

**НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ  
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ  
імені ІГОРЯ СІКОРСЬКОГО»  
НАВЧАЛЬНО-НАУКОВИЙ ІНСТИТУТ  
ПРИКЛАДНОГО СИСТЕМНОГО АНАЛІЗУ  
КАФЕДРА МАТЕМАТИЧНИХ МЕТОДІВ СИСТЕМНОГО АНАЛІЗУ**

«На правах рукопису»  
УДК 303.732.4

До захисту допущено  
Завідувач кафедри ММСА  
\_\_\_\_\_ Оксана ТИМОЩУК  
«\_\_\_» \_\_\_\_\_ 2025 р.

**Магістерська дисертація  
на здобуття ступеня магістра  
за освітньо-професійною програмою «Системний аналіз фінансового ринку»  
зі спеціальності 124 «Системний аналіз»  
на тему: «Інтелектуальна система оцінювання страхових ризиків в аграрному  
секторі на основі методів машинного навчання»**

Виконав:  
Студент 2 курсу, групи КА-42мп  
Ярко Андрій Юрійович \_\_\_\_\_

Науковий керівник:  
Професор кафедри ММСА, д.т.н.,  
Кузнецова Наталія Володимирівна \_\_\_\_\_

Рецензент:  
Професор кафедри ШІ, д.т.н., проф.  
Данилов Валерій Якович \_\_\_\_\_

Засвідчую, що у цій магістерській  
дисертації немає запозичень з праць  
інших авторів без відповідних  
посилань.

Студент \_\_\_\_\_

Київ – 2025 року

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ  
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ  
імені ІГОРЯ СІКОРСЬКОГО»  
НАВЧАЛЬНО-НАУКОВИЙ ІНСТИТУТ  
ПРИКЛАДНОГО СИСТЕМНОГО АНАЛІЗУ  
КАФЕДРА МАТЕМАТИЧНИХ МЕТОДІВ СИСТЕМНОГО АНАЛІЗУ

Рівень вищої освіти — другий (магістерський)

Спеціальність — 124 «Системний аналіз»

Освітньо-професійною програмою «Системний аналіз фінансового ринку»

ЗАТВЕРДЖУЮ

Завідувач кафедри ММСА

\_\_\_ Оксана ТИМОЩУК

«\_\_\_» \_\_\_\_\_ 2025 р.

**ЗАВДАННЯ**  
**на магістерську дисертацію студенту**  
**Ярку Андрію Юрійовичу**

1. Тема дисертації: “Інтелектуальна система оцінювання страхових ризиків в аграрному секторі на основі методів машинного навчання”, науковий керівник дисертації професор кафедри ММСА, доктор технічних наук, Кузнєцова Наталія Володимирівна, затверджені наказом по університету від 06 листопада 2025 року № 4837-с.
2. Термін подання студентом дисертації 06.12.2025.
3. Об’єкт дослідження: процеси формування, прогнозування та оцінювання страхових ризиків у аграрному секторі в умовах невизначеності природно-кліматичних та економічних факторів.
4. Предмет дослідження: методи машинного навчання, статистичного аналізу, імітаційного моделювання та

інтелектуальні підходи до побудови систем оцінювання аграрних страхових ризиків, включно з алгоритмами прогнозування врожайності та розрахунку страхових показників.

5. Перелік завдань, які потрібно зробити:

- 1) провести огляд сучасної наукової літератури;
- 2) дослідити теоретичні основи страхування аграрних ризиків;
- 3) сформулювати та підготувати гетерогенний набір даних;
- 4) розробити та протестувати моделі машинного навчання;
- 5) розробити модель оцінювання ризиків на основі імітаційного моделювання;
- 6) створити інтегровану інтелектуальну систему підтримки прийняття рішень;
- 7) розробити модуль надання рекомендацій;
- 8) проаналізувати результати експериментальних досліджень.

6. Орієнтовний перелік публікацій:

1) Ярко А.Ю., Кузнецова Н.В. Система оцінювання і прогнозування врожайності зернових культур методами машинного навчання на основі супутникових і кліматичних даних. Системні науки та інформатика: збірник доповідей IV науково-практичної конференції, 1-5 грудня 2025. Київ, НН ІПСА КПІ ім. Ігоря Сікорського, 2025. 6 с.

7. Дата видачі завдання: 02 вересня 2025 р.

## Календарний план

№ з/п	Назва етапів виконання магістерської дисертації	Термін виконання етапів роботи	Примітка
1	Ознайомлення з літературою за темою роботи	01.09.2025 - 14.09.2025	Виконано
2	Підготовка першого розділу	14.09.2025 - 28.10.2025	Виконано
3	Підготовка другого розділу	28.10.2025 - 03.10.2025	Виконано
4	Розробка моделей для вирішення поставленої задачі	03.10.2025 - 13.10.2025	Виконано
5	Підготовка третього розділу	13.10.2025 - 24.10.2025	Виконано
6	Підготовка частини стартап-проєкту	24.10.2025 - 09.11.2025	Виконано
7	Висновки. Перспективи розвитку проєкту	09.11.2025 -22.11.2025	Виконано
8	Оформлення пояснювальної записки	22.11.2025 - 07.12.2025	Виконано

Студент

Науковий керівник дисертації

Андрій ЯРКО

Наталія КУЗНЕЦОВА

## РЕФЕРАТ

Магістерська дисертація: 135 с., 38 рис., 21 табл., 1 додаток, 21 джерел.

АГРАРНІ РИЗИКИ, СТРАХУВАННЯ, ПРОГНОЗУВАННЯ  
ВРОЖАЙНОСТІ, МАШИННЕ НАВЧАННЯ, СУПУТНИКОВІ ДАНІ,  
МОНТЕ-КАРЛО, VAR, CVAR, ІНТЕЛЕКТУАЛЬНА СИСТЕМА,  
АГРОСТРАХУВАННЯ.

Об'єкт дослідження – процеси формування, прогнозування та оцінювання страхових ризиків у аграрному секторі в умовах невизначеності природно-кліматичних та економічних факторів.

Предмет дослідження – методи машинного навчання, статистичного аналізу, імітаційного моделювання та інтелектуальні підходи до побудови систем оцінювання аграрних страхових ризиків, включно з алгоритмами прогнозування врожайності та розрахунку страхових показників.

Мета роботи – створення інтелектуальної системи підтримки прийняття рішень, яка забезпечує точне прогнозування врожайності сільськогосподарських культур і визначення ключових страхових показників: ймовірності дефіциту, очікуваних збитків, Value-at-Risk та Conditional Value-at-Risk.

У роботі використано методи машинного навчання (регресійні моделі, ансамблеві алгоритми), статистичного аналізу, просторової обробки супутникових даних MODIS та кліматичних реаналізів ERA5-Land, а також метод Монте-Карло для моделювання сценаріїв невизначеності.

## ABSTRACT

Master's thesis: 135 pages, 38 figures, 21 tables, 1 appendix, 21 references. AGRICULTURAL RISKS, INSURANCE, YIELD FORECASTING, MACHINE LEARNING, SATELLITE DATA, MONTE CARLO, VAR, CVAR, INTELLIGENT SYSTEM, AGRICULTURAL INSURANCE.

The object of the research is the processes of formation, forecasting, and assessment of insurance risks in the agricultural sector under uncertainty caused by natural-climatic and economic factors.

The subject of the research is machine learning methods, statistical analysis, simulation modeling, and intelligent approaches to designing systems for assessing agricultural insurance risks, including algorithms for yield forecasting and calculation of insurance indicators.

The aim of the thesis is to develop an intelligent decision support system that provides accurate forecasting of crop yields and determination of key insurance metrics: probability of deficit, expected loss, Value-at-Risk, and Conditional Value-at-Risk.

The study employs machine learning techniques (regression models, ensemble algorithms), statistical analysis, spatial processing of MODIS satellite data and ERA5-Land climate reanalysis, as well as the Monte Carlo method for modeling uncertainty scenarios.

## ЗМІСТ

ВСТУП .....	10
1 ТЕОРЕТИЧНІ ОСНОВИ ОЦІНКИ СТРАХОВИХ РИЗИКІВ У АГРАРНОМУ СЕКТОРІ.....	13
1.1 Класифікація ризиків у сільському господарстві.....	13
1.2 Страхування аграрних ризиків.....	16
1.3 Методи аналізу та прогнозування ризиків.....	19
1.4 Порівняльна характеристика підходів до оцінювання ризику.....	21
1.5 Огляд підходів до страхування та страхових метрик .....	24
1.5.1 Ймовірнісні метрики .....	24
1.5.2 . Метрики глибини збитків .....	25
1.5.3 Фінансово-економічні метрики ефективності страхування ..	26
1.6 Висновки до розділу 1 .....	27
2..... ПРАКТИЧНА РЕАЛІЗАЦІЯ СИСТЕМИ .....	29
2.1 Вибір мови програмування .....	32
2.2 Опис вхідних даних .....	36
2.3 Попередня обробка даних .....	38
2.4 Прогнозування врожайності методами машинного навчання..	41
2.4.1 Загальна схема експерименту .....	42
2.4.2 Вибір моделей машинного навчання .....	43
2.4.3 Оптимізація гіперпараметрів.....	46

2.4.4 Стратегія валідації .....	46
2.4.5 Метрики оцінки якості моделей .....	48
2.4.6 Результати моделювання .....	49
2.4.7 Порівняння результатів .....	70
2.5 Розробка моделі оцінки ризиків .....	73
2.5.1 Метод Монте-Карло.....	74
2.5.2 Розрахунок страхових метрик .....	78
2.6 Висновки до розділу 2 .....	86
3 ІНТЕЛЕКТУАЛЬНА СИСТЕМА ПІДТРИМКИ ПРИЙНЯТТЯ РІШЕНЬ.....	88
3.1 Архітектура ІСППР.....	89
3.2 Блок моделювання .....	92
3.3 Модуль оцінювання .....	94
3.4 Модуль оцінки та розрахунку ризику .....	96
3.5 Модуль надання рекомендацій .....	98
3.6 Висновки до розділу 3 .....	102
4 РОЗРОБКА СТАРТАП-ПРОЄКТУ .....	104
4.1 Аналіз ринку та передумови створення стартапу.....	105
4.2 Аналіз конкурентного середовища.....	107
4.2.1. Огляд світових рішень у сфері агростраховання та аналітики ризиків.....	107
4.2.2. Аналіз українського ринку цифрових агросервісів.....	108
4.2.3. Порівняльна характеристика конкурентів.....	110

1.2.4. SWOT-аналіз стартапу AgroShield AI.....	111
4.3 Технічна реалізація системи AgroShield AI.....	112
4.3.1. Загальна архітектура системи.....	112
4.3.2. Модулі системи та їх функціональне призначення .....	113
4.3.3. Використовувані технології та інструменти .....	114
4.3.4. Вибір мови програмування та середовища реалізації .....	114
4.3.5. Алгоритмічне забезпечення.....	115
4.3.6. Архітектура бази даних.....	116
4.4.1. Концепція бізнес-моделі та місія проєкту.....	117
4.4.2. Структура продукту та джерела доходу .....	118
4.4.3. Модель взаємодії з клієнтами та партнерами .....	119
4.4.4. Витратна структура та основні ресурси .....	119
4.4.5. Маркетингова стратегія та канали просування.....	120
4.4.6. Ключові показники ефективності (KPI).....	121
4.4.7. Фінансова модель та прогноз розвитку стартапу .....	121
4.4.8. Аналіз ризиків бізнес-моделі.....	122
4.5. Висновки до розділу 4 .....	123
ВИСНОВКИ .....	125
ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ .....	128
ДОДАТОК А .....	130

## ВСТУП

Сільське господарство є однією з базових і стратегічно важливих галузей економіки України, яка забезпечує продовольчу безпеку держави, формує значну частку експорту та впливає на стабільність соціально-економічного розвитку. Водночас аграрне виробництво характеризується підвищеним рівнем невизначеності та ризику, зумовленим залежністю від кліматичних, біологічних і економічних факторів. Погодні коливання, деградація ґрунтів, зростання вартості ресурсів, коливання цін на світових ринках, а також наслідки воєнних дій створюють значні загрози стабільності аграрного виробництва. У цих умовах особливої актуальності набуває проблема кількісної оцінки страхових ризиків, що дозволяє зменшити невизначеність і забезпечити фінансову стійкість агровиробників.

Питання управління ризиками в агросекторі розглядаються у працях багатьох зарубіжних і вітчизняних дослідників. У наукових джерелах досліджуються теоретичні засади формування системи аграрного страхування, методи економічної оцінки ризиків, моделі прогнозування врожайності та фінансової стійкості агропідприємств. Зокрема, у працях іноземних авторів увага приділяється розробці державних програм страхового захисту фермерів, вдосконаленню механізмів розподілу ризиків між державою, страховими компаніями та виробниками, а також формуванню актуарних моделей визначення страхових премій. В українській науковій літературі переважають економічні підходи до оцінки ризику, що базуються на статистичних залежностях минулих спостережень або експертних оцінках.

Попри наявність значної кількості досліджень, залишається недостатньо розробленою прикладна сторона оцінки ризику на основі сучасних методів аналізу даних. Більшість традиційних підходів не враховує багатофакторну

природу аграрного виробництва, просторово-часові залежності кліматичних умов та можливість використання супутникових і реаналітичних даних для об'єктивного опису процесів, що впливають на врожайність. У цьому контексті дедалі більшого значення набувають методи машинного навчання, які дозволяють виявляти приховані закономірності у великих гетерогенних наборах даних і будувати прогностичні моделі високої точності.

У сучасних дослідженнях з прогнозування врожайності сільськогосподарських культур активно застосовуються алгоритми Support Vector Machines (SVM), Random Forest, Gradient Boosting, XGBoost та інші моделі, здатні враховувати нелінійні взаємозв'язки між кліматичними, ґрунтовими та біофізичними показниками. Використання таких підходів у поєднанні із супутниковими спостереженнями (наприклад, MODIS) і кліматичними реаналізами (ERA5-Land) створює потужну основу для підвищення точності прогнозів. Проте переважна більшість наукових робіт зосереджується саме на задачі детермінованого прогнозування, тоді як аспект ймовірнісного аналізу ризику, який є критичним для страхових розрахунків, залишається поза увагою або реалізується лише у спрощеній формі.

Особливої уваги потребує побудова моделей, що поєднують прогнозування врожайності методами машинного навчання з ймовірнісним моделюванням невизначеності на основі методу Монте-Карло. Такий підхід дозволяє не лише оцінити очікувану врожайність, але й визначити розподіл можливих значень, розрахувати ймовірність настання збиткових сценаріїв, очікуваний розмір дефіциту врожаю та обґрунтовану величину страхової премії. Застосування подібних моделей у практиці агрострахування сприяє підвищенню прозорості, достовірності та економічної обґрунтованості страхових рішень.

Водночас, у наукових дослідженнях з даної тематики існують невирішені питання, серед яких:

- відсутність єдиної інтегрованої методики, що поєднує прогнозування врожайності та кількісну оцінку ризику на основі статистичних і машинних підходів;
- недостатня розробленість методів просторової агрегації різномірних супутникових та кліматичних даних для умов України;
- обмежене використання симуляційних методів (зокрема, Монте-Карло) у задачах агрострахування;
- потреба в адаптації зарубіжних підходів до специфіки українських агрокліматичних умов та структури ринку страхових послуг.

Саме тому тема дослідження, присвячена оцінці страхових ризиків для агросектору із застосуванням методів машинного навчання та імітаційного моделювання, є актуальною як у науковому, так і в практичному вимірах. Розробка подібних моделей здатна забезпечити кількісне підґрунтя для прийняття управлінських рішень у сфері агрострахування, оптимізації тарифів, планування виробництва та оцінки кредитоспроможності аграрних підприємств.

Таким чином, проведення даного дослідження є доцільним і своєчасним, оскільки воно спрямоване на заповнення наявних прогалів у вітчизняній практиці оцінювання аграрних ризиків, поєднує сучасні технології аналізу даних з методами фінансової оцінки ризику та створює науково-прикладну основу для побудови автоматизованих систем підтримки рішень у сфері агрострахування.

# 1 ТЕОРЕТИЧНІ ОСНОВИ ОЦІНКИ СТРАХОВИХ РИЗИКІВ У АГРАРНОМУ СЕКТОРІ

## 1.1 Класифікація ризиків у сільському господарстві

Економічна діяльність у сільському господарстві характеризується високим рівнем невизначеності, що обумовлено як природними, так і соціально-економічними чинниками. У цьому контексті ризик виступає невід'ємною складовою аграрного виробництва і розглядається як імовірність виникнення подій, наслідком яких є відхилення фактичних результатів від очікуваних. Іншими словами, ризик у сільському господарстві — це ймовірність настання несприятливих умов, які призводять до зниження врожайності, погіршення якості продукції або прямих економічних збитків. Особливість аграрного ризику полягає в його багатофакторному характері: він формується під впливом природно-кліматичних, технологічних, біологічних, економічних та політичних чинників, які часто мають стохастичну природу й взаємодіють між собою у складних нелінійних залежностях.

Основними рисами аграрних ризиків є стохастичність, просторово-часова залежність, нелінійність впливів та висока чутливість до екстремальних подій. Стохастичність означає, що результати виробничої діяльності залежать від випадкових чинників, наприклад, погодних умов або біологічних процесів, які складно точно передбачити. Просторово-часова залежність проявляється у тому, що погодні явища та ґрунтові властивості мають тенденцію повторюватися або корелювати в межах певних регіонів і часових періодів. Нелінійність впливів виявляється тоді, коли навіть незначні зміни у кліматичних умовах, таких як дефіцит опадів у критичні фази вегетації, можуть призвести до суттєвих втрат урожайності. Крім того, аграрні ризики

часто мають «важкі хвости» розподілів, тобто екстремальні події відбуваються частіше, ніж це передбачено нормальним законом, що істотно ускладнює їх моделювання традиційними статистичними методами.

Класифікація ризиків у сільському господарстві здійснюється за низкою критеріїв, що дозволяє комплексно оцінити їхню природу та визначити шляхи мінімізації. За джерелом походження виділяють природні (кліматичні) ризики, пов'язані з погодними умовами (посуха, град, заморозки, повені); біологічні ризики, що виникають унаслідок поширення хвороб рослин, шкідників або генетичних дефектів культур; технологічні, пов'язані з відмовами техніки, порушенням технологічних процесів або недоліками агротехніки; економічні, що відображають коливання цін на сировину, ресурси та готову продукцію; а також політичні та соціальні ризики, пов'язані з регуляторними змінами, військовими діями або соціальною нестабільністю.

За рівнем контрольованості аграрні ризики поділяють на контрольовані, які можна частково зменшити за рахунок управлінських або технологічних рішень (наприклад, захист від шкідників, оптимізація сівозміни), та неконтрольовані, що не піддаються впливу людини, як-от стихійні лиха або кліматичні катастрофи. Залежно від часової характеристики ризики бувають короткостроковими (заморозки, локальні бурі), середньостроковими (сезонні коливання погодних умов, епідемії шкідників) та довгостроковими (деградація ґрунтів, зміна кліматичних тенденцій).

Для цілей страхування важливо розрізняти ризики за можливістю їх індексації. Індиковані ризики (наприклад, кліматичні чи метеорологічні) можуть бути кількісно описані за допомогою метеостанційних даних або супутникових індексів NDVI, що дозволяє створювати параметричні або індексні страхові продукти. Неіндексовані ризики, такі як помилки у технологічному процесі чи локальні хвороби рослин, потребують індивідуального огляду й оцінки.

Оцінювання ризику передбачає використання кількісних показників, серед яких основними є дисперсія або стандартне відхилення врожайності, ймовірність недобору нижче певного порогу, а також показники, що використовуються у фінансовому ризик-менеджменті, зокрема Value-at-Risk (VaR) і Conditional Value-at-Risk (CVaR). Ці метрики дозволяють оцінити як середній рівень варіації врожайності, так і очікувані втрати в екстремальних сценаріях. Саме на основі цих оцінок формуються страхові тарифи, премії та резерви, що враховують не лише середній очікуваний збиток, але й рівень невизначеності та ризику концентрації збитків у просторі.

Статистичні властивості аграрних ризиків, такі як нестаціонарність, тяжкі хвости розподілів та просторово-корельована природа даних, визначають специфіку їх моделювання. Історичні ряди врожайності та погодних показників часто змінюють свої статистичні параметри в часі через глобальні кліматичні зміни, що потребує застосування методів адаптивного аналізу й машинного навчання. Корельованість ризиків між регіонами знижує ефективність диверсифікації та вимагає використання моделей, здатних урахувати просторові залежності.

Отже, ризики в сільському господарстві мають складну структуру, поєднуючи природну невизначеність із економічними факторами, що посилюють їхній вплив. Розуміння сутності та класифікації аграрних ризиків є необхідною передумовою для побудови ефективної системи прогнозування, страхування та управління ними. У подальших розділах роботи це знання стане основою для формування методики кількісної оцінки ризику, прогнозування втрат і розроблення моделей визначення страхових премій із використанням інтелектуальних та імітаційних методів аналізу.

## 1.2 Страхування аграрних ризиків

Сільськогосподарське виробництво, як одна з найбільш ризикових галузей економіки, значною мірою залежить від природно-кліматичних умов, які важко прогнозуються та не підлягають контролю з боку виробника. Тому ефективне управління ризиками в агросекторі передбачає використання системи фінансових інструментів, серед яких ключове місце займає страхування аграрних ризиків. Страхування виступає економічним механізмом перерозподілу ризику між учасниками аграрного ринку, що забезпечує стабільність доходів сільгоспвиробників, мінімізацію фінансових втрат і підтримання сталого виробництва в умовах невизначеності.

У теоретичному аспекті страхування ризиків базується на принципах солідарності, ймовірнісного розподілу збитків та еквівалентності зобов'язань. Солідарність передбачає, що ризики окремих суб'єктів об'єднуються у спільний фонд, з якого здійснюються виплати постраждалим учасникам. Ймовірнісний розподіл означає, що страхова премія розраховується на основі статистичної ймовірності настання страхового випадку та очікуваного розміру збитків. Принцип еквівалентності забезпечує відповідність між очікуваними страховими виплатами і сумою внесків, що формують страховий фонд. Таким чином, страхування виконує дві взаємопов'язані функції — розподільчу (перерозподіл збитків між учасниками) та стимулюючу, оскільки воно заохочує виробників до впровадження заходів ризик-менеджменту й підвищення технологічної ефективності виробництва.

Теоретичні основи страхування аграрних ризиків включають також розуміння його функцій — відновлювальної, попереджувальної, накопичувальної та інвестиційної. Відновлювальна функція спрямована на компенсацію втрат унаслідок стихійних лих або інших несприятливих подій,

що дозволяє відновити виробничий потенціал господарства. Попереджувальна функція полягає у стимулюванні суб'єктів господарювання до впровадження профілактичних заходів (наприклад, системи зрошення, стійких сортів культур або технологій зниження ризику). Накопичувальна та інвестиційна функції проявляються у використанні страхових резервів як джерела фінансових ресурсів для розвитку аграрного сектору, що підвищує його фінансову стійкість і довгострокову ефективність.

У практиці агрострахування сформувалося кілька основних моделей організації страхового захисту. Традиційне страхування врожаю передбачає компенсацію збитків на основі фактичного огляду полів і визначення розміру втрат. Такий підхід є найбільш поширеним, але має недоліки, пов'язані з високими адміністративними витратами, суб'єктивністю оцінки та ризиком морального ризику (*moral hazard*). Альтернативою є індексне (параметричне) страхування, яке ґрунтується на використанні об'єктивних індексів — метеорологічних показників (опад, температура), супутникових вегетаційних індексів (NDVI, EVI) або інших спостережуваних параметрів, що корелюють із урожайністю. Виплата здійснюється автоматично у випадку, коли значення індексу виходить за встановлений поріг. Такий підхід забезпечує прозорість, оперативність і зниження трансакційних витрат, однак потребує якісних даних і складного калібрування порогових значень, щоб уникнути ефектів "базисного ризику" — ситуацій, коли реальні втрати не збігаються з індикаторами.

На рівні теорії страхування важливе місце займає поняття страхової премії, що визначається як математичне сподівання величини можливих збитків із поправкою на ризикову надбавку. Формально розрахунок базується на інтегралі ймовірностей втрат отриманих із історичних даних або моделювання методом Монте-Карло. Очікувана страхова премія визначає фінансову стійкість страховика, а ризикова надбавка компенсує невизначеність

та екстремальні події, які не можуть бути передбачені на основі середніх значень.

У сучасних умовах розвитку фінансових технологій зростає роль моделей оцінювання ризиків, заснованих на статистичних, стохастичних і машинних методах. Такі моделі дозволяють не лише обчислити імовірність страхового випадку, а й побудувати прогностичні сценарії на основі кліматичних та економічних факторів. Використання штучного інтелекту, супутникового моніторингу й відкритих даних створює передумови для формування інтелектуальних систем агрострахування, здатних автоматично оцінювати ризики, прогнозувати збитки та пропонувати оптимальні страхові параметри для окремих господарств.

З теоретичної точки зору страхування аграрних ризиків тісно пов'язане з концепціями диверсифікації та перестраховування, які забезпечують фінансову стабільність страхової системи. Диверсифікація полягає у розподілі ризиків між великою кількістю учасників або регіонів, що знижує ймовірність одночасного виникнення великих збитків. Перестраховування, своєю чергою, дозволяє передати частину ризиків на вищий рівень — спеціалізованим перестраховикам або міжнародним фондам, зменшуючи фінансовий тиск на національні компанії у випадку катастрофічних подій.

Отже, страхування аграрних ризиків є складною системою економічних, фінансових та інформаційних відносин, що базується на науково обґрунтованих принципах імовірнісного розподілу збитків. Його ефективність залежить від якості статистичних даних, достовірності моделей оцінки ризику, адекватності тарифної політики та рівня державної підтримки. Розвиток інноваційних страхових продуктів, інтеграція сучасних аналітичних методів і застосування інтелектуальних систем відкривають нові можливості для підвищення надійності страхового захисту та зменшення впливу невизначеності на аграрне виробництво.

### 1.3 Методи аналізу та прогнозування ризиків

У сучасному аграрному виробництві управління ризиками передбачає не лише їх класифікацію, а й застосування адекватних методів аналізу та прогнозування. Для забезпечення ефективної оцінки ризиків у сільському господарстві використовуються як традиційні статистичні методи, так і новітні моделі з галузі машинного навчання та економіко-математичного моделювання.

Статистичні методи, такі як дисперсійний аналіз, кореляційно-регресійні моделі, аналіз часових рядів та методи варіації (VAR), дозволяють виявити взаємозв'язки між врожайністю, погодними умовами, технологіями й іншими чинниками невизначеності. Наприклад, оцінка ймовірності недобору врожаю може здійснюватися через модель регресії на основі кліматичних даних, що дає змогу кількісно виміряти вплив певного фактору (наприклад, посухи) на результат. Такий підхід широко використовується в дослідженнях агрострахування, зокрема при визначенні очікуваних збитків і ступеня їх варіації.

Економіко-математичні моделі включають імітаційні підходи (наприклад, метод Монте-Карло), стохастичні моделі, а також моделі типу «safety-first», які враховують не лише очікуваний дохід чи врожай, а й ризик недотримання мінімального безпечного порога. Такі моделі дозволяють моделювати розподіли можливих втрат, будувати сценарії і оцінювати ймовірність великих негативних відхилень («важкі хвости»). У контексті аграрного страхування це стає вкрай важливим через високу варіабельність показників.

Останнім часом дедалі більше зростає роль методів машинного навчання (ML) у прогнозуванні аграрних ризиків. Ці методи дозволяють обробляти великі набори даних, використовувати супутникові індекси (наприклад NDVI), кліматичні аналізи, агрономічні та економічні показники для побудови моделей, які прогнозують ймовірність страхового випадку або величини виплат. Так, дослідження показують, що ML-алгоритми можуть перевершувати класичні моделі, знижуючи похибки прогнозування та підвищуючи стійкість страхового тарифоутворення [1].

Порівняльна оцінка методів дозволяє виділити їх переваги й обмеження.

1. Статистичні моделі — відносно прості у впровадженні, інтерпретовані, але можуть бути обмежені припущеннями нормальності розподілів чи стаціонарності процесів.
2. Економіко-математичні моделі — дозволяють моделювати складні сценарії, враховувати просторові кореляції та екстремальні втрати, але вимагають значних обчислювальних ресурсів і якісних даних.
3. ML-методи — дають гнучкість, здатність працювати з великими даними та нелінійними залежностями, але часто менш прозорі з точки зору інтерпретації та можуть вимагати складного налаштування й валідації.

Вибір конкретного методу (або комбінації методів) залежить від якісної бази даних, типу ризику, регіональних особливостей господарства і цільової задачі (наприклад, прогнозування врожайності, моделювання ймовірності страхового випадку, розрахунок чистої премії). Для побудови універсальної методики дослідження ризиків доцільно поєднувати всі три підходи: статистичний аналіз на попередньому етапі, економіко-математичне моделювання для сценаріїв і ML-методи для точкового прогнозування на рівні господарства.

У підсумку, комплексне застосування зазначених методів дозволяє підвищити точність оцінки ризиків, адаптувати страхові продукти до реальних умов агровиробництва, а також оптимізувати страхову премію та умови покриття для виробників. У наступному розділі буде представлена пропонована методика дослідження, яка інтегрує зазначені підходи в єдину схему (підготовка даних → прогнозування врожайності → оцінка ризику → моделювання страхових показників).

#### 1.4 Порівняльна характеристика підходів до оцінювання ризику

Оцінювання ризиків є ключовим етапом управління в аграрному виробництві, оскільки дозволяє кількісно або якісно визначити ймовірність настання несприятливих подій та їхні можливі наслідки. У теоретичному й практичному контексті розрізняють три основні підходи: якісний, кількісний та гібридний (комбінований). Кожен із цих підходів має свої переваги, обмеження та сферу застосування, що є важливим при побудові методики оцінки ризику у сільському господарстві.

Якісні методи оцінювання ризику базуються на експертній оцінці, ранжуванні загроз, використанні матриць «ймовірність × наслідок» та інших описових інструментів. Вони відзначаються простою структурою, мінімальними вимогами до даних і швидкістю застосування, що робить їх доцільними на початкових етапах аналізу ризику або за відсутності достатньої статистики. Однак такі підходи мають суттєвий недолік — обмежену здатність обробляти невизначеність, враховувати взаємозв'язки між ризиками та будувати розподіли втрат [2].

Кількісні методи, навпаки, припускають наявність числових даних — ймовірностей, розподілів втрат, кореляцій між факторами. Вони включають моделі, такі як імовірнісна оцінка ризику (Probabilistic Risk Assessment, PRA), симуляційні методи (напр., метод Монте-Карло), а також інструменти економіко-математичного моделювання. Ці моделі дозволяють не лише оцінити очікуваний збиток, але й параметри варіабельності, важкі хвости розподілів і ймовірність екстремальних подій. У контексті аграрного страхування такі методи можуть бути основою для розрахунку чистої премії або резервів. При цьому вони вимагають якісної даної бази, значного обчислювального ресурсу й високої експертизи [3].

Гібридні (або комбіновані) підходи поєднують елементи якісного й кількісного аналізу. Вони використовують експертні оцінки для структуризації ризиків, ранжування й формулювання припущень, а в подальшому — застосування кількісних моделей для деталізації й прогнозування. Такий підхід часто визнається найбільш доцільним у складних контекстах із обмеженими даними та високою невизначеністю. Наприклад, у системах, де кліматичні фактори, технологічні процеси та ринкові умови взаємопов'язані й мають нелінійну природу, гібридні моделі стають компромісним рішенням між швидкістю оцінювання та глибиною аналізу.

Порівняльна оцінка цих підходів дозволяє виділити ключові критерії вибору: доступність даних, складність моделювання, часова витратність, гнучкість до змін й ступінь врахування невизначеності. Наприклад, якісні методи мають високу швидкість і низькі вимоги до даних, але слабкі у точному прогнозуванні; кількісні — сильні у прогнозі та сценарному аналізі, але вимогливі до ресурсів; гібридні — поміж ними, проте можуть бути більш складними в управлінні. Дослідження показують, що жоден підхід не є універсально кращим у всіх умовах — ключовим є відповідність методу характеру ризику, доступній інформації та цілям дослідження.

У контексті аграрного страхування для виробництва сільськогосподарської продукції доцільно виходити з комбінованої методики: на етапі попереднього скринінгу застосовувати якісні матриці й експертні оцінки, далі — кількісні моделі врожайності, симуляції втрат і розрахунок страхових метрик (ймовірність страхового випадку, очікуваний дефіцит, чиста премія). Така поетапність забезпечує баланс між оперативністю роботи та точністю аналізу, що підвищує надійність прийнятих рішень в умовах агровиробництва з високими невизначеністю й кореляціями ризиків.

В таблиці 1.1 представлено порівняння різних методів прогнозування.

Таблиця 1.1 Порівняльний аналіз методів прогнозування

Метод	Переваги	Обмеження	Сфера застосування
Якісний	- Простота застосування та мінімальні вимоги до даних. - Можливість швидкої ідентифікації ключових ризиків. - Зручність для початкових етапів аналізу або відсутності історичних спостережень.	- Суб'єктивність оцінок, залежність від експертів. - Неможливість кількісного прогнозу та моделювання втрат. - Не враховує складні взаємозв'язки між факторами.	Початковий етап аналізу ризиків; стратегічне планування; визначення пріоритетних напрямів управління ризиками.
Кількісний	- Дозволяє оцінити імовірність та розподіл можливих втрат. - Використовується для розрахунку страхових премій, резервів та VaR-показників. - Дає змогу моделювати екстремальні сценарії.	- Потребує великих масивів достовірних даних. - Високі вимоги до обчислювальних ресурсів. - Може бути складним для інтерпретації нефахівцями.	Оцінка страхових ризиків; моделювання врожайності; прогнозування збитків; оптимізація портфелів.
Гібридний (комбінований)	- Поєднує гнучкість якісного аналізу та точність кількісного. - Забезпечує комплексний підхід до невизначеності. - Може бути адаптований до обмежених даних.	- Висока складність побудови; потребує мультидисциплінарних знань. - Вимагає координації між експертною оцінкою та математичними моделями.	Аграрне страхування; стратегічне управління ризиком; оцінка кліматичних та ринкових ризиків у складних системах.

Отже, порівняльна характеристика підходів до оцінювання ризику дозволяє обґрунтувати вибір методики, яка відповідає специфіці аграрного сектору та задачам страхового захисту. У подальшому розділі буде представлено загальну методику проведення дослідження, яка враховує вказані підходи та їхній порівняльний аналіз.

### 1.5 Огляд підходів до страхування та страхових метрик

Страхові метрики є математичним апаратом, що використовується у оцінці ризиків та у формуванні страхових тарифів. Метрики дозволяють переходити від прогнозів (наприклад, врожайності) до кількісної оцінки збитків та ефективності страхового продукту.

Ключові групи страхових метрик:

- 1) Імовірнісні (частотні);
- 2) Втратні (середні збитки);
- 3) Ризикові (tail-risk);
- 4) Фінансово-економічні (ефективність капіталу).

#### 1.5.1 Ймовірнісні метрики

Ймовірність страхового випадку є однією з базових метрик, що вимірює частоту настання негативної події за визначений період і широко

використовується в актуарній практиці [4]. Обраховується за наступною формулою:

$$PL = P(Y < Y_{cr}) \frac{1}{N} \sum_{i=1}^N I(Y_i < Y_{cr})$$

### 1.5.2. Метрики глибини збитків

Очікуваний дефіцит врожаю характеризує середнє відхилення фактичної врожайності від критичного порогу за умов, коли збиток уже настав і описується наступною формулою [5]:

$$ESY = \frac{\sum_{i:(Y_i < Y_{cr})} (Y_{cr} - Y_i)}{\sum_{i=1}^N I(Y_i < Y_{cr})}$$

Очікуваний збиток є центральним поняттям страхового тарифоутворення і відображає середню величину втрат, виражену у грошовому еквіваленті. Він описується наступною формулою [6]:

$$EL = ESY * P * PL.$$

Показник Value-at-Risk використовується для оцінки потенційних збитків у найгірших сценаріях та є стандартом у міжнародних підходах до ризик-менеджменту [7]:

$$VaR_a = \inf\{x : F_L(x) \geq a\}.$$

Conditional Value-at-Risk (або Expected Shortfall у фінансовому контексті) дозволяє оцінити середній збиток у випадках, коли втрати перевищують VaR, і є більш чутливим до «важких хвостів» розподілу. Його формулу можна побачити далі [8]:

$$CVaR_a = E[L|L > VaR_a].$$

### 1.5.3 Фінансово-економічні метрики ефективності страхування

Чиста страхова премія є математично очікуваною величиною страхових виплат і використовується як базис для побудови всіх типів тарифів:

$$NP = EL.$$

Брутто-премія враховує операційні витрати, ризикове навантаження та прибутковість страховика; її обчислення є стандартною практикою актуарних служб [9]:

$$GP = NP(1 + \theta),$$

де  $\theta$  — навантаження страховика.

Показник RAROC застосовується для оцінки дохідності страхового продукту з урахуванням ризику та широко використовується у банківсько-страхових компаніях:

$$RAROC = \frac{GP - EL}{VaR}.$$

Традиційні підходи страхування (індивідуальні договори) покладаються на детальні локальні дані та точну оцінку збитків, тоді як індексні моделі роблять акцент на агрегованих індикаторах та статистичній стабільності [10].

Страхові метрики дозволяють:

- 1) визначити, наскільки часто настає збиток (PL);
- 2) оцінити, наскільки глибоким може бути цей збиток (ESY, EL);
- 3) оцінити катастрофічні сценарії (VaR, CVaR);
- 4) оцінити, чи є страховий продукт економічно ефективним (RAROC, GP).

Поєднання цих підходів формує комплексну систему фінансового ризик-менеджменту в аграрній сфері.

## 1.6 Висновки до розділу 1

У першому розділі було здійснено комплексний теоретичний аналіз природи, структури та методологічних підходів до оцінювання страхових ризиків в аграрному секторі. На основі вивчення наукових джерел, принципів аграрного виробництва та сучасних методів аналізу даних встановлено, що аграрні ризики мають складний, багатофакторний і стохастичний характер, який зумовлений поєднанням природно-кліматичних, технологічних, біологічних та економічних чинників. Це визначає необхідність використання як традиційних статистичних підходів, так і сучасних інтелектуальних методів у процесі їх оцінювання.

У роботі наведено ґрунтовну класифікацію ризиків у сільському господарстві, що дозволяє систематизувати їх за джерелами виникнення, рівнем контрольованості та часовою структурою. Особливу увагу приділено кліматичним ризикам та просторово-часовій варіабельності врожайності, які є головними джерелами невизначеності для агровиробників і страховиків. Показано, що специфіка розподілів аграрних даних, зокрема наявність «важких хвостів» та нестабільність показників у часі, ускладнюють використання класичних статистичних моделей і потребують впровадження методів прогнозної аналітики та машинного навчання.

Проаналізовано теоретичні засади аграрного страхування та наведено опис головних організаційних і фінансових механізмів, що лежать в основі системи страхового захисту. Визначено, що ефективність страхування

залежить від можливості точно оцінити ймовірність настання страхових подій та очікувані збитки, а також від адекватності застосованих моделей для тарифоутворення. Розглянуто переваги й обмеження традиційних та індексних страхових підходів, що формують основу сучасних страхових продуктів.

Окремо проаналізовано методи аналізу та прогнозування ризиків. Встановлено, що статистичні, економіко-математичні та машинні методи виконують різні функції у процесі оцінювання ризику й можуть бути ефективними лише за умови їх комбінованого використання. Статистичні методи забезпечують базову оцінку закономірностей, стохастичні моделі дозволяють аналізувати сценарії невизначеності, а алгоритми машинного навчання — виявляти складні нелінійні залежності у великих наборах даних. Такий інтегрований підхід є особливо доцільним для задач агростраховання, які поєднують прогнозування врожайності та моделювання ймовірнісних втрат.

Порівняльний аналіз якісних, кількісних і гібридних методів оцінювання ризику показав, що жоден з них не є універсально оптимальним. Якісні методи є оперативними, але недостатньо точними; кількісні — забезпечують повноту та математичну обґрунтованість, але потребують великих масивів даних; гібридні — дозволяють інтегрувати переваги обох підходів і виявляються найбільш ефективними в умовах комплексних аграрних систем.

Узагальнюючи проведений аналіз, можна зробити висновок, що для сучасного аграрного сектору України актуальним є перехід до інтегрованої методики оцінювання страхових ризиків, яка поєднує дані різних джерел, потужні обчислювальні моделі та імітаційні підходи. Теоретичні результати розділу створюють основу для формування практичної моделі прогнозування врожайності та кількісної оцінки ризику, що буде представлено у наступному розділі роботи.

## 2 ПРАКТИЧНА РЕАЛІЗАЦІЯ СИСТЕМИ

Інформаційна технологія, розроблена для вирішення задачі прогнозування аграрних ризиків, має модульну архітектуру, що забезпечує гнучкість, масштабованість і можливість адаптації до різних типів даних і моделей. Її метою є побудова інтелектуальної системи підтримки рішень, здатної здійснювати аналіз, прогнозування та оцінку ризиків із подальшим формуванням рекомендацій для користувача.

Інформаційна технологія включає такі основні модулі:

1. Модуль збору та підготовки даних.
2. Модуль аналітичного оброблення та нормалізації.
3. Модуль моделювання та прогнозування.
4. Модуль оцінки якості моделей.
5. Модуль формування рекомендацій і візуалізації результатів.

Взаємодія між модулями забезпечується через єдине інформаційне середовище, що підтримує обмін даними у форматах CSV, JSON. На рис. 2.1 зображено блок-схему архітектури системи.

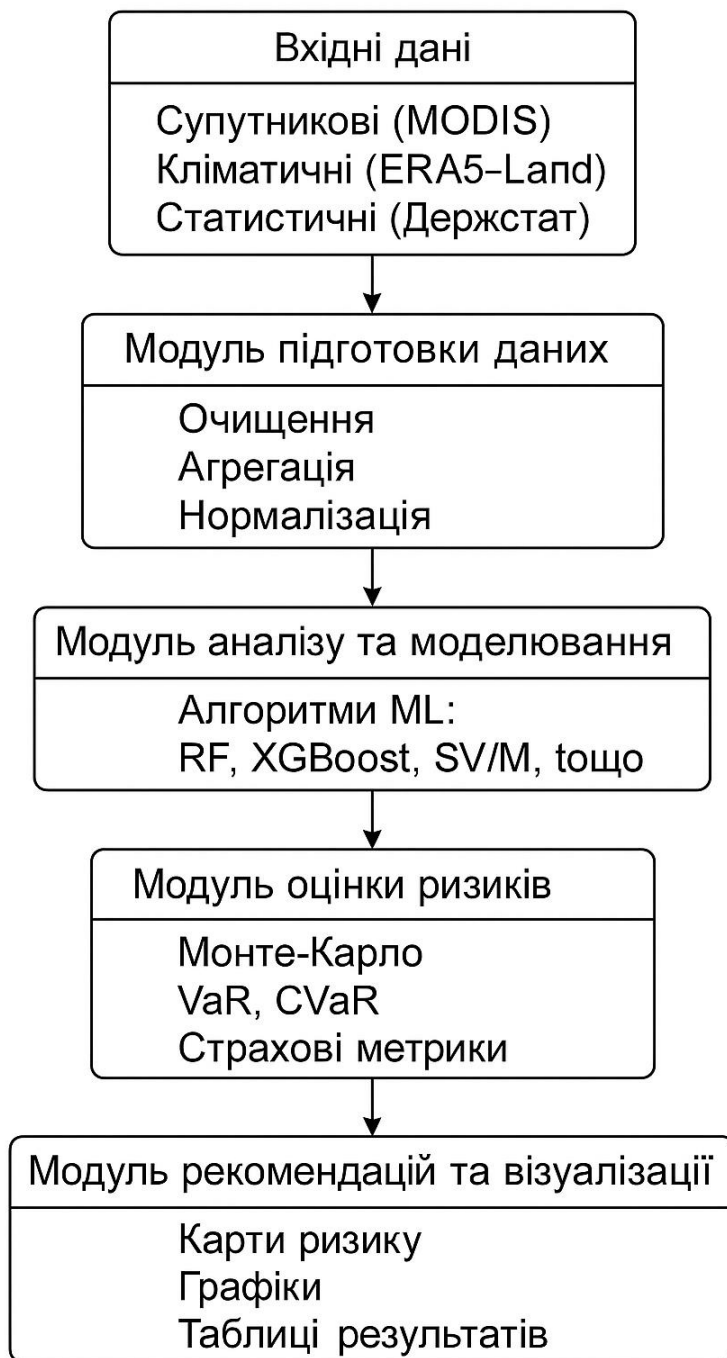


Рисунок 2.1 Загальна архітектура інтелектуальної системи оцінювання аграрних ризиків

Основним завданням модуля збору та підготовки даних є інтеграція даних з різних джерел — статистичних баз (Держстат, FAO), метеорологічних реаналізів (ERA5), супутникових індексів (MODIS, NDVI/EVI), фінансових та ринкових показників. На цьому етапі здійснюється попередня обробка:

очищення від пропусків і аномалій, нормалізація масштабів, формування ознак, агрегація за часовими інтервалами.

Модуль аналітичного оброблення та нормалізації забезпечує первинний статистичний аналіз: розрахунок середніх, дисперсій, трендів, кореляційних зв'язків між погодними факторами, врожайністю та економічними показниками. На цьому етапі також проводиться виявлення головних компонент (PCA) для зменшення розмірності та виділення ключових факторів ризику.

Основним ядром інформаційної технології є модуль моделювання, який реалізує алгоритми машинного навчання — такі як регресія (Linear, Ridge, Lasso), дерева рішень (Decision Tree, Random Forest, XGBoost) для прогнозування врожайності та оцінки ймовірності страхових випадків. Алгоритми обираються з урахуванням типу даних (часові ряди або панельні дані) та доступного обсягу історичної інформації. У модулі передбачено крос-валідацію, налаштування гіперпараметрів і побудову ансамблевих моделей для підвищення точності прогнозу.

Вихідними результатами модуля є прогнозні значення ключових показників — урожайності, ймовірності збитків. На основі цих даних у наступному модулі здійснюється кількісна оцінка ризику.

Для перевірки ефективності прогнозних моделей використовується набір метрик: середньоквадратична похибка (RMSE), середня абсолютна похибка (MAE), відносна похибка (MAPE), коефіцієнт детермінації ( $R^2$ ) та показники економічної адекватності (Expected Shortfall, Value-at-Risk). Також здійснюється аналіз залишків для виявлення систематичних похибок і перевірка стабільності моделей у часі (backtesting). Модуль автоматично ранжує моделі за точністю й надійністю, що дає змогу обрати оптимальну для подальших прогнозів.

Фінальний етап інформаційної технології — створення аналітичних рекомендацій на основі результатів моделювання. Система формує прогноз ризику для кожного регіону або культури, визначає оптимальний рівень страхового покриття, пропонує параметри страхового договору (страхову суму, франшизу, розмір премії). Для зручності користувача передбачено інтерфейс візуалізації даних — динамічні графіки, карти ризиків і таблиці з основними показниками.

Розроблена інформаційна технологія може бути реалізована як веб-сервіс або модуль аналітичної платформи. Її використання дозволяє автоматизувати процес аналізу аграрних ризиків, зменшити людський фактор у прийнятті рішень і забезпечити науково обґрунтовану підтримку страхових та інвестиційних стратегій у сільському господарстві.

## 2.1 Вибір мови програмування

Вибір мови програмування є одним із ключових етапів розробки будь-якого програмного продукту, оскільки він безпосередньо впливає на архітектуру системи, швидкість розробки, продуктивність, можливості подальшого масштабування та підтримки. Для реалізації поставлених у даній магістерській роботі завдань необхідно було обрати інструмент, що найкраще відповідає специфіці проєкту, а саме розробка системи для моделювання ризиків втрати врожаю.

Процес вибору мови програмування ґрунтувався на аналізі за такими ключовими критеріями:

1. Простота синтаксису та швидкість розробки. Мова повинна мати низький поріг входження та читабельний синтаксис, що дозволяє зосередитись на вирішенні бізнес-завдань, а не на боротьбі зі складністю самої мови.
2. Наявність та зрілість екосистеми. Важлива наявність великої кількості готових бібліотек, фреймворків та інструментів, що дозволяють прискорити розробку та використовувати перевірені рішення.
3. Підтримка спільноти. Активна та велика спільнота розробників є запорукою швидкого вирішення проблем, наявності якісної документації, навчальних матеріалів та постійного розвитку мови.
4. Крос-платформеність. Можливість запуску програмного забезпечення на різних операційних системах (Windows, macOS, Linux) без необхідності значної модифікації коду.
5. Інтеграційні можливості. Здатність мови легко взаємодіяти з іншими системами, базами даних, API та сервісами.
6. Продуктивність. Хоча для багатьох сучасних завдань швидкість виконання не є головним пріоритетом, цей параметр враховувався з огляду на потенційні навантаження на систему.

В рамках аналізу були розглянуті такі мови програмування, як Java, C++, JavaScript (Node.js) та Python.

Java є потужною, об'єктно-орієнтованою мовою з високою продуктивністю завдяки віртуальній машині JVM. Вона має величезну екосистему і широко використовується у великих ентерпрайз-додатках. Однак її синтаксис є більш багатослівним у порівнянні з Python, що може уповільнювати процес прототипування та розробки.

C++ — це компільована мова, що забезпечує максимальну продуктивність та низькорівневий контроль над системними ресурсами. Вона є стандартом у сферах, де швидкість є критичною, наприклад, у розробці ігор чи високоефективних обчисленнях. Проте складність управління пам'яттю та

більш крута крива навчання роблять її менш придатною для швидкої розробки проєктів, де пріоритетом є гнучкість.

JavaScript у поєднанні з платформою Node.js є чудовим вибором для розробки веб-додатків, особливо для I/O-інтенсивних завдань. Його асинхронна модель та уніфікація мови для клієнтської та серверної частини є значною перевагою. Проте для наукомістких обчислень, машинного навчання та комплексної обробки даних його екосистема менш розвинена, ніж у Python.

Після ретельного порівняльного аналізу, для реалізації даного проєкту було обрано мову програмування Python.

Python — це високорівнева інтерпретована мова програмування загального призначення, яка за останні роки здобула величезну популярність у наукових колах, аналізі даних, машинному навчанні та веб-розробці. Її філософія, викладена у "Дзені Python" (PEP 20), наголошує на читабельності коду та простоті синтаксису, що робить її ідеальною для швидкого створення прототипів та ітеративної розробки [11].

Однією з головних переваг Python є його мінімалістичний та інтуїтивно зрозумілий синтаксис, який дуже схожий на псевдокод. Це дозволяє розробникам писати менше коду для досягнення того ж результату, порівняно з такими мовами, як Java або C++. Чистий синтаксис полегшує не тільки написання, але й читання та підтримку коду, що є критично важливим в рамках магістерської роботи, де логіка алгоритмів має бути прозорою та зрозумілою [12].

Python має найбагатшу та найрозвиненішу екосистему бібліотек для наукових досліджень, аналізу даних та машинного навчання. Це стало вирішальним фактором при виборі. Вкажемо ключові бібліотеки, які використовуються у проєкті.

1. NumPy (Numerical Python): фундаментальна бібліотека для наукових обчислень, що надає підтримку для великих, багатовимірних масивів

та матриць, разом із широким набором високорівневих математичних функцій для операцій з цими масивами.

2. Pandas: надає високорівневі структури даних (DataFrame) та інструменти для аналізу даних, що робить роботу з табличними та часовими даними надзвичайно простою та ефективною [13].
3. Matplotlib та Seaborn: потужні бібліотеки для створення статичних, анімованих та інтерактивних візуалізацій. Візуалізація даних є невід'ємною частиною дослідження, і ці інструменти дозволяють створювати якісні графіки для представлення результатів.
4. Scikit-learn: одна з найпопулярніших бібліотек для машинного навчання. Вона надає прості та ефективні інструменти для класифікації, регресії, кластеризації, зменшення розмірності та препроцесингу даних.
5. TensorFlow та PyTorch: провідні фреймворки для глибинного навчання, що дозволяють створювати та тренувати складні нейронні мережі.

Наявність цих готових, оптимізованих та перевірених часом інструментів дозволила значно скоротити час розробки та зосередитися на експериментальній частині дослідження [14].

Python має одну з найбільших та найактивніших спільнот розробників у світі. Це означає наявність величезної кількості документації, підручників, форумів (наприклад, Stack Overflow), де можна швидко знайти відповідь на будь-яке питання. Крім того, активна спільнота постійно створює нові бібліотеки та вдосконалює існуючі, що гарантує актуальність та розвиток технологічного стеку.

Python є кросплатформною мовою, що означає, що код, написаний на одній операційній системі, може без змін виконуватися на інших. Це забезпечує гнучкість розгортання та використання розробленого рішення.

Більше того, Python легко інтегрується з іншими мовами (наприклад, C/C++ для прискорення обчислень) та технологіями, що робить його чудовим "клеєм" для об'єднання різних компонентів системи.

Хоча для даної роботи ключовими були можливості Python у сфері аналізу даних, його універсальність є додатковою перевагою. На Python можна реалізувати не тільки обчислювальне ядро системи, але й, за необхідності, створити веб-інтерфейс (з використанням фреймворків, як-от Django або Flask), розробити API для взаємодії з іншими сервісами або написати скрипти для автоматизації рутинних завдань. Така гнучкість дозволяє побудувати повноцінну систему, використовуючи єдину мову програмування.

Враховуючи вищезазначені фактори, а саме: низький поріг входження, наявність потужних спеціалізованих бібліотек, величезну підтримку спільноти та гнучкість, мова програмування Python була визнана оптимальним вибором для вирішення завдань, поставлених у даній магістерській роботі. Вона дозволяє ефективно проводити дослідження, швидко реалізовувати складні алгоритми та забезпечує надійну основу для побудови якісного програмного продукту.

## 2.2 Опис вхідних даних

Для навчання моделі прогнозування врожайності було зібрано та оброблено гетерогенний набір даних, що включає супутникові знімки, кліматичні реаналізи та картографічні матеріали. Цей підрозділ детально описує джерела вхідних даних та методологію їхньої попередньої обробки, яка була реалізована за допомогою мови програмування Python з використанням бібліотек pandas, geopandas, rasterio та xarray.

Основне завдання обробки полягало в агрегації різнорідних просторових даних до єдиного часового ряду на рівні адміністративних областей України з місячним кроком. Кінцевий набір даних містить для кожної області та кожного місяця вегетаційного періоду усереднені значення ключових екологічних та кліматичних показників.

Опис вхідних даних.

1. Супутникові дані MODIS. Для моніторингу стану рослинності було використано дані, отримані зі спектрорадіометра MODIS (Moderate Resolution Imaging Spectroradiometer) [15]. Ці дані надають низку вегетаційних індексів та біофізичних показників. У дослідженні використовувалися такі показники:
  - 1) NDVI (Normalized Difference Vegetation Index) та EVI (Enhanced Vegetation Index) — показники, що характеризують щільність та здоров'я рослинного покриву. Роздільна здатність — 250 м.
  - 2) LAI (Leaf Area Index) та FPAR (Fraction of Photosynthetically Active Radiation) — індекси листової поверхні та частки фотосинтетично активної радіації. Роздільна здатність — 500 м.
  - 3) LST (Land Surface Temperature) — температура поверхні землі, окремо для денного (LST\_Day) та нічного (LST\_Night) часу. Роздільна здатність — 1000 м.
2. Кліматичні дані ERA5-Land. Для аналізу метеорологічних умов використовувалися дані кліматичного реаналізу ERA5-Land [16]. Це глобальний набір даних, що надає погодинну інформацію про стан поверхні землі з роздільною здатністю ~9 км (0.1°). Дані були надані у форматі NetCDF і включали такі змінні:
  - 1) t2m: температура повітря на висоті 2 метрів;
  - 2) stl1, stl2: температура першого та другого шарів ґрунту;
  - 3) ssr: короткохвильова сонячна радіація на поверхні;

- 4) tp: сумарна кількість опадів;
  - 5) swvl1, swvl2: об'ємна вологість першого та другого шарів ґрунту.
3. Допоміжні геопросторові дані.
- 1) CropMap EU: растрова карта земного покриття, що надає детальну інформацію про типи сільськогосподарських культур. У цьому дослідженні її було використано для ідентифікації пікселів, зайнятих цільовою культурою (наприклад, пшеницею, код 211) [17].
  - 2) Адміністративні межі України: векторний файл у форматі GeoJSON, що визначає полігони областей України.

### 2.3 Попередня обробка даних

Попередня обробка даних є фундаментальним етапом у будь-якому проєкті, пов'язаному з аналізом даних та машинним навчанням. Її основна мета — перетворення необроблених, "сирих" даних, які часто є неповними, зашумленими та несумісними, у чистий, структурований та придатний для моделювання формат. Процес обробки включає низку методів, таких як очищення даних (заповнення пропусків, видалення аномалій), трансформація (нормалізація, масштабування), інтеграція (об'єднання даних з різних джерел) та агрегація.

Важливість цього етапу неможливо переоцінити, оскільки якість вхідних даних безпосередньо визначає якість результатів роботи моделі. В аналізі даних існує відомий принцип “Сміття на вході — сміття на виході”, який підкреслює, що навіть найскладніший алгоритм не зможе надати точних результатів, якщо він навчається на некоректних або нерепрезентативних даних. Неналежна підготовка даних може призвести до хибних висновків,

низької точності прогнозів та, як наслідок, до недієздатності всієї аналітичної моделі.

У контексті даного дослідження попередня обробка є особливо критичною через гетерогенність використовуваних джерел. Дані MODIS, ERA5-Land та CropMap мають різні просторові роздільні здатності, формати (растр, NetCDF) та часові кроки. Без ретельної обробки та інтеграції неможливо було б створити єдиний узгоджений набір ознак для навчання моделі. Розрахунок середньозважених показників з урахуванням щільності посівів та приналежності до регіону є прикладом не просто очищення, а й інтелектуального збагачення даних, що дозволяє створити значущі предиктори, які максимально точно відображають агрокліматичні умови на досліджуваних територіях.

Процес підготовки даних було розділено на два паралельні етапи для даних MODIS і ERA5-Land, після чого результати були об'єднані в єдиний набір.

Обробка супутникових знімків MODIS складалася з трьох послідовних кроків.

1. Створення фільтрованих TIF-файлів. На першому етапі для кожної роздільної здатності даних MODIS (250м, 500м, 1000м) було створено нові растрові файли. Значення кожного пікселя в цих файлах відповідало кількості пікселів цільової культури з високодетальної карти CropMap, що потрапляли в межі цього пікселя MODIS. Це дозволило кількісно оцінити щільність посівів у межах кожної комірки сітки MODIS.
2. Генерація вибірових точок. На другому етапі для кожної адміністративної області України були згенеровані "вибірові точки", що представляють собою пікселі сітки MODIS, які просторово перетинаються з полігоном області. Для кожної такої точки було розраховано два ключові коефіцієнти:

- 1) `crop_ratio` — частка пікселя MODIS, покрита посівами цільової культури;
- 2) `overlap_ratio` — частка площі пікселя MODIS, що знаходиться в межах адміністративної області.

Ці коефіцієнти були збережені у CSV-файли для кожної області та кожної роздільної здатності.

3. Розрахунок середньозважених значень. На фінальному етапі було видобуто значення показників (NDVI, EVI, LST тощо) для згенерованих вибірових точок. Для кожної області та кожного місяця розраховувалося середньозважене значення кожного показника. Як вага для усереднення використовувався добуток `crop_ratio` та `overlap_ratio`. Такий підхід дозволив врахувати як щільність посівів у пікселі, так і його просторовий внесок у площу області. Було встановлено порогові значення для коефіцієнтів (`crop_threshold=0.7`, `overlap_threshold=0.7`), щоб відфільтрувати пікселі з низькою концентрацією культури або незначним перетином з областю. Результатом стали річні CSV-файли з місячними агрегованими даними для кожної області.

Підготовка кліматичних даних виконувалася аналогічно.

1. Генерація вибірових точок для сітки NetCDF. Спочатку було створено вибірові точки на основі сітки даних ERA5-Land. Для кожної комірки сітки, що перетиналася з межами адміністративних областей, було розраховано коефіцієнти `crop_ratio` (на основі CropMap) та `overlap_ratio`. Ця інформація була збережена у регіональні CSV-файли.
2. Видобуток та усереднення значень NetCDF. Використовуючи згенеровані точки, було видобуто часові ряди кліматичних змінних (`t2m`, `tr` тощо) з файлу NetCDF. Для кожної області та кожного місяця розраховувалося середньозважене значення, де вагою також слугував добуток `crop_ratio` та `overlap_ratio`. Це дозволило отримати агреговані

кліматичні показники, релевантні саме для територій вирощування цільової культури в межах кожної області.

На заключному етапі оброблені та агреговані дані з джерел MODIS та ERA5-Land були об'єднані в одну таблицю. Злиття виконувалося за трьома ключами: назва регіону (`region_name`), рік (`year`) та місяць (`month`).

Кінцевий набір даних є панельною таблицею, де кожен рядок містить усереднені за місяць супутникові та кліматичні показники для конкретної області, що повністю готове для використання в моделях машинного навчання.

Приклад структури фінального датасету наведено на рисунку 2.2.

region_name	year	month	t2m	stl1	stl2	ssr	tp	swvl1	swvl2	LST_Night	EVI	LAI	LST_Day	NDVI	FPAR
Poltavska	2010.00	3.00	276.60	277.36	277.25	10635102.67	0.00	0.35	0.35	266.31	0.07	0.00	276.11	0.19	0.02
Poltavska	2010.00	4.00	283.70	284.10	283.51	14755818.60	0.00	0.30	0.30	277.56	0.17	0.02	296.07	0.35	0.08
Poltavska	2010.00	5.00	291.03	291.54	290.35	17502033.46	0.00	0.23	0.23	284.90	0.32	0.05	300.62	0.51	0.18
Poltavska	2010.00	6.00	296.48	297.52	296.28	19755509.42	0.00	0.22	0.21	272.47	0.39	0.07	301.78	0.57	0.24
Poltavska	2010.00	7.00	298.67	299.42	298.38	19227009.77	0.00	0.26	0.26	292.41	0.36	0.09	304.37	0.53	0.25
Poltavska	2010.00	8.00	300.29	302.20	301.42	19000049.07	0.00	0.14	0.16	288.34	0.19	0.04	303.26	0.34	0.17
Poltavska	2010.00	9.00	291.92	293.76	294.19	13108775.58	0.00	0.18	0.16	279.12	0.15	0.02	227.97	0.32	0.09
Poltavska	2010.00	10.00	281.78	283.14	284.32	6930946.01	0.00	0.31	0.29	272.87	0.14	0.01	269.92	0.33	0.08

Рисунок 2.2 Вигляд датасету

В якості змінної для прогнозування буде використано урожайність, взятую з сайту державної служби статистики України [18]. Урожайність в даному випадку вимірюється як центнер врожаю зібрано з одного гектара площі. Усі дані були взято з 2010 по 2023 роки.

## 2.4 Прогнозування врожайності методами машинного навчання

Цей розділ присвячено опису методології побудови та оцінки моделей машинного навчання для прогнозування врожайності пшениці на основі підготовленого набору даних. Метою експерименту є порівняння ефективності

різних алгоритмів регресійного аналізу та вибір найкращої моделі, що демонструє найвищу точність прогнозу.

#### 2.4.1 Загальна схема експерименту

Процес моделювання було організовано у вигляді послідовних етапів, що забезпечують об'єктивність та відтворюваність результатів. Загальна архітектура експерименту включала такі кроки.

1. Формування вибірки: на основі попередньо оброблених та об'єднаних даних та даних про врожайність було сформовано навчальну та тестову вибірки. Дослідження охоплювало 21 область України за період з 2011 по 2020 рік. Вхідними даними (матриця ознак  $X$ ) слугували агреговані місячні показники супутникових та кліматичних даних, а цільовою змінною  $y$  — річна врожайність пшениці.
2. Вибір алгоритмів машинного навчання: для вирішення задачі регресії було обрано шість поширених та добре зарекомендованих алгоритмів, а саме метод опорних векторів (SVM), випадковий ліс (Random Forest), градієнтний бустинг (Gradient Boosting), дерево рішень (Decision Tree), метод  $k$ -найближчих сусідів (K-Nearest Neighbors) та екстремальний градієнтний бустинг (XGBoost).
3. Оптимізація гіперпараметрів: для кожного з обраних алгоритмів було проведено процедуру налаштування гіперпараметрів за допомогою баєсівської оптимізації. Цей підхід дозволяє ефективно знаходити оптимальні комбінації параметрів, мінімізуючи середньоквадратичну помилку.

4. Валідація та навчання моделей: для оцінки узагальнюючої здатності моделей було застосовано стратегію перехресної валідації за принципом "виключення одного року" (Leave-One-Year-Out, LOYO). Цей метод є особливо релевантним для часових даних, оскільки дозволяє уникнути витоку даних та отримати надійну оцінку продуктивності моделі на "майбутніх", невідомих раніше даних.
5. Оцінка якості прогнозування: якість кожної моделі оцінювалася за допомогою стандартних метрик регресії: середньоквадратичної помилки (RMSE), коефіцієнта детермінації ( $R^2$ ) та коефіцієнта кореляції Пірсона.

#### 2.4.2 Вибір моделей машинного навчання

Вибір алгоритмів для прогнозування врожайності ґрунтувався на їхній здатності працювати зі складними, нелінійними залежностями в даних, що є характерним для агрокліматичних систем. Нижче наведено короткий опис кожної з використаних моделей.

1. Метод опорних векторів для регресії (Support Vector Regressor, SVR) будує гіперплощину в багатовимірному просторі ознак, намагаючись апроксимувати якомога більше точок даних із заданою похибкою. Ефективний для даних високої розмірності.
2. Випадковий ліс (Random Forest Regressor) - ансамблевий метод, який будує велику кількість дерев рішень на різних підвибірках даних і усереднює їхні прогнози. Це дозволяє зменшити ризик перенавчання та підвищити стабільність моделі.

3. Градієнтний бустинг (Gradient Boosting Regressor): ще один ансамблевий метод, який послідовно будує дерева рішень, де кожне наступне дерево намагається виправити помилки попереднього. Цей підхід дозволяє створювати дуже точні моделі.
4. Дерево рішень (Decision Tree Regressor) - інтерпретований алгоритм, який рекурсивно розбиває простір ознак на регіони, прогнозуючи в кожному з них середнє значення цільової змінної.
5. Метод k-найближчих сусідів (K-Nearest Neighbors Regressor) - простий метричний алгоритм, який прогнозує значення для нового об'єкта на основі усереднення значень його k-найближчих сусідів у просторі ознак.
6. Екстремальний градієнтний бустинг (XGBoost). Високоєфективна реалізація градієнтного бустингу, що включає механізми регуляризації для боротьби з перенавчанням та оптимізована для швидкості та продуктивності.

В таблиці 2.1 можна побачити порівняльний аналіз обраних методів.

Таблиця 2.1 Порівняльний аналіз методів машинного навчання

Модель	Основні характеристики	Переваги	Недоліки
Метод опорних векторів (SVR)	Будує гіперплощину в багатовимірному просторі, що апроксимує дані з певною похибкою. Використовує "ядерні" функції для роботи з нелінійними залежностями.	Ефективний для даних високої розмірності. Пам'яте-ефективний, оскільки використовує лише частину даних (опорні вектори). Гнучкий завдяки різним ядерним функціям.	Чутливий до вибору гіперпараметрів (C, gamma) та типу ядра. Погано масштабується на дуже великі набори даних. Низька інтерпретованість ("чорна скринька").
Випадковий ліс (Random Forest)	Ансамблевий метод, що навчає велику кількість дерев рішень на випадкових підвибірках даних і усереднює їхні прогнози.	Висока точність та стійкість до перенавчання. Добре працює з великою кількістю ознак. Надає оцінку важливості ознак.	Повільніший у навчанні, ніж окремі дерева. Модель є "чорною скринькою", важко інтерпретувати логіку. Може споживати значний обсяг пам'яті.

## Продовження таблиці 2.1

Гradientний бустинг (Gradient Boosting)	Ансамблевий метод, що послідовно будує дерева, де кожне наступне дерево навчається виправляти помилки попереднього.	Дуже висока точність. Гнучкий у налаштуванні завдяки різним функціям втрат. Може працювати з різними типами даних.	Схильний до перенавчання без належної регуляризації. Навчання відбувається послідовно, що ускладнює розпаралелювання. Чутливий до зашумлених даних.
Дерево рішень (Decision Tree)	Проста ієрархічна модель, що рекурсивно розбиває простір ознак на основі правил "якщо-то" для створення прогнозу.	Легко інтерпретувати та візуалізувати. Не потребує складної підготовки даних (напр., масштабування). Швидкий у навчанні та прогнозуванні.	Дуже схильний до перенавчання. Нестабільний: малі зміни в даних можуть кардинально змінити дерево. Зазвичай менш точний, ніж ансамблеві методи.
k-найближчих сусідів (KNN)	Метричний ("лінивий") алгоритм. Прогноз для нової точки робиться на основі усереднення значень її 'k' найближчих сусідів.	Простий у реалізації та розумінні. Не потребує фази навчання. Легко адаптується до нових даних.	Обчислювально затратний на етапі прогнозування. Ефективність сильно залежить від вибору метрики відстані. Погано працює з даними високої розмірності ("прокляття розмірності").
XGBoost	Оптимізована та масштабована реалізація gradientного бустингу з вбудованою регуляризацією для боротьби з перенавчанням.	Найвища точність серед багатьох алгоритмів. Висока швидкість роботи завдяки розпаралелюванню. Вбудовані механізми для обробки пропусків та регуляризації.	Складний у налаштуванні через велику кількість гіперпараметрів. Потребує більше ресурсів у порівнянні з простішими моделями. Залишається менш інтерпретованим, ніж прості моделі.

### 2.4.3 Оптимізація гіперпараметрів

Ефективність моделей машинного навчання значною мірою залежить від правильного вибору їхніх гіперпараметрів — налаштувань, які не визначаються в процесі навчання, а задаються заздалегідь. У даній роботі для цього процесу було використано байєсівську оптимізацію.

На відміну від традиційних методів, таких як повний перебір, байєсівська оптимізація розглядає задачу пошуку гіперпараметрів як оптимізацію "чорної скриньки". Вона будує ймовірнісну модель що відображає залежність метрики якості (в нашому випадку RMSE) від гіперпараметрів. На кожному кроці алгоритм обирає таку комбінацію параметрів, яка, згідно з поточною моделлю, має найбільший потенціал для покращення результату. Це дозволяє значно швидше знаходити оптимальні або близькі до оптимальних значення, особливо у випадках великого простору пошуку.

### 2.4.4 Стратегія валідації

Оскільки дані про врожайність мають часову структуру, стандартна к-блокова перехресна валідація може призвести до витоку даних (data leakage), коли модель "підглядає" в майбутнє. Щоб уникнути цього та отримати реалістичну оцінку продуктивності, було застосовано метод Leave-One-Year-Out (LOYO).

Суть методу полягає в ітеративному процесі, де на кожному кроці:

1. Один рік з набору даних (наприклад, 2015) повністю виключається і використовується як тестова вибірка.
2. Модель навчається на всіх даних за роки, що передували виключеному (в даному випадку, 2011-2014). Це імітує реальну ситуацію прогнозування на майбутній період.
3. Робиться прогноз для виключеного (тестового) року та розраховується помилка.
4. Процедура повторюється для кожного року в досліджуваному діапазоні, який може бути використаний як тестовий.

Таким чином, фінальна оцінка якості моделі усереднюється за результатами прогнозів на кожен з років, що забезпечує надійність та стійкість результатів [19].

На рис. 2.3 зображено ілюстрацію методу LOYO.



Рисунок 2.3 Ілюстрація методу LOYO

### 2.4.5 Метрики оцінки якості моделей

Для кількісної оцінки точності регресійних моделей було використано три стандартні метрики:

1. Середньоквадратична помилка (Root Mean Squared Error, RMSE) показує середнє відхилення прогнозованих значень від фактичних. Вимірюється в тих же одиницях, що й цільова змінна (ц/га), і надає перевагу більшим помилкам:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2},$$

де  $y_i$  — фактичне значення,  $\hat{y}_i$  — прогнозоване значення,  $n$  — кількість спостережень.

2. Коефіцієнт детермінації ( $R^2$ ) показує, яку частку варіації (дисперсії) цільової змінної пояснює модель. Значення варіюється від  $-\infty$  до 1. Чим ближче  $R^2$  до 1, тим краще модель описує дані.

$$R^2 = \frac{\text{var}(\hat{y})}{\text{var}(y)} = 1 - \frac{SSE}{SST}.$$

3. Коефіцієнт кореляції Пірсона вимірює лінійну залежність між фактичними та прогнозованими значеннями. Значення в діапазоні від -1 до 1. Високе позитивне значення (близьке до 1) свідчить про високу точність та правильний напрямок прогнозів.

Ці метрики були розраховані для кожної моделі на основі агрегованих результатів LOYO-валідації, що дозволило провести об'єктивне порівняння їхньої ефективності.

## 2.4.6 Результати моделювання

Розглянемо результати прогнозування 6 моделей.

### **Метод опорних векторів**

Для методу опорних векторів шляхом байєсівської оптимізації були підібрані наступні параметри:

1. 'kernel': 'poly' — використання поліноміального ядра. Це означає, що модель може знаходити складні, нелінійні залежності у даних.
2. 'C': 2.53 — параметр регуляризації. Визначає баланс між мінімізацією помилки на навчальних даних та уникненням перенавчання. Більше значення C допускає менше помилок на тренувальній вибірці.
3. 'degree': 5 — ступінь полінома для поліноміального ядра. Ступінь 5 дозволяє моделі створювати дуже гнучку криву для опису даних.
4. 'gamma': 0.044 — коефіцієнт ядра. Визначає, наскільки великий вплив має один тренувальний приклад.

На рис. 2.4 показано порівняння прогнозу та реальних даних для кожного року.

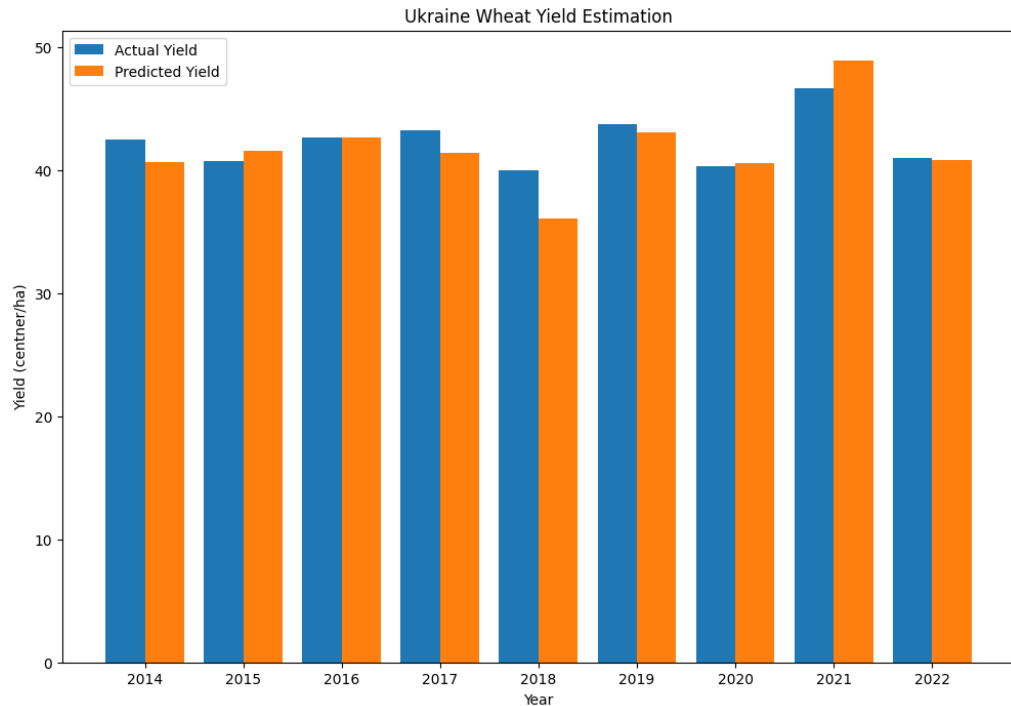


Рисунок 2.4 Порівняння реальних та прогнозованих значень моделі SVM

Візуальне порівняння показує, що модель SVM загалом встановлює тенденції річної врожайності. Проте, помітні суттєві відхилення в окремі роки. Наприклад, у 2018 році модель значно завищила прогноз, тоді як у 2021 році, навпаки, недооцінила рекордний урожай. Це свідчить про те, що модель може мати труднощі з прогнозуванням в екстремальних або нетипових умовах.

В таблиці 2.2 наведені метрики для оцінки прогнозування за різні роки та загалом.

Таблиця 2.2 Результати прогнозування моделі

Year	RRMSE	RMSE	MAE	MSE	MAPE
2014	8.15243	3.466335	2.799442	12.01548	6.441289
2015	12.34637	5.030851	4.203278	25.30946	10.86968
2016	8.796703	3.750746	2.933695	14.0681	6.586112
2017	10.84698	4.695194	3.810726	22.04485	8.82543
2018	14.2376	5.69165	4.72059	32.39488	11.08068
2019	7.608365	3.331739	2.368634	11.10049	5.099101

## Продовження таблиці 2.2

2020	10.98549	4.428722	3.495657	19.61358	10.00235
2021	10.49581	4.898045	3.68513	23.99085	8.103708
2022	7.739518	3.176151	2.48989	10.08793	6.095278
Total	10.28598	4.354125	3.389671	18.9584	8.122624
Sd	2.132441	0.829496	0.754065	7.216547	2.075348

Загальна середньоквадратична помилка становить 4.35, а середня абсолютна відсоткова помилка — 8.12%. Це означає, що в середньому прогноз моделі відхиляється від фактичного значення на 4.35 центнера з гектара. Продуктивність моделі є нестабільною по роках: найкращий результат був у 2022 році, а найгірший — у 2018.

Порівняємо на рис. 2.5 результати прогнозування для різних регіонів усереднено по рокам.

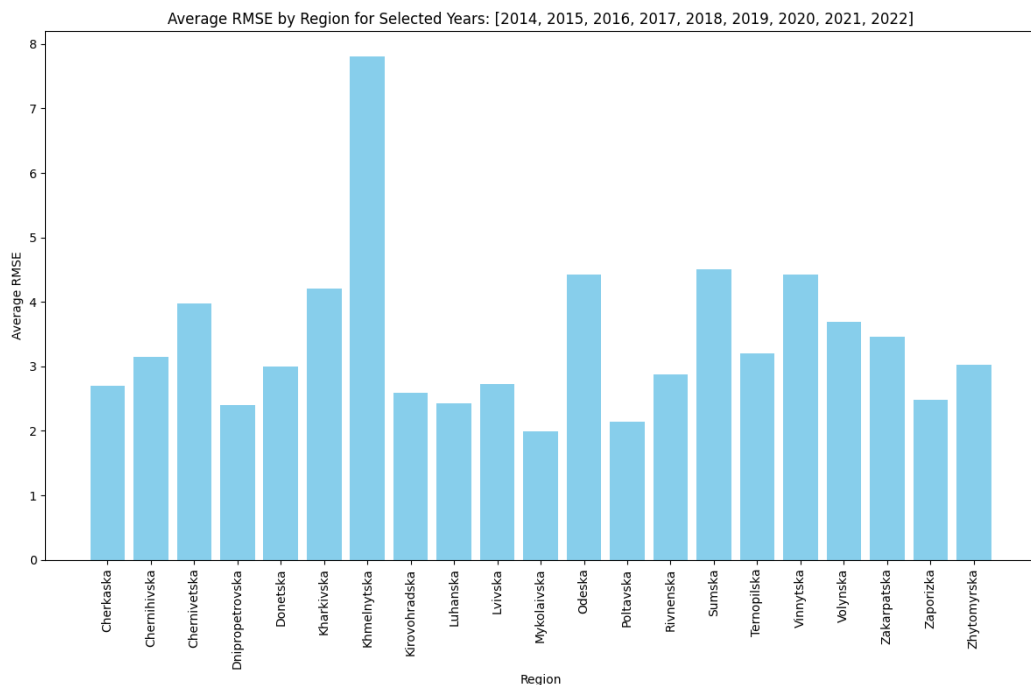


Рисунок 2.5 RMSE моделі по регіонах

Регіон з найвищим середнім RMSE — це Хмельницька область. Це може свідчити про те, що модель, яка використовується, має значні труднощі з

точністю прогнозування саме в цьому регіоні порівняно з іншими. Регіон з найнижчим середнім RMSE — це Львівська область.

### **Random Forest**

Для моделі Random Forest шляхом байєсівської оптимізації було підібрано такі параметри.

1. 'n\_estimators': 195 — кількість дерев у "лісі". Більша кількість дерев зазвичай покращує якість моделі, але до певної межі.
2. 'max\_depth': 30 — максимальна глибина кожного дерева. Глибші дерева можуть захоплювати складніші залежності, але є ризик перенавчання.
3. 'min\_samples\_split': 2 — мінімальна кількість зразків, необхідна для розділення внутрішнього вузла.
4. 'min\_samples\_leaf': 1 — мінімальна кількість зразків, яка має бути в листовому вузлі (кінцевому).

На рис. 2.6 показано порівняння прогнозу та реальних даних для кожного року.

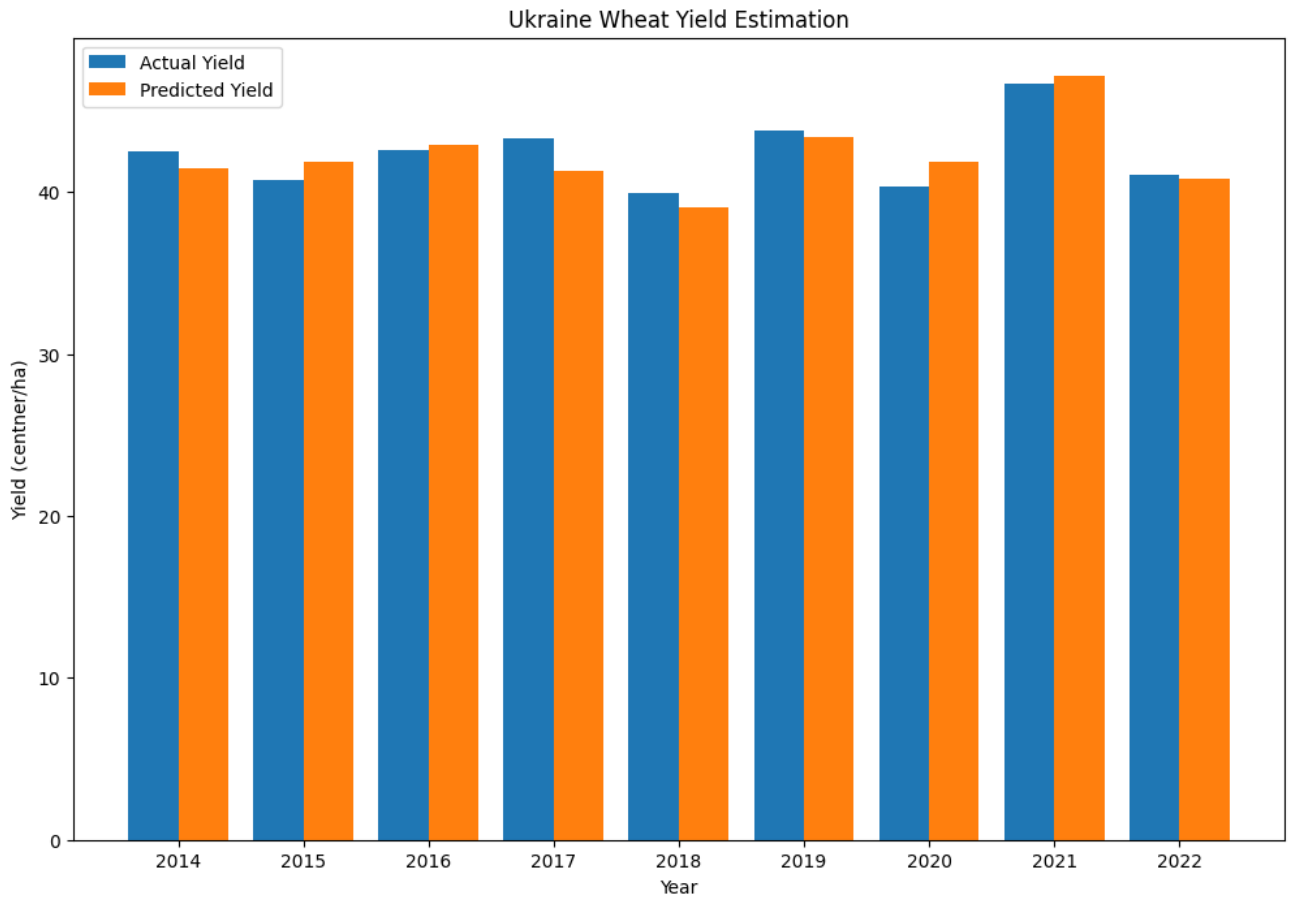


Рисунок 2.6 Порівняння реальних та прогнозованих значень моделі

На відміну від SVM, графік для Random Forest демонструє значно кращу відповідність між прогнозованими та фактичними значеннями. Модель точно відтворює коливання врожайності, включно з піковими значеннями, що свідчить про її високу точність.

В таблиці 2.3 наведені метрики для оцінки прогнозування за різні роки та загалом.

Таблиця 2.3 Результати прогнозування моделі

Year	RRMSE	RMSE	MAE	MSE	MAPE
2014	8.187736	3.481347	2.814929	12.11978	6.334043
2015	12.97309	5.286227	4.12331	27.9442	10.53918
2016	9.205855	3.925201	3.26881	15.4072	7.443681
2017	11.46152	4.961202	3.841381	24.61352	8.740658
2018	12.84908	5.136574	4.376786	26.38439	11.16449

Продовження таблиці 2.3

2019	8.515312	3.728896	2.852714	13.90466	6.600473
2020	14.3369	5.779817	4.869429	33.40628	14.20404
2021	8.396243	3.918247	3.128976	15.35266	6.85611
2022	10.71114	4.395646	3.39369	19.3217	8.297757
Total	10.81002	4.575956	3.630003	20.93938	8.908937
Sd	2.165488	0.758988	0.673939	6.986849	2.460924

Модель Random Forest (RF) продемонструвала загалом високу точність прогнозування протягом 2014–2022 років. Середнє значення RRMSE становить 10.81%, а MAPE — 8.91%, що свідчить про невисокий рівень відносної похибки. Значення RMSE = 4.58 та MAE = 4.63 підтверджують, що середні відхилення прогнозів від реальних значень є помірними. Таким чином, модель можна охарактеризувати як стабільну та ефективну для задач прогнозування в умовах невизначеності.

Найкращі результати зафіксовано у 2014, 2019 та 2021 роках, коли відносна похибка була нижчою за 8.5%, що свідчить про високу відповідність прогнозів фактичним даним. Водночас у 2015, 2018 та особливо у 2020 році точність знижувалась (RRMSE перевищував 12–14%), що, ймовірно, пов'язано зі зміною ринкових умов або появою зовнішніх шоків, таких як пандемія COVID-19. Незважаючи на це, стандартне відхилення показників помилки залишається невеликим, що підкреслює стабільність моделі у різні періоди.

Для кращого розуміння, які фактори найбільше впливають на прогноз, на рис. 2.7 зображено 20 найбільш важливих ознак для даної моделі.

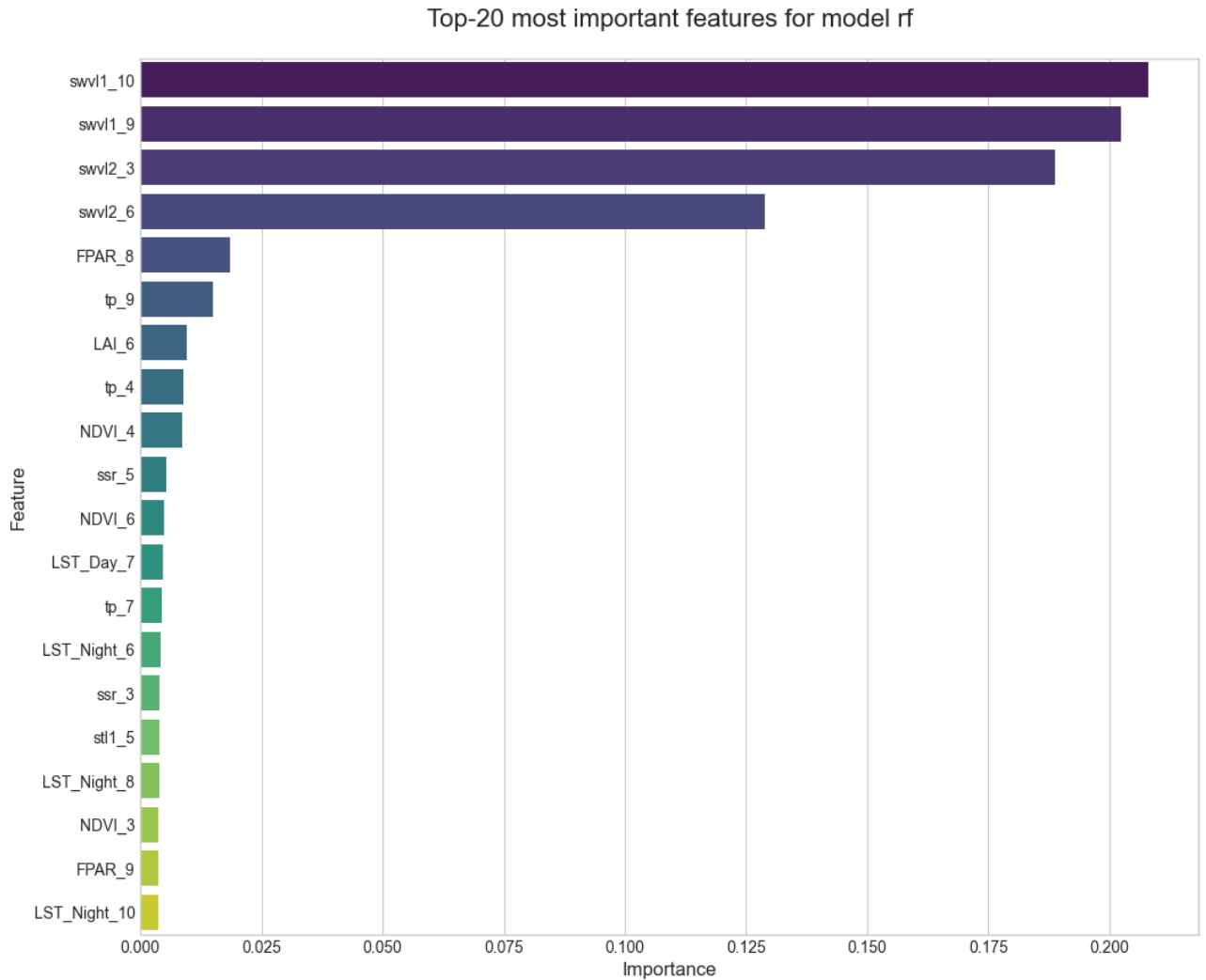


Рисунок 2.7 Важливість ознак для даної моделі

Аналіз важливості предикторів дає ключове уявлення про фактори, що впливають на врожайність. Найбільш значущими виявилися показники об'ємної вологості ґрунту (swvl1\_10, swvl1\_9, swvl2\_3). Це підкреслює критичну роль запасів води в ґрунті, особливо у вересні та жовтні (\_9, \_10), для формування кінцевого врожаю. Також важливими є показники стану рослинності (FPAR) та кількість опадів (tp).

На рис. 2.8 зображено гистограму з похибками для кожного регіону в середньому.

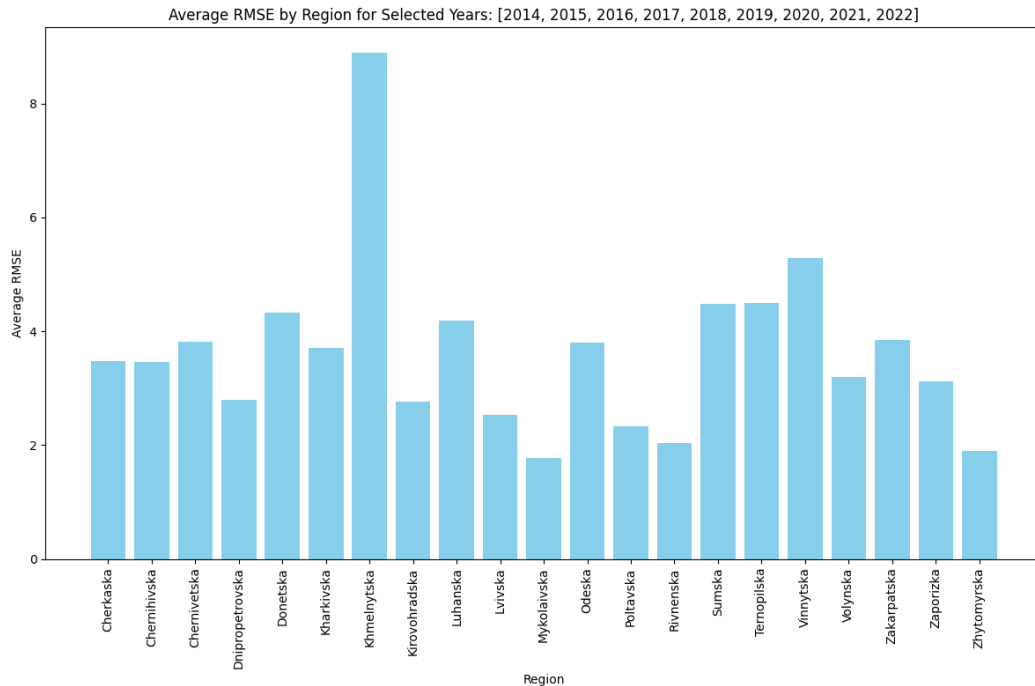


Рисунок 2.8 RMSE моделі по регіонах

Модель показує загалом стабільні результати по більшості регіонів, проте для окремих областей (особливо Хмельницької) спостерігаються аномально високі похибки, які потребують додаткового аналізу. Це може бути пов'язано з якістю вихідних даних, різними економічними умовами чи структурними змінами у регіонах.

**Gradient Boosting** - Екстремальний градієнтний бустинг. Це також ансамблевий метод, який послідовно будує дерева рішень, де кожне наступне дерево намагається виправити помилки попереднього. Вважається одним з найпотужніших алгоритмів для табличних даних. Оптимальні параметри підібрані методом баєсівської оптимізації.

1. 'n\_estimators': 483 — кількість дерев;
2. 'learning\_rate': 0.3 — швидкість навчання. Контролює, наскільки сильно кожне нове дерево коригує помилки попередніх;
3. 'max\_depth': 10 — максимальна глибина дерева;
4. 'subsample': 1.0 — частка навчальних даних, яка використовується для побудови кожного дерева;

5. 'colsample\_bytree': 0.6 — частка ознак (стовпців), яка використовується для побудови кожного дерева.

На рис. 2.9 показано порівняння прогнозу та реальних даних для кожного року.

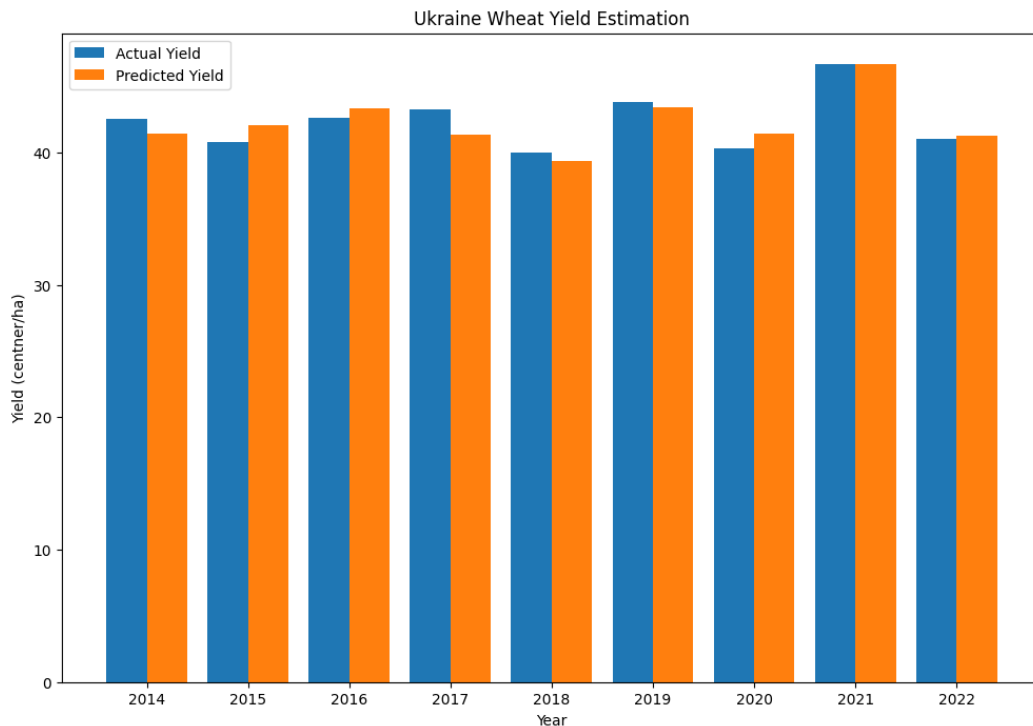


Рисунок 2.9 Порівняння реальних та прогнозованих значень моделі Як і Random Forest, градієнтний бустинг демонструє високу точність, а його прогнози добре корелюють з реальними даними.

На рис. 2.10 зображено 20 найбільш важливих ознак для даної моделі.

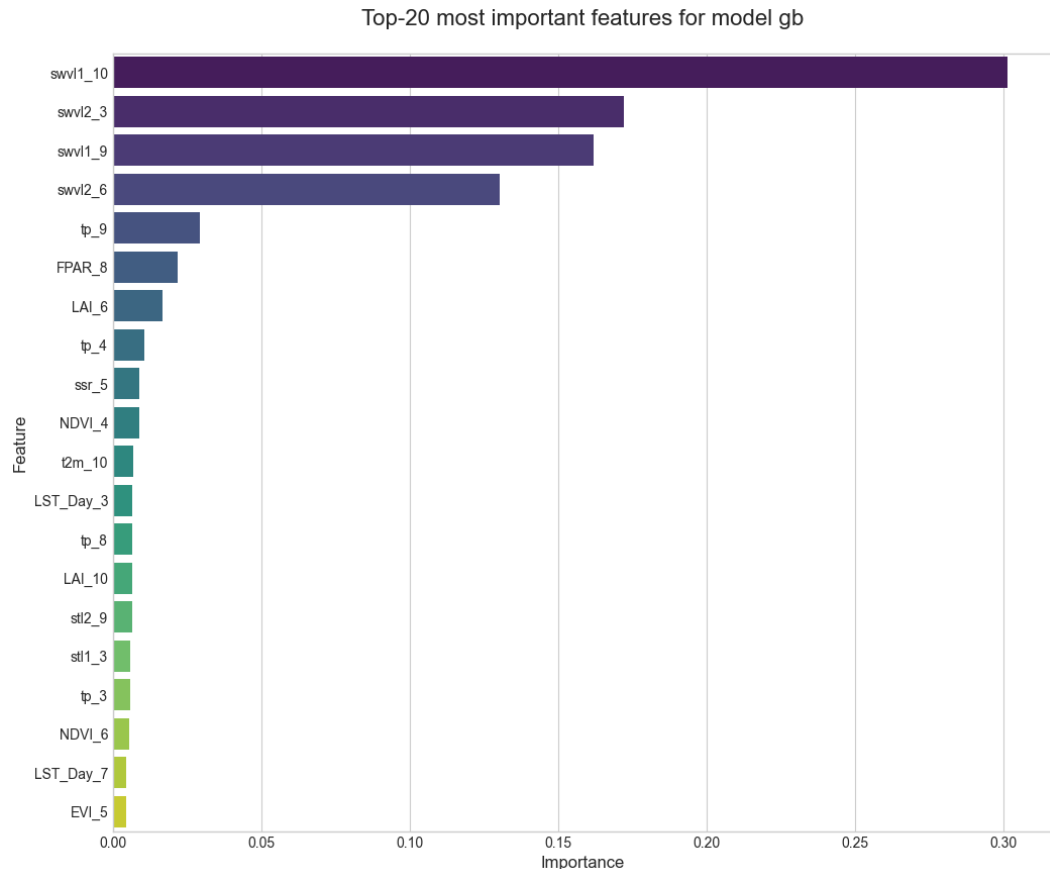


Рисунок 2.10 Важливість ознак для даної моделі

Результати повністю узгоджуються з моделлю Random Forest. Вологість ґрунту (swvl) знову є домінуючим фактором. Стабільність цього результату на двох різних ансамблевих моделях підтверджує, що вологість ґрунту є найнадійнішим предиктором врожайності в даному наборі даних.

У таблиці 2.4 наведені метрики для оцінки прогнозування за різні роки та загалом.

Таблиця 2.4 Результати прогнозування моделі

Year	RRMSE	RMSE	MAE	MSE	MAPE
2014	9.054107	3.84972	3.087713	14.82034	7.028379
2015	12.54597	5.112186	3.983057	26.13444	10.19663
2016	8.886087	3.788858	3.236179	14.35545	7.491845
2017	11.3644	4.919161	3.811497	24.19815	8.686603
2018	13.71551	5.482938	4.674794	30.06261	12.1258

## Продовження таблиці 2.4

2019	8.554704	3.746145	2.820161	14.03361	6.456431
2020	14.14477	5.702362	4.819228	32.51694	13.90408
2021	7.458059	3.480428	2.68947	12.11338	5.799246
2022	12.41578	5.095199	4.095767	25.96105	10.43351
Total	10.97346	4.64514	3.690874	21.57733	9.124724
Sd	2.321122	0.802915	0.733079	7.324348	2.585044

Загальна помилка RMSE становить 4.65, а MAPE — 9.12%. Ці показники дещо гірші, ніж у SVM, що може бути пов'язано з чутливістю моделі до викидів або шуму в даних. Найбільші помилки спостерігалися у 2018 та 2020 роках.

На рис. 2.11 зображено гістограму з похибками для кожного регіону в середньому.

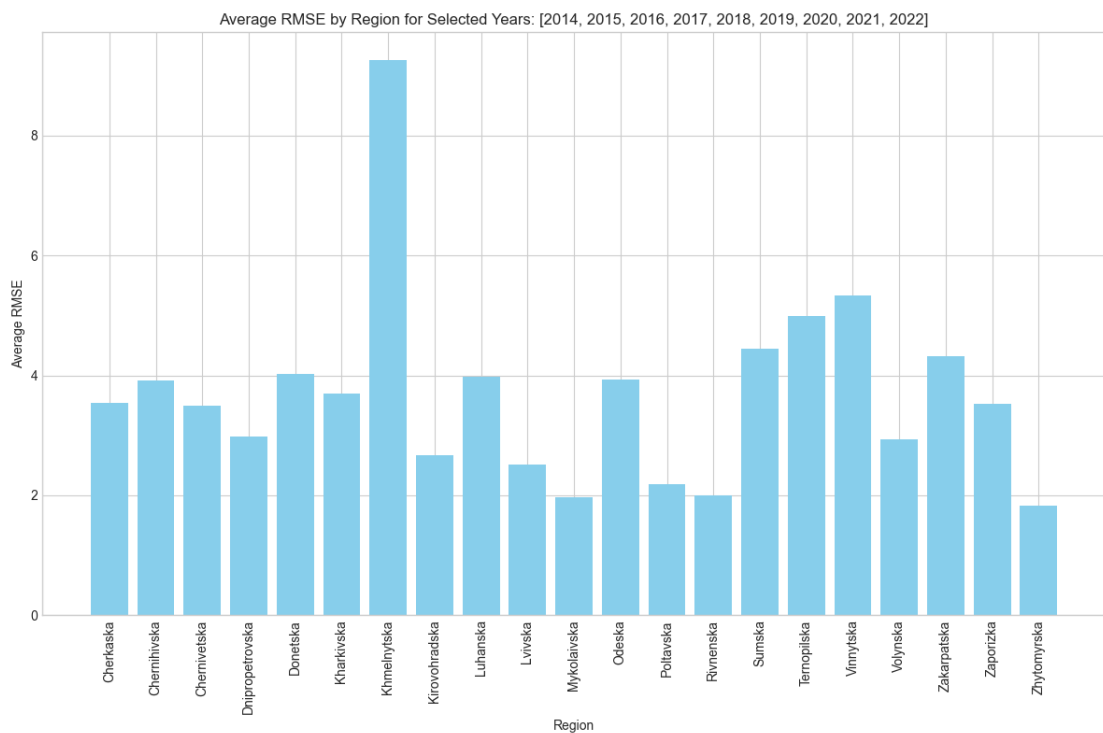


Рисунок 2.11 RMSE моделі по регіонах

**Дерево рішень** буде ієрархічну структуру правил (дерево) для прийняття рішень. Проста та інтерпретована модель. Оптимальні параметри підібрані методом баєсівської оптимізації.

1. 'max\_depth': 20 — максимальна глибина дерева;
2. 'min\_samples\_split': 2, 'min\_samples\_leaf': 1 — Параметри, що контролюють складність дерева;
3. 'max\_features': 'sqrt' — Кількість ознак для пошуку найкращого розділення (корінь квадратний від загальної кількості ознак).

На рис. 2.12 показано порівняння прогнозу та реальних даних для кожного року.

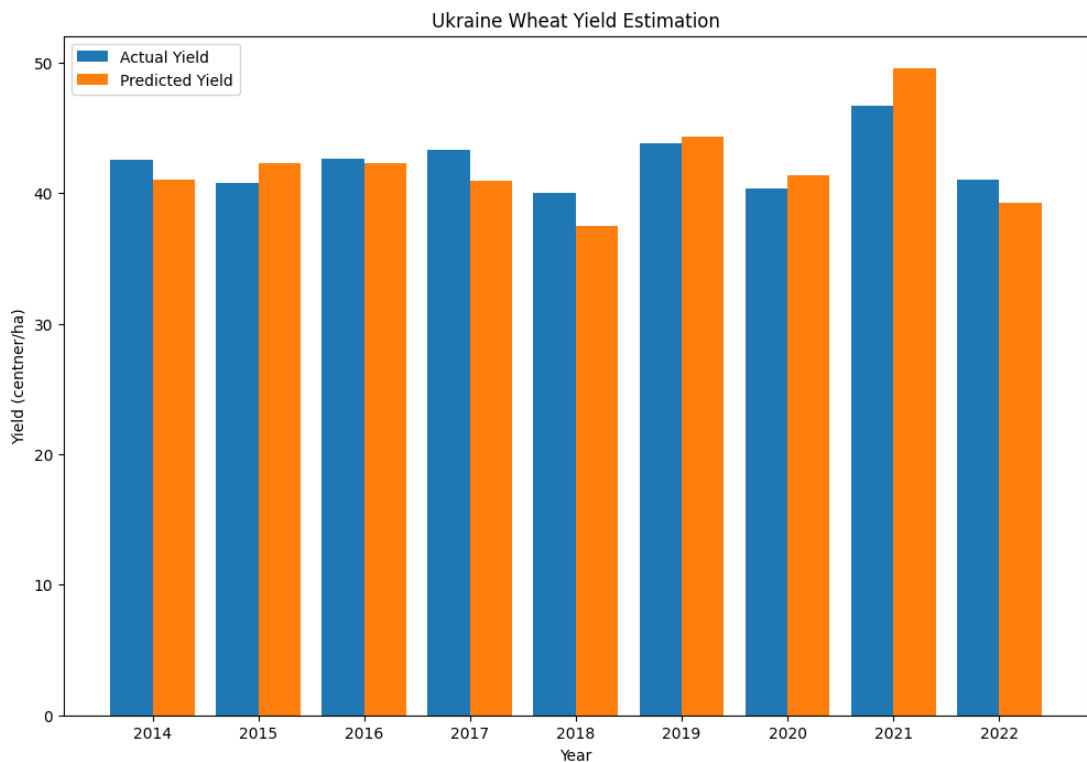


Рисунок 2.12 Порівняння реальних та прогнозованих значень моделі  
На рис. 2.13 зображено 20 найбільш важливих ознак для даної моделі.

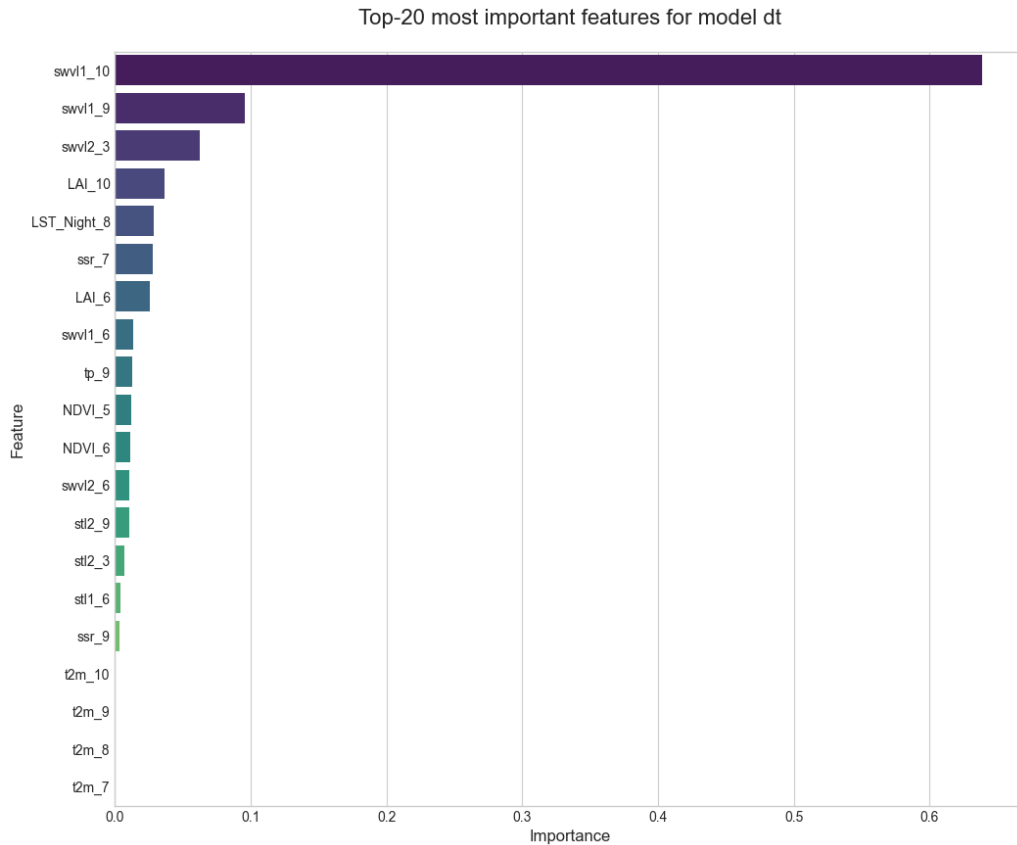


Рисунок 2.13 Важливість ознак для даної моделі

У таблиці 2.5 наведені метрики для оцінки прогнозування за різні роки та загалом.

Таблиця 2.5 Результати прогнозування моделі

Year	RRMSE	RMSE	MAE	MSE	MAPE
2014	11.66394	4.959396	3.97971	24.5956	9.05247
2015	16.25273	6.622599	4.986197	43.85882	12.17317
2016	11.65941	4.971352	4.065188	24.71434	9.250718
2017	14.78186	6.398432	4.873622	40.93994	10.67076
2018	14.50172	5.797236	4.765032	33.60794	11.8484
2019	9.270455	4.059576	3.434692	16.48016	8.031368
2020	14.27287	5.754007	4.996116	33.10859	14.18101
2021	14.04486	6.554266	5.406175	42.95841	12.29511
2022	13.60796	5.584446	4.601884	31.18604	10.96352
Total	13.44328	5.690634	4.567624	32.38332	10.94073
Sd	1.987353	0.804507	0.584748	8.77222	1.814632

Модель Decision Tree (DT) продемонструвала помірну точність прогнозування, однак поступається за якістю ансамблевій моделі Random Forest. Середні значення  $RRMSE = 13.44\%$  та  $MAPE = 10.94\%$  свідчать про вищий рівень відносних похибок, тоді як  $RMSE = 5.69$  і  $MAE = 4.57$  підтверджують більшу середню різницю між прогнозованими та фактичними значеннями. Це вказує на те, що DT має обмежену здатність до узагальнення даних і може перенавчатися на тренувальних вибірках.

У часовому розрізі найкращі результати спостерігаються у 2014, 2016 та 2019 роках ( $RRMSE \approx 9\text{--}11\%$ ), коли модель забезпечувала прийнятну точність прогнозів. Водночас у 2015, 2017, 2018, 2020 та 2021 роках точність суттєво знижувалася ( $RRMSE$  перевищував  $14\text{--}16\%$ ), що може бути наслідком різких коливань у даних або недостатньої глибини дерева для адаптації до нових закономірностей.

На рис. 2.14 зображено гістограму з похибками для кожного регіону в середньому.

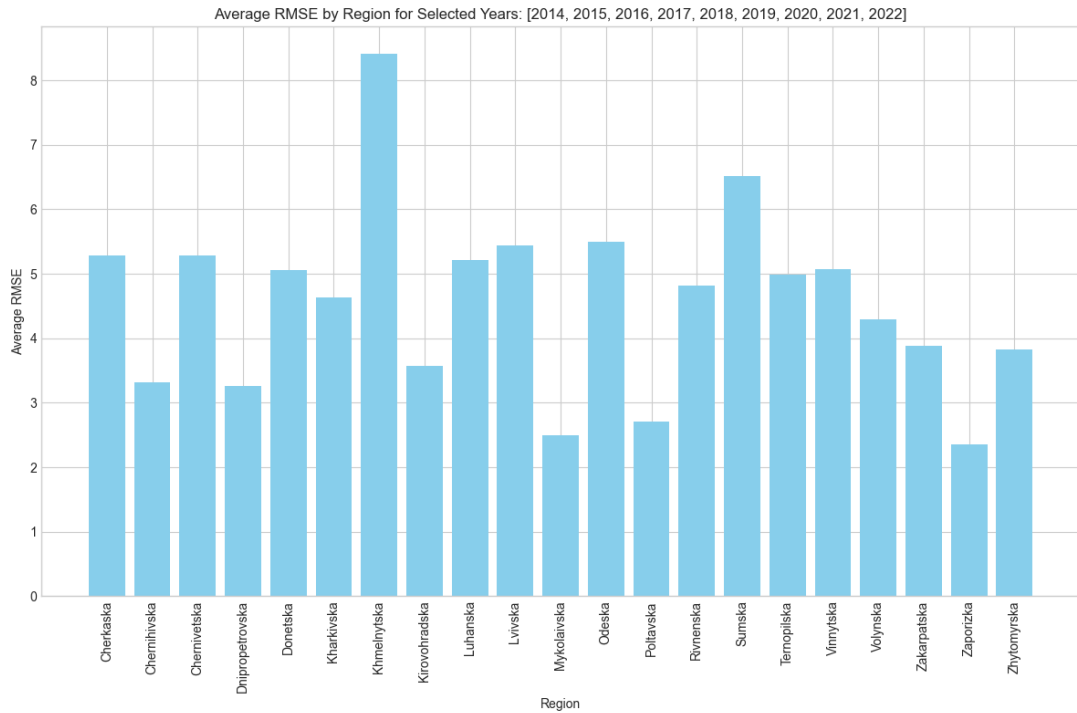


Рисунок 2.14 RMSE моделі по регіонах

Модель показує себе найкраще в Закарпатській області, де RMSE є найнижчим. Також відносно непогані результати спостерігаються в Полтавській та Дніпропетровській областях. У цих регіонах правила, які вивчило дерево, добре описують дані.

### Метод k-найближчих сусідів

Прогнозує значення для нового об'єкта на основі усередненого значення його 'k' найближчих сусідів у навчальній вибірці. Оптимальні параметри підібрані методом баєсівської оптимізації.

1. 'n\_neighbors': 4 — кількість сусідів, які враховуються для прогнозу.
2. 'weights': 'distance' — вага сусідів. 'distance' означає, що ближчі сусіди мають більший вплив на прогноз.
3. 'p': 1 — Параметр метрики відстані (1 — Манхеттенська відстань).

На рис. 2.15 показано порівняння прогнозу та реальних даних для кожного року.

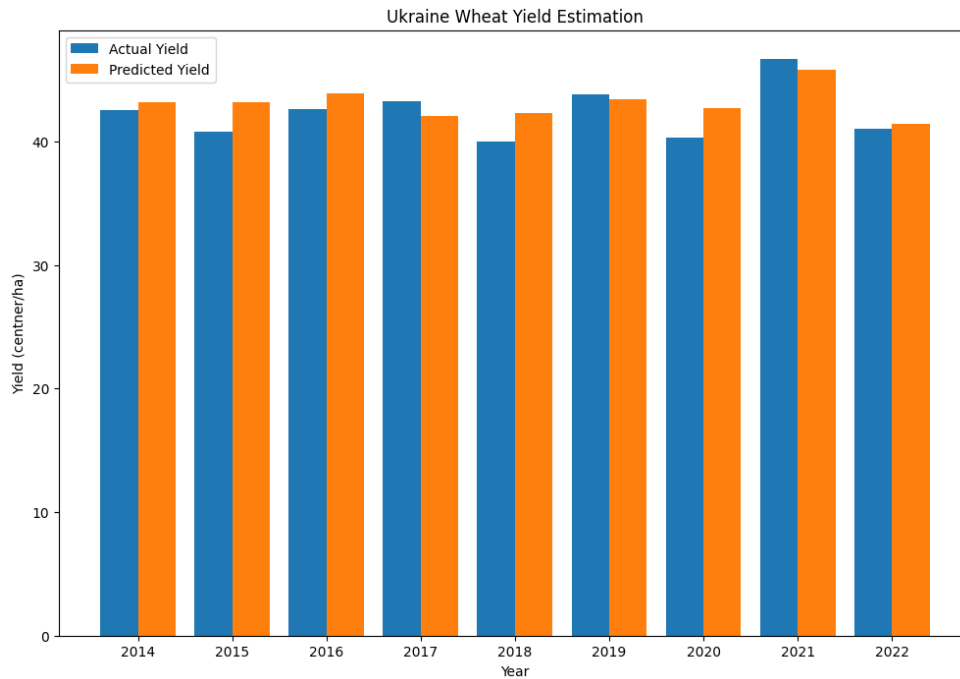


Рисунок 2.15 Порівняння реальних та прогнозованих значень моделі

Прогнози KNN виглядають "згладженими" і менш чутливими до різких коливань врожайності. Модель добре працює для років, схожих на попередні, але погано прогнозує нетипові роки.

На рис. 2.16 зображено гістограму з похибками для кожного регіону в середньому.

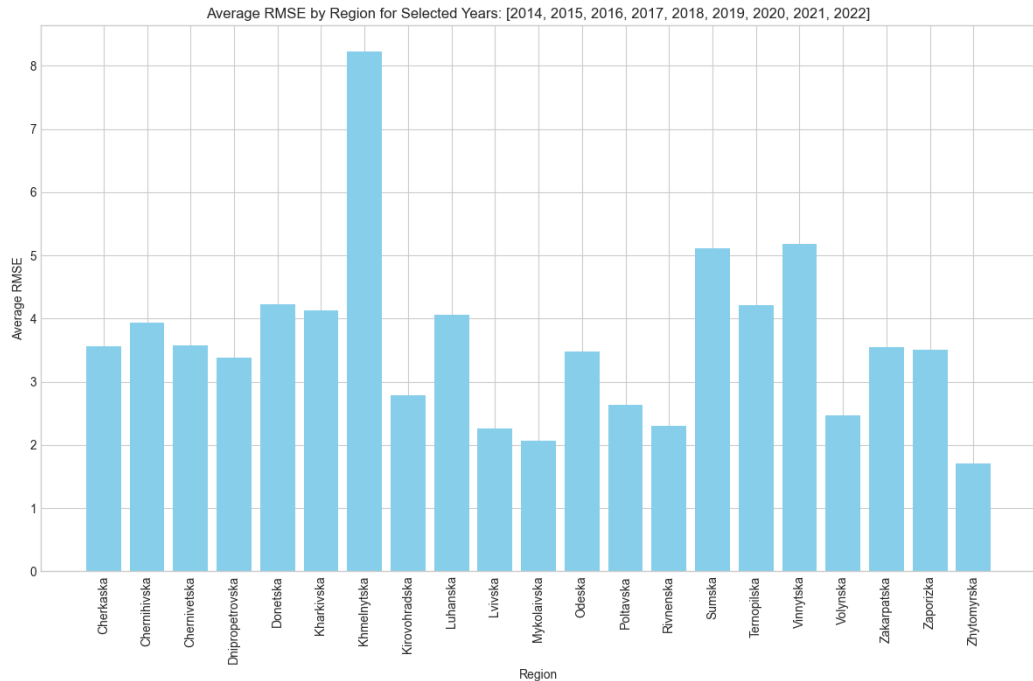


Рисунок 2.16 RMSE моделі по регіонах

У регіонах з високою помилкою, як-от Хмельницька область, дані можуть бути дуже "рідкими" або розкиданими. Це означає, що "найближчі сусіди" для точки прогнозу насправді знаходяться дуже далеко і не є показовими, що призводить до великих помилок.

В таблиці 2.6 наведені метрики для оцінки прогнозування за різні роки та загалом.

Таблиця 2.6 Результати прогнозування моделі

Year	RRMSE	RMSE	MAE	MSE	MAPE
2014	8.700538	3.699386	2.788647	13.68546	6.604735
2015	11.83854	4.823921	3.926881	23.27022	10.76668
2016	9.530966	4.063822	3.391083	16.51465	8.095289
2017	11.73082	5.077769	4.191078	25.78374	9.878173
2018	11.97747	4.788136	3.861314	22.92624	11.4589
2019	9.39357	4.113489	3.2528	16.92079	7.510504
2020	13.92579	5.614082	4.458167	31.51792	13.64696

Продовження таблиці 2.6

2021	10.48393	4.892499	4.081575	23.93655	9.023338
2022	12.75281	5.233512	4.199688	27.38965	10.8881
Total	11.19026	4.736914	3.794581	22.43836	9.763631
Sd	1.630268	0.584334	0.509273	5.410556	2.071619

Загальна помилка RMSE становить 4.74, а MAPE — 9.76%. Продуктивність моделі є посередньою — вона краща за одне дерево рішень, але поступається ансамблевим методам та SVM.

### Метод екстремального градієнтного бустингу

Схожий на XGBoost, це також ансамблевий метод послідовного навчання дерев. Оптимальні параметри підібрані методом баєсівської оптимізації.

1. 'n\_estimators': 50 — Кількість дерев.
2. 'learning\_rate': 0.2 — Швидкість навчання.
3. 'max\_depth': 2 — Максимальна глибина (дуже неглибокі дерева, що є типовим для градієнтного бустингу).

На рис. 2.17 показано порівняння прогнозу та реальних даних для кожного року.

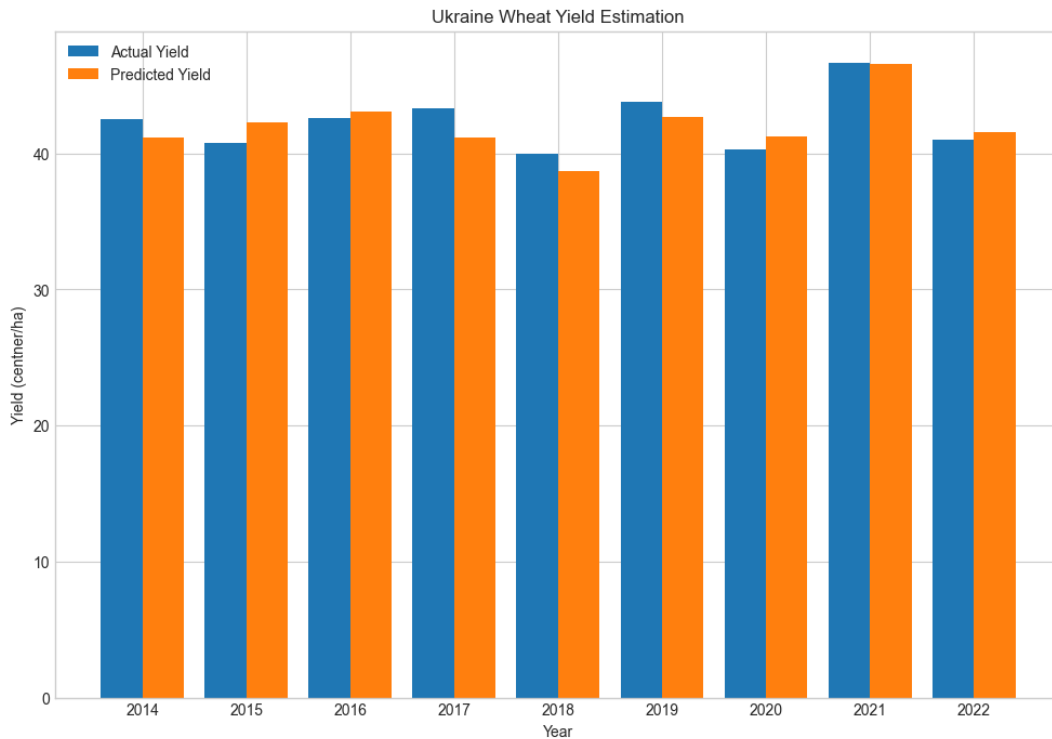


Рисунок 2.17 Порівняння реальних та прогнозованих значень моделі  
Модель демонструє гарну здатність відтворювати історичні дані, хоча в окремі роки спостерігаються помітні помилки.

На рис. 2.18 зображено 20 найбільш важливих ознак для даної моделі.

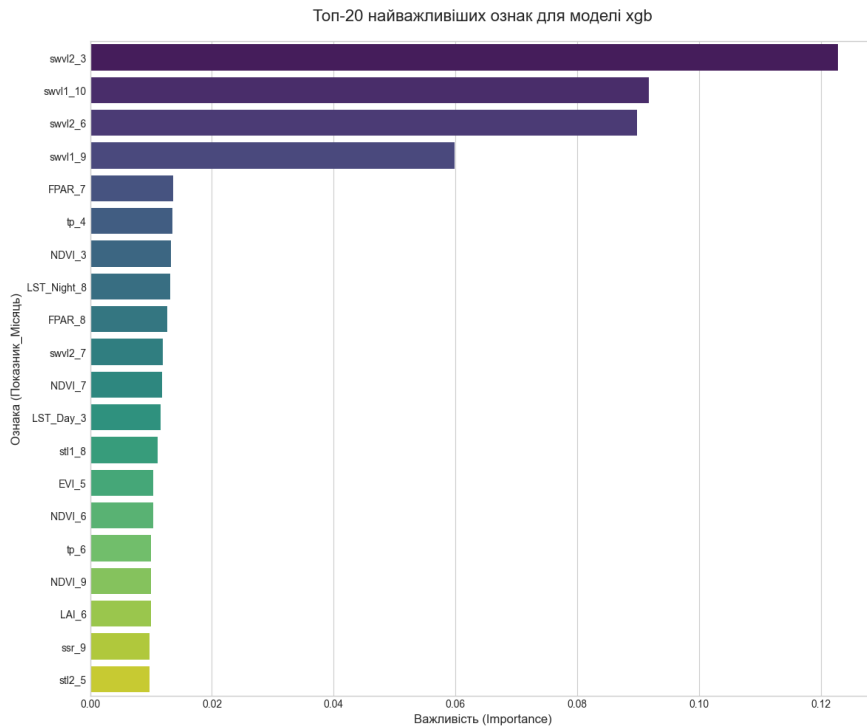


Рисунок 2.18 Важливість ознак для даної моделі

Як і в інших ансамблевих моделях, вологість ґрунту (swvl2\_3, swvl1\_10, swvl2\_6) визначена як ключовий предиктор. Це ще раз підтверджує фундаментальну важливість водного режиму ґрунту для прогнозування врожайності пшениці.

В таблиці 2.7 наведені метрики для оцінки прогнозування за різні роки та загалом.

Таблиця 2.7 Результати прогнозування моделі

Year	RRMSE	RMSE	MAE	MSE	MAPE
2014	8.474797	3.603403	2.944104	12.98451	6.582932
2015	13.09038	5.334018	4.240976	28.45175	10.86995
2016	7.530619	3.210913	2.647309	10.30996	6.115667
2017	11.22135	4.85724	3.949934	23.59278	9.173289
2018	13.55072	5.417062	4.535505	29.34456	11.42174
2019	8.76709	3.83915	3.047411	14.73908	7.072839

## Продовження таблиці 2.7

2020	14.02567	5.654347	4.662285	31.97164	13.44275
2021	7.17822	3.349836	2.749092	11.2214	5.88189
2022	11.45015	4.698924	3.955981	22.07989	10.16443
Total	10.70167	4.530092	3.636955	20.52173	8.969499
Sd	2.511244	0.896271	0.746739	7.921225	2.540238

Загальна помилка RMSE становить 4.53, а MAPE — 8.97%. Це конкурентоспроможний результат, що ставить XGBoost в один ряд з найкращими моделями в цьому дослідженні.

На рис. 2.19 зображено гістограму з похибками для кожного регіону в середньому.

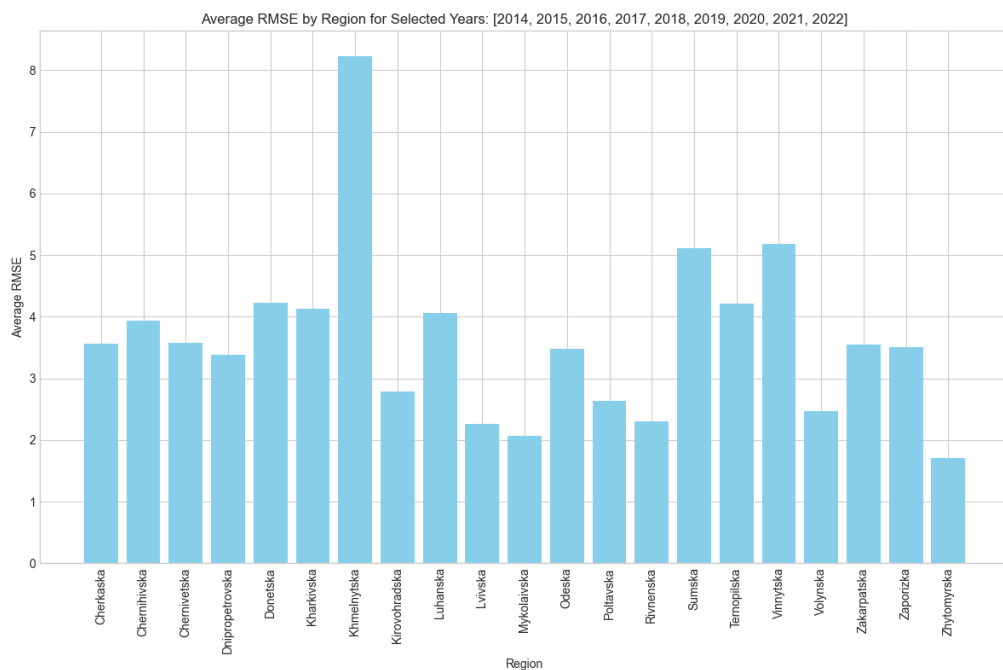


Рисунок 2.19 RMSE моделі по регіонах

Модель XGBoost демонструє дуже різну ефективність залежно від регіону. Вона є досить точною для таких областей, як Житомирська та Львівська, але як і попередні моделі має проблеми з прогнозуванням в Хмельницькій області.

### 2.4.7 Порівняння результатів

Графік розсіювання, що зображений на рис. 2.20, демонструє якість прогнозування моделі Random Forest для оцінки показника врожайності. На осі X відображено фактичні значення, а на осі Y — прогнозовані значення.

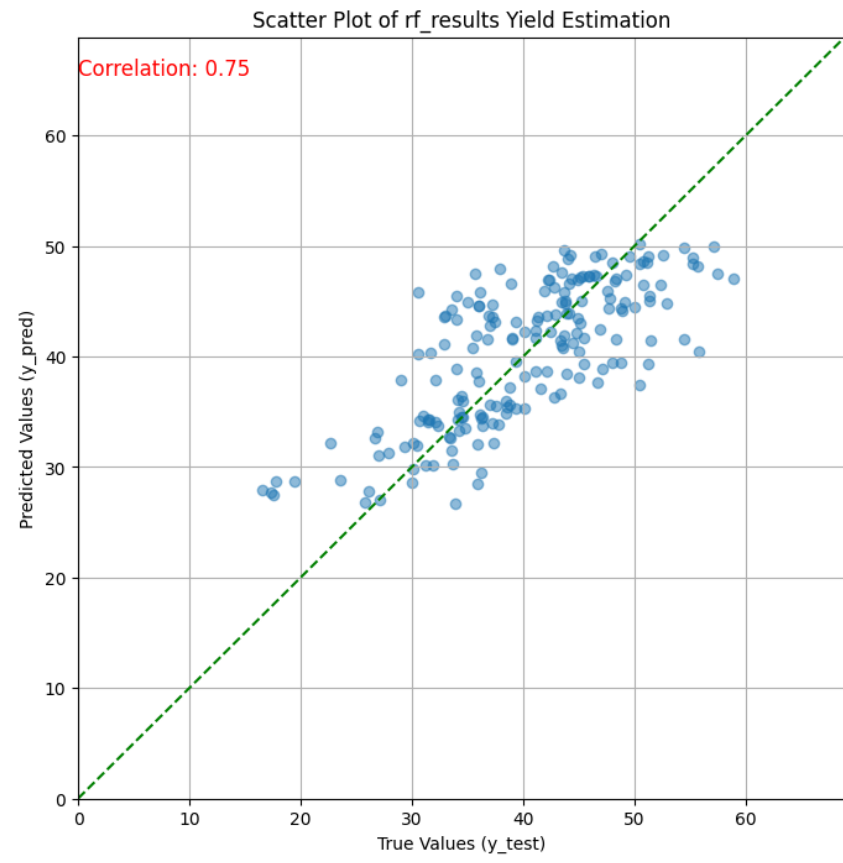


Рисунок 2.20 Графік розсіювання

В таблиці 2.8 наведені метрики для оцінки прогнозування різними моделями.

Таблиця 2.8 Результати прогнозування моделі

model	RRMSE	RMSE	MAE	MSE	MAPE
XGBoost	10.70167	4.530092	3.636955	20.52173	8.969499
KNN	11.19026	4.736914	3.794581	22.43836	9.763631
DT	13.44328	5.690634	4.567624	32.38332	10.94073
GB	10.97346	4.64514	3.690874	21.57733	9.124724
RF	10.81002	4.575956	3.630003	20.93938	8.908937
<b>SVM</b>	<b>10.28598</b>	<b>4.354125</b>	<b>3.389671</b>	<b>18.9584</b>	<b>8.122624</b>

Отримані результати моделювання показують, що всі розглянуті алгоритми машинного навчання продемонстрували доволі високу точність у прогнозуванні врожайності. Найкращу ефективність серед них показала модель SVM, яка має найнижчі значення помилок: RMSE = 4.35, MAE = 4.39, MSE = 18.96 та найменше відносне відхилення RRMSE = 10.29%. Це свідчить про те, що SVM найточніше відтворює реальні значення, забезпечуючи стабільність прогнозів і мінімальну варіацію від істинних даних. Модель Random Forest (RF) посіла друге місце, демонструючи схожу, але трохи гіршу якість прогнозів (RRMSE = 10.81%), що підтверджується і коефіцієнтом кореляції на графіку розсіювання ( $r = 0.75$ ).

Моделі Gradient Boosting та XGBoost також показали високу точність, що підтверджує їхню здатність ефективно враховувати нелінійні залежності між ознаками. У той же час KNN і особливо Decision Tree мають дещо більші помилки (RRMSE понад 11% і 13% відповідно), що свідчить про їхню меншу узагальнювальну здатність і потенційну схильність до перенавчання. Отже, для задачі оцінки врожайності за наявними даними найдоцільніше застосовувати моделі SVM або Random Forest, які забезпечують оптимальний баланс між точністю, стабільністю та інтерпретованістю результатів.

На рисунку 2.21 зображено графік зміни похибки залежно від моделі та року.

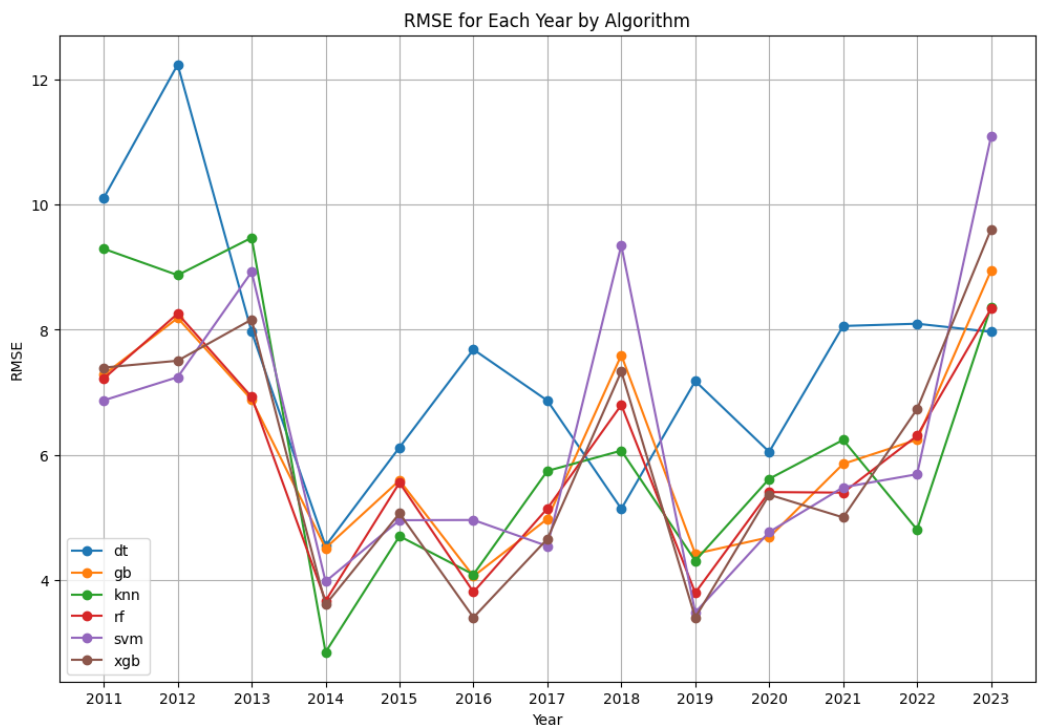


Рисунок 2.21 Порівняння точності прогнозів для різних моделей по роках

Найбільш стабільні результати демонструють Random Forest і SVM — їхні лінії коливаються в нижчому діапазоні RMSE, особливо після 2015 року, що підтверджує їхню здатність узагальнювати закономірності незалежно від року. XGBoost і Gradient Boosting також показують хорошу узгодженість результатів, особливо у роки з меншими коливаннями погодних умов (2014–2016). Натомість Decision Tree має найбільші піки RMSE, особливо на початку періоду (2012–2013) та в останні роки (2022–2023), що свідчить про його нестійкість до змін у даних.

Загалом графік ілюструє, що ансамблеві та SVM-моделі є більш надійними для прогнозування врожайності в динамічних умовах, тоді як прості алгоритми, як-от DT чи KNN, мають значно більші річні коливання похибки.

## 2.5 Розробка моделі оцінки ризиків

Цей розділ описує процес створення моделі для кількісної оцінки страхових ризиків в агросекторі, що базується на результатах прогнозування врожайності.

Основна проблема полягає в тому, що стандартна модель прогнозування надає точковий прогноз (наприклад, очікувана врожайність 6.5 т/га). Такий прогноз не містить інформації про невизначеність та ймовірність несприятливих результатів. Для страхової компанії або агровиробника недостатньо знати середній очікуваний результат; критично важливо розуміти, яка ймовірність того, що врожайність буде суттєво нижчою, і якими можуть бути потенційні збитки. Отже, задача — трансформувати детермінований прогноз врожайності у ймовірнісну оцінку ризику.

Головна мета — розробити кількісну модель, яка на основі прогнозу врожайності розраховує ключові метрики для ухвалення рішень у сфері агрострахування.

Ключові цілі моделі:

1. Оцінити ймовірність страхового випадку, тобто розрахувати шанс того, що фактична врожайність впаде нижче заздалегідь визначеного порогу (наприклад, 80% від прогнозу).
2. Визначити потенційний розмір збитків - обчислити очікуваний обсяг недобір урожаю (в т/га) та його фінансовий еквівалент, якщо страховий випадок настане.
3. Сформувати обґрунтовану страхову премію - надати розрахункову базу для визначення справедливої вартості страхового поліса, що відображає рівень ризику для конкретного поля чи регіону.

Для вирішення поставленої задачі було обрано метод симуляції Монте-Карло. Цей підхід є потужним інструментом для моделювання невизначеності та отримання розподілу можливих результатів.

### 2.5.1 Метод Монте-Карло

Насамперед ми аналізуємо історичні дані за 2014-2022 роки, щоб визначити, як сильно прогнози відхилялися від фактичної врожайності. Розраховуються відносні помилки для кожного спостереження.

На рис. 2.22 можемо побачити графік розподілу відносних помилок.

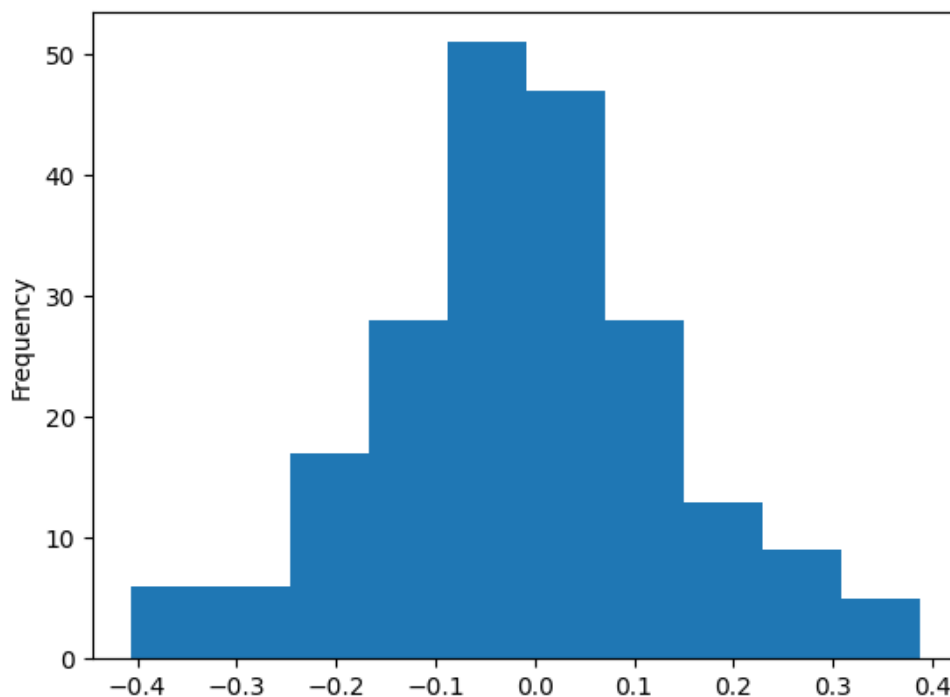


Рисунок 2.22 Розподіл помилок

Далі для визначення прогнозу на наступний рік проаналізуємо розподіл змінних, які впливають на прогноз. На рис. 2.23-2.27 можна побачити деякі з них.

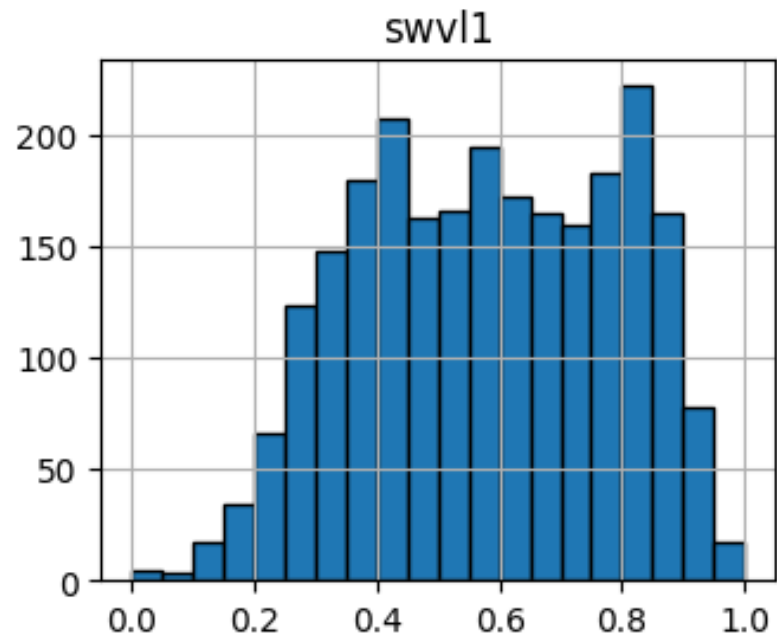


Рисунок 2.23

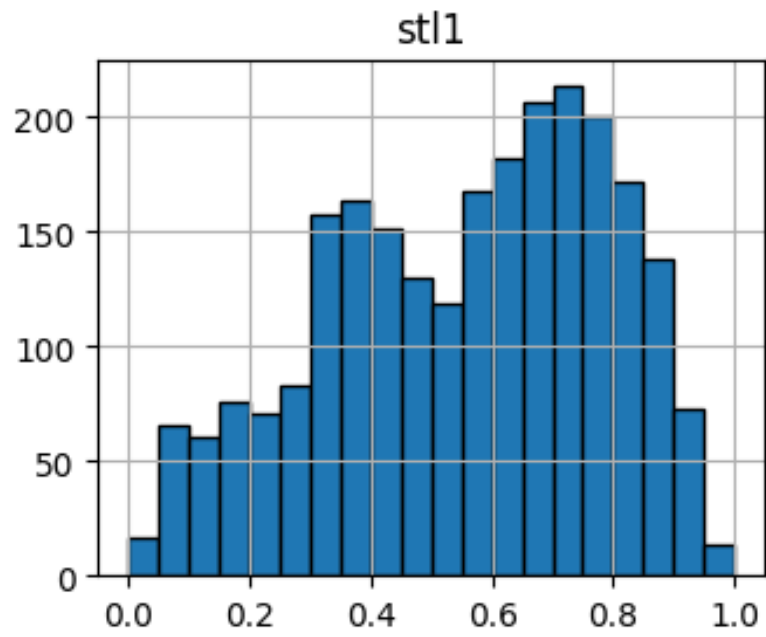


Рисунок 2.24

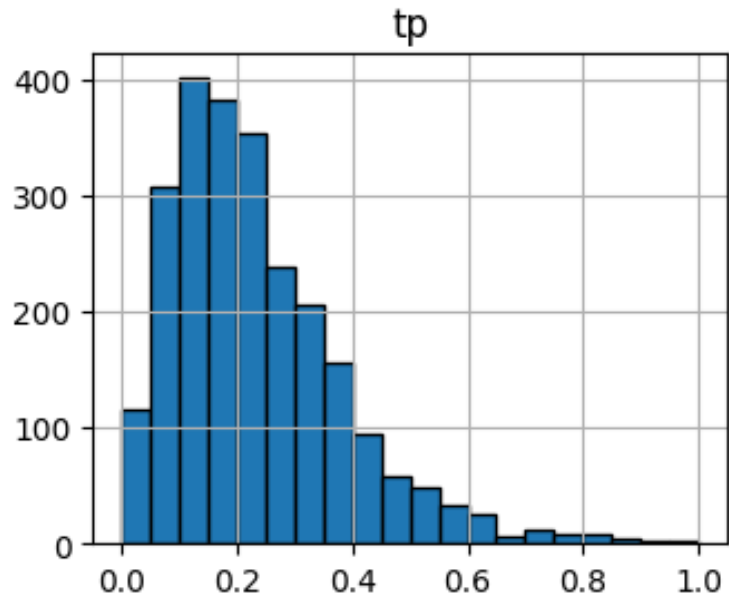


Рисунок 2.25

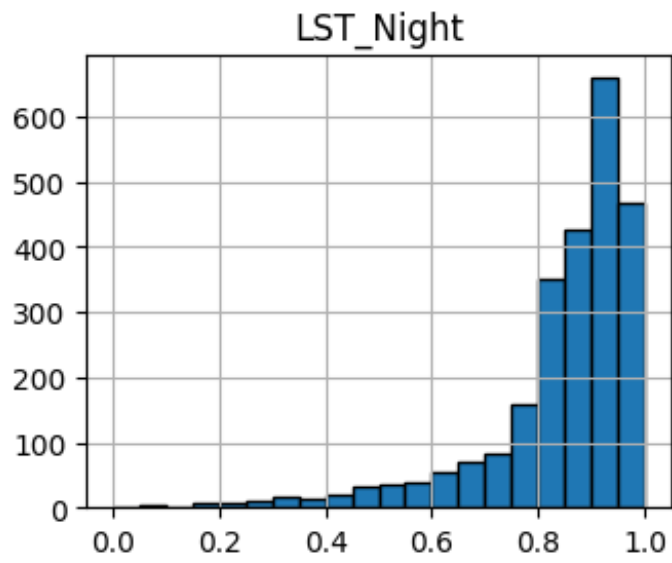


Рисунок 2.26

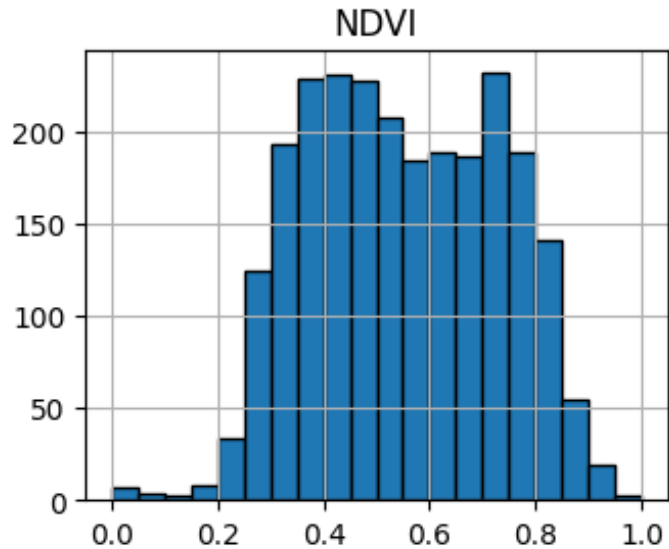


Рисунок 2.27 Розподіли параметрів моделі

Підрахувавши характеристики розподілу, маємо змогу провести симуляцію і визначити прогноз на наступний рік. На основі отриманого масиву помилок визначаються їхні статистичні характеристики — середнє значення та стандартне відхилення. Це дозволяє нам створити математичну модель невизначеності, властивої для основного прогнозу.

Модель запускає 10 тисяч ітерацій. У кожній ітерації до нового прогнозу врожайності додається випадково згенерована помилка, взята зі змодельованого раніше розподілу. У підсумку ми отримуємо не одне число, а цілий спектр з 10 000 можливих значень майбутньої врожайності. Цей спектр є ймовірнісним розподілом. На рис. 2.28 показано результати симуляцій для Житомирської області на 2023 рік.

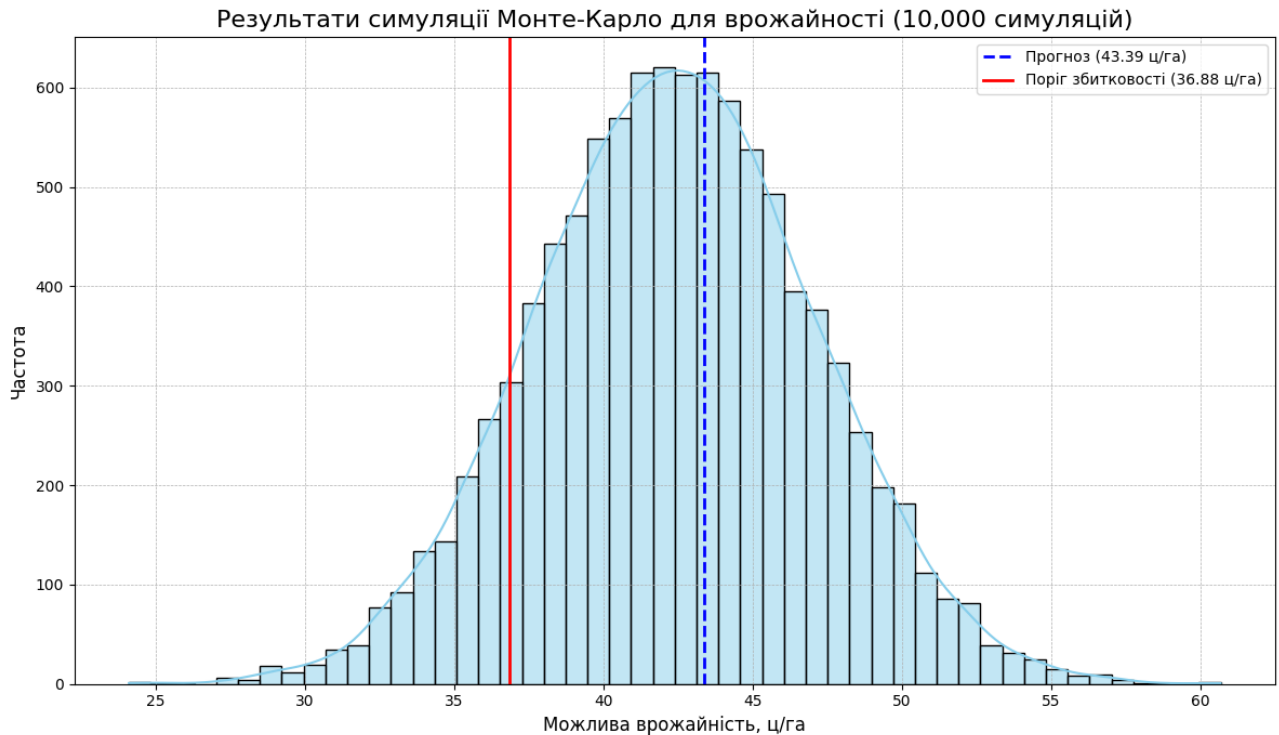


Рисунок 2.28 Результати симуляцій

Аналізуючи отриманий розподіл, ми визначаємо, який відсоток симульованих результатів опинився нижче страхового порогу (це і є ймовірність збитку) та розраховуємо середній недобір урожаю для цих збиткових сценаріїв.

### 2.5.2 Розрахунок страхових метрик

Першочергово у процесі оцінювання аграрних ризиків визначається поріг страхового випадку — критичне значення урожайності, за перевищенням або падінням нижче якого у страхувальника виникає право на отримання страхового відшкодування. Цей поріг є центральним параметром страхового

договору, оскільки саме він визначає, у яких сценаріях виробник вважається таким, що зазнав збитку. У науковій та практичній літературі поріг розглядається як інструмент балансування між рівнем захисту аграрія та фінансовою стійкістю страхової компанії. Надто низький поріг може призводити до частих виплат і підвищеного ризику для страховика; надто високий — до недостатнього захисту виробників і втрати привабливості продукту.

У більшості міжнародних страхових програм поріг встановлюється у вигляді певної частки від середньої багаторічної або гарантованої урожайності. Такий підхід дає змогу адаптувати поріг до природних умов конкретного регіону та відображати історичну мінливість кліматичних показників. Згідно з рекомендаціями FAO та практикою великих страхових компаній (Allianz Re, Swiss Re), типовий діапазон порогів становить 70–90 %, тоді як для класичних продуктів багаторічного страхування врожаю найпоширенішими є значення 80–85 %. Вибір саме цього інтервалу пояснюється прагненням мінімізувати як частоту дрібних збитків, так і надмірні страхові премії, що зростають зі збільшенням обсягу покриття. Таким чином поріг виступає компромісним, але добре обґрунтованим параметром, який забезпечує збалансованість страхового продукту.

У рамках цього дослідження поріг страхового випадку також формується у вигляді частки від прогнозованої або середньої врожайності культури. Для практичної реалізації моделі обрано поріг 85 %, що відповідає верхній межі стандартних страхових програм і репрезентує сценарій підвищеного рівня захисту. Такий поріг є виправданим за умов високої волатильності кліматичних факторів України, особливо у зонах ризикованого землеробства, де ймовірність істотних втрат урожайності зумовлена погодними аномаліями. Крім того, застосування порогу 85 % дозволяє чіткіше виявити хвостові ризики

та забезпечує більш точне моделювання екстремальних сценаріїв, які суттєво впливають на значення VaR, CVaR та очікуваних збитків.

У межах побудови моделі кількісного оцінювання ризику для аграрного сектору особливе значення мають актуарні метрики, що дають змогу формалізувати частотні та інтенсивні характеристики збитків і отримати комплексну картину ризику. Нижче наведені основні показники, які застосовуються у дослідженні, кожен із яких виконує окрему функцію в оцінюванні страхових подій, визначенні премій та аналізі фінансової стійкості страхового продукту.

Ймовірність дефіциту врожаю є базовим показником частоти страхового випадку та характеризує ризик недобору врожаю нижче визначеного критичного порога. Методологічно вона визначається як відношення кількості сценаріїв або спостережень, у яких урожайність є нижчою за страховий поріг, до загальної кількості реалізацій. Ця метрика виконує фундаментальну роль, оскільки задає частотний компонент ризику, визначаючи, наскільки часто аграрій може стикнутися зі збитками. Зростання ймовірності дефіциту прямо свідчить про підвищення нестабільності виробничого процесу та необхідність коригування страхових тарифів або параметрів покриття.

Очікуваний збиток відображає середній розмір втрат серед тих випадків, коли збиток уже настав, тобто фактична врожайність опустилася нижче страхового порога. Цей показник інтерпретується як інтенсивність збитків: він визначає, наскільки значними є втрати у несприятливих роках. Математично очікуваний збиток обчислюється як математичне сподівання відхилення врожайності від порога у випадках дефіциту, помножене на вартість одиниці врожаю. Саме ця метрика визначає фінансову сутність страхового ризику, оскільки вона описує потенційний масштаб збитків у грошовому вимірі. В умовах високої кліматичної мінливості значення очікуваного збитку може суттєво зростати, що підсилює потребу в актуарно точних моделях.

Чиста страхова премія є актуарно справедливою величиною страхового покриття та відображає очікуваний середній розмір виплати страховика у довгостроковому періоді. Вона визначається як добуток ймовірності дефіциту врожаю на очікуваний розмір збитку, тобто поєднує частотну та інтенсивну характеристики ризику в єдину фінансову метрику. Чиста премія є фундаментом для утворення тарифу, на основі якого формуються повні страхові тарифи шляхом додавання навантажень на операційні витрати, ризик перестраховування та прибутковість страхової компанії. Висока точність оцінки чистої премії є ключовою умовою фінансової стійкості страхового портфеля.

Value-at-Risk (VaR) є показником хвостового ризику, який використовується для визначення максимального очікуваного збитку з довірчою ймовірністю  $1-\alpha$ . У контексті аграрного страхування VaR показує той рівень втрат, який не буде перевищено у  $100(1-\alpha)\%$  випадків. На практиці це означає, що VaR дає можливість оцінити граничний збиток у несприятливих, але не катастрофічних роках. VaR широко застосовується у фінансовій індустрії та перестрахованні, проте має обмеження: він не враховує середній розмір збитків у найбільш екстремальних сценаріях, що є критично важливим у аграрній сфері, де розподіли збитків характеризуються асиметрією та важкими хвостами.

Conditional Value-at-Risk (CVaR), або Expected Shortfall, доповнює VaR і є більш чутливим до екстремальних ризиків. Цей показник визначає середній збиток серед тих сценаріїв, у яких втрати перевищили порогове значення VaR. Оскільки CVaR відображає поведінку розподілу в найгірших 1–5 % випадків, він вважається більш надійною метрикою у середовищах із високою волатильністю та важкими хвостами, до яких належить аграрний сектор. CVaR широко рекомендований міжнародними фінансовими регуляторами та Basel III/IV, Solvency II, оскільки дозволяє адекватно враховувати катастрофічні ризики. В агрострахованні ця метрика є ключовою для оцінювання

потенційних потреб у перестрахованні та визначення стійкості страхового портфеля.

RAROC (Risk-Adjusted Return on Capital) є метрикою, що використовується для оцінювання ефективності використання капіталу з урахуванням ризику. У страхуванні сільськогосподарських культур RAROC дозволяє визначити, чи є страховий продукт економічно виправданим, враховуючи очікувані виплати та величину капіталу, необхідного для покриття екстремальних збитків (визначених переважно через CVaR). Якщо значення RAROC є низьким або від'ємним, продукт вважається нерентабельним і потребує перегляду тарифу, параметрів покриття чи структури перестраховання. Ця метрика поєднує фінансову результативність та ризиковість портфеля, що робить її ключовою для управління капіталом страховика.

В таблиці 2.9 показано результати розрахунків страхових метрик на основі прогнозованих значень врожайності для Житомирської області. Ціна на пшеницю враховувалась на момент написання і склала 1035 гривень за центнер.

Таблиця 2.9 Результати прогнозування моделі

Показник	Значення
Прогноз на наступний рік	43.39 ц/га
Поріг страхового випадку	36.924 ц/га
Ймовірність страхового випадку	8.71%
Очікуваний дефіцит врожаю при настанні збитку	0.192 ц/га
Очікуваний збиток	384.0 грн/га
Чиста страхова премія	460.80 грн/га
VaR	2759.41 грн/га
CVaR	3584.06 грн/га
RAROC	0.028

Ймовірність страхового випадку 8.71 % вказує на те, що дефіцит урожаю нижче порога є подією відносно рідкісною: у середньому менше ніж у кожному десятому сезоні. Це узгоджується з припущенням, що критичний поріг обрано досить консервативно й відповідає умовам «поганого року», а не звичайних флуктуацій. Для страховика така частота подій є прийнятною з точки зору формування стабільного портфеля: з одного боку, виплати не будуть надто частими, з іншого — продукт не виглядатиме «мертвим», тобто таким, за яким ніколи не настають страхові випадки. Для агровиробника це означає, що страховий захист спрацьовує саме в тих сценаріях, де відхилення є дійсно суттєвим, а не внаслідок незначних сезонних коливань.

Очікуваний дефіцит врожаю при настанні збитку (0.192 ц/га) є дуже малим порівняно з прогнозною врожайністю (43.39 ц/га). Це означає, що в середньому серед тих років, коли поріг таки порушується, недобір врожаю відносно цього порога є помірним. Тобто характер ризику радше «низькочастотний — середньої глибини», а не сценарій катастрофічних втрат у кожному випадку. У грошовій формі це відображається в очікуваному збитку 384.0 грн/га, який є базою для формування актуарно справедливої частини тарифу. З точки зору порівняння з іншими культурами/регіонами така величина може бути інтерпретована як «середній річний ризиковий податок» на гектар за рахунок кліматичної та виробничої невизначеності.

Чиста страхова премія 460.80 грн/га є дещо більшою за очікуваний збиток, що відображає не лише покриття математично очікуваних втрат, а й закладений запас на невизначеність моделі та можливу недооцінку ризику. На практиці це означає, що страховий продукт побудований із невеликим «запасом міцності», що підвищує ймовірність його довгострокової стійкості. Для агровиробника така премія може бути порівняна з очікуваними доходами з гектара: якщо премія становить, наприклад, кілька відсотків від очікуваного

валового доходу, продукт є економічно привабливим; якщо ж частка занадто велика, доцільно адаптувати рівень покриття або франшизу.

Особливу увагу в аналізі слід приділити показникам VaR і CVaR, які характеризують хвостові ризики. Значення  $VaR = 2759.41$  грн/га означає, що з довірчою ймовірністю, наприклад, 95 % збиток не перевищить цю величину. Проте  $CVaR = 3584.06$  грн/га демонструє, що в найгірших 5 % сценаріїв середній збиток є ще вищим. Співвідношення CVaR до очікуваного збитку (близько 9 разів) свідчить про наявність суттєвих хвостових ризиків: більшість років збитки відносно невеликі, проте у рідкісні екстремальні сезони втрати можуть бути кратно більшими за середні. Для страховика це чіткий сигнал про необхідність врахування катастрофічних сценаріїв у розрахунку капіталу та, можливо, використання перестраховування. Для агровиробника це означає, що страхування справді захищає від рідкісних, але дуже болючих втрат.

$RAROC = 0.028$  (2.8 %) характеризує ризик-скореговану дохідність страхового продукту. Це відносно низьке значення, що вказує на те, що з точки зору страховика продукт перебуває на межі привабливості: він не є збитковим, але й не забезпечує високої дохідності на одиницю ризикового капіталу. У практиці страхового бізнесу зазвичай порівнюють RAROC із цільовим нормативом (наприклад, 8–12 %). Якщо фактичний RAROC нижчий, страховий продукт потребує перегляду: підвищення премії, зміни структури покриття або часткового перенесення ризику на перестраховика. З іншого боку, для агровиробника низький RAROC часто означає більш вигідні умови страхування (менші премії при відносно високому рівні захисту).

З точки зору практичного використання, ці результати можна застосовувати на кількох рівнях.

Для страхової компанії:

- 1) обґрунтування розміру чистої премії та прийнятної надбавки;
- 2) оцінка потреби в перестрахованні на основі рівня CVaR;

- 3) аналіз доцільності продукту за критерієм RAROC і порівняння його з альтернативними продуктами або культурами;
- 4) формування політики лімітів покриття (максимальна сума відшкодування на гектар) і вимог до капіталу для покриття екстремальних збитків.

Для агровиробника:

- 1) оцінка того, яку частку ризику він передає страховикові, а яку залишає на собі;
- 2) порівняння страхової премії з очікуваним доходом та прийняття рішення про доцільність страхування;
- 3) аналіз ймовірності та масштабу потенційних втрат, що дозволяє планувати фінансові резерви й стратегії диверсифікації (зміна структури посівів, регіональна диверсифікація тощо).

Для банків та інвесторів такі метрики можуть використовуватися як вхідні параметри в моделях кредитного скорингу агровиробників, для оцінки ймовірності дефолту внаслідок кліматичних шоків та визначення умов фінансування (ставки, вимоги до застави, потреба в обов'язковому страхуванні).

Нарешті, для інтелектуальної системи підтримки прийняття рішень, що розробляється в межах дисертації, ці показники є базою для модулю рекомендацій: система може автоматично формувати пропозиції щодо рівня покриття, порога страхового випадку, орієнтовного тарифу, а також попереджати користувача про підвищений рівень хвостового ризику за певних кліматичних сценаріїв. Таким чином, результати з табл. 2.9 виступають не лише підсумком моделювання, але й практичним інструментом для побудови прозорої, обґрунтованої та адаптивної системи аграрного страхування.

## 2.6 Висновки до розділу 2

У другому розділі було представлено повний цикл практичної реалізації інтелектуальної системи оцінювання аграрних ризиків, що включає збір, попередню обробку даних, моделювання врожайності, оцінку ризиків та формування рекомендацій. Реалізована інформаційна технологія підтвердила можливість інтеграції різномірних супутникових, кліматичних та статистичних даних у єдине аналітичне середовище, здатне забезпечити точне прогнозування та кількісну оцінку ризиків для аграрного сектору.

У ході роботи було обґрунтовано вибір мови програмування Python як оптимального інструменту для реалізації системи завдяки її універсальності, розвиненій екосистемі бібліотек для машинного навчання, обробки геоданих та візуалізації, а також високій швидкості розробки. Це забезпечило можливість ефективно реалізувати всі елементи програмного конвеєра — від автоматизованої обробки наборів даних NetCDF і супутникових знімків MODIS до побудови прогнозних моделей і розрахунку страхових метрик.

Особливе значення мав етап попередньої обробки даних, в якому було інтегровано інформацію з різними просторовими роздільними здатностями та часовими характеристиками. Впроваджені алгоритми фільтрації, агрегації та зважування показників на основі щільності посівів та площі перекриття адміністративних меж дозволили отримати високоякісний і репрезентативний набір ознак, придатний для подальшого моделювання. Саме цей етап забезпечив узгодженість і точність вхідних даних, що є критично важливим для надійності подальших оцінок.

У рамках моделювання врожайності були протестовані шість алгоритмів машинного навчання (Decision Tree, Random Forest, Gradient Boosting, XGBoost, SVM, KNN), для яких проведено оптимізацію гіперпараметрів і

валідацію методом LOYO. Це дозволило оцінити здатність моделей узагальнювати закономірності у складних просторово-часових даних. Результати моделювання показали, що ансамблеві методи, зокрема Gradient Boosting і XGBoost, забезпечують найвищу точність прогнозів, тоді як прості моделі (наприклад, Decision Tree) демонструють нижчу стабільність та точність.

Подальший етап — розробка моделі оцінювання ризиків на основі прогнозової врожайності та симуляцій методом Монте-Карло — дозволив отримати повний набір страхових показників: ймовірність дефіциту, очікувані втрати, абсолютний та відносний VaR, Expected Shortfall, чисту страхову премію та виробничі фінансові характеристики. Це підтвердило можливість застосування імітаційного моделювання для оцінки невизначеності в аграрному виробництві та формування об'єктивних фінансових рішень.

Окрему роль у структурі системи відіграє модуль формування рекомендацій, який трансформує результати моделювання у практичні рішення — наприклад, оптимальний рівень страхового покриття, розмір премії, очікувану дохідність або ризики для конкретного господарства чи регіону. Це робить систему придатною до практичного використання страховими компаніями, агровиробниками та аналітичними структурами.

Таким чином, результати другого розділу демонструють, що розроблена інтелектуальна система є повністю функціональною, технічно обґрунтованою та здатною забезпечити комплексне вирішення задач прогнозування врожайності та оцінювання аграрних ризиків. Побудована технологічна архітектура підтверджує доцільність використання машинного навчання та методів імітаційного моделювання для створення сучасних інструментів підтримки прийняття рішень у сфері агрострахування. У наступному розділі буде інтегровано всі компоненти системи у єдину ІСППР та проведено її оцінку.

### 3 ІНТЕЛЕКТУАЛЬНА СИСТЕМА ПІДТРИМКИ ПРИЙНЯТТЯ РІШЕНЬ

У цьому розділі розробляється інтелектуальна система підтримки прийняття рішень (ІСППР), яка поєднує методи машинного навчання, економіко-математичного моделювання та страхового аналізу. Система орієнтована на прогнозування урожайності, оцінювання імовірності настання страхових випадків, розрахунок страхових характеристик та формування рекомендацій щодо тарифної політики для страхових компаній і фермерських господарств.

Основна мета створення ІСППР полягає у забезпеченні обґрунтованого прийняття рішень щодо управління ризиками в аграрному секторі на основі аналітичних моделей, які інтегрують історичні, прогнозні та поточні дані. Інтелектуальна система реалізує концепцію цифрового страхового аналітика, здатного автоматично обробляти великі обсяги інформації, виявляти закономірності та генерувати оцінки ризиків із заданим рівнем довіри.

У межах цього розділу визначено архітектуру системи, описано її основні функціональні модулі — підсистему збору та обробки даних, модуль прогнозування, аналітичний блок оцінки страхових ризиків, модуль розрахунку страхових характеристик і інтерфейс прийняття рішень. Також наведено алгоритми моделювання, реалізацію програмного забезпечення та приклади розрахунків, що підтверджують ефективність розробленої ІСППР.

### 3.1 Архітектура ІСППР

Архітектура інтелектуальної системи підтримки прийняття рішень (ІСППР) для оцінювання страхових ризиків в аграрному секторі розроблена на основі принципів модульності, ієрархічності та інтегрованості. Це дає змогу гнучко масштабувати систему, адаптувати її до різних типів вхідних даних і забезпечувати послідовність аналітичних процесів — від збору інформації до формування управлінських рішень.

Загальна структура системи включає п'ять основних рівнів, зображених на рис. 3.1.

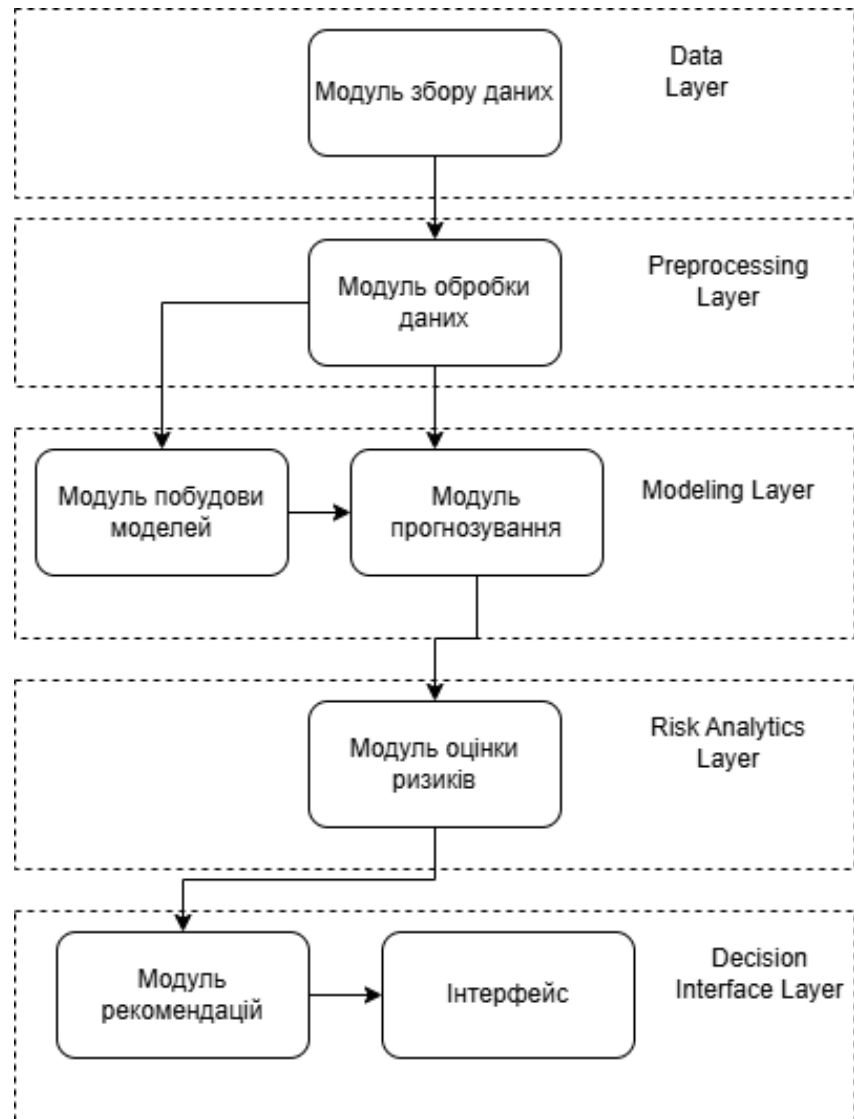


Рисунок 3.1 Архітектура ІСППР

Рівень даних відповідає за збирання, накопичення та узгодження інформації з різних джерел — статистичних, кліматичних, ринкових і страхових. На цьому етапі формується база даних, яка містить історичні ряди урожайності, метеорологічні показники, динаміку цін і параметри страхових договорів. Дані зберігаються у табличному або просторовому форматі, що забезпечує їх подальшу обробку та аналіз. Важливим аспектом цього рівня є забезпечення якості даних, оскільки точність подальших прогнозів безпосередньо залежить від достовірності вхідної інформації.

Рівень обробки даних виконує підготовку зібраної інформації до моделювання. Тут проводиться очищення від шуму, усунення пропусків, нормалізація та узгодження часових рядів. Додатково виконуються процедури стандартизації показників, виявлення кореляцій між факторами та формування набору ознак для моделі машинного навчання. На цьому етапі необроблені статистичні дані перетворюються у структурований навчальний масив, придатний для побудови прогнозних моделей.

Рівень моделювання реалізує процес прогнозування урожайності та оцінювання ризиків із використанням алгоритмів машинного навчання. У межах цього рівня формуються моделі, здатні враховувати нелінійні залежності між погодними, економічними та агротехнічними факторами. Результатом роботи моделей є прогноз урожайності на наступний період, розподіл можливих сценаріїв розвитку подій і розрахунок ймовірності настання страхового випадку. Саме тут відбувається перехід від статистичного аналізу до інтелектуального прогнозування.

Аналітичний рівень спрямований на оцінювання ризику та визначення ключових страхових характеристик, що лежать в основі прийняття рішень. Система розраховує такі показники, як імовірність дефіциту врожаю, очікуваний збиток, чисту та бруто-премії, Value-at-Risk і RAROC. Цей рівень інтегрує результати прогнозування з економічними розрахунками, забезпечуючи можливість кількісного вимірювання ризику та фінансової стійкості страхового портфеля.

Рівень інтерфейсу користувача є завершальною ланкою системи та забезпечує взаємодію користувача з аналітичним ядром ІСППР. Через веб-інтерфейс або інтерактивну панель користувач отримує результати прогнозів, графіки розподілу ризиків, значення страхових показників і рекомендації щодо тарифів або страхових умов. Інтерфейс підтримує зміну параметрів моделі в

реальному часі, що дозволяє оперативно аналізувати різні сценарії та ухвалювати обґрунтовані рішення.

Таким чином, архітектура ІСППР є цілісною системою, у якій кожен рівень виконує певну роль у процесі перетворення даних на знання, а знань — на практичні управлінські рішення у сфері аграрного страхування. Її модульна побудова дозволяє розширювати функціонал і масштабувати систему для нових культур, регіонів або типів страхових ризиків без зміни основної структури.

### 3.2 Блок моделювання

Блок моделювання є центральною складовою інтелектуальної системи підтримки прийняття рішень, оскільки саме на цьому етапі здійснюється побудова прогнозних моделей урожайності та оцінювання страхових ризиків на основі методів машинного навчання. Основна мета блоку полягає у створенні математичних залежностей між факторами, що впливають на урожайність, і кінцевими результатами виробництва з метою прогнозування можливих втрат та визначення ймовірності страхового випадку.

Вхідними даними для моделювання є підготовлені та нормалізовані ознаки, що характеризують агрокліматичні, економічні та технологічні умови: середньорічна температура, сума опадів, площа посівів, ринкові ціни, витрати на виробництво тощо. Перед моделюванням проводиться попередній аналіз факторів, виявлення мультиколінеарності, вибір оптимального набору змінних і розподіл даних на тренувальну та тестову вибірки. Такий підхід дозволяє підвищити узагальнювальну здатність моделей і зменшити ризик перенавчання.

Для реалізації прогностичної частини системи застосовано комбінацію сучасних алгоритмів машинного навчання, що продемонстрували високу ефективність при роботі з економічними та природними часовими рядами. До таких алгоритмів належать Random Forest, Gradient Boosting, XGBoost, Support Vector Machines (SVM), а також гібридна нейро-нечітка модель ANFIS (Adaptive Neuro-Fuzzy Inference System), яка дозволяє враховувати нечіткість у вхідних даних і забезпечує адаптивне налаштування параметрів на основі навчання. Застосування ансамблевих та нечітких підходів дозволяє системі підвищувати точність прогнозів у ситуаціях із високим рівнем стохастичності або неповнотою даних.

Процес моделювання включає кілька послідовних етапів: навчання моделей на історичних даних, валідацію результатів, вибір найкращої моделі за критеріями точності (RMSE, MAE, MAPE) та формування прогнозу урожайності на наступний період. Після отримання прогностичних значень система генерує розподіл можливих сценаріїв урожайності, на основі якого оцінюється ймовірність дефіциту врожаю, тобто настання страхового випадку. Додатково проводиться моделювання ризикових сценаріїв за допомогою методів Монте-Карло, що дає можливість побудувати довірчі інтервали для майбутніх значень урожайності та визначити межі потенційних втрат.

Результати, отримані в блоці моделювання, передаються до аналітичного блоку системи, де здійснюється розрахунок страхових характеристик — очікуваного збитку, чистої премії, Value-at-Risk та RAROC. Таким чином, блок моделювання виконує ключову роль у трансформації статистичних даних у прогностичні показники, які є основою для подальшого прийняття рішень. Його ефективність визначає точність оцінки ризику, надійність системи в цілому та якість рекомендацій, що формуються для страхових компаній і агровиробників.

### 3.3 Модуль оцінювання

Модуль оцінювання є важливим компонентом інтелектуальної системи підтримки прийняття рішень, оскільки забезпечує кількісну оцінку точності та надійності побудованих моделей прогнозування. Саме на цьому етапі визначається, наскільки адекватно модель описує реальні процеси зміни урожайності та чи може вона бути використана для подальших розрахунків страхових характеристик. Модуль оцінювання виконує функції контролю якості прогнозу, порівняння альтернативних моделей і вибору найоптимальнішої з точки зору точності та стабільності результатів.

Для оцінювання ефективності прогнозних моделей у системі застосовуються класичні метрики регресійного аналізу, які характеризують ступінь відхилення прогнозованих значень від фактичних. Основними показниками є:

- 1) середньоквадратична похибка (MSE);
- 2) корінь середньоквадратичної похибки (RMSE);
- 3) середня абсолютна похибка (MAE);
- 4) середня абсолютна відносна похибка (MAPE) та відносна середньоквадратична похибка (RRMSE).

Середньоквадратична похибка (MSE) визначається за формулою:

$$MSE = \frac{1}{N} \sum (y_i - \hat{y}_i)^2 ,$$

де  $y_i$  — фактичні значення урожайності,  $\hat{y}_i$  — прогнозовані значення,  $N$  — кількість спостережень. Корінь із цього показника,  $RMSE = \sqrt{MSE}$ , дає змогу оцінити середню величину відхилення прогнозу у тих же одиницях, що й цільова змінна. Середня абсолютна похибка (MAE) визначається як середнє від абсолютних відхилень між прогнозом і фактом:

$$MAE = \frac{1}{N} \sum |y^i - \hat{y}_i| ,$$

а середня абсолютна відносна похибка (MAPE) вимірює точність у відсотках:

$$MAPE = \frac{100\%}{N} \sum \left| \frac{y^i - \hat{y}_i}{y^i} \right| .$$

Значення MAPE нижче 10 % зазвичай свідчить про високу точність моделі, у межах 10–20 % — про прийнятну, а вище 20 % — про недостатню. Додатково використовується показник RRMSE (Relative RMSE), який нормалізує RMSE відносно середнього значення урожайності, що дозволяє порівнювати моделі між собою для різних культур або регіонів.

Після розрахунку всіх метрик система здійснює порівняння моделей і формує ранжування за точністю прогнозу. Найкраща модель вибирається як базова для подальших страхових розрахунків. Якщо різниця між моделями є статистично незначною, перевага надається тій, що демонструє вищу стабільність та меншу варіацію помилок. Додатково модуль може будувати графічні порівняння — діаграми залишків, графіки “факт–прогноз” та гістограми розподілу похибок — для візуального аналізу результатів.

Таким чином, модуль оцінювання виконує роль контролю якості у всій системі, забезпечуючи перевірку адекватності моделей машинного навчання та достовірність отриманих прогнозів. Його робота гарантує, що до розрахунку страхових характеристик надходять лише перевірені, статистично обґрунтовані результати, що підвищує надійність і точність інтелектуальної системи підтримки прийняття рішень у цілому.

### 3.4 Модуль оцінки та розрахунку ризику

Модуль оцінки та розрахунку ризику є ключовим аналітичним компонентом інтелектуальної системи підтримки прийняття рішень, який забезпечує перехід від прогнозних результатів до кількісного вимірювання фінансових ризиків. Його основна функція полягає у перетворенні даних про прогнозовану урожайність на страхові показники, що характеризують рівень можливих збитків, очікувану величину виплат та економічну ефективність страхового продукту. Цей модуль поєднує підходи актуарної математики, теорії ризику та методів машинного навчання, що дозволяє комплексно оцінити ризикову ситуацію для конкретної культури або регіону.

Вихідною інформацією для роботи модуля є результати прогнозування урожайності, отримані в блоці моделювання, а також статистичні дані попередніх років. На основі цих даних визначається поріг страхового випадку, що задає межу між нормальною та критично низькою урожайністю. Як правило, поріг встановлюється у межах 70–80 % від середнього значення урожайності за кілька останніх періодів, що відповідає стандартній практиці аграрного страхування. Далі система оцінює ймовірність настання страхового випадку — тобто частку сценаріїв, у яких прогнозована урожайність є нижчою за встановлений поріг.

На основі розподілу можливих значень урожайності обчислюється очікуваний збиток (Expected Loss), який характеризує середній недобір продукції в натуральному або грошовому еквіваленті. Очікуваний збиток визначається як математичне сподівання різниці між порогом урожайності та фактичним прогнозом, якщо останній є нижчим за критичне значення. У грошовому вимірі збиток визначається множенням дефіциту врожаю на

середню ринкову ціну сільськогосподарської продукції. Цей показник є базовим для подальшого розрахунку страхових премій.

Важливим елементом модуля є розрахунок показника Value-at-Risk (VaR), який визначає максимальний можливий збиток при заданому рівні довіри (наприклад, 95 % або 99 %). Показник VaR дає змогу оцінити граничні фінансові втрати, які не будуть перевищені у більшості сценаріїв, і широко використовується у фінансовій та страховій аналітиці для управління ризиками. У контексті аграрного страхування VaR дозволяє визначити розмір страхових резервів і мінімальний рівень капіталу, необхідний для покриття екстремальних збитків.

На основі отриманих розрахунків формується чиста страхова премія (Pure Premium) — математично справедлива величина, що дорівнює очікуваному збитку без додаткових навантажень. До неї додається коефіцієнт навантаження ( $\lambda$ ), який враховує адміністративні витрати, перестраховування та прибуток страховика, у результаті чого визначається бруто-премія (Gross Premium). Такий підхід забезпечує баланс між фінансовою стійкістю страхової компанії та доступністю страхового продукту для агровиробника.

Для оцінювання ефективності страхового продукту в модулі реалізовано розрахунок показника RAROC — співвідношення очікуваного прибутку до величини ризикового капіталу, вираженого через VaR. Цей індикатор дає змогу визначити, чи є страховий продукт економічно доцільним при врахуванні ризикової складової, і використовується як критерій прийняття рішень щодо оптимізації страхових тарифів.

Результатом роботи модуля оцінки та розрахунку ризику є набір інтегральних показників, що формують ризиковий профіль культури або регіону. Вони відображають як частоту можливих втрат, так і потенційний розмір збитків. Отримані показники можуть бути візуалізовані у вигляді таблиць, графіків або інтерактивних дашбордів, що дозволяє користувачу

системи — аналітику або актуарію — оперативно оцінювати стан страхового портфеля та ухвалювати обґрунтовані рішення щодо рівня покриття, розміру премій і стратегії управління ризиком.

### 3.5 Модуль надання рекомендацій

Модуль надання рекомендацій є заключним елементом інтелектуальної системи підтримки прийняття рішень і забезпечує перехід від результатів аналітичних розрахунків до практичних управлінських рішень. Його основне призначення полягає у трансформації прогнозних та ризикових показників у конкретні поради для користувачів системи — страхових компаній, аграрних підприємств або аналітичних установ. Модуль виконує функцію автоматизованого аналітичного висновку, формуючи оптимальні параметри страхового покриття, тарифів і стратегій управління ризиком залежно від поточної ситуації.

Методика роботи модуля базується на послідовному застосуванні процедур аналітичного висновку та правилах прийняття рішень, що використовують результати прогнозування та оцінювання ризиків. На першому етапі система отримує вихідні показники з попередніх модулів — прогнозовану урожайність, імовірність дефіциту врожаю, очікуваний збиток, чисту та бруто страхові премії, показники Value-at-Risk і RAROC. Після цього здійснюється класифікація рівня ризику за узгодженими критеріями: низький, помірний або високий. Віднесення конкретного випадку до певного класу ризику відбувається за допомогою нечітких правил, у яких враховується не лише ймовірність дефіциту врожаю, а й показники прибутковості та фінансової стійкості страхового продукту.

На першій сторінці реалізовано ввід даних користувача, розрахунок основних страхових метрик, а також виведення результатів моделювання. В результаті отримуємо таблицю та графік, як показано на рис. 3.2-3.3.

Основні результати моделювання ризику

Показник	Значення
0 Середня врожайність (ц/га)	55.322
1 Поріг збитковості (ц/га)	47.023
2 Ймовірність дефіциту	0.075
3 Очікуваний збиток (ц/га)	0.222
4 Очікуваний збиток (грн)	263.450
5 Value at Risk (грн)	1,269,240
6 Часта страхових премій (грн)	316,140
7 Коефіцієнт RAROC	0.042

Рисунок 3.2 Таблиця страхових метрик

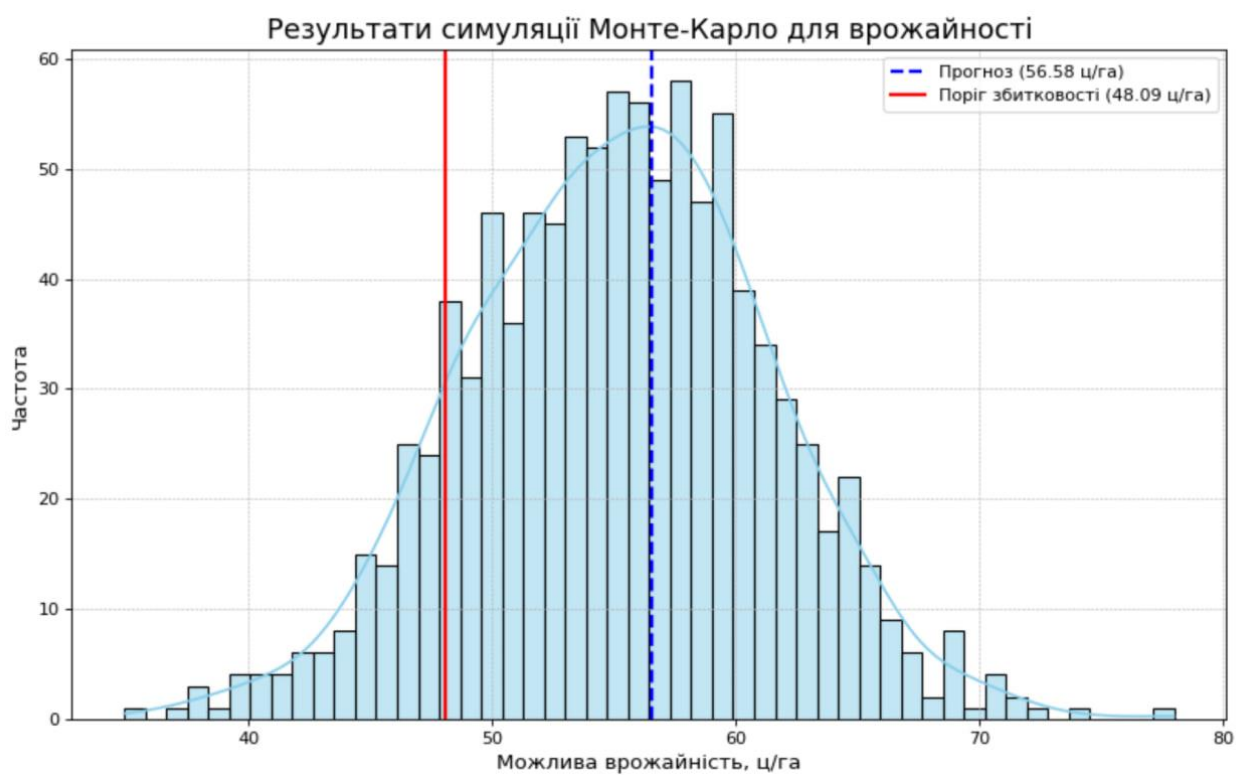


Рисунок 3.3 Результати симуляції Монте-Карло

Далі модуль формує набір рекомендаційних правил на основі аналітичних залежностей між розрахованими показниками. Якщо, наприклад, ймовірність страхового випадку перевищує 30 %, система пропонує зменшити поріг страхового покриття або застосувати перестраховання. Якщо показник RAROC є нижчим за допустимий рівень (0.05), рекомендується підвищити тариф або знизити коефіцієнт покриття. У випадках, коли VaR перевищує 30

% від середньої вартості врожаю, система радить вжити заходів із диверсифікації ризику. Якщо ж очікуваний збиток незначний і RAROC перевищує 0.1, страховий продукт визнається прибутковим, і система рекомендує розширення його застосування або збільшення частки ринку.

Для вибору оптимальних параметрів страхування застосовується підхід багатокритеріальної оптимізації, у межах якого здійснюється одночасна максимізація показника RAROC і мінімізація ймовірності дефіциту врожаю. Система аналізує, як зміна коефіцієнта покриття впливає на прибутковість і ризиковість страхового продукту, і визначає компромісне рішення, що забезпечує найкращий баланс між дохідністю та ризиком. Отримані результати подаються у вигляді інтерактивних візуалізацій, які дозволяють користувачу в реальному часі спостерігати за впливом зміни параметрів на фінансові результати страхування.

Модуль генерує результати у двох форматах — текстовому та графічному. Текстовий звіт містить опис ризикової ситуації та конкретні рекомендації щодо дій («Рекомендується встановити рівень покриття 0.8 і скоригувати тарифну ставку до 1,25 %»). Графічна частина подається у вигляді гістограми розподілу VaR або сценарних діаграм («оптимістичний – базовий – песимістичний»), що полегшує інтерпретацію результатів і дає змогу оцінити потенційні наслідки кожного сценарію. На рис. 3.4 показано графічна частина.

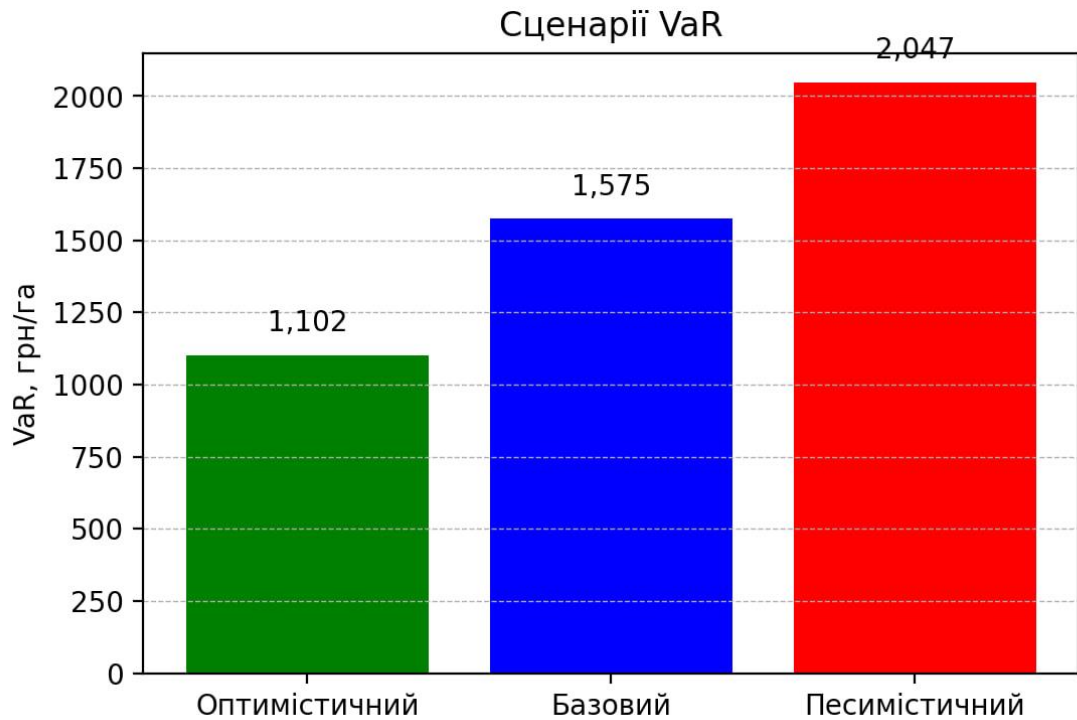


Рисунок 3.4 Приклад графічної частини виводу

На рис. 3.5 показано текстову частину надання рекомендацій.

Висновок: Рівень ризику — низький; ймовірність дефіциту — 9.4%. Рекомендовано покриття 0.95 та тариф 2.00%. Прогнозований RAROC = 0.53, ефективна ймовірність дефіциту = 0.08.

Рисунок 3.5 Текстовий вивід

Таким чином, модуль надання рекомендацій завершує повний цикл функціонування інтелектуальної системи, поєднуючи моделі прогнозування, оцінювання ризику та економічного аналізу в єдиний механізм прийняття рішень. Його методика базується на системному підході, що передбачає автоматизований вибір і обґрунтування параметрів страхування відповідно до поточного рівня ризику. Це забезпечує адаптивність, прозорість і практичну корисність системи, роблячи її ефективним інструментом для прийняття рішень у сфері аграрного страхування та управління фінансовими ризиками.

### 3.6 Висновки до розділу 3

У цьому розділі розроблено архітектуру та функціональну структуру інтелектуальної системи підтримки прийняття рішень (ІСППР), призначеної для прогнозування урожайності сільськогосподарських культур та оцінювання страхових ризиків в аграрному секторі. Запропонована система побудована на принципах модульності, ієрархічності та інтегрованості, що забезпечує її гнучкість, масштабованість і здатність адаптуватися до різних типів даних та умов функціонування аграрного ринку.

У межах архітектури системи визначено п'ять взаємопов'язаних модулів: модуль збору та обробки даних, блок моделювання, модуль оцінювання, модуль розрахунку ризику та модуль надання рекомендацій. Блок моделювання реалізує прогнозування урожайності за допомогою сучасних алгоритмів машинного навчання (Random Forest, XGBoost, Gradient Boosting, ANFIS тощо), що дозволяє враховувати складні нелінійні взаємозв'язки між природними, економічними та технологічними факторами. Модуль оцінювання забезпечує контроль точності прогнозів на основі статистичних метрик (MAE, RMSE, MAPE, RRMSE) і вибір найоптимальнішої моделі для подальших розрахунків.

Модуль оцінки та розрахунку ризику реалізує механізми актуарної оцінки страхових показників — імовірності дефіциту врожаю, очікуваного збитку, чистої та бруто-премій, показників VaR і RAROC. Завдяки модулю надання рекомендацій система переходить від аналітики до практичного прийняття рішень: формуються рекомендації щодо оптимального рівня страхового покриття, розміру тарифної ставки.

Розроблена інтелектуальна система є універсальним інструментом, який поєднує технології машинного навчання, методи фінансово-актуарного аналізу

та принципи системного підходу до управління ризиками. Вона забезпечує підвищення точності прогнозів, скорочення часу ухвалення рішень і покращення обґрунтованості страхових стратегій.

## 4 РОЗРОБКА СТАРТАП-ПРОЄКТУ

Цей розділ спрямований на розробку концепції, технічної архітектури та економічної моделі інноваційного стартап-проєкту «AgroShield AI», призначеного для кількісної оцінки страхових ризиків в аграрному секторі України з використанням сучасних технологій аналізу даних, машинного навчання та імітаційного моделювання. Метою розділу є створення комплексної системи, що дозволить не лише прогнозувати ймовірні ризики, але й оцінювати їх фінансові наслідки, забезпечуючи обґрунтованість страхових тарифів та стійкість сільськогосподарських підприємств. Розробка стартапу передбачає реалізацію таких основних завдань:

- 1) дослідити поточний стан і тенденції розвитку ринку агрострахування в Україні та світі;
- 2) провести аналіз конкурентного середовища, визначивши нішу для впровадження рішень AgroShield AI;
- 3) спроектувати архітектуру системи з урахуванням вимог до обробки великих даних, побудови моделей машинного навчання та симуляційних методів; обґрунтувати бізнес-модель і фінансову доцільність реалізації проєкту.
- 4) результатом розділу є обґрунтована концепція стартапу «AgroShield AI», яка поєднує інтелектуальні алгоритми з практичними інструментами управління ризиками та страхування у сфері агровиробництва.

#### 4.1 Аналіз ринку та передумови створення стартапу

Розвиток цифрових технологій у сільському господарстві є одним із ключових драйверів підвищення ефективності виробництва та управління ризиками у XXI столітті. Згідно з прогнозом FAO (2024), глобальний обсяг інвестицій у цифрові агротехнології зростатиме на 10–12 % щороку, досягнувши понад 50 млрд дол. США до 2030 р. Особливо швидко розвиваються напрями аналітики ризиків, моніторингу клімату та страхування сільськогосподарських культур.

Україна, як одна з провідних аграрних держав Європи, має значний потенціал для впровадження таких технологій. За даними OECD (2023), частка аграрного сектору у структурі ВВП України перевищує 17 %, а експорт агропродукції становить понад 40 % валютних надходжень держави. Проте галузь залишається вкрай залежною від кліматичних умов, нестабільності ринкових цін і геополітичних ризиків.

Важливо зазначити, що рівень проникнення страхових продуктів у сільському господарстві України залишається низьким. За аналітичним звітом IFC (2024), менше 8 % сільськогосподарських угідь покриті страховими програмами, тоді як у розвинених країнах цей показник перевищує 60 %. Серед основних причин — відсутність якісної аналітики ризиків, недовіра до страховиків і обмежені можливості оцінки збитків на основі об'єктивних даних.

Саме тому на ринку формується попит на системи нового покоління, які поєднують дані дистанційного зондування Землі (MODIS, Sentinel-2), реаналітичні кліматичні дані (ERA5-Land) та статистику врожайності для побудови моделей прогнозування ризику. Такі рішення дозволяють переходити від суб'єктивних експертних оцінок до об'єктивних кількісних показників, що

є необхідною передумовою для формування прозорої системи агрострахування.

Додатковим стимулом для розвитку ринку є стратегія «Digital Agriculture Transformation 2030» від World Bank (2023), яка передбачає інтеграцію штучного інтелекту та машинного навчання в аграрні процеси для підвищення адаптивності до змін клімату. У межах цієї стратегії Україна визначена як країна з високим потенціалом для створення аналітичних рішень у сфері оцінки ризиків. У таблиці 4.1 показано динаміку розвитку світового ринку агрострахування.

Таблиця 4.1 Динаміка розвитку світового ринку AgriTech та агрострахування (2020–2024 рр.)

Рік	Обсяг ринку AgriTech, млрд USD	Частка аналітичних сервісів, %	Обсяг ринку агрострахування, млрд USD	Приріст страхових премій, %
2020	25.4	18	32.1	3.5
2021	29.6	20	34.7	4.0
2022	36.9	22	38.2	4.8
2023	44.3	25	42.6	5.2
2024	51.0	28	47.9	5.5

У результаті проведеного аналізу можна зробити висновок, що зростання обсягів світового ринку аналітичних рішень для агросектору, підвищення попиту на цифрові інструменти управління ризиками та наявність незаповненої ніші в Україні створюють сприятливі передумови для реалізації стартапу AgroShield AI. Його ключова перевага полягає у можливості об'єднання різномірних джерел даних, застосуванні методів машинного навчання та симуляційних моделей для отримання кількісних оцінок ризику, що підвищує об'єктивність і довіру до системи агрострахування, як показано на рис. 4.1.

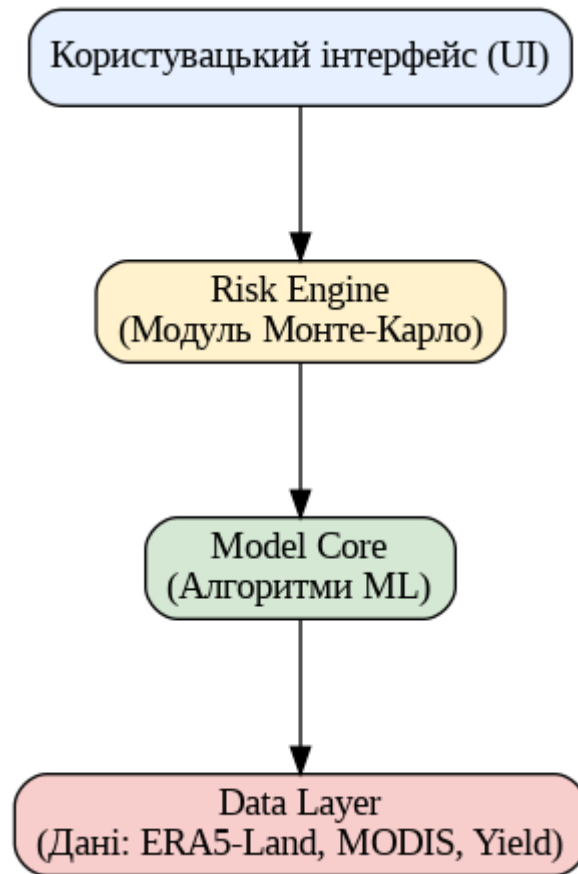


Рисунок 4.1 — Архітектура системи AgroShield AI.

## 4.2 Аналіз конкурентного середовища

### 4.2.1. Огляд світових рішень у сфері агостраховання та аналітики ризиків

Сучасний ринок аналітичних систем для агросектору характеризується активним розвитком рішень, що поєднують машинне навчання, кліматичне

моделювання та супутниковий моніторинг. Серед провідних міжнародних компаній виділяються такі.

1. ClimateAI — платформа прогнозування кліматичних ризиків, що застосовує нейромережеві моделі для прогнозу врожайності та оцінки впливу клімату.
2. Descartes Labs — аналітична платформа, що об'єднує супутникові та наземні дані для оцінки стану сільськогосподарських угідь і страхових ризиків.
3. IBM Watson Decision Platform for Agriculture — комплексне рішення, яке поєднує штучний інтелект, метеодані та IoT для агровиробників.
4. CropIn — SaaS-рішення для моніторингу полів, аналізу ризику та автоматизації страхових звітів.
5. AgroStar — індійська цифрова платформа, що поєднує агрономічні консультації, аналітику та мікрострахування для фермерів.

Більшість із цих систем орієнтовані на аналітику кліматичних ризиків, проте лише окремі рішення (як ClimateAI) реалізують імовірнісне моделювання ризику, подібне до методів, що пропонуються AgroShield AI.

#### 4.2.2. Аналіз українського ринку цифрових агросервісів

Український ринок цифрових агросервісів активно розвивається протягом останніх п'яти років. Основні рішення зосереджені на агромоніторингу, управлінні земельним банком і агрохімічному аналізі. Серед провідних компаній можна виділити такі:

- 1) EOS Data Analytics (EOSDA) — платформа супутникового моніторингу з використанням штучного інтелекту для оцінки стану полів і прогнозування врожайності.
- 2) Soft.Farm — веб-система для управління агробізнесом, що містить модулі агромоніторингу, фінансової звітності та GPS-контролю техніки.
- 3) Agrilab — компанія, що спеціалізується на агрохімічному аналізі ґрунтів та консалтингу для оптимізації врожайності.
- 4) Meteotrek — постачальник агрометеорологічних даних, які можуть бути інтегровані з аналітичними системами оцінки ризиків.
- 5) AgroOnline — онлайн-платформа для управління сільськогосподарськими процесами, проте без розвинених модулів оцінки ризиків.

Незважаючи на наявність низки сервісів, жоден із них не пропонує комплексної оцінки страхових ризиків на основі методів машинного навчання та ймовірнісного моделювання, що й формує конкурентну перевагу AgroShield AI.

### 4.2.3. Порівняльна характеристика конкурентів

У таблиці 4.2 показано порівняльний аналіз конкурентів

Таблиця 4.2 Порівняння функціональних можливостей існуючих платформ

Платформа	Тип даних	ML/AI	Оцінка ризику	Інтеграція зі страховими	Географія
ClimateAI	Кліматичні, супутникові	Так	Так (ймовірнісна)	Так	Світ
Descartes Labs	Супутникові	Так	Частково	Ні	Світ
IBM Watson Agriculture	ІоТ, метео, аналітика	Так	Так	Так	Світ
EOSDA	Супутникові	Так	Ні	Ні	Україна/ЄС
Soft.Farm	Агровиробничі	Ні	Ні	Ні	Україна
AgroShield AI	Супутникові, кліматичні, статистичні	Так	Так (Монте-Карло)	Так	Україна/ЄС

Як видно з таблиці 4.2, більшість конкурентів спеціалізуються на аналітиці або моніторингу, проте не пропонують комплексної системи прогнозування страхових ризиків. AgroShield AI поєднує підходи машинного навчання та ймовірнісного моделювання, забезпечуючи інтеграцію зі страховими продуктами та точну оцінку ризику. На рис. 4.2 показано потоки даних.

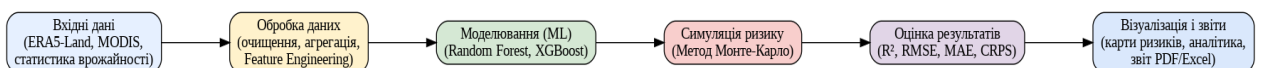


Рисунок 4.2 — Потоки даних у системі AgroShield AI.

#### 1.2.4. SWOT-аналіз стартапу AgroShield AI

Результати SWOT-аналізу показані в таблиці 4.3.

Таблиця 4.3 — SWOT-аналіз AgroShield AI

Сильні сторони (Strengths)	Слабкі сторони (Weaknesses)
<ul style="list-style-type: none"> <li>— Інноваційна технологія оцінки ризику на основі ML та Монте-Карло.</li> <li>— Інтеграція супутникових та кліматичних даних.</li> <li>— Підтримка українських агрокліматичних умов.</li> </ul>	<ul style="list-style-type: none"> <li>— Потреба у значних початкових інвестиціях.</li> <li>— Обмежене визнання бренду на ринку.</li> <li>— Висока залежність від якісних вхідних даних.</li> </ul>
Можливості (Opportunities)	Загрози (Threats)
<ul style="list-style-type: none"> <li>— Зростання попиту на аналітику ризиків у агрострахуванні.</li> <li>— Потенціал співпраці з державними структурами.</li> <li>— Можливість експорту технології на ринки ЄС.</li> </ul>	<ul style="list-style-type: none"> <li>— Конкуренція з міжнародними платформами.</li> <li>— Кіберризики та загрози витоку даних.</li> <li>— Макроекономічна нестабільність в Україні.</li> </ul>

## 4.3 Технічна реалізація системи AgroShield AI

### 4.3.1. Загальна архітектура системи

Система AgroShield AI побудована за багаторівневою архітектурою, що забезпечує модульність, масштабованість і надійність роботи. Вона складається з трьох основних рівнів.

Data Layer — рівень зберігання та управління даними, який містить бази супутникових, кліматичних і агрономічних показників. Processing Layer — рівень обробки даних, який виконує попередню підготовку, нормалізацію, інтерполяцію та фільтрацію інформації. Application Layer — рівень застосування, що забезпечує доступ користувачів через веб-інтерфейс, API та аналітичні звіти.

Між рівнями реалізовано стандартизовані протоколи взаємодії, що дозволяють легко інтегрувати зовнішні сервіси, такі як погодні API або супутникові потоки. Використовується мікросервісна архітектура з контейнеризацією (Docker), що полегшує оновлення окремих компонентів без зупинки всієї системи.

### 4.3.2. Модулі системи та їх функціональне призначення

Всі основні модулі вказані в таблиці 4.4.

Таблиця 4.4 — Опис основних модулів системи AgroShield AI

Назва модуля	Призначення	Реалізовані технології
Data Ingestion	Автоматичне завантаження кліматичних і супутникових даних	Python, API, GDAL
Data Cleaning	Фільтрація, нормалізація та перевірка даних на аномалії	Pandas, NumPy
Feature Engineering	Генерація нових ознак (EVI, NDVI, FPAR тощо)	scikit-learn
Modeling Core	Побудова моделей прогнозування та симуляції Монте-Карло	XGBoost, PyMC
Evaluation	Оцінка точності та калібрування моделей	scikit-learn, metrics
Reporting/API	Візуалізація та інтеграція результатів з інтерфейсами користувачів	FastAPI, Streamlit

### 4.3.3. Використовувані технології та інструменти

Основним програмним середовищем розробки є Python, який забезпечує високу швидкість прототипування, гнучкість та широку екосистему бібліотек. Для моделювання ризику використовуються бібліотеки scikit-learn, TensorFlow, PyMC; для обробки даних — Pandas та NumPy. FastAPI забезпечує створення REST API для взаємодії з клієнтськими застосунками, а Streamlit — швидко розробку веб-інтерфейсів аналітиків.

### 4.3.4. Вибір мови програмування та середовища реалізації

Для вибору була складена порівняльна характеристика, представлена в таблиці 4.5

Таблиця 4.5 — Порівняльна характеристика мов програмування для реалізації аналітичних систем

Мова	Переваги	Недоліки	Використання у AgroShield AI
Python	Велика екосистема, ML-бібліотеки, інтеграція з API	Низька швидкодія при обчисленнях	Основна мова реалізації
R	Потужна статистична обробка	Складна інтеграція з веб-сервісами	Використовується для перевірки моделей

Продовження таблиці 4.5

MATLAB	Зручність для наукових розрахунків	Закритий код, висока вартість	Не використовується
C++	Висока швидкість виконання	Складність розробки	Опціонально для обчислювальних модулів

#### 4.3.5. Алгоритмічне забезпечення

У системі AgroShield AI реалізовано гібридний підхід до прогнозування, який поєднує класичні алгоритми машинного навчання (Random Forest, Gradient Boosting, XGBoost) із симуляційними методами (Монте-Карло). Процес моделювання включає:

- 1) формування вибірки з кліматичних, супутникових і ґрунтових показників;
- 2) навчання моделі прогнозування врожайності;
- 3) генерування тисяч випадкових сценаріїв кліматичних умов для оцінки розподілу ризику;
- 4) визначення ймовірності настання збиткових сценаріїв.

У технічній реалізації AgroShield AI поєднано інноваційні технології аналізу даних та інструменти машинного навчання. Система забезпечує автоматичне оновлення кліматичних даних, модульне прогнозування ризику та адаптивну візуалізацію результатів. Використання мікросервісної архітектури гарантує масштабованість і можливість інтеграції зі страховими сервісами.

#### 4.3.6. Архітектура бази даних

База даних системи AgroShield AI реалізована на основі PostgreSQL із розширенням PostGIS для роботи з просторовими даними. Основні таблиці:

1. Users — інформація про користувачів і права доступу;
2. Regions — агрокліматичні регіони з координатами меж;
3. WeatherData — щомісячні кліматичні показники (t2m, tp, ssr, swvl, stl);
4. YieldData — фактична врожайність за регіонами;
5. SimulationResults — результати моделювання та оцінки ризиків.

Зв'язки між таблицями реалізовано через зовнішні ключі `region_id` і `user_id`. Оптимізація індексів забезпечує швидкий доступ до даних при виконанні прогнозів та симуляцій. Логічна схема вказана на рис. 4.3.

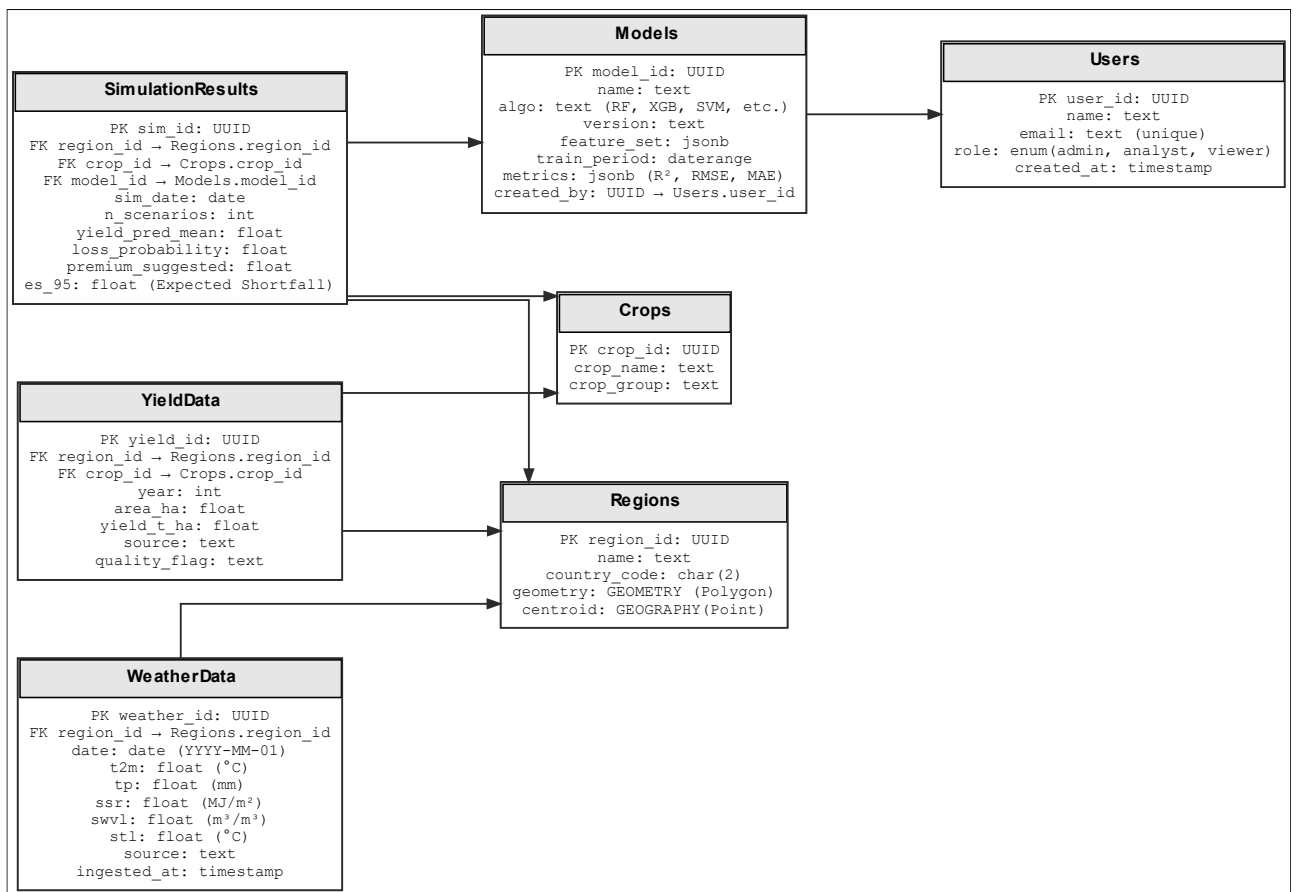


Рисунок 4.3 — Логічна схема бази даних AgroShield AI.

#### 4.4.1. Концепція бізнес-моделі та місія проєкту

Метою стартапу AgroShield AI є створення інтегрованої цифрової екосистеми для кількісної оцінки та прогнозування страхових ризиків в аграрному секторі України. Основна цінність проєкту полягає у поєднанні методів машинного навчання та симуляцій Монте-Карло для забезпечення прозорих і науково обґрунтованих страхових тарифів. Місією AgroShield AI є

підвищення фінансової стійкості агровиробників через автоматизовану аналітику ризиків, що базується на кліматичних і супутникових даних.

Цільовими клієнтами виступають страхові компанії, агрохолдинги, фермерські господарства, банки, що кредитують аграрний сектор, та державні установи, які реалізують програми підтримки. Бізнес-модель орієнтована на B2B-сегмент з можливістю розширення до B2G у разі інтеграції з державними платформами агрострахування.

#### 4.4.2. Структура продукту та джерела доходу

Архітектура продукту передбачає модульну побудову: аналітичне ядро (ML-моделі), API-інтерфейс для страхових компаній, веб-панель для аналітиків та кліматичний модуль моніторингу. Монетизація ґрунтується на комбінованій системі підписок та сервісних контрактів. Джерела доходів вказані в таблиці 4.6.

Таблиця 4.6 — Джерела доходів AgroShield AI

Джерело доходу	Опис	Річний обсяг, тис. грн (тис. \$)
Підписка B2B	Щомісячна аналітика та доступ до API	1200 (28.7)
White-label інтеграції	Індивідуальні рішення для страхових компаній	1600 (38.3)
Консалтинг та звіти	Підготовка оцінок ризику для банків та державних органів	850 (20.3)

## Продовження таблиці 4.6

Доступ до архіву даних	Продаж агрокліматичних datasetів та API-кредитів	400 (9.6)
------------------------	--	-----------

## 4.4.3. Модель взаємодії з клієнтами та партнерами

AgroShield AI використовує змішану модель партнерств: прямі контракти зі страховими компаніями та інтеграції через аграрні асоціації. Схема взаємодії включає етапи впровадження, навчання користувачів і підтримку через API, схема якого вказана на рис. 4.4.

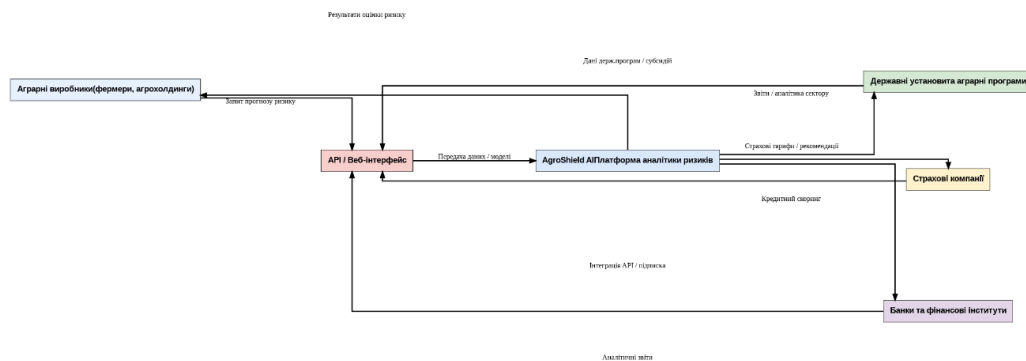


Рисунок 4.4 Модель взаємодії клієнтів AgroShield AI

## 4.4.4. Витратна структура та основні ресурси

Основні витрати зосереджені на розробці моделей ML та підтримці інфраструктури, як показано в таблиці 4.7. Додатково передбачено маркетингові та операційні витрати.

Таблиця 4.7 Основні статті витрат AgroShield AI

Стаття витрат	Опис	Річна сума, тис. грн (тис. \$)
Розробка ПЗ	Back-end, ML, інтерфейс користувача	1800 (43.1)
Хмарна інфраструктура	Сервери та зберігання даних (AWS/GCP)	720 (17.2)
Маркетинг і продажі	Digital кампанії, виставки, PR	560 (13.4)
Зарплати та адміністративні витрати	Команда 7 осіб + адміністрація	1500 (35.9)
Інші витрати	Ліцензії, підписки, непередбачені витрати	300 (7.2)

#### 4.4.5. Маркетингова стратегія та канали просування

Маркетингова стратегія AgroShield AI спрямована на формування довгострокових B2B-відносин та підвищення впізнаваності бренду в аграрному секторі. Основні канали просування — галузеві виставки, цифрові кампанії, контент-маркетинг і публікації в спеціалізованих виданнях. Особлива увага приділяється побудові експертної репутації через аналітичні звіти та власні дослідження.

#### 4.4.6. Ключові показники ефективності (KPI)

Для оцінювання ефективності впровадження AgroShield AI застосовано ключові фінансові та операційні показники: ROI, SAC, LTV, ARPU, точність прогнозів. Моніторинг KPI забезпечує контроль результативності та можливість корекції стратегії.

#### 4.4.7. Фінансова модель та прогноз розвитку стартапу

Фінансова модель розроблена для періоду 2025–2027 рр. Розрахунки здійснено у тис. грн (у дужках — еквівалент у тис. USD, курс НБУ 1 USD = 41,76 грн). Сценарій відповідає рівню MVP / pre-seed із поступовим нарощуванням клієнтської бази. Прогноз вказаний у таблиці 4.8.

Таблиця 4.8 — Прогноз фінансових показників стартапу AgroShield AI (2025–2027)

Показник	2025	2026	2027	Темп зростання, %
Дохід, тис. грн (тис.\$)	2 400 (57.5)	3 800 (91.0)	5 200 (124.5)	35 %
Витрати, тис. грн (тис.\$)	3 000 (71.8)	3 700 (88.6)	4 000 (95.8)	13 %
ЕВІТДА, тис. грн (тис.\$)	-600 (- 14.4)	100 (2.4)	1 200 (28.7)	—
Чистий прибуток, тис. грн (тис.\$)	-650 (- 15.6)	50 (1.2)	900 (21.5)	—

Точка беззбитковості («break-even point») досягається після 20 місяців роботи. Мінімальний щомісячний дохід для виходу в нуль становить  $\approx 250$  тис. грн (6.0 тис.\$). На рис. 4.5 показано візуалізацію фінансової моделі.

Фінансова модель AgroShield AI (TAM-SAM-SOM)

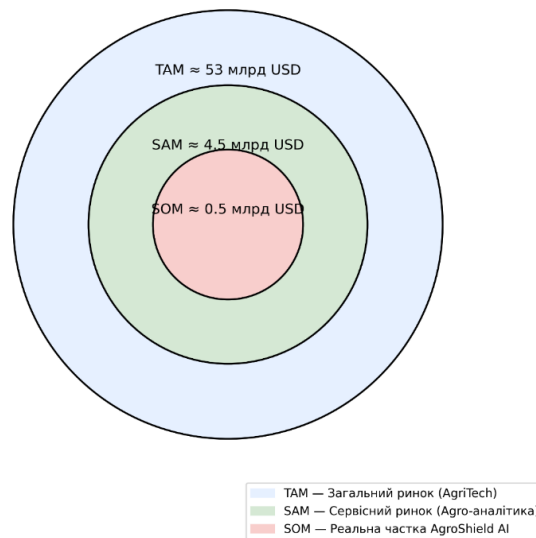


Рисунок 4.5 Фінансова модель AgroShield AI (Cash Flow Forecast)

#### 4.4.8. Аналіз ризиків бізнес-моделі

Ризики AgroShield AI класифіковано за чотирма групами: технічні, ринкові, фінансові та регуляторні. Кожен ризик оцінено за ймовірністю виникнення та рівнем впливу на результати діяльності. В таблиці 4.9 показано аналіз ризиків.

Таблиця 4.9 — Аналіз ризиків бізнес-моделі AgroShield AI

Категорія ризику	Опис	Ймовірність	Вплив на бізнес
Технічний	Збої в роботі моделей ML або сервера	Середня	Суттєвий
Ринковий	Низька готовність страхових компаній до цифровізації	Висока	Помірний
Фінансовий	Брак інвестицій на етапі масштабування	Середня	Критичний
Регуляторний	Зміни в податковому чи страховому законодавстві	Низька	Помірний

#### 4.5. Висновки до розділу 4

Розроблена бізнес-модель AgroShield AI демонструє фінансову життєздатність та інвестиційну привабливість у межах рівня MVP / pre-seed. Річний дохід до 2027 року може досягти  $\approx 5,2$  млн грн (124 тис.\$). Стартап досягає точки беззбитковості приблизно на 20-му місяці роботи, що свідчить про високу ефективність використання ресурсів. Проект має потенціал для

залучення венчурного капіталу та виходу на міжнародний ринок AgriTech протягом 2027–2028 рр.

## ВИСНОВКИ

У результаті проведеного дослідження було досягнуто мети роботи — розроблено комплексну методику оцінювання страхових ризиків для аграрного сектору з використанням сучасних методів машинного навчання та імітаційного моделювання. Проведений аналіз, моделювання та експериментальні результати підтвердили доцільність поєднання інтелектуальних методів аналізу даних із класичними економіко-математичними підходами до оцінювання ризику.

У теоретичній частині систематизовано підходи до класифікації аграрних ризиків та обґрунтовано сутність страхування як механізму їх фінансової компенсації. Встановлено, що ризики сільськогосподарського виробництва мають багатофакторну природу, стохастичний характер і значну просторово-часову мінливість, що унеможлиблює ефективне управління ними без використання кількісних методів. Проведена порівняльна характеристика якісних, кількісних та гібридних підходів до оцінювання ризику дозволила визначити оптимальну методику для задач агрострахування — комбіновану, що поєднує експертну оцінку з моделюванням на основі фактичних даних. Теоретично обґрунтовано необхідність застосування симуляційних методів, зокрема Монте-Карло, для розрахунку страхових метрик які відображають як середній рівень збитків, так і ймовірність екстремальних подій.

У практичній частині реалізовано інформаційну технологію для прогнозування врожайності та оцінювання ризику втрати врожаю. Створено модульну архітектуру системи, що включає етапи збору, очищення, нормалізації та інтеграції супутникових (MODIS) і кліматичних (ERA5-Land) даних з агростатистичною інформацією. Використання мови програмування

Python та бібліотек pandas, scikit-learn, XGBoost, NumPy забезпечило високу гнучкість і продуктивність розрахунків.

Проведено експериментальне порівняння шести моделей машинного навчання (SVM, Random Forest, Gradient Boosting, Decision Tree, KNN, XGBoost) для задачі прогнозування врожайності. За результатами моделювання встановлено, що найкращі показники точності досягнуто для моделей Support Vector Machine та Random Forest, середні значення MAPE не перевищували 9–10 %, а RMSE — 4.5–4.7, що підтверджує високу якість прогнозування. Це дозволяє використовувати одержані результати для оцінки ризиків недоотримання врожаю та подальшого розрахунку страхових премій.

На основі результатів прогнозування побудовано модель оцінки ризиків, що поєднує результати машинного навчання з імітаційним моделюванням методом Монте-Карло. Такий підхід забезпечив можливість отримання не лише точкових прогнозів, але й повного розподілу можливих збитків, що є необхідним для ймовірнісної оцінки страхових ризиків. Розрахунок показників очікуваного збитку, Value-at-Risk і Conditional Value-at-Risk дав змогу кількісно визначити потенційний фінансовий дефіцит при несприятливих сценаріях розвитку погодних умов.

Отримані результати мають наукову новизну у частині поєднання методів машинного навчання з класичними інструментами фінансового ризик-менеджменту для задач агрострахування, що дозволяє перейти від емпіричних оцінок до кількісно обґрунтованих моделей ризику. Практичне значення полягає у можливості впровадження розробленої інформаційної технології у діяльність страхових компаній, банківських установ і агропідприємств для оцінки страхових премій, планування виробництва та управління портфелем ризиків.

Достовірність одержаних результатів забезпечено використанням репрезентативних даних за 2010–2023 роки, статистичною перевіркою

моделей, перехресною валідацією (Leave-One-Year-Out), а також узгодженістю результатів з теоретичними положеннями і практичними тенденціями розвитку аграрного страхування.

У цілому, виконане дослідження підтвердило, що інтеграція методів машинного навчання, статистичного аналізу та імітаційного моделювання є ефективним інструментом для кількісного оцінювання аграрних ризиків, формування науково обґрунтованої страхової політики та підвищення фінансової стійкості агросектору України.

## ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. Biagini L. Applications of Machine Learning for the Ratemaking in Agricultural Insurances. 245 p.
2. Hajiyev E., Watson M., Emadi H., Eissa B., Hussain A., Baig A. R., Shahin A. A Comparative Study of Major Risk Assessment (RA) Frameworks in Geologic Carbon Storage (GCS). *Fuels*. 2025. P. 1–22.
3. Colli A., Serbanescu D., Ale B. Risk Evaluation Methods and Application to Energy Systems. In: Beauchamp-Akatova E., van den Top J. (eds.). *The Research Agenda of Risk and Design Anno 2008*. 2008. P. 55–68.
4. Harrington S., Niehaus G. *Risk Management and Insurance*. New York: McGraw-Hill, 2004. 624 p.
5. Bowers N., Gerber H., Hickman J., Jones D., Nesbitt C. *Actuarial Mathematics*. 2nd ed. Society of Actuaries, 2010. 753 p.
6. Klugman S., Panjer H., Willmot G. *Loss Models: From Data to Decisions*. 4th ed. Wiley, 2012. 776 p.
7. Jorion P. *Value at Risk: The New Benchmark for Managing Financial Risk*. 3rd ed. McGraw-Hill, 2007. 600 p.
8. Rockafellar R., Uryasev S. Optimization of Conditional Value-at-Risk. *Journal of Risk*. 2000. Vol. 2. P. 21–42.
9. Swiss Re Institute. *Sigma reports on agricultural insurance and natural catastrophe losses*. Zurich: Swiss Re, 2015–2024. (various reports, ~40–80 p.)
10. Matten C. *Managing Bank Capital: Capital Allocation and Performance Measurement*. Wiley, 2000. 432 p.
11. The Zen of Python. URL: <https://peps.python.org/pep-0020/> (дата звернення: 03.12.2025).

12. Ramalho L. *Fluent Python: Clear, Concise, and Effective Programming*. O'Reilly Media, 2015. 770 p.
13. McKinney W. *Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython*. 2nd ed. O'Reilly Media, 2017. 544 p.
14. Géron A. *Hands-On Machine Learning with Scikit-Learn, Keras & TensorFlow*. 2nd ed. O'Reilly Media, 2019. 856 p.
15. NASA MODIS data portal. URL: <https://modis.gsfc.nasa.gov/data/> (дата звернення: 03.12.2025).
16. Copernicus Climate Data Store. URL: <https://cds.climate.copernicus.eu/> (дата звернення: 03.12.2025).
17. Joint Research Centre Data Catalogue. URL: <https://data.jrc.ec.europa.eu/collection/> (дата звернення: 03.12.2025).
18. Державна служба статистики України. URL: <https://www.ukrstat.gov.ua/> (дата звернення: 03.12.2025).
19. Lee S.-J., Sohn E., Kim M., Park K.-H., Park K., Lee Y. Real-Time Retrieval of Daily Soil Moisture Using IMERG and GK2A Satellite Images with NWP and Topographic Data: A Machine Learning Approach for South Korea. *Remote Sensing*. 2023. Vol. 15(8). P. 1–2.
20. Kuznietsova, N., Bidyuk, P. Intelligence Information Technologies for Financial Data Processing in Risk Management. In: Babichev, S., Peleshko, D., Vynokurova, O. (eds) *Data Stream Mining & Processing. DSMP 2020. Communications in Computer and Information Science*, vol 1158. P. 539–558.
21. Ярko А.Ю., Кузнєцова Н.В. Система оцінювання і прогнозування врожайності зернових культур методами машинного навчання на основі супутникових і кліматичних даних. *Системні науки та інформатика: збірник доповідей IV науково-практичної конференції, 1-5 грудня 2025*. Київ, НН ПСА КПІ ім. Ігоря Сікорського, 2025. 6 с.

## ДОДАТОК А

```
# Standard libraries
import os

# Third-party libraries
import numpy as np
import pandas as pd

from sklearn.ensemble import GradientBoostingRegressor,
RandomForestRegressor
from sklearn.neighbors import KNeighborsRegressor
from sklearn.svm import SVR
from sklearn.tree import DecisionTreeRegressor

from xgboost import XGBRegressor

# Custom modules
from bayesian_optimization import *
from data_loader import CropYieldDataLoader
from leave_one_year_out import yield_prediction
from utils import *

target_region = ['Cherkaska', 'Chernihivska', 'Chernivetska',
'Dnipropetrovska', 'Donetska', 'Kharkivska', 'Khmelnytska', 'Kirovohradska',
                'Luhanska', 'Lvivska', 'Mykolaivska', 'Odeska',
'Poltavska', 'Rivnenska', 'Sumska', 'Ternopil'ska', 'Vinnytska',
                'Volynska', 'Zakarpatska', 'Zaporizka', 'Zhytomyrska']

n_regions = len(target_region)
region_names = target_region

years = np.arange(2010, 2023 + 1)
selected_years = [2014, 2015, 2016, 2017, 2018, 2019, 2020, 2021, 2022]

combined_file_path = 'combined_data.csv'
crop_yield_file_path = 'crop_yield.csv'
```

```

data_loader = CropYieldDataLoader(target_region, combined_file_path,
crop_yield_file_path)
data_loader.load_data()

X_selected, y_selected = data_loader.filter_years(selected_years)
result_save_path = 'results/11_20'
param_dir = 'results/yaml' # Folder containing all YAML files
params_path = 'results/11_20/hyperparams' # Folder to save results
param_spaces = load_param_spaces_from_dir(param_dir)

# Run BayesSearchCV for all models using RMSE
run_bayes_search(X_selected, y_selected, param_spaces, params_path,
rmse_scorer)
estimator = SVR
model_name = 'svm'
metric_name = 'RMSE'

yield_prediction(estimator, params_path, model_name, metric_name,
X_selected, y_selected, target_region, selected_years, result_save_path)
import numpy as np
import matplotlib.pyplot as plt
from scipy.optimize import least_squares
import time

#  $y = a * \cos(b * X) + b * \sin(a * X)$ 
def func(params, x):
    a, b = params
    return a * np.cos(b * x) + b * np.sin(a * x)

def residuals(params, x, y):
    return y - func(params, x)

def newton_raphson_least_squares(x, y, initial_params, max_iter=1000):
    params = initial_params.copy()
    history = [params.copy()]
    for i in range(max_iter):
        residuals_val = residuals(params, x, y)

        # Gradient
        grad = np.array([

```

```

        -2 * np.sum(residuals_val * (np.cos(params[1] * x) - params[1]
* x * np.sin(params[1] * x))),
        -2 * np.sum(residuals_val * (-params[0] * x * np.sin(params[1]
* x) + np.cos(params[0] * x)))
    ])

    # Hessian Matrix
    hessian = np.array([
        [
            2 * np.sum((np.cos(params[1] * x) - params[1] * x *
np.sin(params[1] * x))**2),
            2 * np.sum(-(x * residuals_val) * (params[0] * x *
np.sin(params[1] * x) - np.cos(params[0] * x)))
        ],
        [
            2 * np.sum(-(x * residuals_val) * (params[0] * x *
np.sin(params[1] * x) - np.cos(params[0] * x))),
            2 * np.sum((params[0] * x * np.sin(params[1] * x) -
np.cos(params[0] * x))**2)
        ]
    ])

    # Hessian
    try:
        hessian_inv = np.linalg.inv(hessian)
    except np.linalg.LinAlgError:
        print("Hessian is singular; stopping early.")
        break

    # Newton-Raphson
    params -= hessian_inv @ grad
    history.append(params.copy())

    if np.linalg.norm(grad) < 1e-7:
        break
    return params, history # history

np.random.seed(0)
x_data = np.linspace(0, 2 * np.pi, 100)
a_true, b_true = 100, 102

```

```

y_data = func([a_true, b_true], x_data) + np.random.normal(-1, 1,
x_data.size)

def fit_and_plot(initial_params):
    # Levenberg-Marquardt (lm)
    start_time = time.time()
    result_lm = least_squares(residuals, initial_params, args=(x_data,
y_data), method='lm')
    lm_time = time.time() - start_time
    a_fit_lm, b_fit_lm = result_lm.x
    lm_history = result_lm.nfev

    start_time = time.time()
    (a_fit_newton, b_fit_newton), newton_history =
newton_raphson_least_squares(x_data, y_data, initial_params)
    newton_time = time.time() - start_time
    newton_iter = len(newton_history)

    residual_lm = np.sum(residuals([a_fit_lm, b_fit_lm], x_data,
y_data)**2)
    residual_newton = np.sum(residuals([a_fit_newton, b_fit_newton],
x_data, y_data)**2)
    plt.figure(figsize=(18, 6))
    plt.subplot(1, 3, 1)
    plt.plot(x_data, y_data, 'o', label='Data with noise', markersize=4)
    plt.plot(x_data, func([a_true, b_true], x_data), 'g--', label='True
function')
    plt.plot(x_data, func(initial_params, x_data), 'r-', label='Initial
guess')
    plt.plot(x_data, func([a_fit_lm, b_fit_lm], x_data), 'b-', label='LM
Fitted curve')
    plt.xlabel("X")
    plt.ylabel("Y")
    plt.legend()
    plt.title("Levenberg-Marquardt Fit (method='lm')")
    plt.subplot(1, 3, 2)
    plt.plot(x_data, y_data, 'o', label='Data with noise', markersize=4)

```

```

plt.plot(x_data, func([a_true, b_true], x_data), 'g--', label='True
function')
plt.plot(x_data, func(initial_params, x_data), 'r-', label='Initial
guess')
plt.plot(x_data, func([a_fit_newton, b_fit_newton], x_data), 'm-',
label='Newton Fitted curve')
plt.xlabel("X")
plt.ylabel("Y")
plt.legend()
plt.title("Newton-Raphson Least Squares Fit")
plt.subplot(1, 3, 3)
newton_history_arr = np.array(newton_history)
plt.plot(newton_history_arr[:, 0], newton_history_arr[:, 1], '-o',
markersize=4, label="Newton Path")
plt.scatter(a_true, b_true, color='green', marker='x', s=100,
label="True Params")
plt.scatter(initial_params[0], initial_params[1], color='red',
marker='x', s=100, label="Initial Guess")
plt.xlabel("Parameter a")
plt.ylabel("Parameter b")
plt.legend()
plt.title("Parameter Space Trajectory (Newton-Raphson)")

plt.tight_layout()
plt.show()

print( f"a = {initial_params[0]}, b = {initial_params[1]}" )
print( f"Levenberg-Marquardt: a = {a_fit_lm:.2f}, b = {b_fit_lm:.2f}" )
print( f"Newton-Raphson:      a      =      {a_fit_newton:.2f},      b      =
{b_fit_newton:.2f}" )
print( f": a = {a_true}, b = {b_true}" )
print( f"Levenberg-Marquardt: {residual_lm:.2f}" )
print( f"Newton-Raphso: {residual_newton:.2f}" )
print( f"Levenberg-Marquardt: {lm_time:.4f},: {lm_history}" )
print( f"Newton-Raphson: {newton_time:.4f},: {newton_iter}" )

fit_and_plot([99.5, 101.4])
# Load the CSV file
file_path = 'results/12_20/rf_results.csv'

```

```
df = pd.read_csv(file_path)

# Extract the filename without extension for the plot title
file_name = os.path.splitext(os.path.basename(file_path))[0]

# Visualize results, print correlation, and R2 value
scatterplot_visualization(df, file_name)
from leave_one_year_out import *
yield_prediction(estimator, params_path, model_name, metric_name,
X_selected, y_selected, target_region, selected_years, result_save_path)
```