма демонструє складну ієрархію залежностей, реалізовану через відношення включення, що дозволяє модульно структурувати бізнес-логіку системи:

1. Централізація прогнозування: варіант використання «Predict Crop» є стрижневим компонентом системи. Він автоматично активується (включається) як при запиті користувача на прогноз («Get Prediction»), так і в процесі валідації під час навчання ансамблевої моделі. Це гарантує використання ідентичних обчислювальних механізмів на різних стадіях життєвого циклу системи.

2. Технологічний стек навчання: процес «Train Ensemble Model» неможливий без попередньої підготовки масивів даних. Відповідно, він включає прецедент «Process Data», що передбачає очищення, нормалізацію та структурування вхідних агроекологічних показників перед подачею на вхід ансамблевих алгоритмів.

					БР.ІП – 06.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		49

3. Ієрархія доступу: чітке розмежування повноважень між акторами (Фермером та Розробником) забезпечує безпеку та стабільність системи, де користувач отримує готові рішення, а експерт — інструменти для вдосконалення математичного апарату.

Таким чином, представлена діаграма описує замкнений цикл функціонування інтелектуального рішення: від адміністрування та навчання моделей дослідником до отримання практичної цінності кінцевим споживачем в особі фермера.

3.1.3. Розробка діаграми послідовностей

Діаграма послідовностей (рисунок 3.3) ілюструє часове впорядкування подій та механізми передачі повідомлень між об'єктами системи. Сценарій взаємодії починається з автентифікації користувача, після чого здійснюється перехід до інтерфейсу введення параметрів.

					БР.ІП – 06.00.00.000 ПЗ	Арк.
						50
Змн.	Арк.	№ докум.	Підпис	Дата		

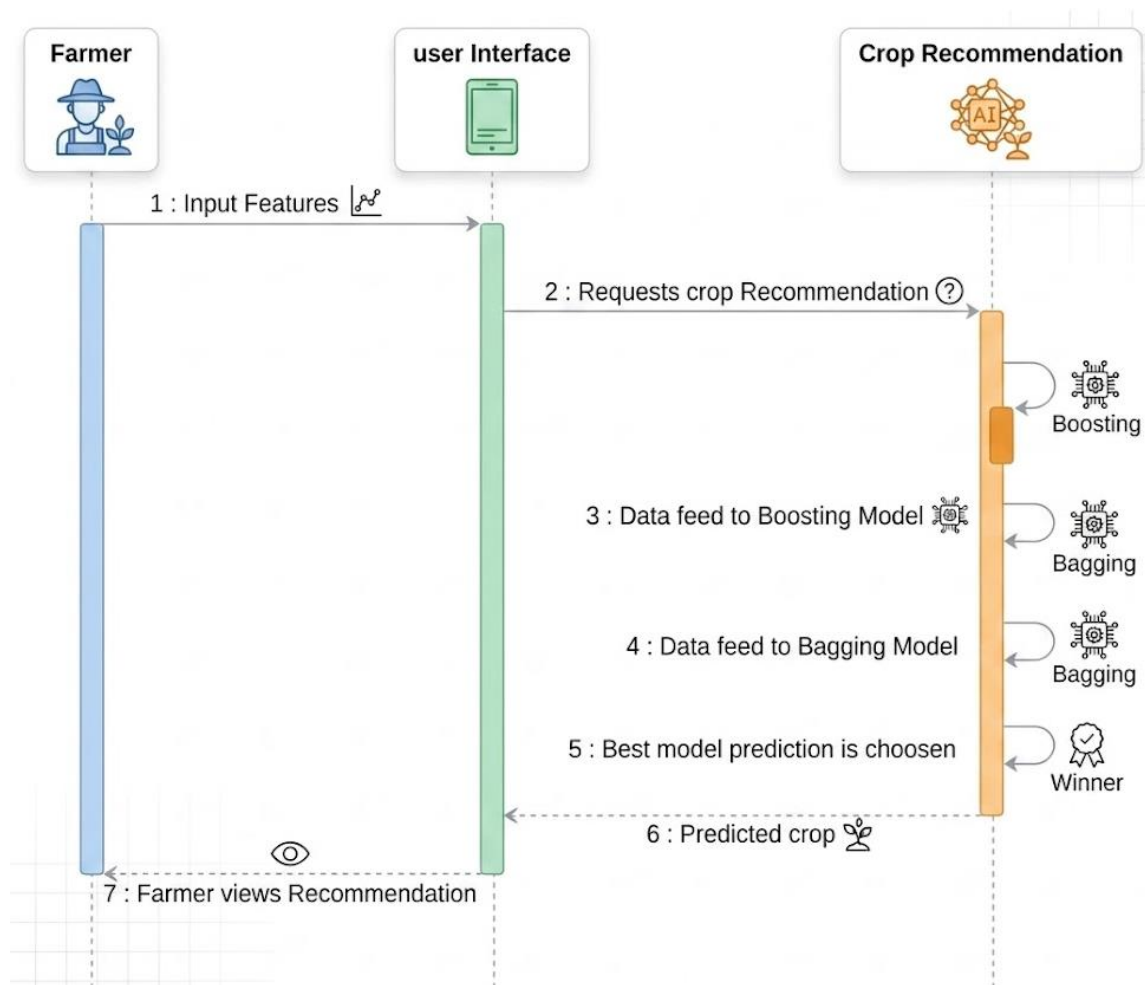


Рисунок 3.3 – Діаграма послідовностей

Передача значень вмісту макроелементів (азоту, фосфору, калію) та метеорологічних показників активує обчислювальний модуль. Система ініціює виклик ансамблевої моделі, яка виконує класифікацію вхідного вектора даних для визначення оптимальної сільськогосподарської культури. Завершальним етапом є повернення результату та його відображення на сторінці результатів разом із вихідними параметрами запиту.

Більш детально розглянемо етапи функціонування та потоки даних. Процес реалізується через взаємодію трьох ключових сутностей: актора (Farmer), системного інтерфейсу (User Interface) та аналітичного модуля (Crop Recommendation).

1. Ініціалізація та передача ознак. Процес починається з ініціації запиту користувачем (Фермером), який через інтерфейс вводить вектор ознак (Input Features), що включає хімічні показники ґрунту та метеорологічні параметри.

Графічний інтерфейс трансформує ці дані у формалізований запит на отримання рекомендації, спрямовуючи його до обчислювального ядра системи.

2. Обчислювальний цикл ансамблевого моделювання. На рівні модуля «Crop Recommendation» відбувається ітераційна обробка даних за допомогою ансамблевих стратегій навчання. Спершу дані інтегруються в модель бустингу (Boosting Model), після чого паралельно або послідовно здійснюється їх обробка моделлю беггінгу (Bagging Model). Такий підхід дозволяє врахувати різні аспекти нелінійних залежностей у даних.

3. Селекція оптимального прогнозу. Після завершення роботи обох ансамблевих моделей система виконує етап порівняльного аналізу. На основі встановлених критеріїв точності здійснюється вибір найкращого предиктивного значення (Best model prediction is chosen), що забезпечує максимальну релевантність рекомендації.

4. Зворотний зв'язок та візуалізація. Результируючий показник (Predicted crop) повертається від обчислювального модуля до інтерфейсу користувача. Завершальним етапом є візуалізація рекомендації, що дозволяє фермеру ознайомитися з прогнозом та прийняти рішення щодо подальшої сільськогосподарської діяльності.

Таким чином, діаграма демонструє замкнений цикл обробки інформації, де вхідні параметри навколишнього середовища проходять через багаторівневий фільтр ансамблевих моделей для формування точного агрономічного прогнозу.

3.1.4. Реалізація процедурної логіки за допомогою діаграми діяльності

Потік керування та логіка прийняття рішень у межах системи візуалізовані на діаграмі діяльності (рисунок 3.4). Вона описує послідовність кроків: від входу в систему з обов'язковою верифікацією облікових даних до генерації фінальної рекомендації.

					БР.ІП – 06.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		52

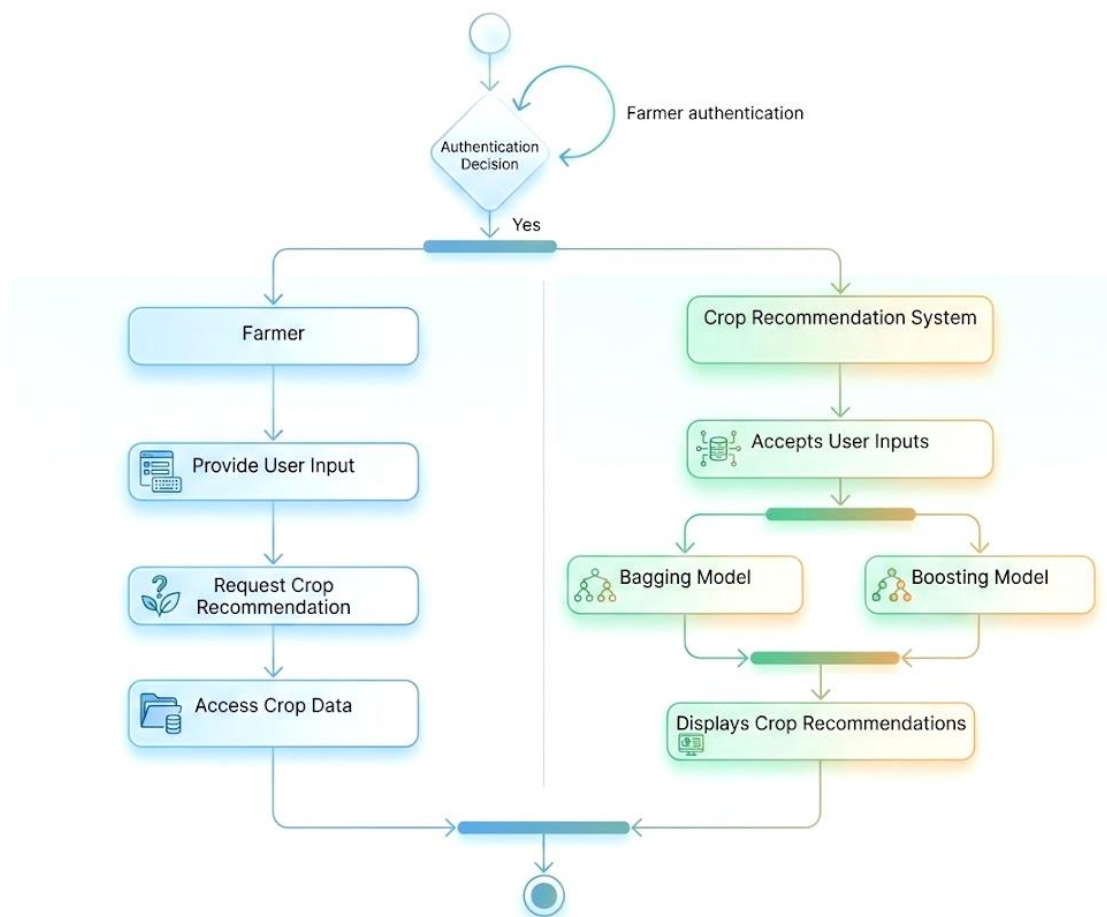


Рисунок 3.4 – Діаграма діяльності

Особливістю даного процесу є паралельна або послідовна обробка даних моделями беггінгу (Bagging) та бустингу (Boosting), що забезпечує високу точність прогнозу. Після завершення обчислювального циклу система переходить до стану формування висновку, надаючи фермеру науково обґрунтовані поради щодо вибору культури для вирощування у заданих умовах.

Початковий стан системи визначається входом у вузол автентифікації. Контрольний потік передбачає використання вузла прийняття рішень (Decision Diamond) для верифікації облікових даних фермера. У разі успішної перевірки (умова «Yes») система переходить до стадії паралельного виконання операцій, що реалізується через механізм розділення потоку (Fork Node).

Даний етап поділяє життєвий цикл активності на дві паралельні гілки, що відображають взаємодію клієнтської та серверної частин системи:

- потік операцій користувача: Описує послідовність дій фермера, яка включає безпосереднє введення агрономічних параметрів (Provide user input), формування запиту на розрахунок (Request crop Recommendation) та отримання доступу до структурованих даних про культури (Access Crop Data).

- потік обчислювальної системи: Відображає логіку роботи бекенд-модуля. Після отримання вхідних даних від користувача (Accepts User Inputs) система ініціює вторинне розгалуження для паралельної обробки інформації.

Ключовою особливістю алгоритмічної структури є одночасна активація двох ансамблевих стратегій машинного навчання:

- Bagging Model - реалізація методу бэггінгу для зменшення дисперсії прогнозу.

- Boosting Model - застосування бустингу для мінімізації систематичної похибки.

Паралельне виконання цих процесів дозволяє оптимізувати часові витрати на обчислення та забезпечує багатоаспектний аналіз вхідних параметрів. Після завершення роботи обох моделей потоки синхронізуються у вузлі злиття (Join Node), що передує фазі візуалізації — формуванню та виведенню фінальних рекомендацій (Displays Crop Recommendations).

Кінцевим етапом функціонування системи є консолідація результатів користувацького запиту та системної відповіді. Об'єднання потоків керування гарантує, що діяльність завершується лише після успішного виконання всіх паралельних підпроцесів. Фінальний вузол (Activity Final Node) сигналізує про успішне надання послуги та перехід системи в стан очікування нового запиту.

Таким чином, діаграма діяльності демонструє високий рівень модуляризації та автоматизації процесів, де інтелектуальні обчислення інтегровані в логічну структуру взаємодії з кінцевим користувачем.

Застосування UML-інструментарію дозволило формалізувати архітектуру інтелектуальної системи, забезпечивши прозорість процесів обробки даних та взаємодії компонентів. Сформовані моделі слугують

дорожньою картою для програмної реалізації та подальшої масштабованості системи в умовах реального аграрного виробництва.

3.2. Аналіз результатів класифікації та оцінка прогностичної здатності моделі

Для всебічної оцінки ефективності розробленої системи та верифікації точності прогнозів було проведено детальний аналіз результатів на тестовій вибірці. Основними інструментами оцінки виступили матриця невідповідностей (Confusion Matrix) та розширений звіт про класифікацію (Classification Report), що дозволяють ідентифікувати слабкі місця моделі та оцінити її стійкість до специфічних помилок.

3.2.1. Аналіз матриці невідповідностей

На рисунку 3.5 представлена матриця невідповідностей, яка візуалізує кореляцію між фактичними та прогнозованими класами сільськогосподарських культур. Даний інструмент дозволяє кількісно оцінити кількість істинно позитивних (TP) та хибно позитивних (FP) результатів у розрізі кожного класу.

Аналіз графічних даних свідчить про високу роздільну здатність моделі для переважної більшості культур. Проте виявлено поодинокі випадки міжкласового змішування:

- сочевиця (Lentils) у ряді випадків була помилково ідентифікована як джут (Jute).
- кавуни (Watermelons) були класифіковані як банани (Bananas).

Такі похибки зазвичай зумовлені подібністю агроекологічних векторів (схожі діапазони значень NPK, температури та вологості) для зазначених пар культур, що створює труднощі для побудови ідеальної роздільної гіперплощини між ними. Розуміння цих паттернів є критичним для подальшого вдосконалення алгоритмів шляхом введення додаткових ознак

					БР.ІП – 06.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		55

або тонкого налаштування ваг.

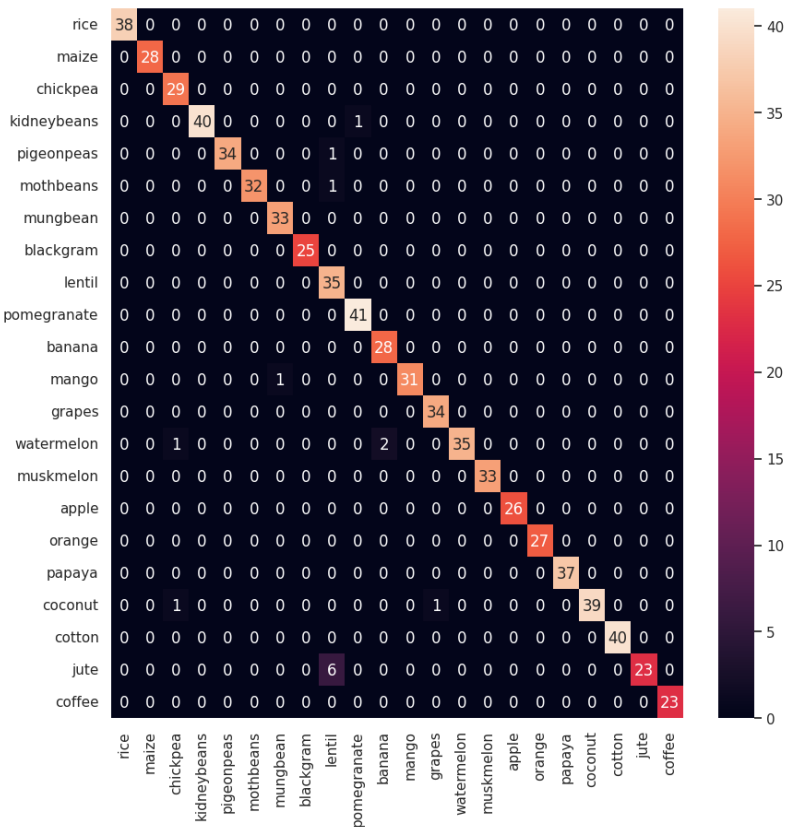


Рисунок 3.5 – Матриця невідповідностей

3.2.2. Верифікація за метриками звіту про класифікацію

Для отримання статистично обґрунтованого висновку щодо якості роботи класифікатора було сформовано звіт, наведений на рисунку 3.6. Звіт включає такі ключові метрики:

- Точність (Precision) - відношення правильно спрогнозованих об'єктів класу до загальної кількості об'єктів, віднесених моделлю до цього класу.
- Повнота (Recall) - здатність класифікатора знаходити всі об'єкти заданого класу.
- F1-міра (F1-score) - середнє гармонійне точності та повноти, що забезпечує збалансовану оцінку ефективності.
- Підтримка (Support) - фактична кількість екземплярів кожного класу у тестовому наборі даних.

Результати статистичного аналізу:

- джут (Jute): зафіксовано найнижчий показник точності (Precision), що вказує на схильність моделі до гіпердіагностики цього класу (віднесення до нього інших культур).

- рис (Rice): характеризується мінімальними значеннями повноти та F1-міри. Це свідчить про те, що певна частка реальних посівів рису не була розпізнана системою.

- кавун (Watermelon): має найменший показник підтримки (Support) у вибірці, що може впливати на статистичну значущість результатів для даного конкретного класу.

Клас	Precision	Recall	F1-score	Support
apple	1.0	1.0	1.0	38
banana	1.0	1.0	1.0	28
blackgram	0.94	1.0	0.97	29
chickpea	1.0	0.98	0.99	41
coconut	1.0	0.97	0.99	35
coffee	1.0	0.97	0.98	33
cotton	0.97	1.0	0.99	33
grapes	1.0	1.0	1.0	25
jute	0.81	1.0	0.9	35
kidneybeans	0.98	1.0	0.99	41
lentil	0.93	1.0	0.97	28
maize	1.0	0.97	0.98	32
mango	0.97	1.0	0.99	34
mothbeans	1.0	0.92	0.96	38
mungbean	1.0	1.0	1.0	33
muskmelon	1.0	1.0	1.0	26
orange	1.0	1.0	1.0	27
papaya	1.0	1.0	1.0	37
pigeonpeas	1.0	0.95	0.97	41
pomegranate	1.0	1.0	1.0	40
rice	1.0	0.79	0.88	29
watermelon	1.0	1.0	1.0	23
accuracy	nan	nan	0.98	726
macro avg	0.98	0.98	0.98	726
weighted avg	0.98	0.98	0.98	726

Рисунок 3.6 – Звіт щодо якості роботи класифікатора

Незважаючи на виявлені поодинокі випадки помилкової класифікації, загальна точність (Accuracy) інтелектуальної моделі становить 98%. Це підтверджує високу адекватність обраного ансамблевого підходу та його придатність для використання в реальних системах підтримки прийняття рішень у сільському господарстві. Зважене середнє всіх метрик вказує на стабільність прогнозів та здатність моделі ефективно оперувати багатовимірними агрономічними даними.

3.3. Розробка структури інтерфейсу рекомендаційної системи

На рисунку 3.7 показано комплексну багатoshарову архітектуру та програмну структуру системи рекомендації сільськогосподарських культур.

Проектна модель базується на принципах модульності, масштабованості та чіткого розподілу відповідальності між функціональними рівнями.

Загальна логіка функціонування системи детермінується вертикальною ієрархією рівнів, що забезпечують повний цикл обробки даних: від інтерфейсної взаємодії до персистентного зберігання інформації.

Рівень представлення (Frontend Layer): реалізований на базі фреймворків React/Next.js. Він забезпечує візуалізацію компонентів, маршрутизацію та управління глобальним станом системи за допомогою бібліотеки Redux.

Проміжний рівень (Middleware & Logic): включає API Gateway, що виступає єдиною точкою входу для REST/GraphQL запитів, та модуль Authentication, який гарантує безпеку доступу через протоколи JWT та OAuth 2.0.

Ядро системи (Service Layer): складається з двох ключових модулів:

1. Data Processing: відповідає за життєвий цикл даних (Data Pipeline), включаючи етапи валідації, очищення та нормалізації вхідних параметрів.

2. Prediction Service: інкапсулює математичний апарат системи. Тут розміщені моделі машинного навчання, зокрема ансамблева модель, що

виконує класифікацію культур на основі оброблених ознак.

Рівень даних (Database Layer): базується на реляційній СКБД PostgreSQL, де забезпечується збереження профілів користувачів, історії прогнозів, параметрів навчених моделей та вихідних наборів даних.

Нижня частина рисунка демонструє фізичну організацію файлової структури проєкту, що відображає імплементацію вищеописаної архітектури.

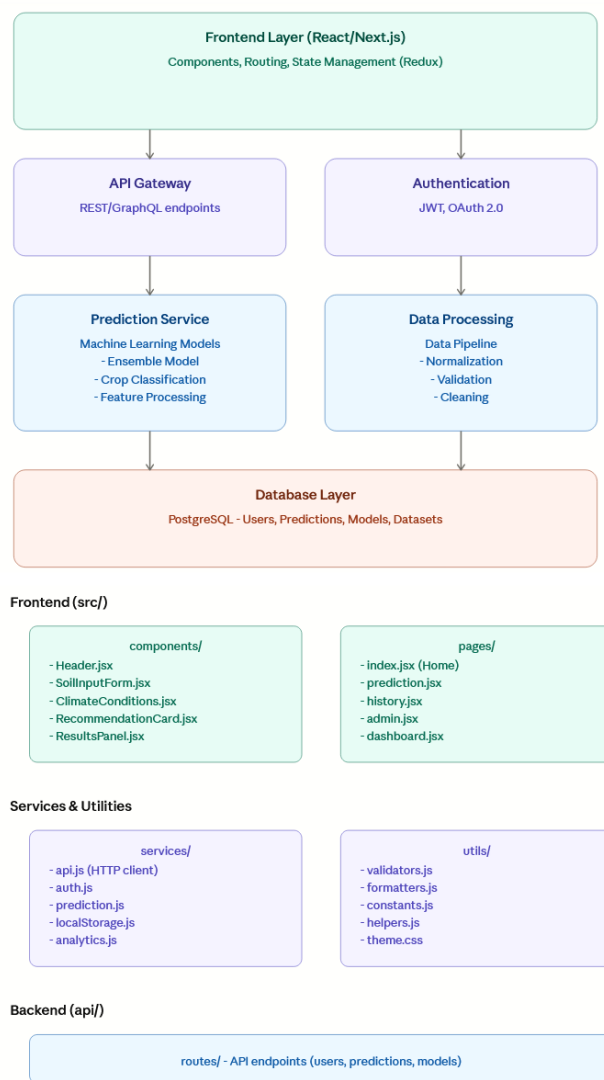


Рисунок 3.7 - Програмна структура системи рекомендацій

Фронтенд-структура (src/):

- Директорія components/ містить атомарні елементи інтерфейсу, такі як

форми введення показників ґрунту (SoilInputForm.jsx), метеорологічних умов (ClimateConditions.jsx), а також панелі візуалізації результатів.

- Директорія pages/ детермінує логіку маршрутизації, включаючи домашню сторінку, модулі прогнозування, перегляду історії та адміністративну панель керування моделями.

Сервіси та утиліти:

- Модуль services/ виступає як прошарок між фронтом та API, інкапсулюючи логіку HTTP-клієнтів, автентифікації та взаємодії з аналітичним сервісом.

- Модуль utils/ містить допоміжні скрипти: валідатори даних, формати, константи та глобальні налаштування стилів (theme.css).

Бекенд-інтерфейс (api/):

- Директорія routes/ визначає кінцеві точки API (endpoints), що забезпечують обмін даними між клієнтською частиною та серверною логікою щодо користувачів, прогнозів та конфігурацій моделей.

Дана структура демонструє системний підхід до розробки СППР, де кожна компонента — від нормалізації даних до ансамблевого моделювання — інтегрована в єдиний технологічний стек. Це забезпечує високу відтворюваність результатів та надійність системи в умовах реального експлуатаційного навантаження.

3.4. Розробка інтерфейсу системи надання рекомендацій для оптимального вибору культур та функціональна логіка взаємодії

Програмний інтерфейс системи AgroSense спроектований як високорівнева оболонка (Front-end), що забезпечує взаємодію кінцевого користувача з аналітичним ядром предиктивної моделі. Дизайн інтерфейсу базується на принципах модульності та когнітивної ергономіки, що дозволяє

					БР.ІП – 06.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		60

мінімізувати помилки при введенні вхідних даних та забезпечує прозорість процесу прийняття рішень.

Ліва панель інтерфейсу призначена для формування багатовимірного вектора ознак, який згодом передається на вхід ансамблевої моделі. Вона структурована за двома основними категоріями (рис. 3.8):

- Параметри ґрунту: блок містить поля для введення кількісних показників концентрації макроелементів — азоту (N), фосфору (P) та калію (K) у вимірі кг/га, а також водневого показника (pH) ґрунту.

- Кліматичні умови: блок параметризації екзогенних чинників, що включає середньодобову температуру (°C), відносну вологість повітря (%) та сумарну кількість опадів (мм).

Для управління станом системи передбачено два функціональні елементи: кнопка ініціалізації розрахунку «Згенерувати Рекомендацію» та кнопка скидання значень для очищення вхідних полів.

Рисунок 3.8 – Графічний інтерфейс системи для введення вхідних даних

					БР.ІП – 06.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		61

У нижній частині панелі керування розміщено інформаційний блок, який верифікує параметри використовуваної математичної моделі. Вказано, що система базується на Ансамблевій моделі (Ensemble Model) із зафіксованим показником точності на рівні 98%. Це забезпечує високий рівень довіри користувача до отриманих прогнозів.

Права (центральна) область інтерфейсу є динамічним полем відображення результатів. У початковому стані вона знаходиться у режимі очікування («Готово до аналізу»), надаючи користувачеві інструкції щодо подальших дій. Після активації обчислювального модуля ця область трансформується у панель візуалізації, де відображається рекомендована культура з відповідним агрономічним обґрунтуванням.

Графічна оболонка системи AgroSense поєднує в собі інструменти прецизійного введення даних та потужний аналітичний інструментарій, що робить її ефективною СППР для аграрного сектору.

3.5. Аналіз результатів моделювання та структура вихідних даних

Нижче наведено опис функціонального стану інтерфейсу системи AgroSense після завершення обчислювального циклу та виведення результатів предиктивного аналізу.

Після активації обчислювального модуля права панель інтерфейсу трансформується з режиму очікування у розгорнутий аналітичний звіт, що містить ієрархію рекомендацій та обґрунтування обраного рішення.

Центральним елементом звіту є блок ідентифікації найбільш підходящої культури. У даному сценарії система рекомендує Пшеницю озиму із показником довіри моделі на рівні 66.6%. Опис результату включає три рівні деталізації:

- Текстова анотація: надає стислу характеристику сумісності культури з введеним рН (7) та високим вмістом калію (45 кг/га).

					БР.ІП – 06.00.00.000 ПЗ	Арк.
						62
Змн.	Арк.	№ докум.	Підпис	Дата		

- Оптимальні умови: блок порівняльної верифікації, де вказано цільові діапазони для вегетації (рН: 6.5–7.5, температура: 12–25°C, опади: 400–600 мм).
- Агрономічні переваги: перелік ключових якостей культури, таких як висока врожайність та стійкість до морозів.

АгроSense

Система інтелектуальних рекомендацій для оптимального вибору культур

ПАРАМЕТРИ ҐРУНТУ

Азот (N), kg/ha

55

Фосфор (P), kg/ha

35

Калій (K), kg/ha

45

pH ґрунту

7

КЛІМАТИЧНІ УМОВИ

Температура, °C

35

Вологість, %

70

Опади, mm

500

Згенерувати
Рекомендацію

Скинути

ВИКОРИСТАНА МОДЕЛЬ

Ансамблева модель (Ensemble Model)
Точність моделі: 98%

Пшениця озима

Довіра моделі: 66.6%

Опис

Добре підходить для ґрунтів з pH 6.8 та високим вмістом калію. Потребує помірних опадів та температури.

✓ Оптимальні умови:

pH: 6.5-7.5, Температура: 12-25°C, Опади: 400-600 mm

☀ Переваги:

Висока врожайність, стійка до морозів, добре зберігається

Оцінки моделі для всіх культур

Пшениця озима

6658.1%

Соняшник

4806.7%

Кукурудза на зерно

3825.8%

Резюме введених даних

Азот (N)

55

kg/ha

Фосфор (P)

35

kg/ha

Калій (K)

45

kg/ha

pH

7

Температура

35

°C

Вологість

70

%

Опади

500

mm

Рисунок 3.9 – Надання рекомендації системою

Під блоком основної рекомендації розміщено розділ «Оцінки моделі для всіх культур», що дозволяє користувачеві оцінити ймовірнісну перевагу обраного варіанта над конкурентними. Система виділяє три пріоритетні позиції:

- Пшениця озима (найвищий бал за внутрішньою шкалою класифікатора).

- Соняшник (альтернативний варіант).
- Кукурудза на зерно (третій за релевантністю результат).

Оцінки моделі для всіх культур

Соняшник	4370.6%
Пшениця озима	4032.1%
Кукурудза на зерно	683.6%

Рисунок 3.10 – Оцінки моделі для всіх культур

Візуалізація реалізована за допомогою динамічних прогрес-барів, що полегшує порівняльний аналіз вихідних значень моделі.

Завершальним компонентом звіту є «Резюме введених даних», яке дублює параметри ґрунту (N: 55, P: 35, K: 45, pH: 7) та кліматичні показники (Температура: 35°C, Вологість: 70%, Опادي: 500 мм). Це забезпечує функцію зворотного контролю, дозволяючи користувачеві впевнитися, що прогноз базується на коректних вхідних величинах.

Резюме введених даних

Азот (N) 45 kg/ha	Фосфор (P) 25 kg/ha	Калій (K) 35 kg/ha	pH 8
Температура 20 °C	Вологість 6 %	Опади 500 mm	

Рисунок 3.11 - Резюме введених даних

Сформований звіт забезпечує комплексний підхід до надання рекомендацій, поєднуючи статистичну імовірність з практичними агрономічними настановами, що відповідає вимогам до сучасних інтелектуальних систем підтримки прийняття рішень.

					БР.ІП – 06.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		64

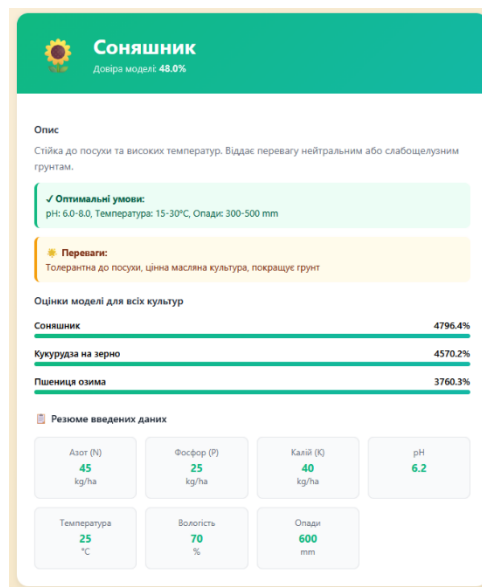


Рисунок 3.11 – Вигляд інтерфейсу при зміні розміру екрану (закінчення)

На основі проведеного дослідження та розробки предиктивної системи AgroSense, можна сформулювати наступні узагальнюючі висновки щодо її ефективності та практичної значущості:

1. Технологічна ефективність та точність моделювання

Розроблена система продемонструвала високу прогностичну здатність завдяки застосуванню гетерогенного ансамблю методів машинного навчання. Поєднання стратегій беггінгу та бустингу дозволило досягти загальної точності класифікації на рівні 98%. Це підтверджує гіпотезу про те, що ансамблеві методи є більш стійкими до викидів у агрономічних даних та здатні ефективно апроксимувати нелінійні залежності між вхідними параметрами навколишнього середовища та врожайністю культур.

2. Комплексність аналізу ознакового простору

Система базується на багатовимірному векторі ознак, що охоплює як ендегенні характеристики ґрунту (N, P, K, pH), так і екзогенні кліматичні чинники (температура, вологість, опади). Такий підхід забезпечує всебічний моніторинг умов вегетації, що дозволяє надавати не просто статистично імовірні, а біологічно обґрунтовані рекомендації для конкретних ділянок.

3. Функціональність та архітектурна гнучкість

Архітектура системи, побудована на стеку React/Next.js для фронтенду та PostgreSQL для збереження даних, забезпечує високу швидкість обробки запитів та масштабованість. Реалізований графічний інтерфейс успішно виконує функцію інтелектуального посередника, трансформуючи складні математичні обчислення у зрозумілі для кінцевого користувача (фермера) агрономічні звіти, що включають:

- ідентифікацію домінуючої культури з вказанням рівня довіри моделі;
- ієрархічне ранжування альтернативних варіантів посіву;
- детальне обґрунтування вибору на основі відповідності вхідних параметрів оптимальним умовам росту.

Впровадження СППР AgroSense у практику прецизійного землеробства дозволяє значно знизити виробничі ризики, пов'язані з некоректним вибором культур. Система забезпечує науковий підхід до планування сівоzmіни, що сприяє раціональному використанню земельних ресурсів та підвищенню економічної ефективності агровиробництва.

3.6. Технічна реалізація системи надання рекомендацій

Для технічної реалізації системи AgroSense, архітектура якої була описана раніше, наведено ключові фрагменти програмного коду на мовах Python (бекенд та ML-ядро) та JavaScript/React (фронтенд).

1. Реалізація гетерогенного ансамблю за стратегією стекінгу

Згідно із запропонованою архітектурою, ядро системи використовує дворівневу модель: базові алгоритми (SVM, Logistic Regression, Naive Bayes) та мета-класифікатор для фінального прогнозу.

Лістинг 3.1. Реалізація гетерогенного ансамблю

```
# Імпорт необхідних бібліотек версій, вказаних у специфікації
import pandas as pd # версія 2.0.3
from sklearn.svm import SVC
from sklearn.linear_model import LogisticRegression
from sklearn.naive_bayes import GaussianNB
from sklearn.ensemble import StackingClassifier
```

					БР.ІП – 06.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		68

```

from sklearn.preprocessing import StandardScaler
from sklearn.model_selection import train_test_split

# 1. Попередня обробка даних
def prepare_data(df):
    X = df[['N', 'P', 'K', 'temperature', 'humidity', 'ph', 'rainfall']]
    y = df['label']

    scaler = StandardScaler()
    X_scaled = scaler.fit_transform(X)
    return train_test_split(X_scaled, y, test_size=0.2, random_state=42)

# 2. Побудова архітектури ансамблю
base_learners = [
    ('svm', SVC(probability=True, kernel='rbf')),
    ('logit', LogisticRegression(max_iter=1000)),
    ('gnb', GaussianNB())
]

# Використання Logistic Regression як мета-класифікатора
stacking_model = StackingClassifier(
    estimators=base_learners,
    final_estimator=LogisticRegression(),
    cv=5
)

# Навчання моделі
stacking_model.fit(X_train, y_train)

```

2. Серверний API-ендпоінт для отримання прогнозів

Для інтеграції ML-моделі з інтерфейсом користувача використовується FastAPI. Цей фрагмент відповідає за обробку вхідного вектора ознак та повернення результату класифікації.

Лістинг 3.2. Серверний API-ендпоінт

```

from fastapi import FastAPI
from pydantic import BaseModel

app = FastAPI()

class SoilFeatures(BaseModel):
    N: float
    P: float
    K: float
    temperature: float
    humidity: float
    ph: float
    rainfall: float

@app.post("/predict")
async def get_crop_recommendation(data: SoilFeatures):
    # Трансформація вхідних даних у масив NumPy
    input_vector = [[

```

					БР.ІП – 06.00.00.000 ПЗ	Арк.
						69
Змн.	Арк.	№ докум.	Підпис	Дата		

```

        data.N, data.P, data.K,
        data.temperature, data.humidity,
        data.ph, data.rainfall
    ]]

    # Виклик навченої моделі (приклад)
    prediction = stacking_model.predict(input_vector)
    confidence = stacking_model.predict_proba(input_vector).max()

    return {
        "recommended_crop": prediction[0],
        "confidence": f"{round(confidence * 100, 1)}%",
        "status": "success"
    }

```

3. Фронтенд-компонент: Форма введення даних (React)

Компонент SoilInputForm.jsx відповідає за збір параметрів ґрунту та клімату від користувача, забезпечуючи візуалізацію, представлену в AgroSense.

Лістинг 3.3. Форма введення даних

```

import React, { useState } from 'react';

const SoilInputForm = ({ onPredict }) => {
    const [features, setFeatures] = useState({
        N: 50, P: 30, K: 40, ph: 6.8,
        temperature: 20, humidity: 60, rainfall: 500
    });

    const handleChange = (e) => {
        setFeatures({ ...features, [e.target.name]: e.target.value });
    };

    const handleSubmit = async (e) => {
        e.preventDefault();
        // Виклик функції прогнозування через API
        onPredict(features);
    };

    return (
        <form onSubmit={handleSubmit} className="p-4 bg-white rounded-lg shadow">
            <h3 className="text-lg font-bold mb-4">ПАРАМЕТРИ ГРУНТУ</h3>
            <div className="grid grid-cols-1 gap-4">
                <input
                    name="N"
                    type="number"
                    onChange={handleChange}
                    placeholder="Азот (N), kg/ha"
                    className="border p-2 rounded"
                />

```

					БР.ІП – 06.00.00.000 ПЗ	Арк.
						70
Змн.	Арк.	№ докум.	Підпис	Дата		

```

    { /* Аналогічні поля для P, K, pH та кліматичних умов */ }
    <button
      type="submit"
      className="bg-teal-500 text-white font-bold py-2 rounded hover:bg-teal-
    >
      ⚡ Згенерувати Рекомендацію
    </button>
  </div>
</form>
);
};

export default SoilInputForm;

```

4. Робота з базою даних (PostgreSQL)

Опис схеми даних для збереження історії прогнозів у таблиці predictions, що є частиною Database Layer.

Лістинг 3.4. Створення таблиці БД

```

CREATE TABLE predictions (
  id SERIAL PRIMARY KEY,
  user_id INT REFERENCES users(id),
  n_value FLOAT NOT NULL,
  p_value FLOAT NOT NULL,
  k_value FLOAT NOT NULL,
  ph_value FLOAT NOT NULL,
  temp_value FLOAT NOT NULL,
  hum_value FLOAT NOT NULL,
  rain_value FLOAT NOT NULL,
  predicted_crop VARCHAR(100),
  confidence_score FLOAT,
  created_at TIMESTAMP DEFAULT CURRENT_TIMESTAMP
);

```

Зазначені фрагменти коду забезпечують цілісність системи: від первинної обробки даних у Pandas до візуалізації результатів у React. Кожна частина коду відповідає за конкретний рівень (Layer), описаний у попередніх підпунктах в архітектурній схемі.

3.7 Висновки по розділу

У третьому розділі виконано проектування архітектури системи та реалізовано програмне забезпечення для підтримки прийняття рішень у сфері

					БР.ІП – 06.00.00.000 ПЗ	Арк.
						71
Змн.	Арк.	№ докум.	Підпис	Дата		

агрономії. За допомогою UML-моделювання сформовано структурні та поведінкові моделі системи, що дозволило формалізувати логіку взаємодії компонентів і забезпечити цілісність програмної архітектури. Проведений аналіз результатів класифікації показав, що реалізована модель характеризується високими показниками точності та здатністю ефективно прогнозувати оптимальні агрономічні рішення.

Розроблений користувацький інтерфейс забезпечує зручну взаємодію із системою, автоматизує процес обробки аграрних даних і спрощує формування рекомендацій щодо вибору культур. Отримані результати підтвердили практичну цінність створеної системи та можливість її подальшого застосування й масштабування в сучасних агроінформаційних середовищах.

					БР.ІП – 06.00.00.000 ПЗ	Арк.
						72
Змн.	Арк.	№ докум.	Підпис	Дата		

ВИСНОВКИ

В бакалаврській роботі було проведено дослідження сучасних методів інтелектуального аналізу аграрних даних, здійснено проектування архітектури програмної системи підтримки прийняття рішень та реалізовано програмний засіб, орієнтований на підвищення ефективності агрономічної діяльності шляхом використання ансамблевих моделей машинного навчання.

У першому розділі роботи виконано ґрунтовний аналіз предметної області побудови класифікаційних моделей для агрономії. Дослідження показало, що традиційні підходи до планування посівів, оцінювання стану ґрунтів та прогнозування агроекологічних параметрів значною мірою залежать від емпіричних знань і суб'єктивних оцінок спеціалістів, що ускладнює забезпечення стабільності та точності прийняття рішень. Встановлено, що застосування методів машинного навчання дозволяє автоматизувати процес аналізу великих масивів агроданих та забезпечити формування більш обґрунтованих рекомендацій. Особливу увагу приділено ансамблевим методам, які демонструють підвищену стійкість до шуму, кращу узагальнювальну здатність та вищу точність класифікації у порівнянні з окремими моделями.

У другому розділі роботи розглянуто інтелектуальні технології аналізу агроекологічних факторів та реалізацію предиктивних моделей для агросистем. Проведено обґрунтування вибору інструментальних засобів і програмного середовища, які забезпечують ефективну реалізацію алгоритмів машинного навчання та обробки даних. Визначено, що використання сучасних бібліотек аналізу даних і фреймворків машинного навчання дозволяє забезпечити гнучкість, масштабованість та адаптивність системи.

Значну увагу приділено дослідженню набору даних, який використовувався для навчання та тестування моделей. Виконано статистичний аналіз ознакового простору, досліджено розподіл ознак, рівень

					БР.ІП – 06.00.00.000 ПЗ	Арк.
						73
Змн.	Арк.	№ докум.	Підпис	Дата		

асиметрії та взаємозв'язки між параметрами. Проведений кореляційний аналіз дозволив виявити найбільш інформативні характеристики, що мають суттєвий вплив на результати класифікації. Також здійснено ідентифікацію аномальних значень та оцінено стійкість моделей до впливу шумових даних. Отримані результати підтвердили доцільність попередньої обробки даних як важливого етапу побудови високоточної класифікаційної системи.

У третьому розділі здійснено проєктування архітектури програмної системи та її програмну реалізацію. За допомогою UML-моделювання розроблено діаграми класів, варіантів використання, послідовностей та діяльності, що дозволило формалізувати структуру системи, логіку взаємодії компонентів і функціональні процеси. Використання уніфікованої мови моделювання забезпечило чітке представлення архітектурних рішень і спростило подальшу реалізацію програмного продукту.

Під час оцінювання результатів класифікації проведено аналіз матриці невідповідностей та ключових метрик якості моделей. Встановлено, що реалізована класифікаційна система характеризується високим рівнем точності прогнозування та здатна ефективно ідентифікувати оптимальні варіанти культур для конкретних агроекологічних умов. Аналіз метрик класифікаційного звіту підтвердив збалансованість моделі та її придатність до практичного використання в умовах реального агровиробництва.

У межах програмної реалізації розроблено інтерфейс системи надання рекомендацій, орієнтований на забезпечення зручної взаємодії користувача з програмним засобом. Реалізовано функціональну логіку введення агрономічних параметрів, обробки даних та формування рекомендацій щодо вибору культур. Розроблена система дозволяє автоматизувати процес прийняття рішень, скоротити час аналізу даних та підвищити ефективність управління аграрними процесами.

Таким чином, у роботі досягнуто поставленої мети — розроблено класифікаційну модель для програмного забезпечення потреб агрономії, яка

					БР.ІП – 06.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		74

поєднує методи ансамблевого машинного навчання, інструменти статистичного аналізу та сучасні підходи до проектування програмних систем. Отримані результати підтверджують перспективність застосування інтелектуальних технологій у сфері агрономії та створюють основу для подальшого розвитку адаптивних систем підтримки прийняття рішень у сільському господарстві.

					БР.ІП – 06.00.00.000 ПЗ	Арк.
						75
Змн.	Арк.	№ докум.	Підпис	Дата		

ПЕРЕЛІК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Breiman L. Random Forests // Machine Learning. – New York : Springer, 2001. – Vol. 45, No. 1. – P. 5–32.
2. Cortes C., Vapnik V. Support-Vector Networks // Machine Learning. – Boston : Kluwer Academic Publishers, 1995. – Vol. 20, No. 3. – P. 273–297.
3. Quinlan J. R. Induction of Decision Trees // Machine Learning. – Boston : Kluwer Academic Publishers, 1986. – Vol. 1, No. 1. – P. 81–106.
4. Friedman J. H. Greedy Function Approximation: A Gradient Boosting Machine // Annals of Statistics. – Hayward : Institute of Mathematical Statistics, 2001. – Vol. 29, No. 5. – P. 1189–1232.
5. Kotsiantis S. B. Decision Trees: A Recent Overview // Artificial Intelligence Review. – Dordrecht : Springer, 2013. – Vol. 39, No. 4. – P. 261–283.
6. Witten I. H., Frank E., Hall M. A. Data Mining: Practical Machine Learning Tools and Techniques. – Burlington : Morgan Kaufmann Publishers, 2016. – 654 p.
7. Hastie T., Tibshirani R., Friedman J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. – New York : Springer, 2009. – 745 p.
8. Bishop C. M. Pattern Recognition and Machine Learning. – Singapore : Springer, 2006. – 738 p.
9. Russell S., Norvig P. Artificial Intelligence: A Modern Approach. – Upper Saddle River : Pearson Education, 2021. – 1136 p.
10. Goodfellow I., Bengio Y., Courville A. Deep Learning. – Cambridge : MIT Press, 2016. – 775 p.
11. Kamilaris A., Prenafeta-Boldú F. X. Deep Learning in Agriculture: A Survey // Computers and Electronics in Agriculture. – Amsterdam : Elsevier, 2018. – Vol. 147. – P. 70–90.

					БР.ІІІ – 06.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		76

12. Liakos K. G., Busato P., Moshou D., Pearson S., Bochtis D. Machine Learning in Agriculture: A Review // Sensors. – Basel : MDPI, 2018. – Vol. 18, No. 8. – P. 2674–2698.
13. Wolfert S., Ge L., Verdouw C., Bogaardt M.-J. Big Data in Smart Farming – A Review // Agricultural Systems. – Amsterdam : Elsevier, 2017. – Vol. 153. – P. 69–80.
14. Sharma A., Jain A., Gupta P., Chowdary V. Machine Learning Applications for Precision Agriculture: A Comprehensive Review // IEEE Access. – New York : IEEE, 2021. – Vol. 9. – P. 4843–4873.
15. Chlingaryan A., Sukkariéh S., Whelan B. Machine Learning Approaches for Crop Yield Prediction and Nitrogen Status Estimation in Precision Agriculture // Computers and Electronics in Agriculture. – Amsterdam : Elsevier, 2018. – Vol. 151. – P. 61–69.
16. Jeong J. H., Resop J. P., Mueller N. D., Fleisher D. H., Yun K., Butler E. E., Timlin D. J., Shim K.-M., Gerber J. S., Reddy V. R., Kim S.-H. Random Forests for Global and Regional Crop Yield Predictions // PLoS ONE. – San Francisco : Public Library of Science, 2016. – Vol. 11, No. 6. – P. 1–15.
17. Crisci C., Ghattas B., Perera G. A Review of Supervised Machine Learning Algorithms and Their Applications to Ecological Data // Ecological Modelling. – Amsterdam : Elsevier, 2012. – Vol. 240. – P. 113–122.
18. Biau G., Scornet E. A Random Forest Guided Tour // TEST. – Heidelberg : Springer, 2016. – Vol. 25, No. 2. – P. 197–227.
19. Zhang C., Ma Y. Ensemble Machine Learning: Methods and Applications. – New York : Springer, 2012. – 329 p.
20. Han J., Kamber M., Pei J. Data Mining: Concepts and Techniques. – Waltham : Morgan Kaufmann, 2011. – 703 p.
21. Mohri M., Rostamizadeh A., Talwalkar A. Foundations of Machine Learning. – Cambridge : MIT Press, 2018. – 505 p.

					БР.ІІІ – 06.00.00.000 ІІЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		77

22. Marsland S. Machine Learning: An Algorithmic Perspective. – Boca Raton : CRC Press, 2015. – 457 p.
23. Kelleher J. D., Namee B. M., D’Arcy A. Fundamentals of Machine Learning for Predictive Data Analytics. – Cambridge : MIT Press, 2015. – 624 p.
24. Aggarwal C. C. Data Mining: The Textbook. – Cham : Springer, 2015. – 734 p.
25. Kuhn M., Johnson K. Applied Predictive Modeling. – New York : Springer, 2013. – 595 p.
26. Géron A. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow. – Sebastopol : O’Reilly Media, 2022. – 851 p.
27. Pedregosa F., Varoquaux G., Gramfort A., Michel V., Thirion B., Grisel O., Blondel M., Prettenhofer P., Weiss R., Dubourg V., Vanderplas J., Passos A., Cournapeau D., Brucher M., Perrot M., Duchesnay É. Scikit-learn: Machine Learning in Python // Journal of Machine Learning Research. – Cambridge : MIT Press, 2011. – Vol. 12. – P. 2825–2830.
28. Chen T., Guestrin C. XGBoost: A Scalable Tree Boosting System // Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. – New York : ACM Press, 2016. – P. 785–794.
29. Ke G., Meng Q., Finley T., Wang T., Chen W., Ma W., Ye Q., Liu T.-Y. LightGBM: A Highly Efficient Gradient Boosting Decision Tree // Advances in Neural Information Processing Systems. – Red Hook : Curran Associates, 2017. – Vol. 30. – P. 3146–3154.
30. Dorogush A. V., Ershov V., Gulin A. CatBoost: Gradient Boosting with Categorical Features Support // Proceedings of the Workshop on ML Systems at NIPS. – Long Beach : NeurIPS Foundation, 2018. – P. 1–7.
31. Jordan M. I., Mitchell T. M. Machine Learning: Trends, Perspectives, and Prospects // Science. – Washington : American Association for the Advancement of Science, 2015. – Vol. 349, No. 6245. – P. 255–260.

					БР.ІІІ – 06.00.00.000 ІІЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		78

32. Provost F., Fawcett T. Data Science for Business. – Sebastopol : O'Reilly Media, 2013. – 414 p.
33. James G., Witten D., Hastie T., Tibshirani R. An Introduction to Statistical Learning. – New York : Springer, 2021. – 607 p.
34. Sarker I. H. Machine Learning: Algorithms, Real-World Applications and Research Directions // SN Computer Science. – Cham : Springer, 2021. – Vol. 2, No. 3. – P. 1–21.
35. Pathan M., Patel N., Yagnik H., Shah M. Artificial Intelligence in Agriculture: Challenges, Opportunities, and Future Trends // Progress in Artificial Intelligence. – Cham : Springer, 2020. – Vol. 9, No. 4. – P. 1–19.
36. Ray P. P. Internet of Things for Smart Agriculture: Technologies, Practices and Future Direction // Journal of Ambient Intelligence and Smart Environments. – Amsterdam : IOS Press, 2017. – Vol. 9, No. 4. – P. 395–420.

					БР.ІП – 06.00.00.000 ПЗ	Арк.
						79
Змн.	Арк.	№ докум.	Підпис	Дата		

БІБЛІОГРАФІЧНА ДОВІДКА

Тема бакалаврської роботи: “Розроблення класифікаційної моделі для програмного забезпечення потреб агрономії ”

Обсяг пояснювальної записки: 79 аркушів.

Дата закінчення роботи: 9 червня 2026 р.

Підпис студента _____