

**МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ ЕКОНОМІЧНИЙ УНІВЕРСИТЕТ ІМЕНІ
ВАДИМА ГЕТЬМАНА**

**Навчально-науковий інститут
«Інститут інформаційних технологій в економіці»**

Кафедра математичного моделювання та статистики

Освітньо-професійна програма «Економічна кібернетика»

Галузь знань 05 «Соціальні та поведінкові науки»
Спеціальність 051 «Економіка»

Форма здобуття освіти: очна (денна)

КВАЛІФІКАЦІЙНА БАКАЛАВРСЬКА РОБОТА

на тему **«Аналіз та прогнозування урожайності зернових культур за
допомогою методів машинного навчання»**

(назва теми)

здобувача Ус Ельвіни Юріївни
(ПІБ)

_____ *(підпис)*

Науковий керівник: кандидат економічних наук,
доцент Тетяна КМИТЮК

_____ *(підпис)*

**Робота допущена до захисту перед екзаменаційною
комісією з атестації здобувачів вищої освіти (ЕК)**

Завідувач кафедри: кандидат фізико-математичних наук,
професор Галина ВЕЛИКОІВАНЕНКО

_____ *(підпис)*

Київ 2025

**МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ ЕКОНОМІЧНИЙ УНІВЕРСИТЕТ
ІМЕНІ ВАДИМА ГЕТЬМАНА**

**Навчально-науковий інститут
«Інститут інформаційних технологій в економіці»**

Кафедра математичного моделювання та статистики

Освітньо-професійна програма «Економічна кібернетика»

Галузь знань 05 «Соціальні та поведінкові науки»
Спеціальність 051 «Економіка»

ПОГОДЖЕНО

Керівник проектної групи (гарант)
освітньо-професійної програми
Ольга ПРИТОМАНОВА

(підпис)
_____ 2025 р.

ЗАТВЕРДЖУЮ

Завідувач кафедри математичного
моделювання та статистики
Галина ВЕЛИКОІВАНЕНКО

(підпис)
_____ 2025 р.

ІНДИВІДУАЛЬНЕ ЗАВДАННЯ

здобувача вищої освіти

Ус Ельвіни Юріївни

(прізвище, ім'я, по батькові)

_____ очної (денної)

форми здобуття освіти

очної (денної), заочної, дистанційної

на підготовку кваліфікаційної бакалаврської роботи

на тему

«Аналіз та прогнозування урожайності зернових культур за допомогою методів машинного навчання»

Тему затверджено наказом ректора Університету від «___» _____ 20__ р. № _____

Кваліфікаційна бакалаврська робота виконується на матеріалах
сайтів Статистика сільського господарства України від AgroStats
<https://agrostats.uhmi.org.ua/>, Урожайність культур сільськогосподарських за регіонами
<https://data.gov.ua/dataset/188450fa-3a63-43b5-a783-503b52b3f847/resource/a4cc75cb-3a9d-4c97-9cfb-00f3241416be>, сайт Державної служби статистики України <https://www.ukrstat.gov.ua/> та інших відкритих джерел.

План кваліфікаційної бакалаврської роботи

Розділ 1	Теоретичні основи аналізу та прогнозування урожайності зернових культур
Розділ 2	Моделювання та прогнозування урожайності зернових культур

Об'єкт дослідження:	Урожайність зернових культур України.
Предмет дослідження:	Методи машинного навчання для аналізу та прогнозування урожайності зернових культур в Україні.
Мета виконання кваліфікаційної бакалаврської роботи:	Аналіз та математичне моделювання урожайності зернових культур в Україні з використанням методів машинного навчання з метою покращення точності прогнозів і оптимізації аграрного виробництва.

Конкретні завдання, які здобувач повинен виконати для досягнення поставленої мети:

У розділі 1: 1)Опис основних понять, пов'язаних з урожайністю зернових культур. 2)Аналіз факторів, що впливають на урожайність зернових культур.3)Дослідження впливу війни на аграрний сектор та урожайність.4) Дослідити традиційні методи прогнозування урожайності та їхні обмеження

У розділі 2: 1)Опис, підготовка та первинний аналіз вхідних даних.2)Застосування традиційних статистичних методів аналізу (ANOVA, кореляція, лінійна регресія).3)Побудова моделей прогнозування на основі машинного навчання (Random Forest, XGBoost, нейронні мережі, ETS, ARIMA).4)Порівняння точності моделей прогнозування та інтерпретація результатів.

Завдання підготував
науковий керівник

(підпис)

Тетяна КМИТЮК

(ім'я, прізвище)

«__» _____ 20__ р.

Завдання одержав
здобувач

(підпис)

Ельвіна УС

(ім'я, прізвище)

«__» _____ 20__ р.

Реферат

Кваліфікаційна бакалаврська робота складається зі вступу, двох розділів, висновків, переліку використаних джерел, що містить 34 найменувань на 4 сторінках, та 2 додатка на 9 сторінках. Загальний обсяг роботи складає 81 сторінку друкованого тексту, з яких 68 сторінок основної частини. Робота містить 67 рисунків та 31 таблиць.

«Аналіз та прогнозування урожайності зернових культур за допомогою методів машинного навчання»

Об'єктом дослідження є урожайність зернових культур у регіонах України у період з 2000 до 2023 року.

Предметом дослідження виступають методи машинного навчання для аналізу та прогнозування урожайності зернових культур в Україні.

Мета та завдання виконання дослідження. Метою виконання кваліфікаційної бакалаврської роботи є аналіз та математичне моделювання урожайності зернових культур в Україні з використанням методів машинного навчання з метою покращення точності прогнозів і оптимізації аграрного виробництва.

Для досягнення поставленої мети було виконано такі *завдання*:

- Опис основних понять, пов'язаних з урожайністю зернових культур.
- Аналіз факторів, що впливають на урожайність зернових культур.
- Дослідження впливу війни на аграрний сектор та урожайність.
- Дослідити традиційні методи прогнозування урожайності та їхні обмеження
- Опис, підготовка та первинний аналіз вхідних даних.
- Застосування традиційних статистичних методів аналізу (ANOVA, кореляція, лінійна регресія).
- Побудова моделей прогнозування на основі машинного навчання (Random Forest, XGBoost, нейронні мережі, ETS, ARIMA).
- Порівняння точності моделей прогнозування та інтерпретація результатів.

Теоретична, методична та практична значущість отриманих результатів. У роботі обґрунтовано доцільність використання машинного навчання в аграрній аналітиці. Розглянуто вплив різних факторів на урожайність, а також наслідки повномасштабної війни для аграрного сектору. Запропоновано методики оцінки регіональних відмінностей урожайності, виявлено найбільш продуктивні області України. Побудовано й порівняно декілька моделей прогнозування, серед яких найточнішими виявились нейронні мережі (nnetar). Здійснено багаторівневу візуалізацію результатів, зокрема коробчасті діаграми для аналізу урожайності за регіонами.

Практичні результати дослідження полягають у створенні прогнозних моделей, які можуть бути використані у плануванні посівних кампаній, формуванні аграрної політики та оцінці ризиків виробництва. Запропонований інструментарій дає змогу підвищити ефективність агробізнесу через покращення управлінських рішень, що стосуються урожайності в регіональному розрізі.

Рік виконання кваліфікаційної бакалаврської роботи – 2025.

Рік захисту роботи – 2025.

Ключові слова: урожайність, зернові культури, машинне навчання, аграрна аналітика, регресія, нейронні мережі, прогноз, коробчаста діаграма, RMSE

ЗМІСТ

ВСТУП	3
РОЗДІЛ 1 ТЕОРЕТИЧНІ ОСНОВИ АНАЛІЗУ ТА ПРОГНОЗУВАННЯ УРОЖАЙНОСТІ ЗЕРНОВИХ КУЛЬТУР	6
1.1 Поняття урожайності та її роль у сільському господарстві.....	6
1.2 Фактори, що впливають на урожайність зернових культур.....	9
1.3 Економіко-математичні методи аналізу та прогнозування урожайності	13
РОЗДІЛ 2 МОДЕЛЮВАННЯ ТА ПРОГНОЗУВАННЯ УРОЖАЙНОСТІ ЗЕРНОВИХ КУЛЬТУР	21
2.1. Опис, підготовка та первинний аналіз даних.....	21
2.2. Моделювання урожайності традиційними економіко-математичними методами	25
2.3. Застосування методів машинного навчання для прогнозування урожайності зернових культур по областях України.....	32
2.4. Порівняння точності моделей прогнозування	40
ВИСНОВКИ	67
СПИСОК ЛІТЕРАТУРНИХ ДЖЕРЕЛ	69
ДОДАТКИ	73

ВСТУП

Сільське господарство - це одна з ключових галузей економіки України, гарантує продовольчу стабільність і становить вагомую частку національного експорту та має значний вплив на розвиток регіонів. Показники урожайності зернових культур є одним з головних індикаторів успішності сільськогосподарського виробництва та постійно контролюється державними органами, науковими інститутами та самими аграріями.

В умовах кліматичних змін, економічної нестабільності та збройного конфлікту, що серйозно вплинув на українське сільське господарство, нагальною є потреба у ефективних способах аналізу й передбачення урожайності. Паралельно, бурхливий розвиток інформаційних технологій та математичного інструментарію машинного навчання відкриває широкі перспективи для створення моделей, які враховують як внутрішні показники аграрного сектору, так і зовнішні фактори впливу.

Актуальність теми полягає в тому, що виробництво зерна є наріжним каменем агропромислового комплексу України, що гарантує продовольчу безпеку країни. В умовах кліматичних змін, військового стану, фінансової нестабільності та глобальних викликів аграрна сфера потребує швидких рішень, особливо щодо планування та прогнозування урожайності. Використання методів машинного навчання для моделювання урожайності - це інноваційний підхід, що значно збільшує точність прогнозів, порівнюючи з традиційними методами. Це надзвичайно важливо для України, де аграрний сектор відіграє стратегічну роль.

Аналіз попередніх досліджень та публікацій

Питання прогнозування урожайності досліджувалося у роботах таких українських та закордонних науковців, як В. Месель-Веселяк, С. Колодізев, І. Бурковський, J. W. Jones, D. Lobell, R. Fischer. У їхніх працях розглядаються впливи кліматичних і економічних чинників на урожайність, застосовуються регресійні, симуляційні та агрономічні моделі. Однак у більшості з них обмежене

використання сучасних алгоритмів машинного навчання, які здатні виявляти складні нелінійні взаємозв'язки у великих обсягах даних. Тому актуальним є створення моделей прогнозування урожайності саме з використанням машинного навчання на основі фактичних даних регіонів України, що й становить головну новаторську складову цієї роботи.

Метою кваліфікаційної бакалаврської роботи є аналіз та математичне моделювання урожайності зернових культур в Україні з використанням методів машинного навчання з метою покращення точності прогнозів і оптимізації аграрного виробництва.

Для досягнення поставленої мети визначено такі завдання:

- Опис основних понять, пов'язаних з урожайністю зернових культур.
- Аналіз факторів, що впливають на урожайність зернових культур.
- Дослідження впливу війни на аграрний сектор та урожайність.
- Дослідити традиційні методи прогнозування урожайності та їхні обмеження
- Опис, підготовка та первинний аналіз вхідних даних.
- Застосування традиційних статистичних методів аналізу (ANOVA, кореляція, лінійна регресія).
- Побудова моделей прогнозування на основі машинного навчання (Random Forest, XGBoost, нейронні мережі, ETS, ARIMA).
- Порівняння точності моделей прогнозування та інтерпретація результатів.

Об'єкт дослідження є урожайність зернових культур України.

Предметом дослідження є методи машинного навчання для аналізу та прогнозування урожайності зернових культур в Україні.

Методи дослідження

У роботі застосовано описові статистичні методи для аналізу структури та варіації урожайності (середнє значення, дисперсія, коробчасті діаграми), класичні економетричні підходи (лінійна регресія, ANOVA), методи часових рядів (ETS, ARIMA), а також алгоритми машинного навчання (Random Forest, XGBoost, нейронні мережі). Порівняння моделей здійснено на основі похибки прогнозування (RMSE), що дало можливість оцінити точність і придатність кожного підходу.

Теоретичне, методичне та практичне значення отриманих результатів

Теоретична значущість роботи полягає в систематизації знань щодо застосування методів машинного навчання в аграрному прогнозуванні. Методична значущість - у формуванні підходу до порівняльного аналізу моделей прогнозування урожайності. Практична цінність результатів полягає в тому, що розроблені моделі можуть бути використані агропідприємствами, регіональними адміністраціями та державними установами для точнішого прогнозування урожайності, планування посівних площ, оптимізації ресурсів та підвищення ефективності агровиробництва навіть в умовах військового часу.

Інформаційною базою дослідження стали статистичні дані про урожайність зернових культур по регіонах України за період 2000–2023 років, зібрані з відкритих джерел – зокрема з офіційних сайтів Державної служби статистики України, а також з наукових публікацій та баз даних.

РОЗДІЛ 1

ТЕОРЕТИЧНІ ОСНОВИ АНАЛІЗУ ТА ПРОГНОЗУВАННЯ УРОЖАЙНОСТІ ЗЕРНОВИХ КУЛЬТУР

1.1 Поняття урожайності та її роль у сільському господарстві

Сільське господарство є однією з ключових галузей економіки, оскільки воно забезпечує населення продуктами харчування, а промисловість — сировиною. Від ефективності аграрного сектора залежить продовольча безпека країни та її експортний потенціал. Сільське господарство традиційно є важливою складовою вітчизняного господарського комплексу. Тенденції та перспективи розвитку галузі визначаються загальним станом національної економіки, на якій, у свою чергу, значною мірою позначається динаміка основних показників аграрної діяльності. Специфіка сільськогосподарського виробництва, а саме: характер виробленої продукції, приналежність до первинного сектору економіки, залежність від погодних умов, визначає в той чи інший період і ступінь впливу галузі на макроекономічну ситуацію в Україні. На сьогодні сільське господарство займає частку у 17,0% у структурі валової доданої вартості України і є другою за величиною галуззю, яка наповнює державний бюджет країни (див. рис. 1.1) [13].



Рисунок 1.1 - Структура валової доданої вартості у III кварталі 2021 року, %

Джерело: розроблено автором на основі [13]

Найбільшу частку серед суб'єктів господарювання агросектору займають підприємства з вирощування зернових та зернобобових культур – 44,3%. На другому місці – ті, що вирощують соняшник (29,8%). Далі йдуть підприємства з вирощування сої (12%), ріпаку озимого та кользи (6,8%), і цукрового буряку (0,7%). Це пов'язано з тим, що зернові та зернобобові культури є основною складовою експорту, а Україна входить до трійки найбільших експортерів зерна у світі.

Урожайність – це кількість рослинницької продукції, отриманої з одиниці площі. Для культур відкритого ґрунту її вираховують в центнерах з гектара (ц/га), а для тепличного виробництва — в кг з 1 м² [21].

У плануванні, обліку та економічному аналізі виділяють такі види урожайності [21]:

1) потенційна урожайність - максимальна кількість продукції, яку можна отримати з 1 га за умови повної реалізації продуктивних можливостей сільськогосподарської культури або сорту. Обчислюється до ідеальних і звичайних умов науково-дослідними установами. Використовується для визначення структури землеробства, набору сортів і сільськогосподарських культур;

2) планова урожайність - кількість продукції, яку можна отримати з 1 га в конкретних умовах господарства. Визначається перед посівом з урахуванням потенціалу сорту, досягнутого рівня урожайності, родючості ґрунту, забезпеченості технікою, мінеральними добривами тощо;

3) очікувана урожайність - передбачуваний збір продукції, який визначається в окремі періоди вегетації за густотою стеблостою та загальним станом рослин. Вимірюється в ц з 1 га або оціночно: висока, середня, низька. Використовується для планування агротехнічних заходів;

4) урожайність на корені (біологічна урожайність) - кількість вирощеної продукції, встановлену вибірковим методом або методом проб. Використовується в економічному аналізі для виявлення резервів зниження втрат при збиранні;

5) фактичний збір - урожайність, що визначається за оприбуткованою або чистою (після обробки) вагою продукції на 1 га посівної, весняної або фактично зібраної площі.

2) Тип ґрунту – якісні характеристики та родючість ґрунтів (вміст гумусу, структура ґрунту) можуть визначати рівень виробництва зерна. Ґрунти, що характеризуються низьким вмістом поживних елементів, потребуватимуть збільшеної кількості добрив для підтримання високих показників урожайності.

Як приклад, довготривале використання суперабсорбуючого полімеру (SAP) та органічного добрива (OM) сприяє збільшенню кількості макропор (>1000 мкм) у ґрунті на 5–20 см, що покращує водостійкість агрегатів та здатність утримувати вологу. Це, в свою чергу, веде до збільшення урожаю пшениці на 20–30%.

3) Агротехнічні заходи – правильний вибір сортів зернових культур, використання добрив, сівозміна, обробка ґрунтів, а також заходи захисту рослин від шкідливих організмів та хвороб – усі ці аспекти можуть значно вплинути на рівень урожайності.

Для ілюстрації, метааналіз практик ведення сільського господарства в посушливих умовах, проведений у США та Китаї, показав, що поєднання мінімального обробітку ґрунту з покривними бобовими культурами (LCC) та органічними добривами сприяє збільшенню урожайності озимої пшениці на 58% при використанні низьких доз азотних добрив. Це пояснюється покращенням структури ґрунту та сидерацією.

4) Технології у землеробстві - застосування передових агротехнологій, таких як автоматизація збирання врожаю, використання точного землеробства та впровадження інноваційних систем зрошення, здатне суттєво покращити результативність вирощування зернових культур.

Для прикладу, застосування точного землеробства (диференційоване внесення мінеральних добрив, ґрунтові сенсори) призвело до зростання урожайності кукурудзи на 6% у Сполучених Штатах, одночасно зменшивши обсяг використання гербіцидів на 9%. Застосування зображень із супутників для спостереження за станом посівів дозволяє оперативно виявляти ознаки стресу у рослин.

5) Економічні та соціальні фактори - вартість добрив, пестицидів, сільськогосподарської техніки, а також державне фінансування та кредитні програми можуть відігравати вирішальну роль у підвищенні урожайності.

Наприклад, витрати на генетично модифіковане насіння кукурудзи зросли вдвічі між 1996 і 2016 роками, проте середні витрати на бушель знизилися на 25% завдяки масштабуванню виробництва та підвищенню продуктивності. Державні програми підтримки фермерів у посушливих регіонах сприяють переходу на ресурсозберігаючі технології.

6) Воєнні дії, що почалися на українській землі з лютого 2022 року, призвели до масштабного нищення інфраструктури, що є критично важливою для аграрного сектору. Одними з найбільш уражених об'єктів стали сільськогосподарські землі та виробничі потужності, які потрапили в зону бойових дій. Вибухи, артилерійські обстріли та мінування спричинили знищення значних площ орних земель, садів, виноградників та інших сільськогосподарських територій. Це не лише унеможливило поточні та майбутні посівні кампанії, а й зруйнувало екосистеми, що підтримували сільське господарство [11].

Окремої уваги потребує мінування аграрних земель. Україна нині є однією з найбільш замінованих країн світу. За оцінками, сотні тисяч гектарів орних земель забруднені мінами та боєприпасами, що не розірвалися. Це унеможливує їхнє використання для сільськогосподарських робіт, ставить під загрозу життя фермерів та працівників і потребує значних зусиль з розмінування. Міни та інші вибухонебезпечні предмети не лише обмежують доступ до цих територій, а й серйозно шкодять екосистемі, завдаючи збитків ґрунтам та біорізноманіттю.

Не можна також забувати, що, окрім руйнування фізичної інфраструктури, війна спровокувала серйозні збої в ланцюгах поставок аграрної продукції. Блокада портів, пошкодження транспортних шляхів та загальна небезпека пересування ускладнили транспортування як у середині країни, так і за її межами. Це створило суттєві проблеми для експорту української сільськогосподарської продукції, особливо зернових, які традиційно займають провідні позиції на міжнародних ринках. Через блокаду морських портів, що є ключовими для експорту, аграрії були

вимушені шукати альтернативні шляхи доставки продукції. Відсутність можливості використовувати морські маршрути призвела до значного зменшення обсягів експорту і збільшення витрат на транспортування. Це, в свою чергу, негативно позначилося на економічній стабільності аграрного сектору та доходах фермерів. Внутрішні ланцюги поставок також зазнали серйозних збоїв. Багато фермерів зіткнулися з проблемами постачання необхідних ресурсів, як-от добрива, насіння, паливо та інші матеріали, потрібні для ведення сільського господарства. Ці труднощі ускладнюють процес виробництва та знижують ефективність аграрних підприємств.

В Україні, у 2022 році, доходи від експорту сільськогосподарської продукції зросли до 52% від усіх експортних надходжень у порівнянні з доходами 2021 року (41%). Це пояснюється тим, що ми зменшили асортимент або кількість експорту інших основних продуктів (метали, руда, обладнання та інше). Але в абсолютних цифрах вартість експортованої сільськогосподарської продукції у 2022 році склала 21 мільярд доларів США проти 27 мільярдів доларів США у 2021 році (тобто спостерігається зниження експорту на 22,2% за рік) [2].

До початку повномасштабної війни Україна експортувала 95% зернових та олійних культур чорноморськими портами. З лютого 2022 року ця продукція перевозилась до країн ЄС через дунайські порти, залізницею та автоперевезеннями, що значно підвищило витрати на транспортування та перевалку. Часткове відновлення перевезення агропродукції Чорним морем з серпня 2022 року підтримало аграріїв і дало їм змогу реалізовувати продукцію з меншими витратами на логістику. Але до довоєнних обсягів експорту навіть не наблизило.

Сільське господарство України зазнало збитків на понад 40 млрд дол. США у 2022 році, перш за все, внаслідок втрати посівних площ, зменшення урожайності, збільшення витрат на логістику, а також скорочення експорту агропродукції. Величезний вплив на здатність нашої держави виробляти продовольство справили такі чинники, як втрата та ушкодження засобів виробництва, нестача робочих рук, стрімке подорожчання паливно-мастильних матеріалів, добрив та засобів захисту рослин. Особливо відчутно це було під час проведення весняної посівної кампанії

2022 року. Тоді ранніми зерновими було засіяно приблизно 79% площ, а восени – лише 60% полів було засіяно озимими культурами.

Не зважаючи на те, що у 2023 році Україна звільняла окуповані території та проводила розмінування земель, декілька мільйонів гектарів залишились необробленими через неможливість здійснювати там будь-яку виробничу діяльність. Однак, завдяки сприятливим погодним умовам для сільського господарства урожайність вирощуваних культур у 2023 році значно зросла та компенсувала втрати площ. Урожайність зернових – особливо ячменю та пшениці – в деяких областях сягнула рекордного рівня. Згідно зі статистичними даними, в середньому з 1 гектара у 2022 році було зібрано 45,8 центнера зернових, а у 2023 році – 51,8 центнера з гектара [2].

1.3 Економіко-математичні методи аналізу та прогнозування урожайності

У дослідженні використовували ряд звичних статистичних підходів для аналізу та передбачення урожайності, кожен з яких відрізняється своїми перевагами та недоліками.

Перевірка даних на нормальність

Опис:

Задля оцінювання гіпотези про нормальність розподілу урожайності було застосовано два статистичні методи: тест Шапіро–Уїлка та критерій Колмогорова–Смірнова [10].

Критерій К.-С. (K-S test) - Критерій Колмогорова-Смирнова обчислюється при відомому середньому і середньоквадратичному відхиленні генеральної сукупності. Якщо обчислена D-статистика значуща, то гіпотеза про те, що дані мають нормальний розподіл, відкидається.

Переваги:

- Легкий для виконання розрахунків та інтерпретації результатів.

- Підходить як для зіставлення розподілу з теоретичним, так і для порівняння двох емпіричних розподілів між собою.
- Має асимптотичний закон, що не залежить від типу розподілу, дозволяючи використовувати критерій навіть за відсутності інформації про точний розподіл даних, при великих вибірках.
- Корисний для попередньої оцінки відповідності розподілів у масивних вибірках.

Обмеження:

- Вимагає достатньо великого розміру вибірки (понад 50 спостережень).
- Теоретичний розподіл має бути повністю визначений, включаючи всі його параметри.
- Не передбачає коригувань на втрату ступенів свободи, у випадку коли параметри розподілу оцінюються на основі вибірових даних.
- Найбільш чутливий до змін у центральній частині розподілу, а не на його краях.
- Менш ефективний у виявленні невеликих відхилень, особливо на кінцях розподілу.

Критерій Шапіро-Уїлка (Shapiro – Wilk's) W test – W - за даними спостережень обчислюється W -статистика, і якщо вона значуща, гіпотеза про нормальний характер розподілу відкидається, інакше – приймається.

Переваги:

- Висока чутливість до відхилень від нормального розподілу, особливо у вибірках невеликого та середнього розміру.
- Статистично обґрунтований підхід, який показує добрі результати при обсязі спостережень до 2000.
- Широко доступний у різних статистичних програмних пакетах (R, Python, SPSS та інші).
- Використовується для формальної перевірки припущень, що передують застосуванню параметричних методів.

Обмеження:

- Чутливий до аномальних значень, які часто зустрічаються в сільськогосподарських даних.
- При великих вибірках навіть незначні відхилення від нормальності можуть визначатись як статистично значущі.
- Не підходить для мультимодальних розподілів або розподілів із значною асиметрією.
- Не бере до уваги структурну залежність у часових рядах (сезонність, тренд).

Кореляційний аналіз

Його суть полягає у визначенні ступеня тісноти зв'язку між двома і більше випадковими величинами (факторами). Кореляційний аналіз є корисним на попередньому етапі обробки даних [3].

Переваги:

- Легкість інтерпретації результатів.
- Добре підходить для первинного аналізу зв'язків між змінними.
- Сприяє виявленню потенційних змінних для подальшого моделювання.

Обмеження:

- Розкриває лише лінійні взаємозалежності – складні, нелінійні зв'язки залишаються невидимими.
- Кореляція не дорівнює причинно-наслідковому зв'язку.
- Може призводити до хибних зв'язків (спуріозна кореляція), які обумовлені наявністю іншої прихованої змінної.
- Чутливий до аномалій та пропущених даних.

Дисперсійний аналіз (ANOVA)

Це група методів для обробки даних, які залежать від якісних факторів. Ці методи оцінюють суттєвість впливу факторів на результати спостережень. Дисперсійний аналіз використовують на попередньому етапі аналізу даних при перевірці однорідності даних, адекватності моделі регресії і т.п [3].

Переваги:

- Ефективний для порівняння декількох груп.

- Легко піддається інтерпретації.
 - Стандартний інструмент для аналізу категоріальних факторів.
- Обмеження:
- Вимагає нормального розподілу у кожній групі даних.
 - Потребує гомогенності дисперсій – рівності варіацій між групами.
 - Не враховує взаємодії між кількома факторами (наприклад, рік * регіон).
 - Погано працює з даними, які містять нестабільність або тренди у часі.

Лінійна регресія

Даний метод полягає у пошуку лінії найкращого підходу, яка може допомогти у прогнозуванні результату для безперервних залежних змінних. В основі методу лежить лінійна функція обчислення зв'язку між однією незалежною та однією залежною змінною. У множинній лінійній регресії є більше двох незалежних змінних [30].

Переваги:

- Простота побудови та розуміння моделі.
- Можна легко візуалізувати результат.
- Підходить для швидкої оцінки впливу окремих змінних.

Обмеження:

- Припускає лінійний взаємозв'язок між змінними.
- Не враховує взаємодій та складних структур у даних.
- Проблема мультиколінеарності (взаємної залежності предикторів) може спотворити результати.
- Не адаптована до змінної динаміки або сезонності без попереднього перетворення даних.

Візуалізація даних

Коробчасті діаграми (Boxplots)

Опис:

Коробчаста діаграма (ящик з вусами) відображає п'ять основних статистичних показників: мінімум, перший квартиль, медіану, третій квартиль і

максимум. Також виявляються потенційні викиди, які позначаються окремими точками [14].

Переваги:

- Наочно показує розкид даних, симетрію та наявність викидів.
- Зручно порівнювати розподіли між різними групами (наприклад, регіонами).
- Не залежить від форми розподілу, корисний для попередньої діагностики.

Обмеження:

- Не відображає кількість спостережень у кожній частині діаграми.
- Може бути недостатньо інформативним для малих вибірок.
- Не показує тренду у часових рядах.

Трендові графіки

Опис:

Графік, який демонструє динаміку змін, наприклад, урожайності з часом. По горизонтальній осі можуть бути відображені роки, по вертикальній, наприклад, урожайність [15].

Переваги:

- Дозволяють візуалізувати тренд, сезонність та коливання у часі.
- Легко відслідковувати зміни в окремих регіонах або загальні тенденції.
- Корисний для виявлення структурних зламів або змін у даних.

Обмеження:

- Може бути зашумленим при великій кількості значень.
- Не показує розподіл даних (асиметрію, викиди).
- Вимагає додаткового аналізу для точного виявлення причин змін.

Графіки розсіювання (Scatter Plots)

Опис:

Двовимірний графік, де кожна точка репрезентує пару значень – наприклад, рік і урожайність, або два регіони для порівняння [23].

Переваги:

- Відмінно підходять для виявлення кореляцій і взаємозв'язків.
- Дозволяють побачити кластеризацію, лінійність або асиметрію.
- Можна додати лінії регресії, що покращує інтерпретацію.

Обмеження:

- Ефективність знижується при великій кількості точок – графік може бути «захаращений».
- Може бути візуально неінформативним, якщо дані сильно коливаються або мають тренди.
- Не показує розподілу окремо, як гістограми чи boxplot'и.

Порівняння із сучасними методами прогнозування

Не зважаючи на те, що традиційні статистичні підходи мають тривалу історію застосування, сучасні методи машинного навчання (ML), які також використовувались у даному дослідженні, демонструють значно більшу гнучкість та точність.

Дерева рішень та випадкові ліси (Random Forest):

Дерева рішень у машинному навчанні використовуються як передбачувальні моделі, що відображають знання про об'єкт (представлені гілками) у множині рішень. Це один з підходів до передбачувального моделювання у статистиці, добуванні даних та машинному навчанні [5].

В аналізі рішень «дерево рішень» використовуються як візуальний і аналітичний інструмент підтримки ухвалення рішень, де розраховуються очікувані значення (або очікувана корисність) конкуруючих альтернатив.

Дерево рішень складається з трьох типів вузлів:

Вузли рішення — зазвичай представлені квадратами.

Імовірнісні вузли — представляються у вигляді кола.

Замикаючі вузли — представляються у вигляді трикутника.

Випадковий ліс - це підхід ML для вирішення проблем класифікації та регресії. Він використовує ансамблеве навчання - метод вирішення складних завдань шляхом інтеграції багатьох класифікаторів [25].

Алгоритм визначає результат на основі прогнозів дерева рішень. Він прогнозує шляхом усереднення результатів різних дерев. Зі збільшенням кількості дерев точність прогнозу зростає.

Недоліків алгоритму дерева рішень можна уникнути, використовуючи метод випадкового лісу. Це підвищує точність, зменшуючи при цьому надмірне використання набору даних. Він генерує прогнози, не вимагаючи великої кількості налаштувань пакета.

- Здатні враховувати нелінійні залежності та взаємодії між змінними.
- Можуть перенавчатись, особливо при недостатньому налаштуванні параметрів.

XGBoost:

Extreme Gradient Boost (XGBoost) - це масштабований інструментарій для навчання на основі узагальненого дерева рішень з градієнтним прискоренням (GBDT). Це найкраща бібліотека машинного навчання для задач регресії, класифікації та ранжування, яка включає в себе паралельний бустинг дерев [26].

- Забезпечує високу точність завдяки багатьом ітеративним покращенням моделі.
- Потребує ретельного налаштування гіперпараметрів та значних ресурсів.

Нейронні мережі:

Нейронні мережі – це сімейство архітектур моделей, розроблених для пошуку нелінійних закономірностей у даних. Під час тренування нейронної мережі модель автоматично вивчає, які поєднання ознак будуть оптимальними для вхідних даних і мінімізують втрати [9].

- Ефективно моделюють складні, багаторівневі залежності, особливо при наявності великої кількості даних.
- Вимагають великого обсягу обчислювальних ресурсів і часу для навчання.

ETS та ARIMA:

Моделі ETS — це сімейство моделей часових рядів із базовою моделлю простору станів, що складається з компонента рівня, компонента тенденції (Т), сезонного компонента (S) і терміну помилки (E) [29].

ARIMA (англ. autoregressive integrated moving average) є узагальненням моделі авторегресійної ковзної середньої (ARMA). Обидві ці моделі адаптуються до даних часових рядів або для кращого розуміння даних, або для прогнозування. Моделі ARIMA застосовуються в деяких випадках, коли дані демонструють докази нестаціонарності. Коли сезонність відображається в часовому ряді, можна застосувати сезонну різницю, щоб усунути сезонний компонент [29].

- Потужні для моделювання часових рядів, особливо коли маємо сезонність або тренди.
- Не враховують зовнішні пояснюючі змінні (наприклад, агротехнічні або економічні фактори).

РОЗДІЛ 2

МОДЕЛЮВАННЯ ТА ПРОГНОЗУВАННЯ УРОЖАЙНОСТІ ЗЕРНОВИХ КУЛЬТУР

2.1. Опис, підготовка та первинний аналіз даних

Дані щодо урожайності зернових культур, що використовувалися для аналізу та прогнозування, отримані з офіційного статистичного сайту «Статистика сільського господарства України від AgroStats» та Державної служби статистики України [31,33]. Набір даних містить 576 спостережень за період з 2000 по 2023 роки по всіх областях України. У таблиці (див. додаток А) наведені ключові показники, зокрема рік спостереження, область та урожайність.

Початковим етапом став статистичний аналіз даних з метою оцінки базових характеристик розподілу урожайності. З використанням описової статистики було виявлено середні значення, дисперсію та стандартне відхилення для кожної області та року.

Таблиця 1.1 – Результати описової статистики

Область	Рік	Урожайність
Length:576	Min. :2000	Min. :0.00
Class :character	1st Qu.:2006	1st Qu.:25.50
Mode :character	Median :2012	Median :32.70
	Mean :2012	Mean :36.45
	3rd Qu.:2017	3rd Qu.:45.73
	Max. :2023	Max. :77.20

Джерело: розроблено автором

З табл. 2.1, як ми бачимо, впливає: змінна область нараховує 576 спостережень та має текстовий формат. Діапазон значень змінної Рік коливається від 2000 до 2023 року. Медіана (2012) та середнє (2012) збігаються, що натякає на симетричний розподіл. Дані рівномірно розподілені по роках. Змінна Урожайність

демонструє мінімальне значення (0.00), що може свідчити про відсутність врожаю в певні роки. Медіана (32.7) показує, що 50% даних не перевищують 32.7. Середнє значення (36.45) дещо перевищує медіану, що може бути ознакою впливу кількох високих значень, які зміщують середнє. 75-й квантиль (45.73) означає, що тільки 25% значень перевищують 45.73. Максимальне значення (77.2) суттєво вище за 75-го квантилю, що потенційно вказує на наявність аномальних (надзвичайно високих) показників урожайності.

Додатково, в процесі аналізу було визначено такі характеристики, як середнє, медіана та стандартне відхилення. Результати: середнє значення – 36.4, медіана – 32.7, стандартне відхилення – 14.3.

Для аналізу нормальності розподілу даних було застосовано тести Шапіро-Вілка та Колмогорова-Смірнова, результати яких наведені у табл. 2.2, а також побудовано гістограму розподілу урожайності, як показано на рисунку 2.1.

Таблиця 2.2 - Результати тестів

Тест	Результат	p-value
Шапіро-Уїлкова	W = 0.95191	9.439e-13
Колмогорова Смірнова	D = 0.11431	5.809e-07

Джерело: розроблено автором

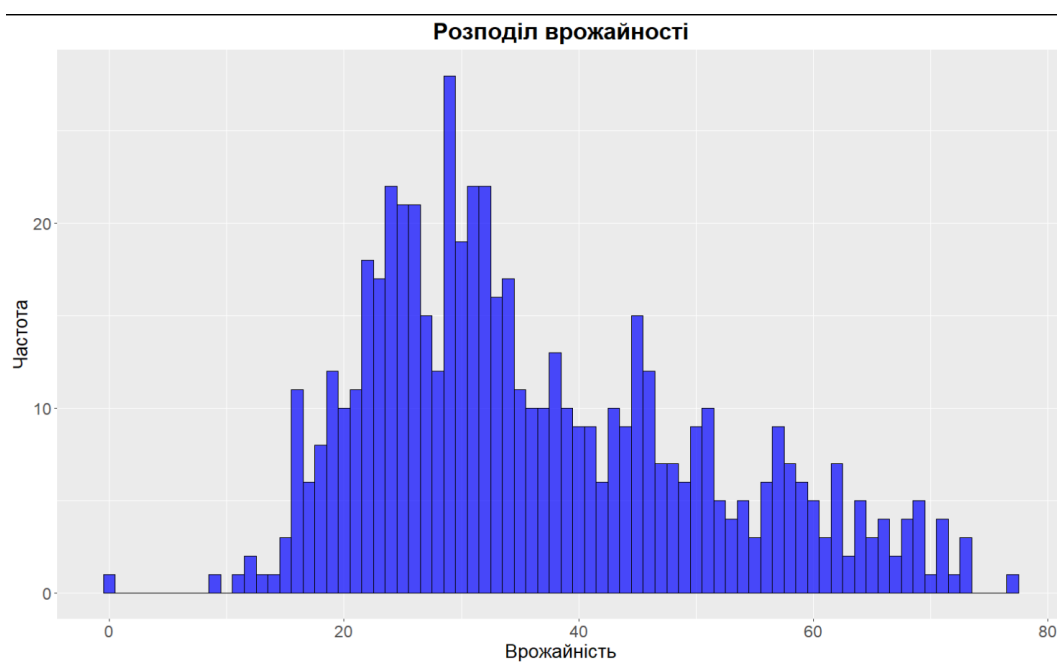


Рисунок 2.1- Розподіл урожайності зернових культур

Джерело: розроблено автором

З даних, що представлені на рис. 2.1, стає очевидним асиметричний розподіл з правим "хвостом". Це означає, що у вибірці є значна частка високих значень урожайності. Крім того, спостерігаються декілька піків, що потенційно свідчить про мультимодальність - наявність декількох кластерів з різними середніми показниками урожайності. Основна частина даних зосереджена в діапазоні від 20 до 50 ц/га. Найбільш ймовірні значення урожайності знаходяться в проміжку 25-35 ц/га, що узгоджується з медіаною (32.7) та середнім значенням (36.4), які отримано в результаті описової статистики. Дані характеризуються значним розкидом (від 0 до 77.2), що підтверджується високим значенням стандартного відхилення (14.3). Виявлені поодинокі низькі значення (близько 0 ц/га), які, можливо, є аномальними випадками. З результатів тестування, представлених у табл. 2.3, випливає, що розподіл даних не є нормальним, оскільки значення p-value за обома тестами менші за 0.05. Відтак, ми відхиляємо нульову гіпотезу та констатуємо, що розподіл урожайності суттєво відрізняється від нормального.

Окрім цього, було побудовано кореляційну матрицю на основі аналізованих даних, графічне відображення якої можна побачити на рис. 2.2.

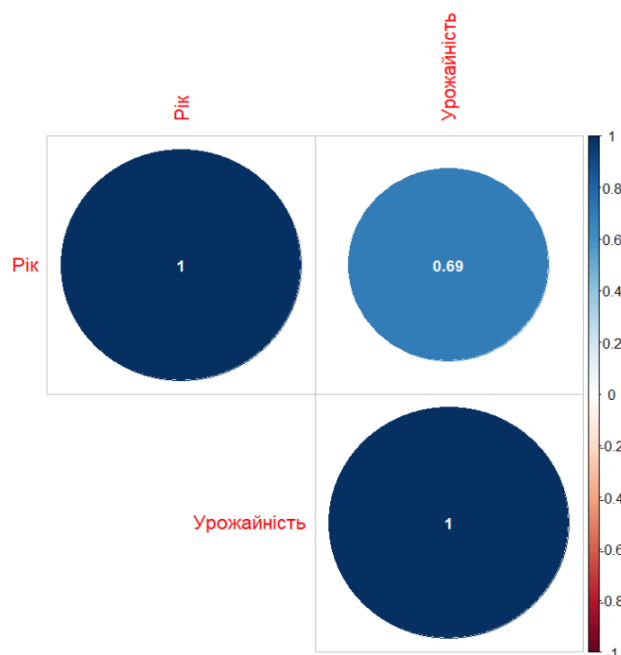


Рисунок 2.2 - Результати кореляційного аналізу

Джерело: розроблено автором

З отриманих результатів на рис. 2.2 , кореляція між показниками "Рік" та "Урожайність" є суттєвою, сягаючи 0,69, що свідчить про виражену тенденцію до покращення продуктивності у сільськогосподарському секторі.

Під час дослідження було проведено зіставний аналіз урожайності між роками 2000, 2023 та 2022, 2023. За отриманими показниками: середня урожайність у 2022 році склала 45.79; середня урожайність у 2023 році: 51.48. Різниця середніх значень становить приблизно 5.7 одиниць (урожайність 2023 року є вищою). t-статистика: -1.2711. Ступені свободи: 41.685. p-value: 0.2107. Оскільки p-value має значення 0.2107, тобто перевищує 0,05, це свідчить про відсутність статистично значущої різниці між середніми показниками урожайності у 2022 та 2023 роках, інакше кажучи, отримані дані не надають достатньо підстав стверджувати, що урожайність у 2023 році істотно відрізняється від показників 2022 року. При порівнянні урожайності за 2000 та 2023 роки: середня урожайність у 2000 році: 19.87, що значно менше, ніж у 2023 році. Різниця середніх значень: ≈ 31.6 одиниць (урожайність значно збільшилася у 2023 році). t-статистика: -8.4984. Ступені свободи: 24.962. p-value: $7.766e-09$ (приблизно 0.000000007766). Оскільки значення p-value надзвичайно мале, це вказує на статистично значущу різницю між середньою урожайністю у 2000 та 2023 роках, тому ми відкидаємо нульову гіпотезу та констатуємо, що урожайність у 2023 році значно вища. Таким чином, можна зробити висновок, що за 23 роки урожайність в Україні суттєво зросла.

Для порівняння середніх значень кількох груп і визначення, чи є статистично значущі відмінності між ними було використано ANOVA (Analysis of Variance) (див. табл. 2.3).

Таблиця 2.3 – Результат ANOVA

	Df	Sum sq	Mean Sq	F value	Pr(>F)
Область	23	28360	1233	7.612	<2e-16***
Residuals	552	89414	162		

Джерело: розроблено автором

Результати наведені у табл. 2.3 , презентують нам наступне: Df (Degree of Freedom, ступінь свободи) – для кожної з змінних. Для фактору "Область" це 23, а це значить, що було досліджено 24 групи (тобто, стільки областей). Для решти

варіацій (Residuals) цей показник дорівнює 552. Sum Sq (Сума квадратів) дає нам інформацію про те, як розподіляється варіація: між різними групами та в межах кожної з них. Для фактору "Область" сума квадратів дорівнює 28360 – що відображає різницю між областями. Для залишкової варіації це число становить 89414. Mean Sq (Середнє квадратичне значення) для "Область" становить 1233, а для відхилень, що лишились, – 162. F-значення (співвідношення середнього квадрату для "Область" до середнього квадрату для відхилень) дорівнює 7.612. Значне F-значення свідчить про те, що різниця між групами набагато більша, ніж різниця всередині груп. $P(>F)$ являє собою р-значення для перевірки гіпотези, що всі групи мають однакове середнє. Р-значення менше 0.001 (або $2e-16$) вказує на статистичну значущість результату. Отже, можна припустити, що середній урожай суттєво розрізняється між різними областями. Зважаючи на те, що р-значення значно менше 0.05, дозволено стверджувати про статистично важливі відмінності у урожайності між різними областями.

2.2. Моделювання урожайності традиційними економіко-математичними методами

У цьому підрозділі здійснено моделювання урожайності зернових культур із використанням традиційних економіко-математичних методів. Зокрема, було побудовано модель лінійної регресії для оцінки впливу часу та регіонального чинника на урожайність. Для візуалізації результатів застосовано графічні інструменти: коробчасту діаграму, графіки розсіювання, трендові графіки по кожній області, а також розсіювання з кольоровими точками. Це дозволило глибше проаналізувати просторово-часові особливості та варіативність урожайності в регіональному розрізі.

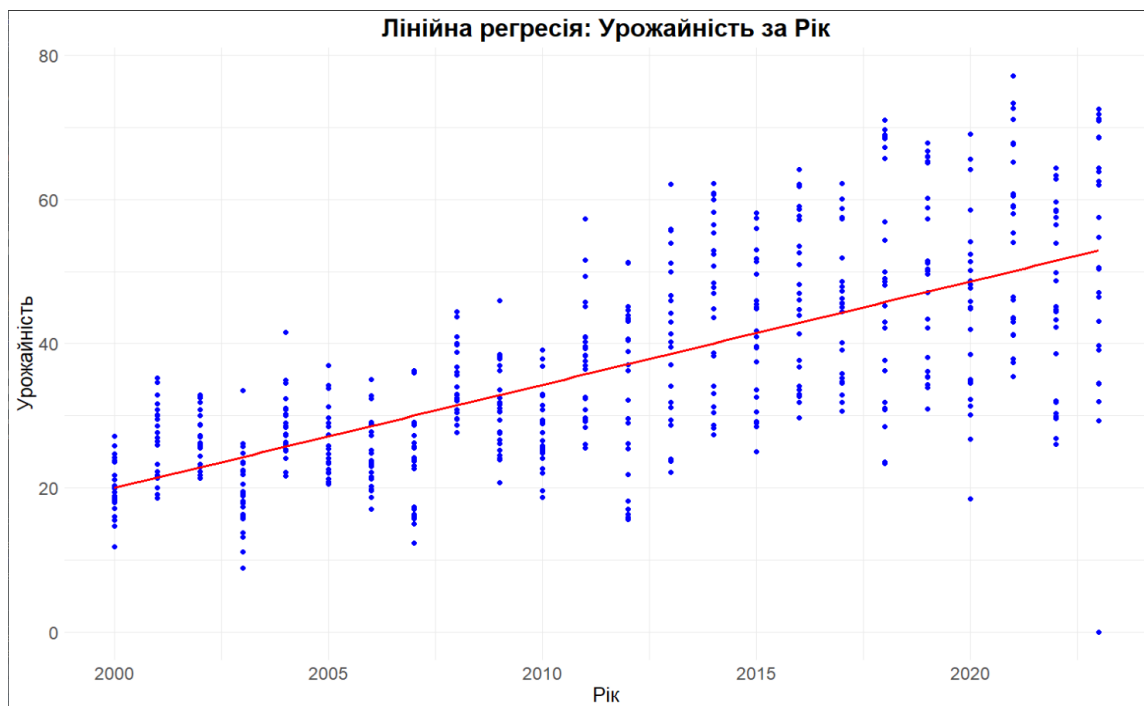


Рисунок 2.3 – Графік лінійної регресії

Джерело: розроблено автором

Таблиця 2.4 – Результати Лінійної регресії

Coefficients:	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-2.837e+03	9.341e+01	-30.373	< 2e-16 ***
Рік	1.432e+00	4.643e-02	30.851	< 2e-16 ***
ОбластьВолинська	-1.173e+01	2.227e+00	-5.268	1.99e-07 ***
ОбластьДніпропетровська	-1.474e+01	2.227e+00	-6.619	8.61e-11 ***
ОбластьДонецька	-1.668e+01	2.227e+00	-7.489	2.78e-13 ***
ОбластьЖитомирська	-6.629e+00	2.227e+00	-2.977	0.003037 **
ОбластьЗакарпатська	-6.638e+00	2.227e+00	-2.981	0.003001 **
ОбластьЗапорізька	-1.878e+01	2.227e+00	-8.434	2.95e-16 ***
ОбластьІвано-Франківська	-4.317e+00	2.227e+00	-1.939	0.053059 .
ОбластьКиївська	6.417e-01	2.227e+00	0.288	0.773324
ОбластьКіровоградська	-8.283e+00	2.227e+00	-3.720	0.000220 ***
ОбластьЛуганська	-1.852e+01	2.227e+00	-8.318	7.07e-16 ***
ОбластьЛьвівська	-6.533e+00	2.227e+00	-2.934	0.003484 **
ОбластьМиколаївська	-1.711e+01	2.227e+00	-7.683	7.14e-14 ***
ОбластьОдеська	-1.652e+01	2.227e+00	-7.419	4.48e-13 ***
ОбластьПолтавська	-5.958e-01	2.227e+00	-0.268	0.789117
ОбластьРівненська	-8.146e+00	2.227e+00	-3.658	0.000278 ***
ОбластьСумська	-1.363e+00	2.227e+00	-0.612	0.540858

Продовження таблиці 2.4

Область Тернопільська	-3.004e+00	2.227e+00	-1.349	0.177837
Область Харківська	-1.005e+01	2.227e+00	-4.513	7.80e-06 ***
Область Херсонська	-1.829e+01	2.227e+00	-8.215	1.52e-15 ***
Область Хмельницька	-1.500e-01	2.227e+00	-0.067	0.946316
Область Черкаська	3.542e+00	2.227e+00	1.591	0.112283
Область Чернівецька	-2.471e+00	2.227e+00	-1.110	0.267635
Область Чернігівська	-1.075e+00	2.227e+00	-0.483	0.629443

Джерело: розроблено автором

Створено математичні моделі, зокрема лінійну регресію. За результатами, наведеними у табл. 2.4, рік виявився суттєвим фактором з коефіцієнтом 1.432, що вказує на щорічне збільшення урожайності на 1.43 ц/га. Для року та переважної більшості областей (наприклад, Волинської, Дніпропетровської, Донецької, Луганської, Миколаївської тощо) р-значення нижче 0.05, що засвідчує статистично значущий вплив цих факторів на урожайність. Щодо окремих областей, таких як Київська ($p = 0.773$), Сумська ($p = 0.541$) та Хмельницька ($p = 0.946$), коефіцієнти не мають статистичної значущості, тобто не чинять суттєвого впливу на показник урожайності. Модель демонструє також добру пояснювальну здатність, оскільки Multiple R-squared: 0.7216 – приблизно 72% коливань урожайності пояснюються моделлю. Adjusted R-squared: 0.7095 – це скориговане значення, яке враховує кількість змінних у моделі. F-статистика: 59.52 з р-значенням меншим за $2.2e-16$ свідчить про статистичну значущість моделі в цілому, тобто варіація, яку модель пояснює, є значущою.

Для глибшого аналізу даних було побудовано коробчасту діаграму між змінною Урожайність та Область, графік розсіювання для кожної області, графік трендів та графік розсіяння з кольоровими точками для різних областей.

Результати побудови коробчастої діаграми наведені нижче на рис. 2.4.

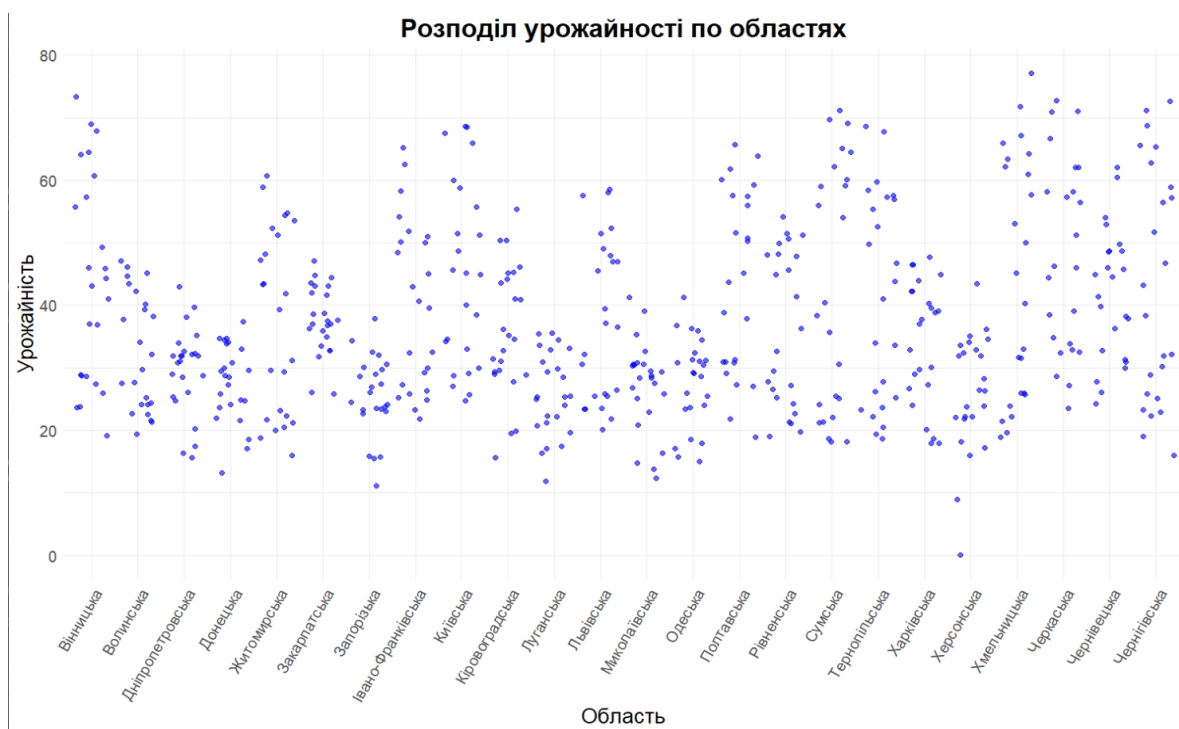


Рисунок 2.5 – Результат графіку розсіювання для різних областей України

Джерело: розроблено автором

На рис. 2.5 зображено розподіл урожайності по регіонах України через точкову діаграму. З графіка стає очевидним, що найбільші показники урожайності (вище 65) зафіксовані в таких областях: Черкаська, Хмельницька, Сумська, Тернопільська, Чернігівська та Вінницька. Це, ймовірно, зумовлено високою родючістю ґрунтів, сприятливими кліматичними умовами та застосуванням передових агротехнологій. Також, відносно високі показники (60–70) демонструють Київська, Полтавська, Івано-Франківська та Житомирська області. З іншого боку, найнижчі значення урожайності (менше 20) спостерігаються в Херсонській, Запорізькій, Миколаївській та Одеській областях, що, імовірно, пояснюється посушливим кліматом, меншою інтенсивністю агровиробництва або дією інших негативних факторів. Підсумовуючи, варто наголосити на значних міжрегіональних відмінностях у показниках урожайності, що підкреслює важливість регіонального підходу до розробки аграрної політики та розвитку сільського господарства.

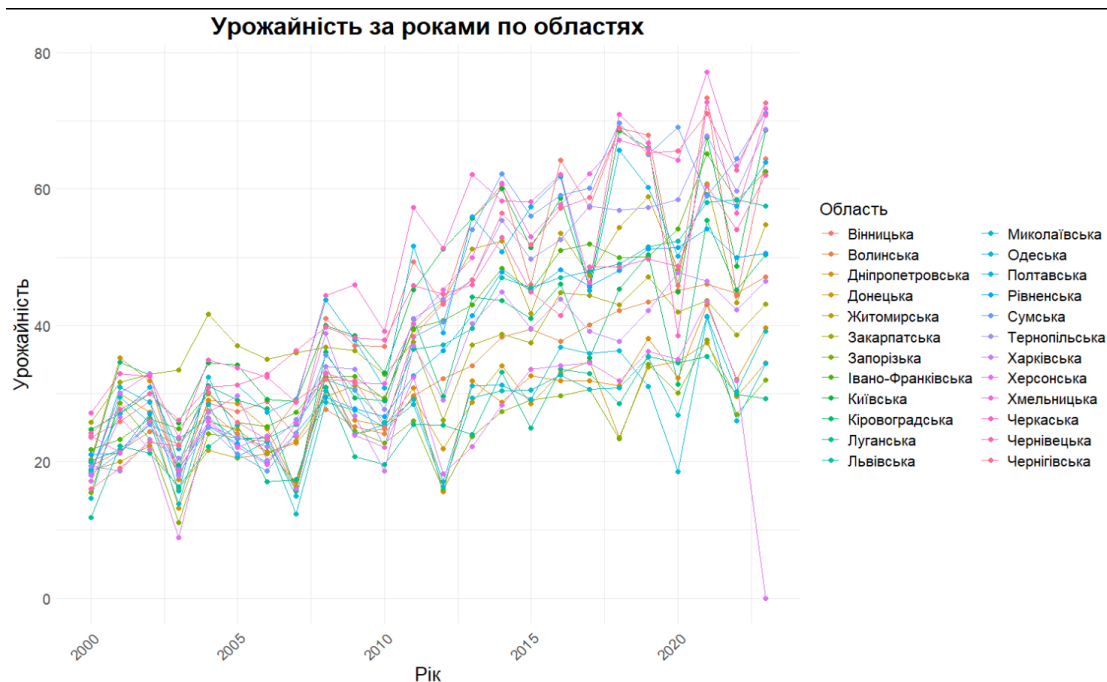


Рисунок 2.6 – Тренди урожайності за регіонами України

Джерело: розроблено автором

Згідно з графіком, представленим на рис. 2.6, спостерігається загальна тенденція до збільшення, хоча й з певними змінами. Зокрема, у проміжки часу 2003–2004, 2010–2011, 2015–2016 та 2020–2021 роки відзначаються значні спади та підйоми в урожайності, що може бути пов'язано з впливом економічних або погодних факторів.

Окремі регіони (зокрема, ті лінії, що розташовані вище) демонструють суттєво більшу середню урожайність порівняно з іншими, припускається, що це можуть бути західні або центральні області.



Рисунок 2.7 – Графік розсіяння з кольоровими точками для різних областей

Джерело: розроблено автором

На рис. 2.7 спостерігається явне збільшення показників урожайності в період з 2000 по 2023 роки. Особливо помітним є приріст після 2010 року. Цей тренд може вказувати на поступове впровадження передових сільськогосподарських технологій, вдосконалення аграрної політики та збільшення фінансування сільського господарства. Існує певна різниця між регіонами: хоча всі області демонструють позитивну динаміку, діапазон урожайності залишається суттєвим. У деякі роки варіації становлять від 20 до понад 70 ц/га, що свідчить про нерівномірний прогрес агросектору в різних областях. Найвищі значення урожайності (більше 70 ц/га) фіксуються переважно після 2015 року, що може бути наслідком сприятливіших кліматичних умов, підвищеної уваги до точного землеробства або оновлення сільськогосподарської техніки. Після 2020 року зафіксовано незначні коливання, які, ймовірно, зумовлені впливом зовнішніх факторів, зокрема економічною нестабільністю, пандемією COVID-19 або військовими діями, що могли частково вплинути на обсяги виробництва.

2.3. Застосування методів машинного навчання для прогнозування урожайності зернових культур по областях України

Для глибокого аналізу даних урожайності зернових культур по областях України було використано сучасні методи машинного навчання.

Насамперед, щоб збудувати моделі, дані поділили на навчальну й тестову вибірки, дотримуючись пропорції 80:20. Згодом, спираючись на ці вибірки, було створено лінійну регресію, дерева рішень, випадкові ліси, метод векторів, XGBoost (екстремальне градієнтне бустування) та нейронні мережі. Крім того, були побудовані ETS модель, ARIMA, а найкраща модель визначалася за значенням RMSE кожної.

Таблиця 2.5 – Значення RMSE кожної моделі

Модель	RMSE
Лінійна регресія	7.836789
Дерева рішень	9.403207
Випадкові ліси	6.338285
SVM	7.26723
XGBoost	5.217825
Нейронні мережі однокрокова (nnetar)	1.729982
	18.452
Нейронні мережі багат шарова (nnfor)	9.428916
	16.89594
ETS	25.48221
ARIMA	23.66923

Джерело: розроблено автором

З аналізу даних, що містяться в табл. 2.5, очевидно, що модель з найкращими показниками RMSE – це одноступенева нейронна мережа. Вона продемонструвала найнижчий рівень RMSE на навчальному наборі даних серед всіх розроблених моделей.

Під час процесу створення моделей деякі з них супроводжувалися графічною візуалізацією, зокрема, це стосувалося дерев рішень та нейронних мереж.



Рисунок 2.8 – Графічне зображення дерев рішень

Джерело: розроблено автором

На рис. 2.8 зображено дерево рішень, корінь якого (верхня вершина) демонструє середнє значення урожайності 36, спираючись на 463 спостереження ($n=463$, що охоплює 100% даних). Початкове розгалуження відбувається на основі критерію "Рік < 2013". Ліва гілка (Рік < 2013) характеризується середньою урожайністю 28 та часткою спостережень 53% ($n=246$). Далі відбувається розподіл за критерієм "Рік < 2008". Лівий вузол (Рік < 2008) демонструє середню урожайність 24, що базується на 146 спостереженнях (32%), і це найнижчий показник середньої урожайності у дереві. Правий вузол ($2008 \leq \text{Рік} < 2013$) має середню урожайність 33, спираючись на 100 спостережень (22%). Права гілка (Рік ≥ 2013) представляє середню урожайність 46, спираючись на 217 спостережень (47%), що є найвищим середнім показником урожайності. Відтак, спостерігається тенденція зростання урожайності з часом, що підтверджується поступовим збільшенням середніх значень у пізніші роки: до 2008 року — низька урожайність (24), у період 2008-2012 — середня (33), після 2013 року — висока (46). Ключовою

змінною, що визначає прогноз урожайності у цій моделі, є рік. Це може бути наслідком: впровадження сучасних агротехнологій, удосконалення техніки, змін клімату, реалізації державних програм підтримки сільського господарства тощо.

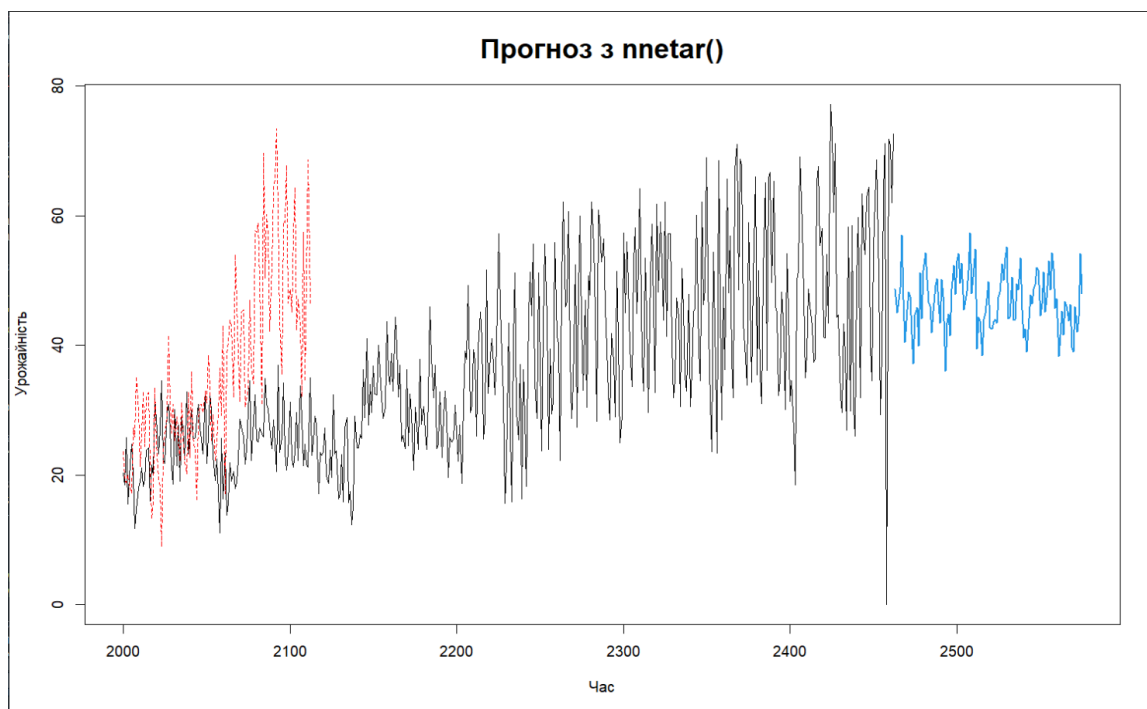


Рисунок 2.9 – Графік прогнозу урожайності з використанням нейронної мережі авторегресії (NNAR)

Джерело: розроблено автором

На рис. 2.9 зображено графік передбачення урожайності, сформований за допомогою нейронної мережі авторегресії, сконструйованої з використанням `nnetar`, використовуючи задані параметри. На цьому графіку, чорним кольором виділено основну частину часового ряду. Ця лінія використовується як навчальний набір. Вона охоплює більшість часового проміжку, демонструючи поступове збільшення урожайності з часом, і включає в себе помітну варіативність (можливо, сезонну чи кліматичну), а також один різкий аномальний спад майже до нуля. Це може бути пов'язано з помилкою збору даних, посухою чи іншими екстремальними умовами. Червона пунктирна лінія — це тестова вибірка, яка використовується для перевірки якості моделі. Вона показує досить високу волатильність, зокрема в ділянці навколо точки 2100, де урожайність коливається в діапазоні 20–60. Синя лінія є прогнозом на майбутнє, модель передбачає урожайність у межах 40–50 одиниць, відносно стабільність, без різких злетів чи падінь та легку хвильову структуру, тобто збереження коливань, але в більш згладженій формі.

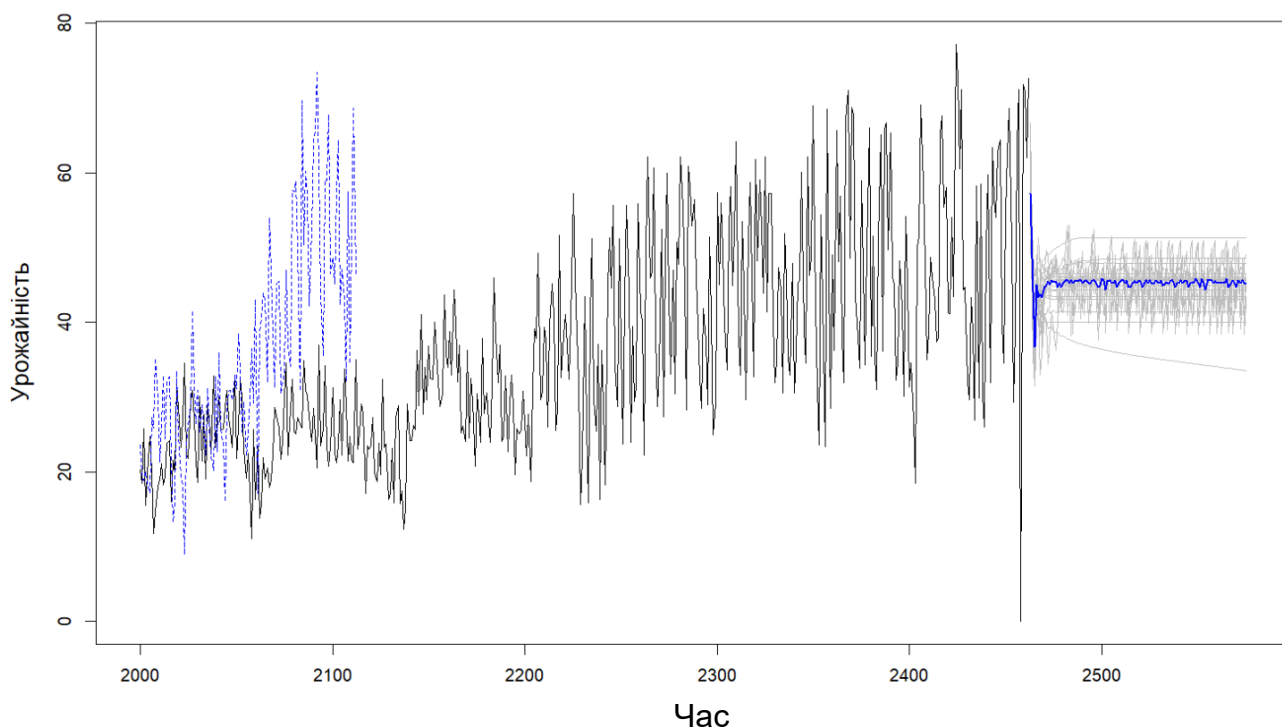


Рисунок 2.10 – Прогноз урожайності з використанням моделі MLP (Multi-Layer Персерptron)

Джерело: розроблено автором

На рис. 2.10 чорна лінія — це реальні (історичні) величини урожайності. Як і в попередньому графіку, вони демонструють загальний висхідний тренд з часом, помітну варіативність, особливо в крайні роки та один аномальний спад до майже нуля, ймовірно спричинений екстремальними умовами або помилкою в даних, синя лінія — це передбачені значення урожайності на майбутнє, одержані за допомогою моделі MLP. Вона вказує на стабілізацію урожайності в межах приблизно 40 одиниць, враховує легкі коливання, але не має різких піків або спадів, має вужчий діапазон змін порівняно з фактичними величинами. Сіра зона навколо синьої лінії є довірчими інтервалами прогнозу (ймовірно, 80% чи 95%), які вказують на можливий діапазон змін урожайності і чим ширша сіра зона, тим менш певна модель у точності прогнозу, також на цьому графіку довірчий інтервал є помірно широким, але не надто — що свідчить про середню впевненість моделі.

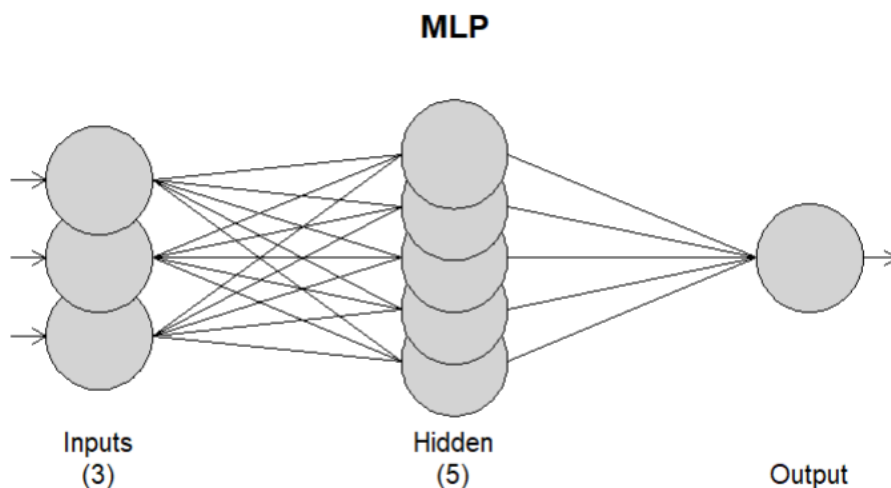


Рисунок 2.11 – Архітектура багатошарового перцептрона (MLP — Multi-Layer Perceptron)

Джерело: розроблено автором

На рис. 2.11 на початку вхідні дані надходять у мережу, потому вони зважуються, проходять крізь прихований шар, далі передаються у вихідний шар, де формується підсумкове значення прогнозу. Також під час навчання мережа налаштовує ваги сполук між нейронами, щоб мінімізувати похибку прогнозу. З отриманих результатів бачимо, що Inputs (вхідний шар) має 3 нейрони, що означає, що модель бере три вхідні змінні (наприклад: рік, температура, кількість опадів) і кожен нейрон передає своє значення усім нейронам прихованого шару. Hidden (прихований шар) складається з 5 нейронів, кожен нейрон отримує зважені суми всіх вхідних даних, проходить через активаційну функцію (наприклад, ReLU або sigmoid). Output (вихідний шар) це один нейрон, що виробляє прогнозну величину (наприклад, прогноз урожайності на певний рік).

Також для детального аналізу було графічно відтворено порівняння прогнозів з фактичними значеннями.



Рисунок 2.12 – Порівняння прогнозованих та фактичних значень для лінійної регресії

Джерело: розроблено автором

На рис. 2.12 кожна точка відповідає одній спостережуваній парі “факт-передбачення”. З результатів можливо мовити, що загальна кореляція позитивна бо є виразна лінійна залежність між фактичними й прогнозованими значеннями. Це засвідчує, що модель діє загалом добре, адже передбачення співпадають з реальною урожайністю. Наявні також певні відхилення від ідеалу, хоча багато точок розташовані поряд з червоною лінією, помітно також розкид — особливо в середньому діапазоні фактичних значень (20–40), це свідчить про те, що модель має похибки: в окремих випадках вона або недооцінює, або перевищує урожайність. Помітно ще те, що модель краще справляється з прогнозами при високих значеннях урожайності, де точки ближчі до червоної лінії, при низьких значеннях (10–25) спостерігається більше шуму — точки сильніше розкидані.



Рисунок 2.13 – Порівняння прогнозованих та фактичних значень для дерев рішень

Джерело: розроблено автором

На рис. 2.13 за здобутими наслідками можливо мовити, що модель дерева ухваленень дає "східчасті" передбачення, адже всі показники прогнозів (ось Y) обмежуються кількома рівнями: ~25, ~32, ~35, це характерний ефект для дерева ухваленень — модель "рівнями" ухвалює рішення в певних діапазонах, без поступових переходів. Наявний також високий рівень дискретності, не дивлячись на широкий розподіл дійсних значень, передбачення зосереджуються тільки у 3–4 фіксованих величинах, це свідчить, що модель занадто проста чи перенавчена на окремих вузлах дерева. Є недооцінка та переоцінка, приміром, для фактичних величин у діапазоні 35–60 модель часто занижує прогноз — більшість прогнозів зупиняються на рівні ~47, що гірше, аніж лінійна регресія, для низьких фактичних величин (менше 25) — модель трохи краще справляється, проте знову ж таки надає тільки обмежену кількість значень (~24–25). Також наявне віддалення від лінії ідеалу, що вказує на високу похибку, адже чимало точок значно віддалені від червоної діагоналі, особливо при середніх і високих фактичних показниках урожайності.

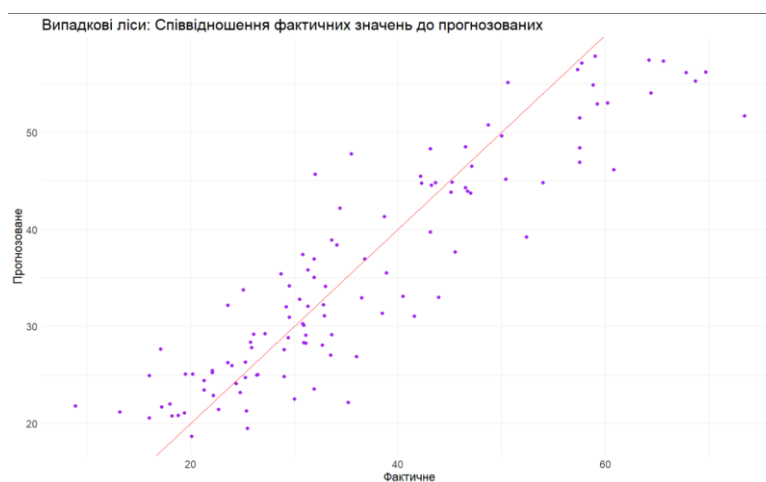


Рисунок 2.14 – Порівняння прогнозованих та фактичних значень для випадкових лісів

Джерело: розроблено автором

На рис. 2.14 з одержаних результатів видно, що точки розміщені доволі щільно навколо діагоналі, що вказує на добру точність моделі. На відміну від дерева рішень передбачення більш плавні та без "східчастого" ефекту. Також модель досить адекватно відтворює як низькі, так і високі величини урожайності. Помітно ще те, що прогнози трохи "розмиваються" навколо лінії ідеалу. З результатів видно, що наявна висока кореляція між фактичними та

передбачуваними значеннями, є гнучкість моделі, яка віддзеркалює як низькі, так і високі величини та є менше систематичних зсувів адже немає стабільної переоцінки або недооцінки по всьому діапазону. Видно ще й певні недоліки, приміром, у нижній частині (до 30) є кілька точок, де модель трохи переоцінює урожайність, розкид точок зростає в середньому діапазоні (~35–45), тобто тут модель менш точна, а для окремих точок спостерігається відхилення понад 10 одиниць, що може бути критичним у практичному застосуванні (наприклад, для планування врожаю).

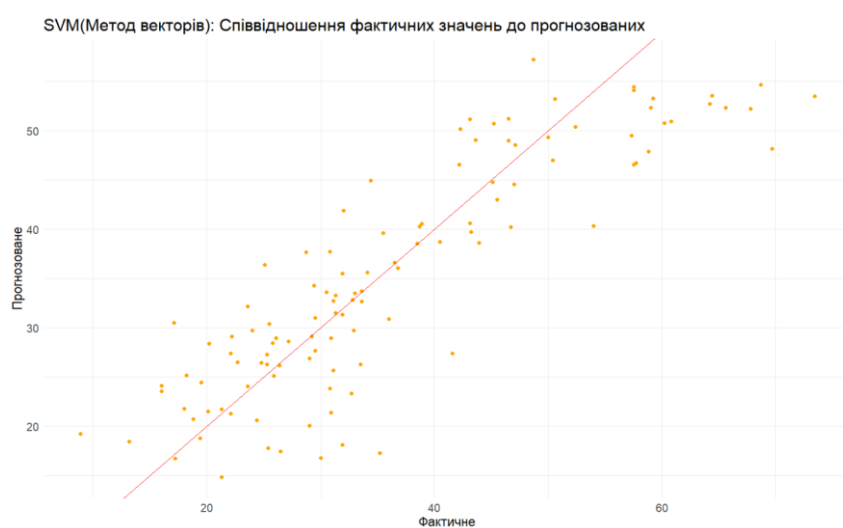


Рисунок 2.15 – Порівняння прогнозованих та фактичних значень для методу векторів

Джерело: розроблено автором

На рис. 2.15 за одержаними результатами можна мовити, що точки переважно розміщені поблизу червоної лінії, що говорить про гарну точність моделі SVM. Наявна також лінійна залежність між фактичними та прогнозованими показниками, що свідчить про адекватність моделі. Проте в отриманих результатах є деякі вади, наприклад, розкиди точок по обидва боки від лінії, особливо в діапазоні Actual = 20–40, це може вказувати на неадекватне узгодження моделі з даними в цьому сегменті. Також у верхньому правому кутку деякі точки відстають від лінії, тобто модель недооцінює високу урожайність.



Рисунок 2.16 – Порівняння прогнозованих та фактичних значень для XGBoost (екстримальний градієнтний бустинг)

Джерело: розроблено автором

На рис. 2.16, виходячи з отриманих даних, можна стверджувати, що переважна кількість точок згрупована біля лінії, вказуючи на високий рівень точності моделі. У верхньому діапазоні (Фактичні показники > 40) модель адекватно відтворює високі значення урожайності, що є надзвичайно важливим. Крім того, спостерігається менше розкиду точок, ніж на попередньому графіку, збудованому за допомогою SVM, особливо в середньому діапазоні, але все ж деякі точки виходять за межі червоної лінії, тобто прогнози містять певні похибки.

2.4. Порівняння точності моделей прогнозування

Для прогнозування урожайності на наступні 4 роки було побудовано моделі ARIMA та ETS. За показниками AIC та RMSE було визначено, яка з моделей для кожної з областей є кращою. Нижче наведені графічні зображення та табличні результати (див. додаток Б) результатів прогнозованих моделей для кожної області України.

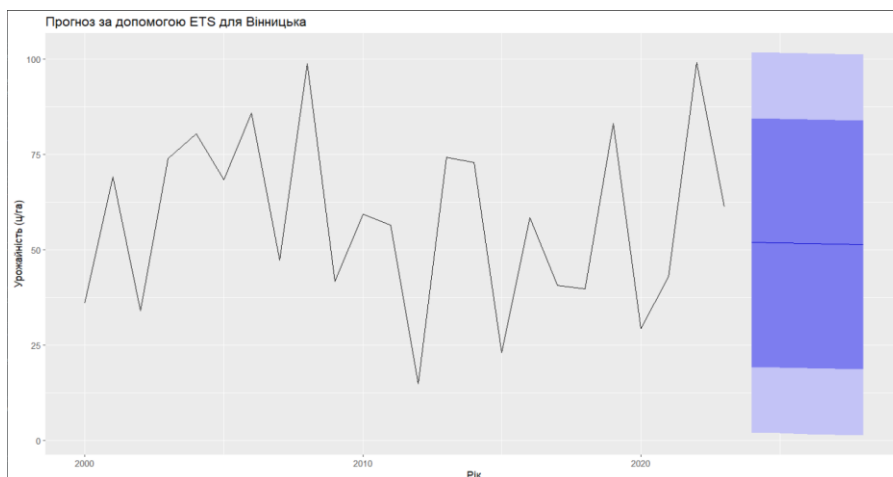


Рисунок 2.17 – Модель ETS для Вінницької області

Джерело: розроблено автором

Графік, зображений на рис. 2.17 показує, що історичні дані мають значну мінливість адже урожайність варіювалася від приблизно 20 до понад 100 центнерів з гектара. Спостерігається тенденція до зростання з певними коливаннями, що, можливо, мало вплив на вибір ETS-моделі з адитивним трендом. Прогноз демонструє відносну стабільність урожайності на рівні приблизно 70–80 ц/га протягом наступних 5 років, хоча й із досить широким інтервалом невизначеності (зокрема, в межах 95%).

ETS-модель забезпечує прогноз, що враховує як тренд у даних, так і невизначеність майбутніх значень. Однак значна варіативність у минулому зумовлює широкі довірчі інтервали, що слід враховувати під час планування чи ухвалення агрополітичних рішень.

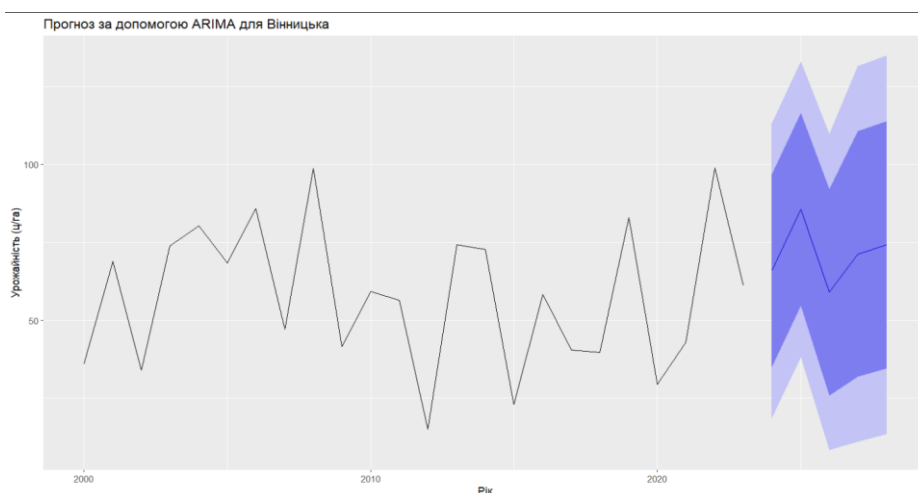


Рисунок 2.18 – Модель ARIMA для Вінницької області

Джерело: розроблено автором

Графік на рис. 2.18, ARIMA-модель показує зростання урожайності на майбутні 4 роки на відміну від моделі ETS, яка показує більш стабільний прогноз. Широкі довірчі інтервали (особливо після 2026 року) вказують на високу невизначеність прогнозу.

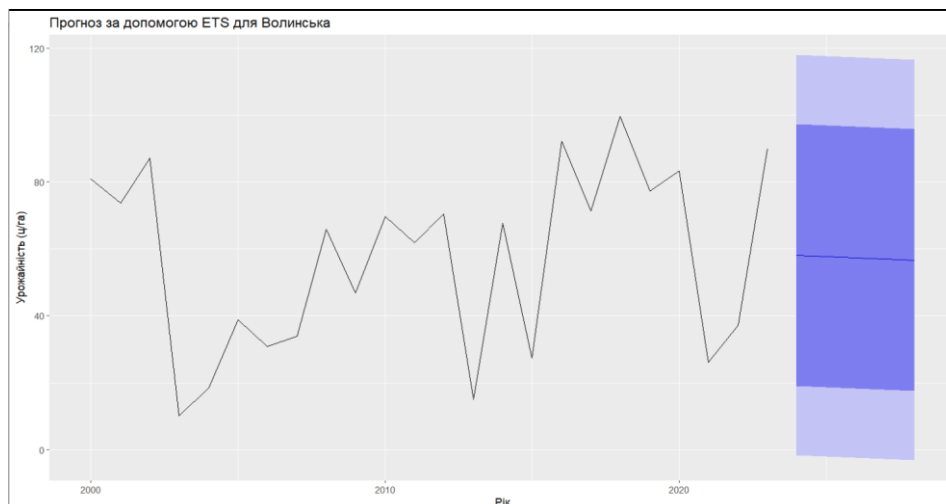


Рисунок 2.19 – Модель ETS для Волинської області

Джерело: розроблено автором

Графік, зображений на рис. 2.19 показує помітні щорічні коливання урожайності на Волині за попередні періоди. Прогноз, сформований за допомогою моделі ETS, вказує на тенденцію до стабілізації або незначного зменшення урожайності в найближчому майбутньому. Проте чим далі у часі відбувається прогнозування, тим ширше розширюються інтервали, що вказує на збільшення невизначеності.

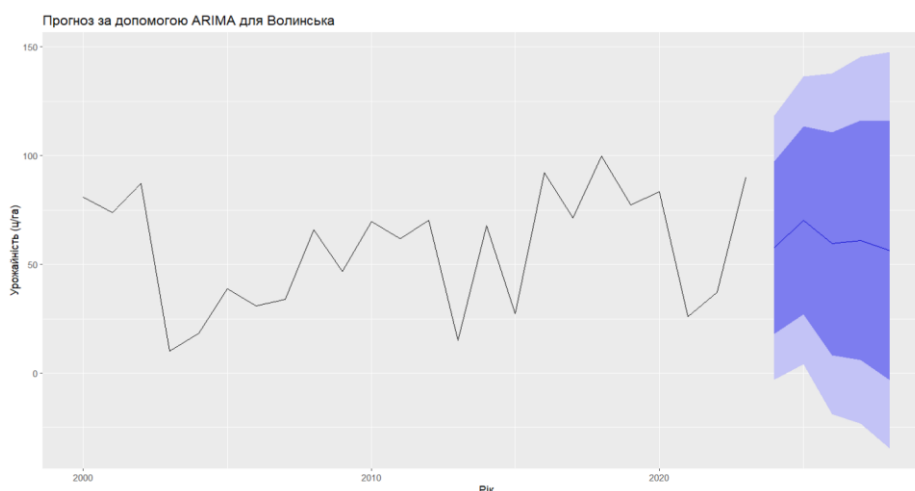


Рисунок 2.20 – Модель ARIMA для Волинської області

Джерело: розроблено автором

Графік, зображений на рис. 2.20 створений на основі моделі ARIMA, аналогічно відображає високу мінливість урожайності в Волинській області у минулому періоді. На відміну від прогнозу, згенерованого ETS-моделлю, яка демонструвала стабільність або слабке зростання, прогноз ARIMA для Волині передбачає спочатку невелике підвищення показників урожайності, а потім поступове зниження показників в найближчому майбутньому. Інтервали прогнозу, як і очікувалося для моделей часових рядів, розширюються пропорційно до віддаленості періоду прогнозування. Порівняння прогнозів, отриманих з використанням моделей ETS та ARIMA для Волинської області, виявляє суттєві відмінності у передбачуваній траєкторії розвитку урожайності.

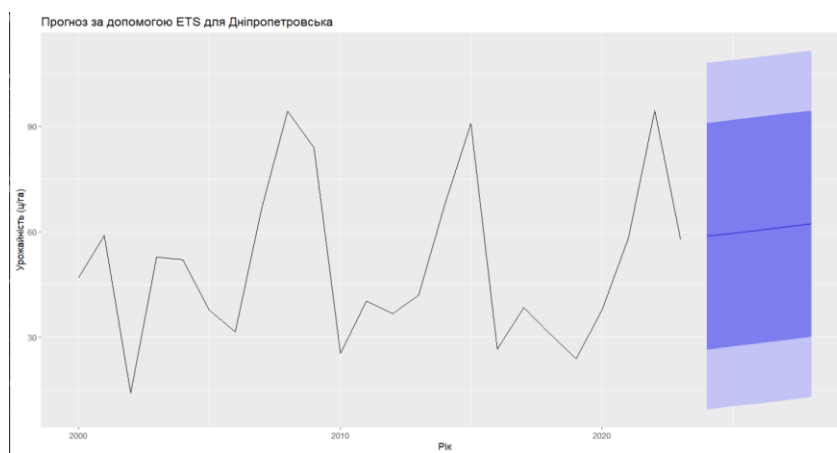


Рисунок 2.21 – Модель ETS для Дніпропетровської області

Джерело: розроблено автором

Графік на рис. 2.21 демонструє відчутну мінливість урожайності на Дніпропетровщині в минулому періоді. Прогнозування, зроблене на основі ETS-моделі, свідчить про ймовірне усталення або поступове збільшення урожайності у майбутньому. Також прогноз містить в собі невизначеність, що відображається довірчими проміжками, котрі розширюються зі збільшенням терміну прогнозування.

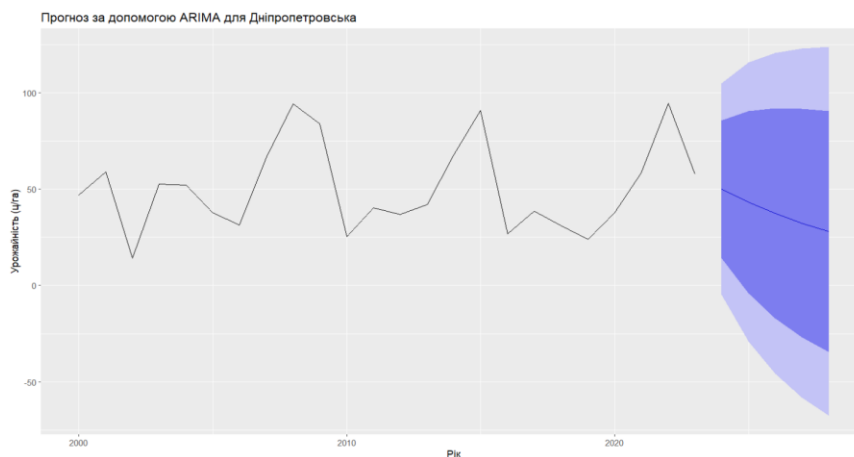


Рисунок 2.22 – Модель ARIMA для Дніпропетровської області

Джерело: розроблено автором

Графік, на рис. 2.22 побудований на основі ARIMA-моделі, теж демонструє велику нестабільність показників урожайності на Дніпропетровщині у минулому. Однак, на відміну від прогнозу ETS, який вказував на стабільність чи незначне збільшення, прогноз ARIMA для Дніпропетровської області передбачає різке зменшення урожайності у майбутньому. Інтервали передбачення також розширюються зі збільшенням періоду прогнозу, що є звичною особливістю для прогнозів часових рядів. У порівнянні прогнозів, отриманих за допомогою ETS та ARIMA моделей для Дніпропетровської області, спостерігаємо, що різні методи прогнозування можуть призводити до різних висновків щодо майбутньої тенденції. У даному випадку, ARIMA-модель вказує на дещо більш песимістичний розвиток подій, ніж модель ETS.

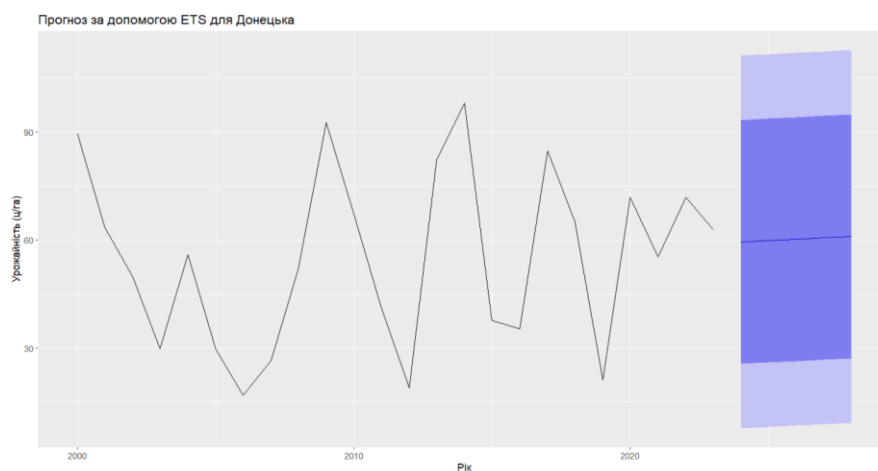


Рисунок 2.23 – Модель ETS для Донецької області

Джерело: розроблено автором

Графік, на рис. 2.23 демонструє помітну мінливість урожайності у Донецькій області за попередні періоди. Прогноз, розрахований за допомогою

моделі ETS, передбачає невелике зростання урожайності в майбутньому. Помітно також і те, що прогноз містить певну невизначеність, яка відображається довірчими інтервалами, що є досить широкими.

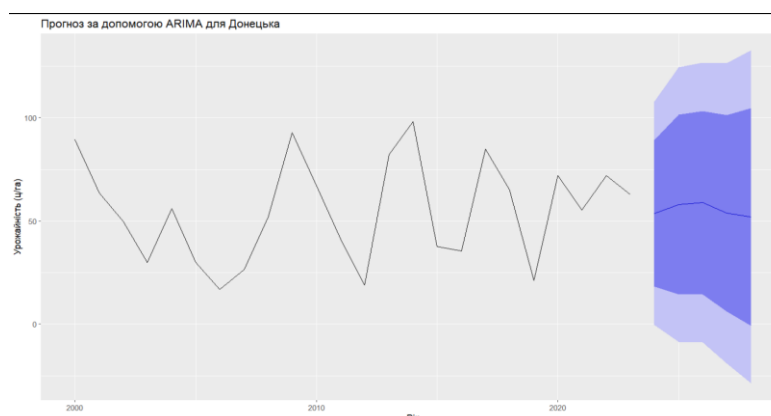


Рисунок 2.24 – Модель ARIMA для Донецької області

Джерело: розроблено автором

Графік, зображений на рис. 2.24 побудований на базі моделі ARIMA, також демонструє помітну мінливість показників урожайності в Донецькій області у минулому періоді. Однак, у порівнянні з прогнозом ETS, котрий показував невелике зростання, передбачення ARIMA для Донеччини показує поступове зменшення урожайності в майбутньому. Прогнозовані інтервали на цьому графіку є доволі широкими, що може вказувати на значну невизначеність прогнозу, особливо враховуючи теперішню непросту ситуацію в регіоні. Співставляючи прогнози, одержані за допомогою моделей ETS та ARIMA для Донецької області, спостерігається розходження у передбачуваній майбутній тенденції.

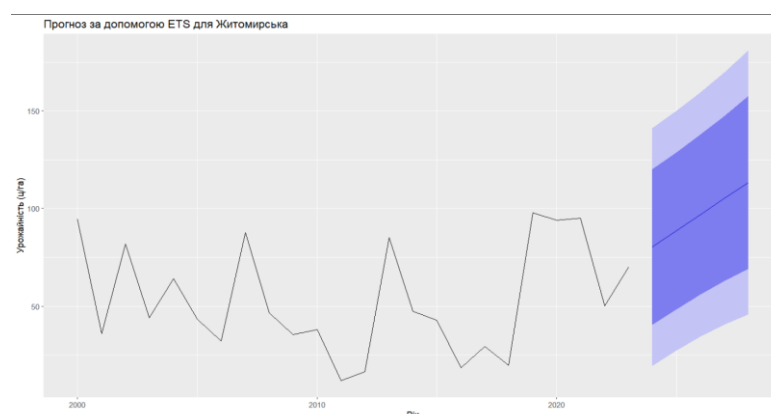


Рисунок 2.25 – Модель ETS для Житомирської області

Джерело: розроблено автором

Графік на рис. 2.25 демонструє відчутну мінливість урожайності на Житомирщині в минулому періоді. Прогнозна модель ETS вказує на чіткий тренд

до збільшення урожайності в майбутньому. Також помітно ймовірну невизначеність прогнозу, котра візуалізується розширенням довірчих інтервалів із віддаленням в часі.

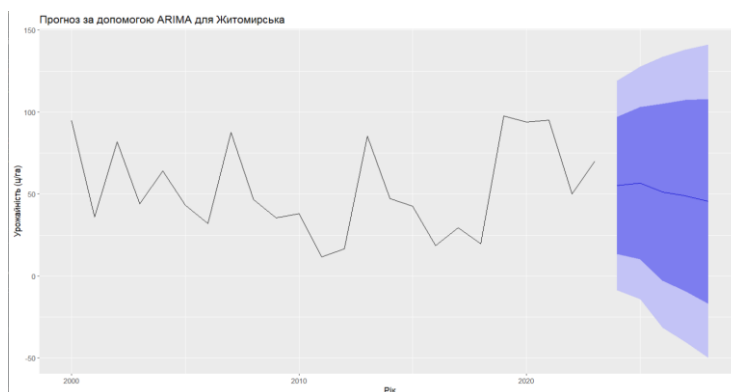


Рисунок 2.26 – Модель ARIMA для Житомирської області

Джерело: розроблено автором

Графік, на рис. 2.26 побудований на основі моделі ARIMA, аналогічно відображає значну нестабільність урожайності на Житомирщині у минулому. Втім, на відміну від прогнозу ETS, котрий передбачав різке зростання урожайності в майбутньому, прогноз ARIMA для Житомирської області демонструє виражену тенденцію до спадання урожайності у майбутньому періоді. Інтервали передбачення також розширюються з віддаленістю у часі. Порівнюючи передбачення, одержані за допомогою моделей ETS та ARIMA для Житомирщини, спостерігається різниця в передбачуваній майбутній траєкторії. ARIMA прогнозує більш негативний сценарій по урожайності, тоді як ETS передбачає зростання показників.

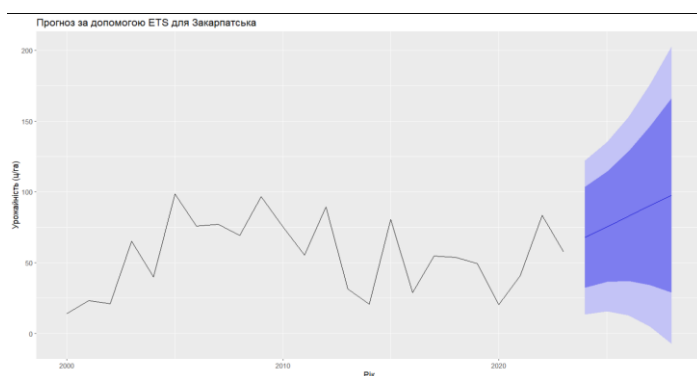


Рисунок 2.27 – Модель ETS для Закарпатської області

Джерело: розроблено автором

Графік на рис. 2.27 демонструє суттєві коливання урожайності на Закарпатті у минулі періоди. Прогноз, що спирається на модель ETS, передбачає різке

зростання показників. Помітно, що прогноз має рівень невизначеності, котра відображається у довірчих інтервалах, які розширюються з віддаленням у часі.

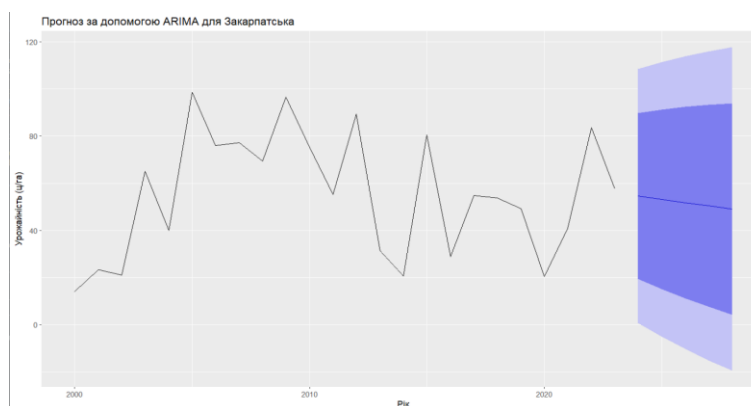


Рисунок 2.28 – Модель ARIMA для Закарпатської області

Джерело: розроблено автором

Даний графік, зображений на рис. 2.28, побудований на основі моделі ARIMA, демонструє коливання урожайності в Закарпатській області, що вже спостерігалася раніше. Але, на відміну від прогнозу ETS, який вказував на різке збільшення показників, ARIMA для Закарпаття пророкує поступове зменшення урожайності в майбутньому. При зіставленні прогнозів, що впливають з моделей ETS та ARIMA для Закарпатської області, бачимо розбіжність у передбачуваному тренді майбутнього.

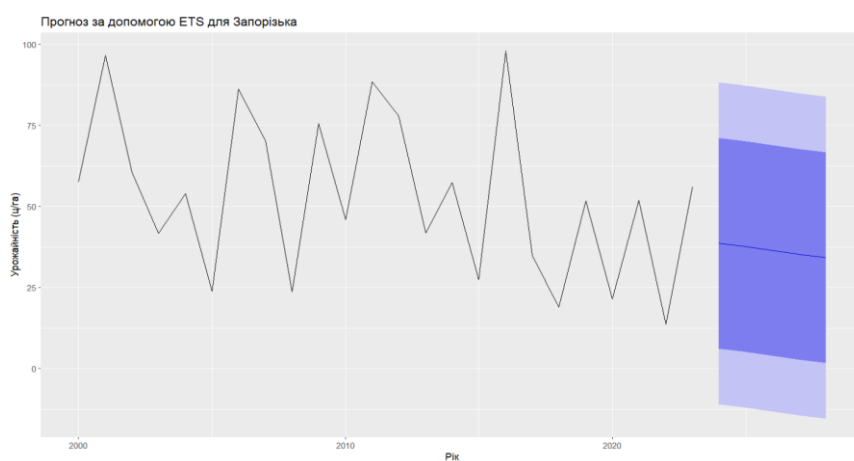


Рисунок 2.29 – Модель ETS для Запорізької області

Джерело: розроблено автором

Графік на рис. 2.29 демонструє істотну мінливість урожайності сільськогосподарських культур Запорізької області у минулому періоді. Прогнозна модель ETS передбачає тенденцію до незначного зменшення показників урожайності найближчим часом. Також прогноз містить невизначеність, яка ілюструється довірчими інтервалами, котрі збільшуються з часом прогнозу.

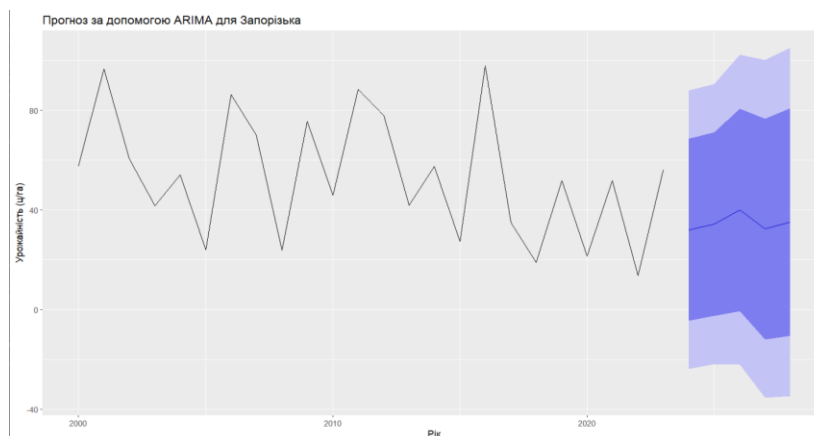


Рисунок 2.30 – Модель ARIMA для Запорізької області

Джерело: розроблено автором

На графіку зображеному на рис. 2.30 помітно, що в майбутньому показники урожайності в Запорізькій області поступово збільшуються порівняно з графіком на рисунку 2.29, на якому прогнозується поступове зменшення урожайності в майбутньому. Помітно також, що наявна певна невизначеність в прогнозі адже на графіку можна побачити широкі довірчі інтервали, що засвідчують це.

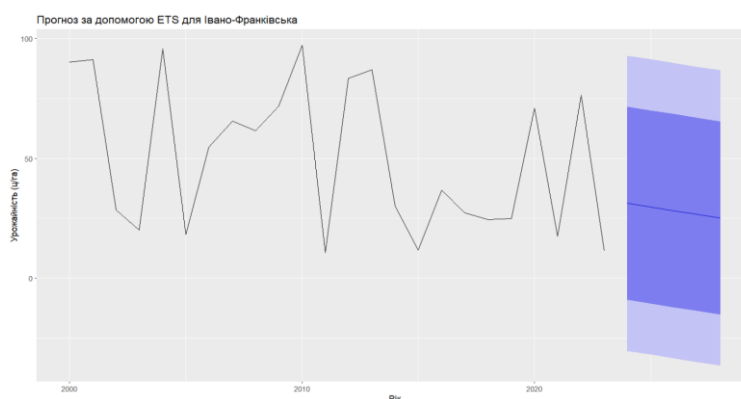


Рисунок 2.31 – Модель ETS для Івано-Франківської області

Джерело: розроблено автором

Графік, зображений на рис. 2.31, демонструє істотні коливання урожайності на Івано-Франківщині в минулому періоді. Прогноз, складений на основі моделі ETS, прогнозує виражену тенденцію до зниження показників урожайності у майбутньому. Також даний прогноз містить певну ступінь невизначеності, котра виражається розширенням довірчих інтервалів по мірі віддалення від теперішнього часу.

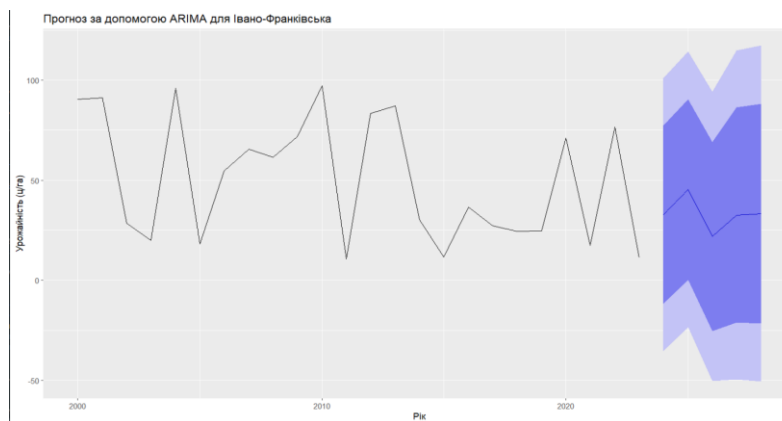


Рисунок 2.32 – Модель ARIMA для Івано-Франківської області

Джерело: розроблено автором

Даний графік, зображений на рис. 2.32 та побудований на основі моделі ARIMA, також показує суттєві коливання урожайності на Івано-Франківщині в минулому. Однак, на відміну від передбачення ETS, яке віщувало помітне зменшення, прогноз ARIMA демонструє більш усталену картину з незначним трендом до спаду. Інтервали передбачення також розширюються з плином часу, що є звичайним явищем для прогнозів часових рядів. Зіставляючи передбачення, здобуті за допомогою моделей ETS та ARIMA для Івано-Франківської області, спостерігаємо певну різницю у прогнозованому майбутньому тренді. ARIMA надає менш похмурий прогноз, аніж ETS.

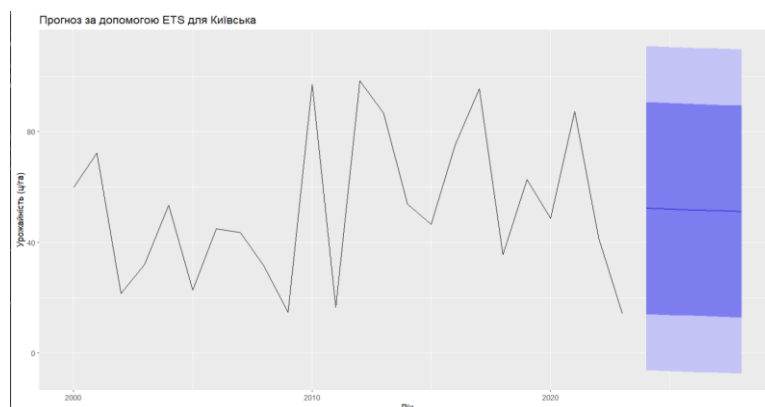


Рисунок 2.33 – Модель ETS для Київської області

Джерело: розроблено автором

Графік на рис. 2.33 показує, що урожайність має різко коливальний характер, особливо у проміжку 2010–2020 років. Присутні як високі піки (приблизно до 95 ц/га), так і глибокі спади — це ознака високої волатильності. Прогноз демонструє стабільність показників або ж невелике спадання, але довірчі

інтервали досить широкі, що вказує на високу невизначеність щодо майбутніх значень.

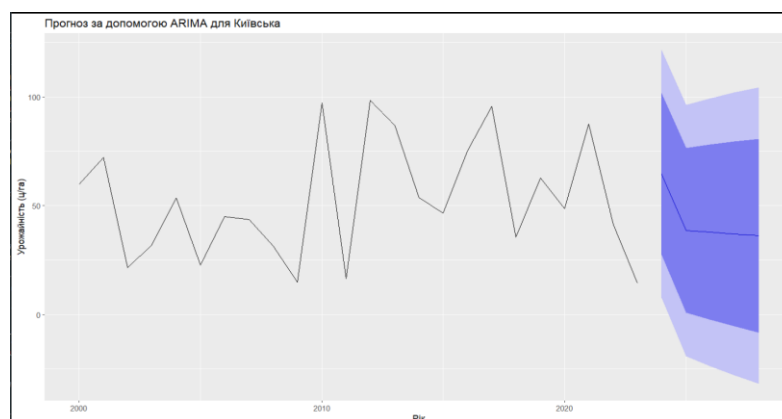


Рисунок 2.34 – Модель ARIMA для Київської області

Джерело: розроблено автором

Графік, зображений на рис. 2.34 та побудований на основі моделі ARIMA, демонструє відчутні коливання урожайності на Київщині у минулому. Прогнозна лінія показує, що урожайність у найближчі роки, ймовірно, залишиться на рівні десь 50–60 ц/га з тенденцією до незначного зниження. Інтервал невизначеності досить широкий, що свідчить про значний рівень непередбачуваності майбутніх значень урожайності.

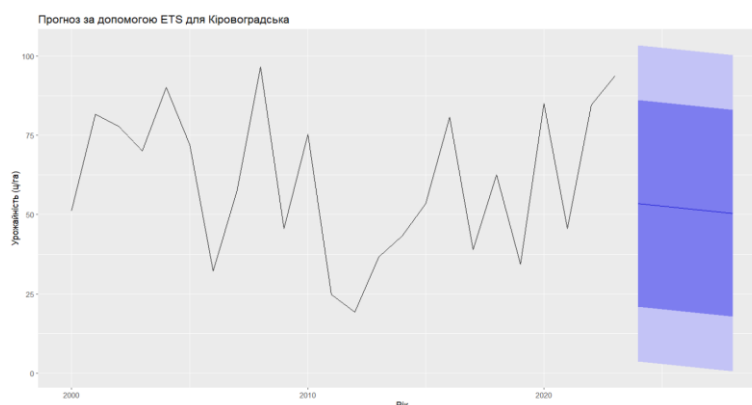


Рисунок 2.35 – Модель ETS для Кіровоградської області

Джерело: розроблено автором

Графік, зображений на рис. 2.35 демонструє по історичних даних значні коливання з року в рік. Є періоди різкого зростання (наприклад, близько 2010 та 2020 років), а також суттєві падіння (особливо помітне падіння після 2010 року). В останні роки спостерігається зростання урожайності, яка досягає максимуму на момент побудови прогнозу. Прогноз вказує на можливе зниження урожайності в

майбутньому, проте слід враховувати значну невизначеність, що відображена у прогностичних інтервалах.

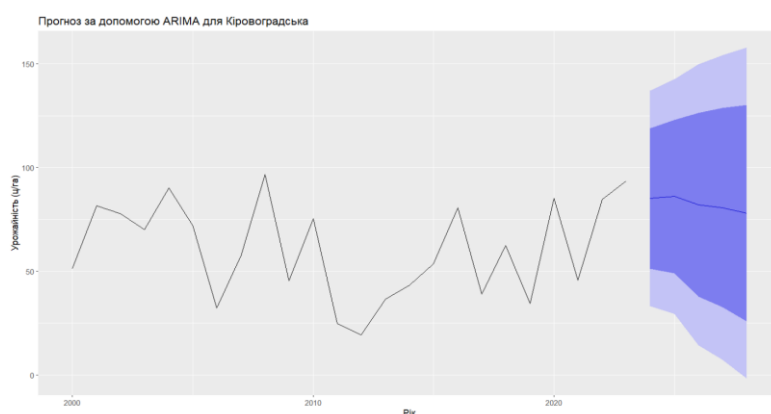


Рисунок 2.36 – Модель ARIMA для Кіровоградської області

Джерело: розроблено автором

Графік зображений на рис. 2.36 ілюструє історичні зміни урожайності в Кіровоградській області та прогноз на майбутнє, отриманий за допомогою моделі ARIMA. На відміну від попереднього прогнозу, сформованого з використанням ETS, цей передбачує більш стабільну урожайність у майбутньому, з ледь помітною тенденцією до спаду. Проте, як і раніше, необхідно брати до уваги інтервали невизначеності, котрі демонструють можливий коридор майбутніх значень.

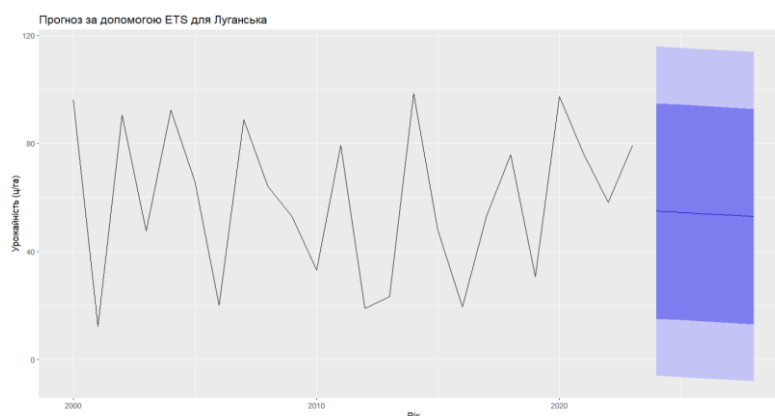


Рисунок 2.37 – Модель ETS для Луганської області

Джерело: розроблено автором

Графік, зображений на рис. 2.37 ілюструє суттєві коливання урожайності на Луганщині у минулому періоді. Прогнозна модель ETS вказує на можливість стабілізації або незначного спадання обсягів врожаю у майбутньому. Помітно також, що прогноз містить елемент невизначеності, який демонструється довірчими інтервалами, що розширюються пропорційно до віддаленості часового горизонту.

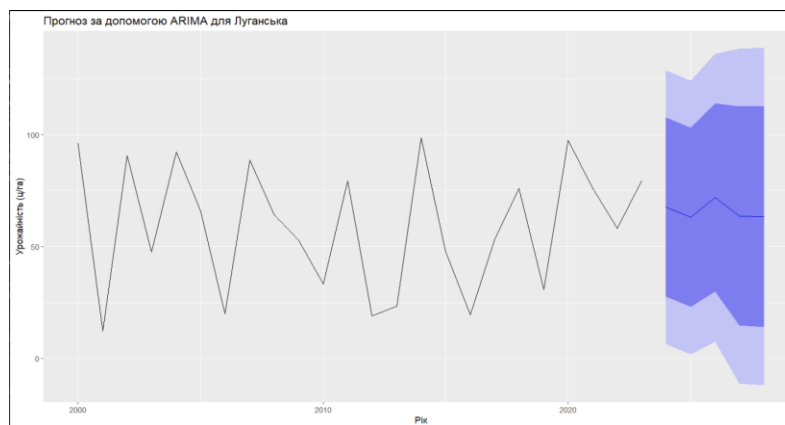


Рисунок 2.38 – Модель ARIMA для Луганської області

Джерело: розроблено автором

На графіку, що представлений на рис. 2.38, по історичним даним як і на попередньому графіку помітні значні коливання показників урожайності. Прогноз побудований за допомогою моделі ARIMA показує невелике зростання та в кінці невелике спадання показників на відміну від прогнозу побудованому за допомогою ETS.

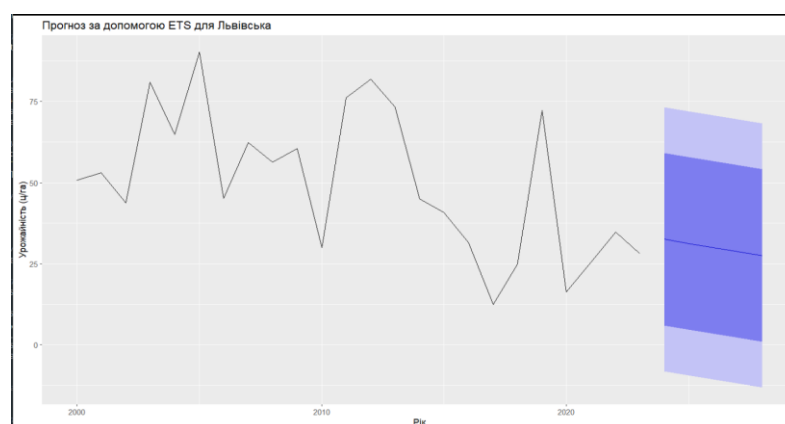


Рисунок 2.39 – Модель ETS для Львівської області

Джерело: розроблено автором

На графіку, зображеному на рис. 2.39 історичні дані про урожайність у Львівській області показують значні коливання урожайності з року в рік, але в останні роки спостерігається тенденція до спадання. Прогнозовані значення урожайності на майбутній період за моделлю ETS демонструють продовження тенденції до спадання урожайності. Також прогноз має певну невизначеність, яка відображається довірчими інтервалами, що розширюються з віддаленням у майбутнє.

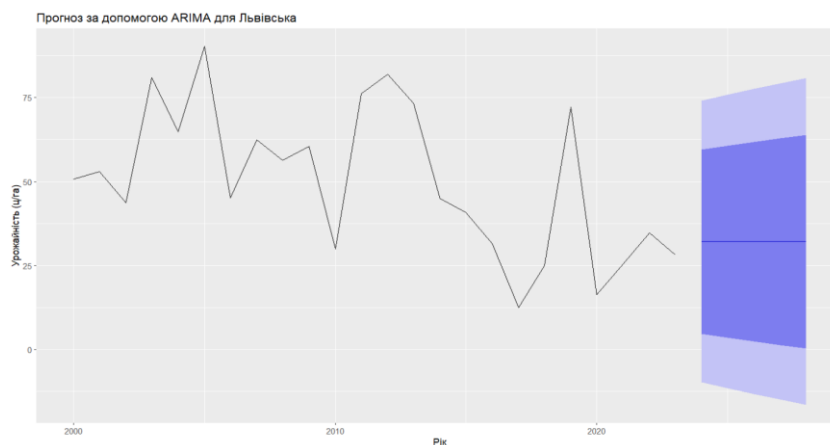


Рисунок 2.40 – Модель ARIMA для Львівської області

Джерело: розроблено автором

На графіку, зображеному на рис. 2.40 помітно, що історичні дані як і на попередньому графіку коливаються. Прогноз вказує на відносно стабільну урожайність з невеликою тенденцією до зниження. Порівнюючи з попереднім прогнозом для Львівської області, виконаним за допомогою моделі ETS, обидві моделі вказують на можливе зниження урожайності, хоча ETS показував більш виражену тенденцію до спаду.

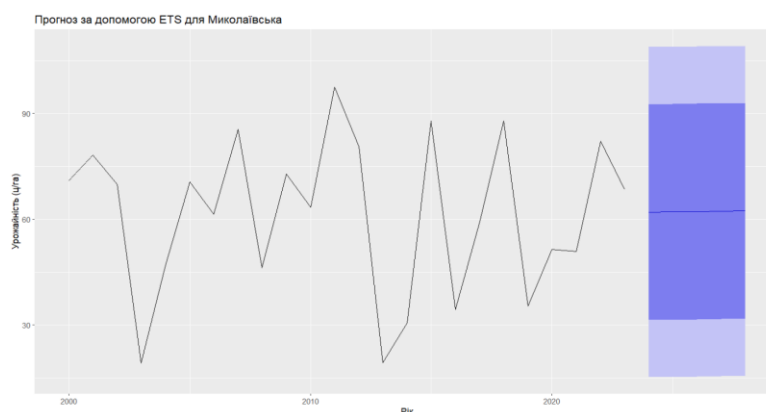


Рисунок 2.41 – Модель ETS для Миколаївської області

Джерело: розроблено автором

На графіку зображеному на рис. 2.41 історичні дані коливаються з року в рік. Прогнозовані значення урожайності на майбутній період за моделлю ETS демонструють певну стабілізацію або незначне зростання урожайності. Також прогноз має певну невизначеність, яка відображається довірчими інтервалами, що розширюються з віддаленням у майбутнє.

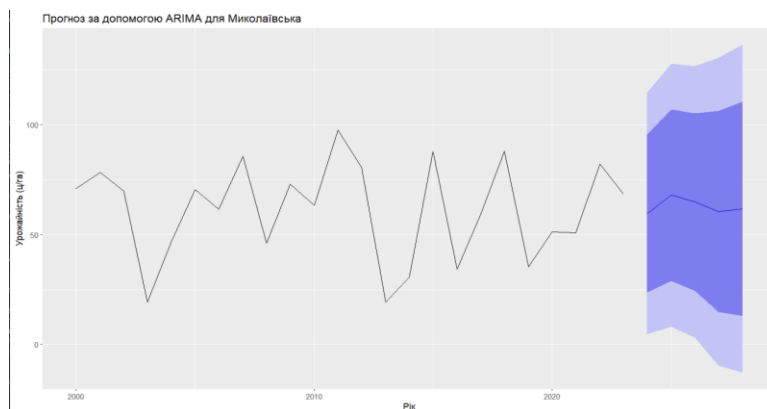


Рисунок 2.42 – Модель ARIMA для Миколаївської області

Джерело: розроблено автором

На графіку, що представлено на рис. 2.42, прогнозовані значення урожайності на майбутній період, отримані за моделлю ARIMA, також засвідчують зростання показники як і у прогнозі ETS для цієї ж території, котрий демонстрував стійкість чи незначне збільшення показників. Проміжки прогнозу тут значні, особливо у майбутньому, та досягають від'ємних показників урожайності, що є невідповідним у контексті сільськогосподарської діяльності.

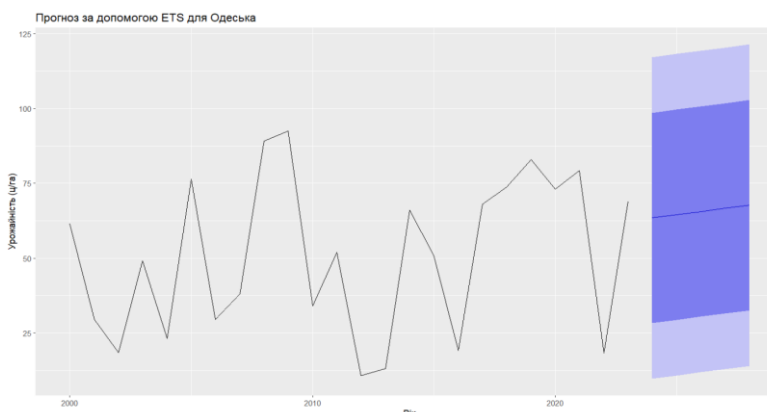


Рисунок 2.43 – Модель ETS для Одеської області

Джерело: розроблено автором

На графіку зображеному на рис. 2.43 помітно, що історичні дані коливаються з року в рік, проте в останні роки є велика тенденція до зростання. Прогнозовані значення урожайності на майбутній період за моделлю ETS демонструють продовження зростання, проте не такого різкого. Також прогноз має певну невизначеність, що підтверджується широкими довірчими інтервалами.

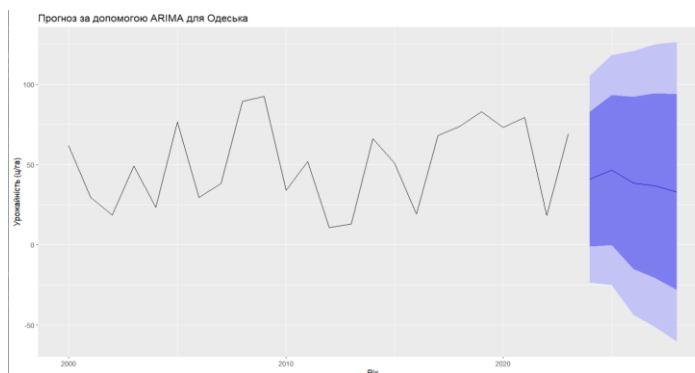


Рисунок 2.44 – Модель ARIMA для Одеської області

Джерело: розроблено автором

На графіку зображеному на рис. 2.44 прогнозовані значення за моделлю ARIMA демонструють невелике підвищення в одному з років у майбутньому та подальше спадання урожайності. Проміжки прогнозу є значними та сягають від'ємних значень, що не відповідає сільськогосподарській діяльності.

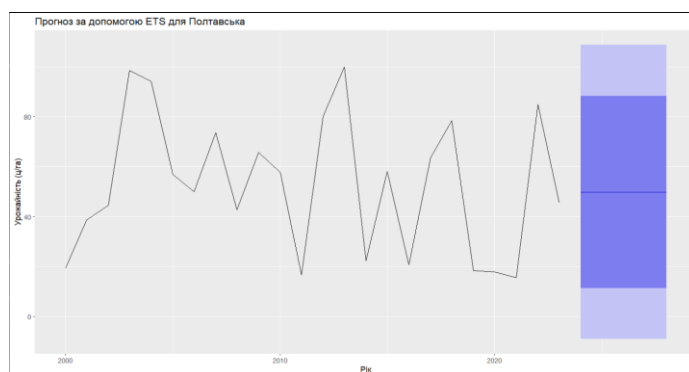


Рисунок 2.45 – Модель ETS для Полтавської області

Джерело: розроблено автором

Графік зображений на рис. 2.45 демонструє відчутні коливання урожайності на Полтавщині в минулому періоді. Прогнозування, здійснене на основі моделі ETS, передбачає стабільність урожайності або ж невелике зростання показників в найближчому майбутньому.

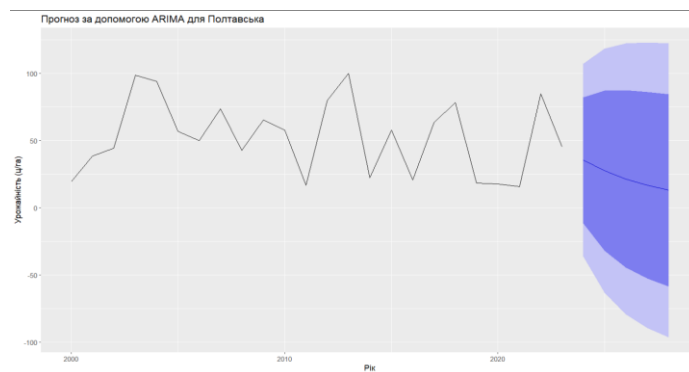


Рисунок 2.46 – Модель ARIMA для Полтавської області

Джерело: розроблено автором

На графіку, зображеному на рис. 2.46, прогнозовані дані вказують на значне зниження урожайності в майбутньому, можливо навіть до від'ємних показників, що, звісно, неможливо для реальних врожаїв. Таке може означати, що модель ARIMA в даному випадку не є найоптимальнішою для передбачення майбутнього. Якщо порівнювати з попередніми прогнозами для Полтавщини, зробленими з використанням моделі ETS, котрі демонстрували незначне збільшення, результати ARIMA значно різняться.

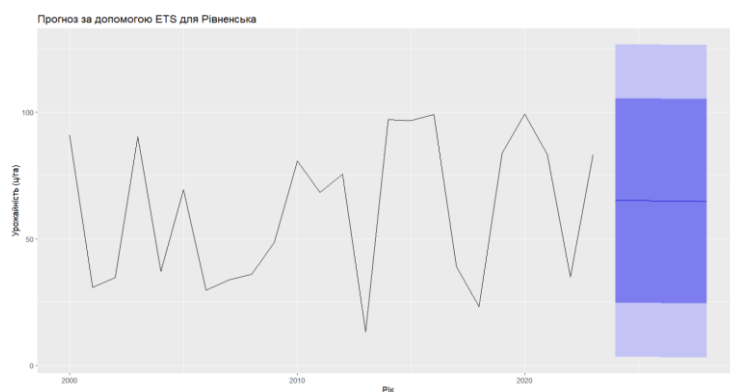


Рисунок 2.47 – Модель ETS для Рівненської області

Джерело: розроблено автором

Графік, зображений на рис. 2.47, демонструє історичну мінливість урожайності в Рівненській області та прогноз на майбутнє, розрахований за допомогою моделі ETS. Прогноз вказує на можливе невелике зростання урожайності в майбутньому. Прогноз також має певні невизначеності, про що вказують широкі довірчі інтервали.

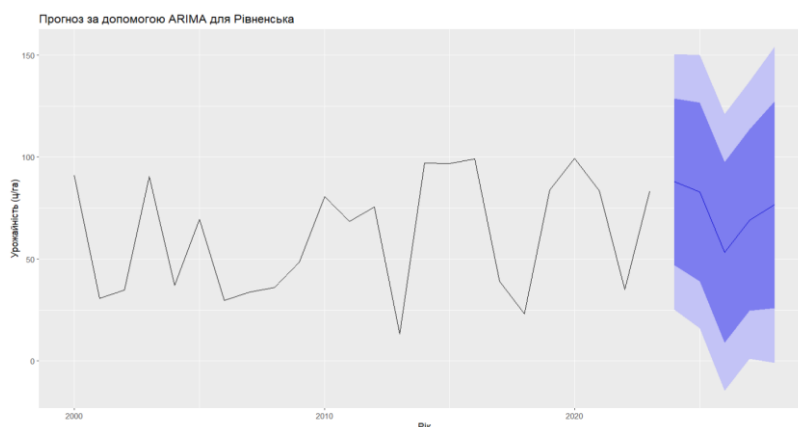


Рисунок 2.48 – Модель ARIMA для Рівненської області

Джерело: розроблено автором

На графіку, зображеному на рис. 2.48, прогноз вказує на відносно стабільний рівень урожайності з тенденцією до зростання. Порівнюючи з попереднім

прогнозом для Рівненської області, виконаним за допомогою моделі ETS, обидві моделі показують на зростання показників, хоча ARIMA демонструє дещо більшу волатильність у прогнозному періоді ніж ETS.

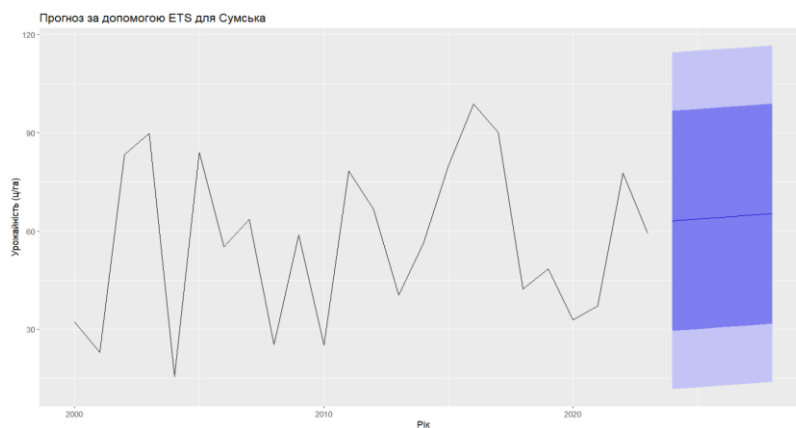


Рисунок 2.49 – Модель ETS для Сумської області

Джерело: розроблено автором

Графік, зображений на рис. 2.49, демонструє історичну мінливість урожайності в Сумській області та прогноз на майбутнє, розрахований за допомогою моделі ETS. Прогноз вказує на можливе зростання урожайності в майбутньому. Довірчі інтервали є досить широкими, що говорить про наявність невизначеності.

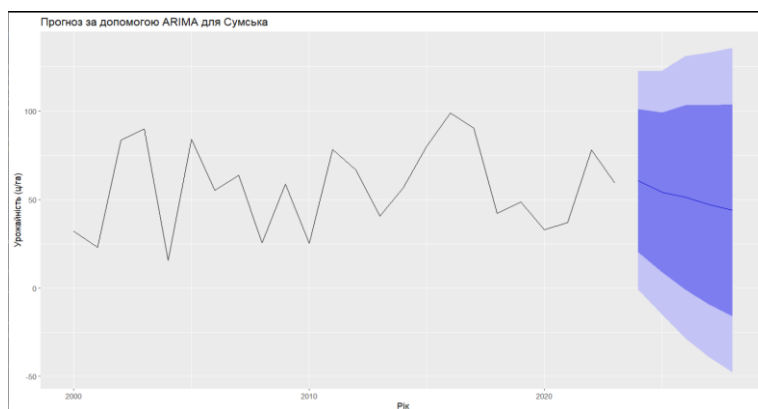


Рисунок 2.50 – Модель ARIMA для Сумської області

Джерело: розроблено автором

На графіку, зображеному на рис. 2.50, прогноз вказує на помітне різке зниження урожайності в майбутньому. Порівнюючи з попереднім прогнозом для Сумської області, виконаним за допомогою моделі ETS, який показував невелике зростання, результати ARIMA є протилежними.

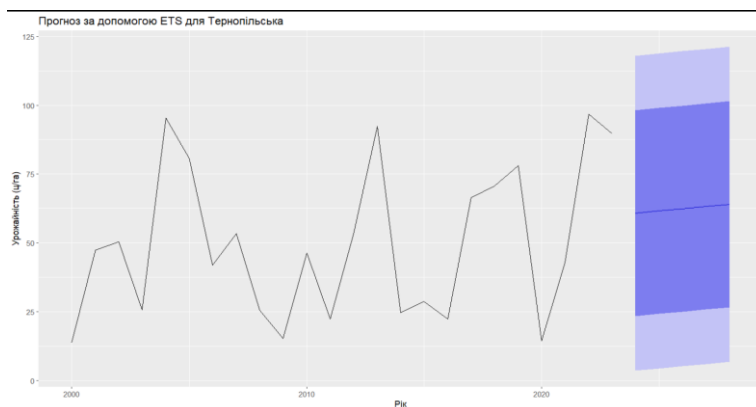


Рисунок 2.51 – Модель ETS для Тернопільської області

Джерело: розроблено автором

На графіку, зображеному на рис. 2.51, урожайність у Тернопільській області демонструє значні міжрічні коливання. Є періоди різкого зростання (наприклад, близько 2005 та 2020 років), а також суттєві падіння (наприклад, близько 2016 та 2019 років). Загалом, тренд не є чітко зростаючим чи спадним - урожайність змінюється хвилеподібно. Прогноз побудований за допомогою моделі ETS показує, що показники урожайності в майбутніх роках будуть поступово зростати.

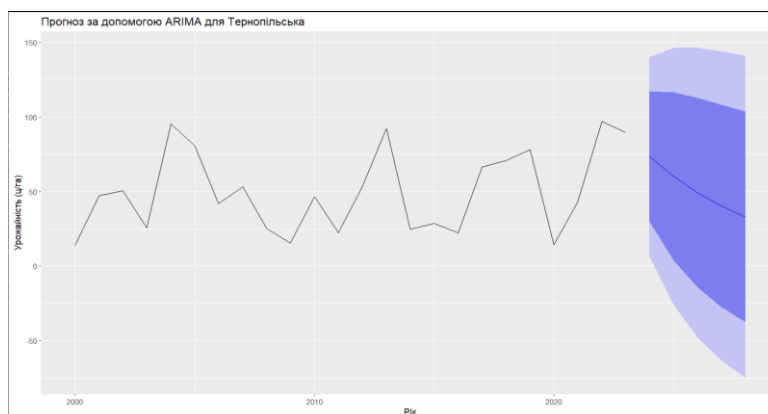


Рисунок 2.52 – Модель ARIMA для Тернопільської області

Джерело: розроблено автором

На графіку, зображеному на рис. 2.52 прогнозовані показники урожайності показують різку тенденцію до зниження урожайності у найближчі роки. Інтервал невизначеності прогнозу досить широкий: чим далі в майбутнє, тим більша невизначеність. Нижня межа прогнозу навіть виходить за нуль, що свідчить про дуже високу невпевненість моделі щодо майбутніх значень. Порівнюючи з попереднім прогнозом для Тернопільської області, виконаним за допомогою моделі ETS, який показував невелике зростання, результати ARIMA є протилежними.

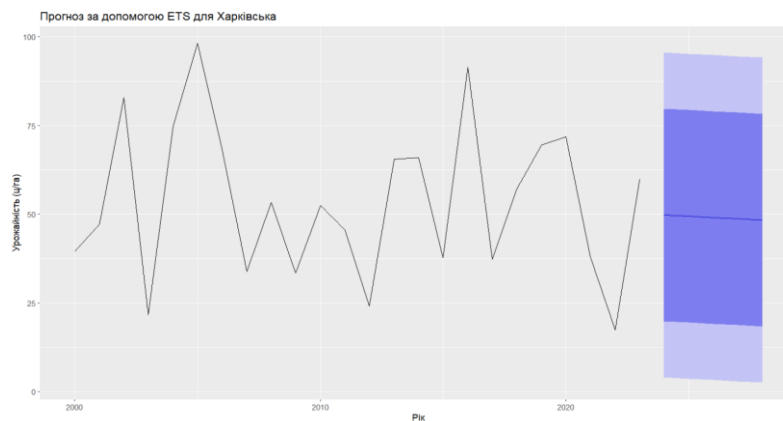


Рисунок 2.53 – Модель ETS для Харківської області

Джерело: розроблено автором

На графіку, зображеному на рис. 2.53, урожайність у Харківській області має значні коливання з року в рік. Спостерігаються як різкі підйоми (наприклад, близько 2005 та 2015 років), так і суттєві падіння (особливо помітно близько 2002, 2012 та 2019 років). Загалом, тренд не є стабільним - урожайність змінюється хвилеподібно. Прогнозовані значення побудовані за допомогою моделі ETS демонструють повільне спадання показникі у майбутньому. Також на графіку помітні широкі довірчі інтервали, що доводять про наявність невизначеності.

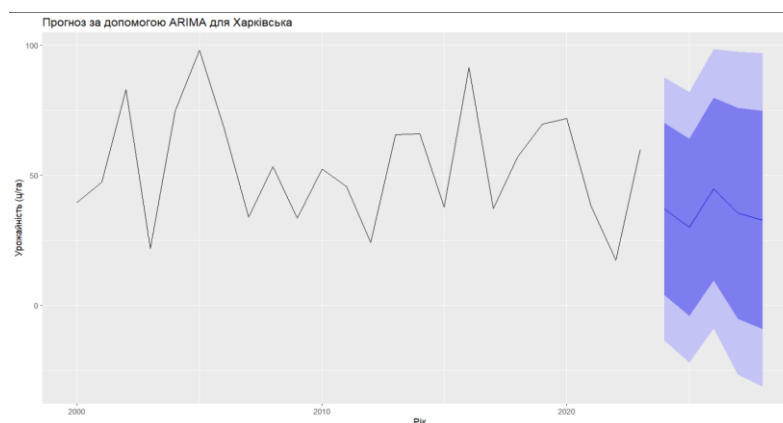


Рисунок 2.54 – Модель ARIMA для Харківської області

Джерело: розроблено автором

На графіку, зображеному на рис. 2.54, Прогноз передбачає відносно стабільну урожайність у майбутньому з деякими коливаннями. Порівнюючи результати прогнозів, обидва в майбутньому показують спадання показників урожайності, проте в прогнозі побудованому завдяки моделі ARIMA урожайність у майбутньому спадає не так стрімко, як у прогнозі ETS.

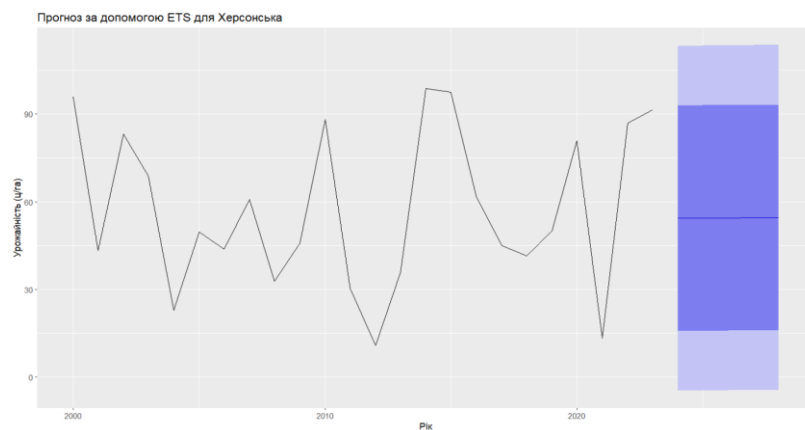


Рисунок 2.55 – Модель ETS для Херсонської області

Джерело: розроблено автором

На графіку, зображеному на рис. 2.55, урожайність у Херсонській області демонструє дуже значні коливання з року в рік. Є періоди різких падінь (наприклад, близько 2003, 2012 та 2020 років) і стрімких підйомів (2009, 2017, 2021). Видно, що жодної стабільної тенденції до зростання чи зниження немає – урожайність змінюється хвилеподібно. Прогноз вказує на стабільність урожайності в майбутньому або ж на можливе невелике зростання показників.

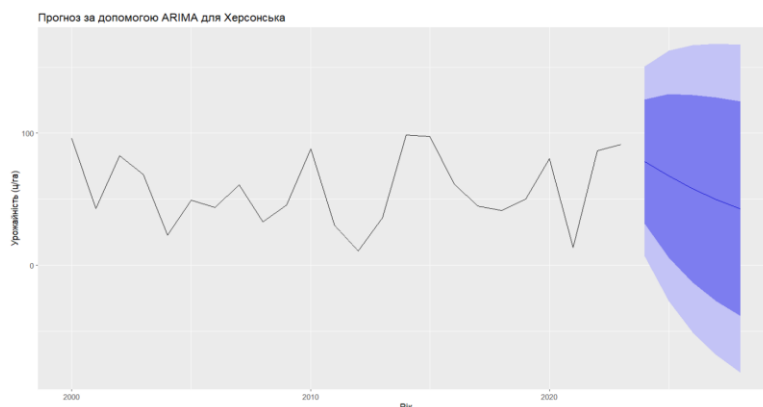


Рисунок 2.56 – Модель ARIMA для Херсонської області

Джерело: розроблено автором

Графік, зображений на рис. 2.56, показує, що прогноз передбачає різке зниження урожайності в майбутньому. Порівнюючи прогнози, помітно велику різницю в обох моделях адже у прогнозованих значення через ETS урожайність поступово зростає, а в прогнозі через ARIMA - досить стрімко спадає.

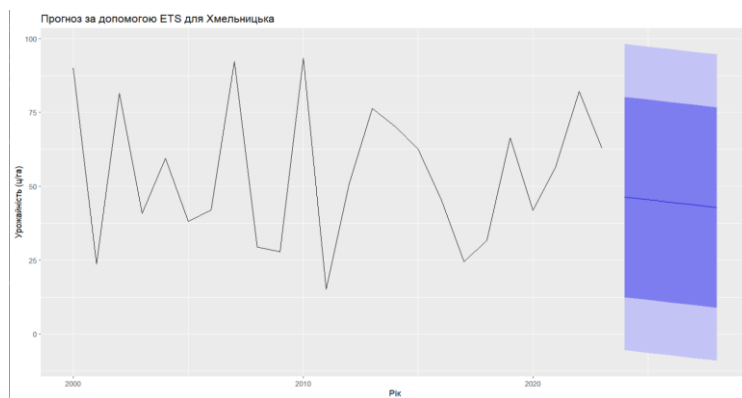


Рисунок 2.57 – Модель ETS для Хмельницької області

Джерело: розроблено автором

На графіку, зображеному на рис. 2.57, урожайність у Хмельницькій області дуже мінлива: спостерігаються як різкі підйоми, так і суттєві падіння. Особливо помітні різкі зниження урожайності (наприклад, близько 2012 року), а також періоди різкого зростання (наприклад, 2009, 2018). Загальної стабільної тенденції не простежується - урожайність змінюється хвилеподібно. Прогнозовані значення урожайності вказують на поступове спадання у майбутньому.

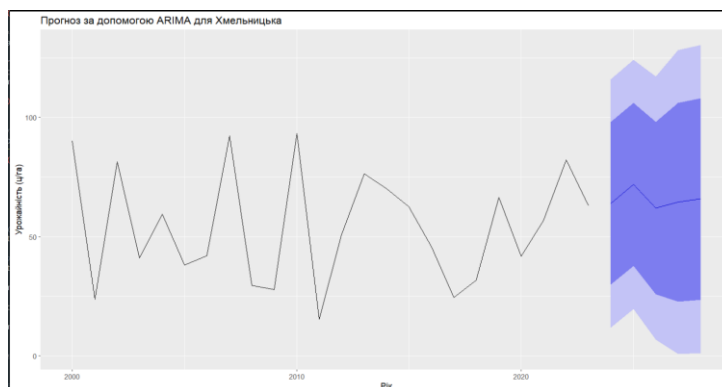


Рисунок 2.58 – Модель ARIMA для Хмельницької області

Джерело: розроблено автором

Графік, зображений на рис. 2.58, показує, що прогноз передбачає поступове зростання показників у майбутньому. Порівнюючи з попереднім прогнозом для Хмельницької області, виконаним за допомогою моделі ETS, обидві моделі різняться прогнозами, ARIMA вказує на невелике зростання, а ETS - на невелике спаадння урожайності.

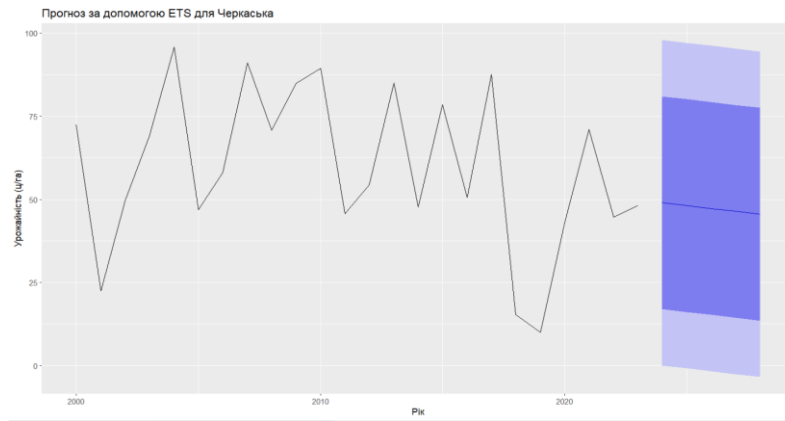


Рисунок 2.59 – Модель ETS для Черкаської області

Джерело: розроблено автором

На графіку, зображеному на рис. 2.59, урожайність у Черкаській області має значні коливання з року в рік. Спостерігаються як різкі підйоми (наприклад, на початку 2010-х років), так і суттєві падіння (особливо помітне різке зниження після 2015 року). Загалом, тренд не є стабільним – урожайність змінюється хвилеподібно. Прогнозовані значення вказують на невелике зниження показників у майбутньому.

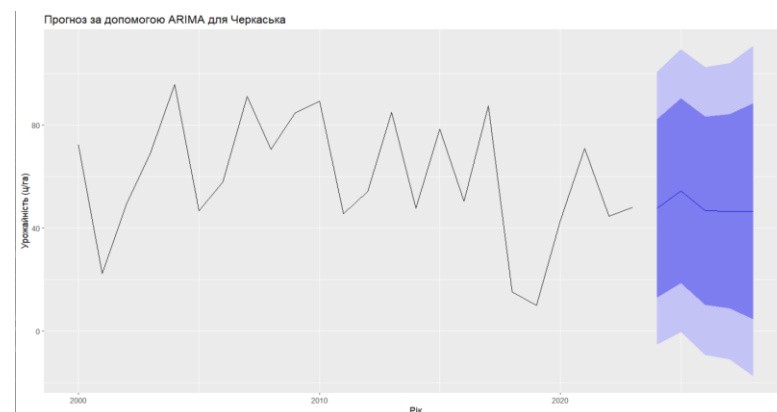


Рисунок 2.60 – Модель ARIMA для Черкаської області

Джерело: розроблено автором

Графік, зображений на рис. 2.60, показує, що прогноз передбачає відносно стабільну урожайність. Порівнюючи з попереднім прогнозом для Черкаської області, виконаним за допомогою моделі ETS, обидві моделі вказують на можливе невелике зниження урожайності в майбутньому, проте прогноз ARIMA є більш стабільним ніж ETS.

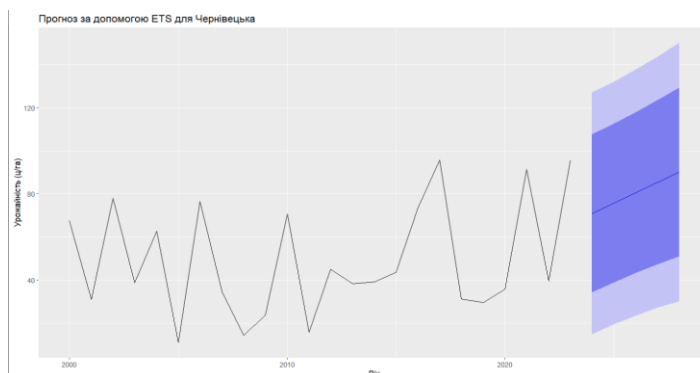


Рисунок 2.61 – Модель ETS для Чернівецької області

Джерело: розроблено автором

На графіку, зображеному на рис. 2.61, помітно наявність історичної мінливості урожайності в Чернівецькій області. Прогноз вказує на чітку тенденцію до зростання урожайності в майбутньому.

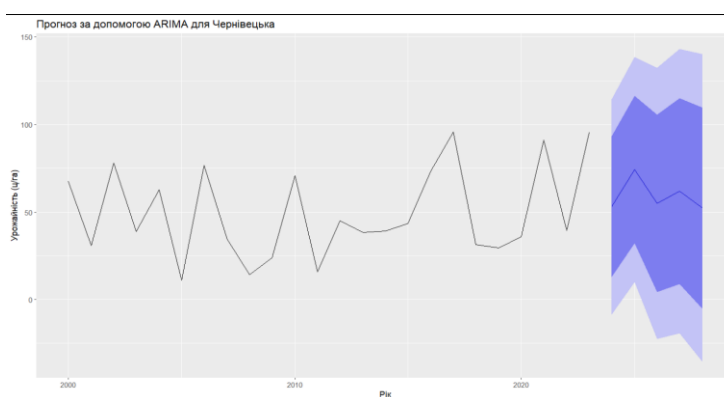


Рисунок 2.62 – Модель ARIMA для Чернівецької області

Джерело: розроблено автором

На графіку, зображеному на рис. 2.62, прогноз передбачає відносно стабільну урожайність у майбутньому з деякими коливаннями та спаданням показників. Порівнюючи з попереднім прогнозом для Чернівецької області, виконаним за допомогою моделі ETS, ARIMA показує спадання урожайності в майбутньому, тоді як ETS вказував на тенденцію до зростання.

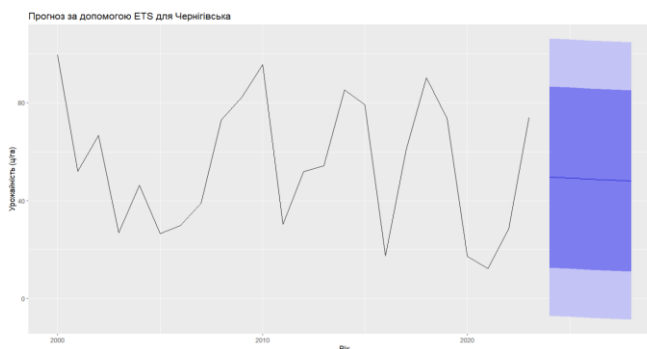


Рисунок 2.63 – Модель ETS для Чернігівської області

Джерело: розроблено автором

чином, вибір моделі повинен базуватися на індивідуальних особливостях часових рядів для кожного регіону, з урахуванням точності, складності та інтерпретованості результатів.

ВИСНОВКИ

У кваліфікаційній бакалаврській роботі було здійснено всебічне дослідження урожайності зернових культур в Україні на основі статистичних даних за 2000–2023 роки. Робота поєднала теоретичні аспекти аграрної аналітики, класичні статистичні підходи та сучасні методи машинного навчання для побудови ефективних моделей прогнозування урожайності. Проведений аналіз підтвердив, що точність прогнозів може бути суттєво покращена за рахунок використання гнучкіших та адаптивніших моделей, зокрема нейронних мереж.

У процесі дослідження було виконано низку завдань, а саме охарактеризовано основні поняття щодо урожайності зернових культур; проаналізовано фактори впливу на урожайність; досліджено вплив війни на аграрний сектор та урожайність; оцінено традиційні методи прогнозування урожайності та визначено їхні обмеження; проведено опис, підготовку та попередній аналіз вхідних даних; застосовано моделі прогнозування на основі машинного навчання (Random Forest, XGBoost, нейронні мережі) та методів часових рядів (ETS, ARIMA); проведено порівняння точності моделей за метрикою RMSE.

Характеристика основних понять щодо урожайності зернових культур допомогло зрозуміти, що сільське господарство є важливою сферою. Завдяки аналізу термінів, таких як урожайність, врожай, валовий збір, посівні площі, стало можливим розмежувати поняття, які часто плутають, і правильно їх застосовувати в подальшому аналізі.

Аналіз факторів впливу на урожайність допоміг визначити ключові групи чинників: природно-кліматичні (опади, температура), економічні (фінансування, технології), соціальні (кадровий потенціал), а також політичні. Встановлено, що найсильніший вплив мають погодні умови та рівень технічного забезпечення.

Дослідження впливу війни на аграрний сектор та урожайність допомогло проаналізувати регіональні особливості у урожайності після 2022 року. Виявити

суттєве зниження показників у південних та східних областях через бойові дії, мінування полів та порушення логістики.

Оцінка традиційних методів прогнозування урожайності та визначення їх обмежень допомогло з'ясувати точність моделей, взаємозв'язки між змінними та інше.

Проведення опису, підготовки та попереднього аналізу вхідних даних допомогло виявити тренди, аномалії та розподіл урожайності.

Застосування моделей прогнозування на основі машинного навчання (Random Forest, XGBoost, нейронні мережі) та методів часових рядів (ETS, ARIMA) допомогло визначити ефективність тої чи іншої моделі. Методи часових рядів допомогли спрогнозувати дані на майбутні 4 роки та побачити, яка ж область демонструє найкращі результати.

Проведення порівняння точності моделей за метрикою RMSE допомогло визначити при аналізі, яка з моделей демонструє найкращі результати. Найкращий результат (найнижче значення похибки) показала модель на основі одношарової нейронної мережі (nnetar). ETS та XGBoost також продемонстрували високу точність, тоді як ARIMA та лінійна регресія поступилися за ефективністю.

Загалом, дослідження підтвердило, що сучасні методи машинного навчання мають значний потенціал у сфері аграрної аналітики. Результати цієї роботи мають теоретичне значення як база для подальших досліджень та розвитку моделей прогнозування, методичну цінність як приклад поєднання різних підходів, та практичну важливість для впровадження в систему підтримки рішень аграрного менеджменту.

Таким чином, поставлена мета — побудова ефективної системи прогнозування урожайності зернових культур із використанням методів машинного навчання — була досягнута в повному обсязі. Отримані результати можуть стати основою для створення прикладних програмних продуктів у сфері аграрної аналітики, а також інструментом прийняття стратегічних рішень у сільському господарстві України.

[005_%D0%9C%D0%B0%D1%85%D0%B0%D0%BD%D0%B5%D1%86%D1%8C%20%D1%82%D0%B0%20%D1%96%D0%BD_Statistica%2012_%D0%B%D0%B0%20%D1%81%D0%B0%D0%B9%D1%82.pdf?sequence=1&isAllowed=y](https://doi.org/10.32782/TNU-2663-6468/2024.5/04)

11. Кузьменко Г. О., Телендй А. А. (2024). Вплив війни на аграрний сектор України: виклики та роль державного управління у відновленні галузі. *Вчені записки ТНУ ім. В. І. Вернадського. Серія: Державне управління*, № 5. <https://doi.org/10.32782/TNU-2663-6468/2024.5/04>.
12. Морозова М. М. Структура нейронної мережі для прогнозування урожайності та контролю параметрів мікроклімату. Сорок третя всеукраїнська практично-пізнавальна інтернет-конференція. URL: <https://naukam.triada.in.ua/index.php/konferentsiji/74-sorok-tretya-vseukrajinska-praktichno-piznavalna-internet-konferentsiya/973-struktura-nejronnoji-merezhi-dlya-prognozuvannya-vrozhajnosti-ta-kontrolyu-parametriv-mikroklimatu>
13. Пасько С. (2022). Стан та перспективи розвитку аграрного сектору економіки України. *Вісник УНУ. Міжнародні економічні відносини та світова економіка*, 6. <https://doi.org/10.32782/2413-9971/2022-42-30>
14. Переваги та недоліки коробчастої діаграми. URL: <https://www.sciencing.com/advantages-disadvantages-box-plot-12025269/>
15. Переваги та обмеження лінійних графіків для аналізу трендів. URL: <https://fastercapital.com/topics/benefits-and-limitations-of-line-graphs-for-trend-analysis.html>
16. Сільське господарство. URL: <https://sd4ua.org/golovni-temi-stalogo-rozvitku/silske-gospodarstvo/>
17. Статистика вирощування зернових в Україні 2007 – 2024. URL: <http://shareupotential.com/ru/BE/ukrainian-zerno-statistika-vyraschivania-2020.html>
18. Статистика сільського господарства України від AgroStats. URL: <https://agrostats.uhmi.org.ua/>

19. Урожай зернових та олійних в Україні. URL: <https://voxukraine.org/urozhaj-zernovyh-ta-olijnyh-v-ukrayini>
20. Урожайність культур сільськогосподарських за регіонами. URL: <https://data.gov.ua/en/dataset/188450fa-3a63-43b5-a783-503b52b3f847>
21. Урожайність сільськогосподарських культур. URL: <https://superagronom.com/slovník-agronoma/urozhaynist-silskogospodarskih-kultur-id18874>
22. Яровий А. Т., Страхов Є. М. (2019). Аналіз часових рядів : навч.-метод. посіб. для студентів математичних та економічних спеціальностей. – Одеса : Освіта України. С.109. URL: https://onu.edu.ua/pub/bank/userfiles/files/fmfit/nachalni_materiali/navch_mat_111_mag/analiz_chasovykh_ryadiv.pdf
23. A Quick Guide to Scatter Diagrams in Project Management. URL: <https://www.wrike.com/blog/quick-guide-scatter-diagrams/>
24. Adil, M., Lv, F., Cao, L., Lu, H., Lu, S., Gul, I., Bashir, S., Wang, Z., Li, T., & Feng, W. (2024). Long-term effects of agronomic practices on winter wheat yield and NUE in dryland regions of USA and China: a long-term meta-analysis. *Scientific Reports*, 14(1), 24777. <https://doi.org/10.1038/s41598-024-74910-7>
25. Breiman, L. (2001). Random Forests. *Machine Learning*, 45, 5–32. <https://doi.org/10.1023/A:1010933404324>
26. Chen, T., Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. <http://dx.doi.org/10.1145/2939672.2939785>.
27. Hyndman, R. J., Athanasopoulos, G. (2018). *Forecasting: principles and practice*, 2nd ed. OTexts: Melbourne, Australia. OTexts.com/fpp2
28. Increasing Farm Yield Through AgTech Solutions. URL: <https://www.agrivi.com/blog/improvements-in-agricultural-technology-increase-farm-yields/>
29. Kmytiuk, T., Majore, G., & Bilyk, T. (2024). Time series forecasting of price of the agricultural products using data science. *Agricultural and Resource*

<https://doi.org/10.51599/are.2024.10.03.01>

30. Linear Regression in Machine learning. URL: <https://www.geeksforgeeks.org/ml-linear-regression/>
31. McGill, R., Tukey, J. W., Larsen, W. A. (1978). Variations of Box Plots. *The American Statistician*, 32(1), 12–16. <https://doi.org/10.2307/2683468>
32. Mukaka, M. M. (2012). Statistics corner: A guide to appropriate use of correlation coefficient in medical research. *Malawi Medical Journal*, 24(3), 69–71.
33. Using econometric models for managing distribution processes. (2024). *Economic Journal of Lesya Ukrainka Volyn National University*, 2(38), 134–144.
34. Yang, Y., Wu, J., Du, Y.-L., Gao, C., Tang, D. W. S., & van der Ploeg, M. (2022). Effect on soil properties and crop yields to long-term application of superabsorbent polymer and manure. *Frontiers in Environmental Science*, 10, 859434. doi: 10.3389/fenvs.2022.859

