

*Електронний журнал «Ефективна економіка» включено до переліку наукових фахових видань України з питань економіки (Категорія «Б», Наказ Міністерства освіти і науки України № 975 від 11.07.2019). Спеціальності – 051, 071, 072, 073, 075, 076, 292. Ефективна економіка. 2025. № 10.*

**DOI: <http://doi.org/10.32702/2307-2105.2025.10.110>**

**УДК 633.1:303:338**

*Т. Л. Кмитюк,*

*к. е. н., доцент, доцент кафедри штучного інтелекту,*

*моделювання та статистики,*

*Київський національний економічний університет імені Вадима Гетьмана*

*ORCID ID: <https://orcid.org/0000-0001-5262-856X>*

*Е. Ю. Ус,*

*здобувач другого (магістерського) рівня вищої освіти, освітньої програми*

*«Економічна кібернетика і Дата Сайнс»*

*Київський національний економічний університет імені Вадима Гетьмана*

*ORCID ID: <https://orcid.org/0009-0005-5555-550X>*

## **ПРОГНОЗУВАННЯ УРОЖАЙНОСТІ ЗЕРНОВИХ КУЛЬТУР В УКРАЇНІ ЯК ІНСТРУМЕНТ ОЦІНЮВАННЯ ПРОДОВОЛЬЧОЇ БЕЗПЕКИ**

*Т. Кмытиук,*

*PhD in Economics, Associate Professor, Associate Professor of the Department of*

*Artificial Intelligence, Modeling and Statistics,*

*Kyiv National Economic University named after Vadym Hetman*

*E. Us,*

*Student of the second (master's) level of higher education, educational program*

*"Economic Cybernetics and Data Science",*

*Kyiv National Economic University named after Vadym Hetman*

## **FORECASTING GRAIN YIELDS IN UKRAINE AS A TOOL FOR ASSESSING FOOD SECURITY**

*У статті здійснено комплексне дослідження можливостей прогнозування урожайності зернових культур в Україні з метою підвищення ефективності аграрного управління та забезпечення продовольчої безпеки держави. Проаналізовано широкий спектр методів прогнозування, включаючи як традиційні статистичні підходи (тест Шапіро–Уїлка, тест Колмогорова–Смірнова, кореляційний аналіз, дисперсійний аналіз (ANOVA), методи лінійної регресії), так і сучасні методи машинного навчання, зокрема Random Forest, XGBoost, штучні нейронні мережі, а також часові моделі ARIMA та ETS. Особливу увагу приділено порівнянню точності прогнозів, які генерують різні моделі. За результатами емпіричного моделювання, яке базується на офіційних статистичних даних за період 2000–2023 років, встановлено, що нейронні мережі демонструють найвищу точність у прогнозуванні урожайності. Отримані результати можуть слугувати основою для подальшого вдосконалення системи моніторингу сільськогосподарського виробництва та розробки рекомендацій для аграрної політики в умовах змін клімату та економічної нестабільності.*

*This article presents a detailed and multifaceted analysis of grain crop yield forecasting in Ukraine, underlining its strategic relevance for enhancing national food security, optimizing agricultural policy, and mitigating risks associated with environmental and economic instability. As global agriculture faces mounting challenges driven by climate change, resource scarcity, geopolitical tensions, and market volatility, the ability to forecast crop productivity accurately has become a cornerstone of sustainable development and evidence-based governance. The study adopts an integrated methodological framework that combines conventional statistical and econometric techniques—such as the Shapiro–Wilk and Kolmogorov–Smirnov normality tests, correlation analysis, ANOVA (analysis of variance), and linear regression—with cutting-edge machine learning algorithms, including Random Forest, XGBoost, artificial neural networks, and time series forecasting models such as ARIMA and ETS. The research is based on official data from 2000 to*

2023, enabling robust model training, testing, and validation across diverse climatic and regional conditions. Forecast performance was evaluated using standard predictive accuracy metrics, particularly RMSE, ensuring a reliable and transparent comparison between statistical and algorithmic methods. The results demonstrate that artificial neural networks consistently outperform traditional and machine learning counterparts in terms of predictive precision and long-term stability. This highlights the transformative potential of deep learning-based models as core components of modern agri-analytical systems and food policy planning tools. Moreover, the study's findings can be adapted for implementation in other countries with comparable agroecological structures and institutional capacities. The integration of data-driven approaches into agricultural forecasting is essential for developing resilient food systems capable of withstanding future uncertainties and ensuring sustainable rural development on both regional and global scales.

**Ключові слова:** продовольча безпека, урожайність, зернові культури, машинне навчання, математична модель, прогнозування, аграрний сектор.

**Keywords:** food security, yield, grain crops, machine learning, mathematical model, forecasting, agricultural sector.

**Постановка проблеми у загальному вигляді та її зв'язок із важливими науковими чи практичними завданнями.** Проблема забезпечення продовольчої безпеки залишається актуальною у контексті зростання населення світу, змін клімату та нестабільності на аграрних ринках. Україна, як один із провідних експортерів зернових культур, відіграє ключову роль у глобальній продовольчій системі. З огляду на значення аграрного сектора для економіки держави, прогнозування урожайності є необхідною умовою для формування ефективної політики управління сільським господарством. Актуальність дослідження зумовлена необхідністю побудови надійних моделей прогнозування урожайності, здатних забезпечити точні оцінки майбутніх показників та запобігти продовольчим кризам. Попри наявність значної

кількості досліджень щодо прогнозування аграрних показників, низка аспектів, зокрема регіональна деталізація моделей, інтеграція кліматичних змінних та використання новітніх алгоритмів машинного навчання, залишаються недостатньо вивченими. Новизна цієї роботи полягає у комплексному підході до моделювання урожайності зернових культур на основі відкритих статистичних даних та порівнянні результативності різних моделей.

***Аналіз останніх досліджень і публікацій.*** Дослідження в галузі прогнозування урожайності активно розвиваються завдяки досягненням у сфері аналізу даних. Зокрема, у роботі [1] розглянуто причинно-наслідкові зв'язки між урожайністю та кліматичними умовами, підкреслено важливість зменшення розриву між потенційною та реальною урожайністю. У роботі [2] продемонстровано ефективність застосування глибоких нейронних мереж на основі супутникових даних для прогнозування урожайності, тоді як у дослідженнях [5] і [6] рекурентні нейронні мережі успішно використано для прогнозування цін на сільськогосподарську продукцію. Проте більшість моделей орієнтовані на ціноутворення, тоді як модельний апарат прогнозування урожайності з урахуванням просторово-часових залежностей досі потребує розвитку. Крім того, існуючі дослідження рідко використовують порівняння класичних та сучасних ML-методів в одному аналізі.

***Формулювання цілей статті (постановка завдання).*** Мета дослідження полягає в аналізі, побудові та тестуванні моделей прогнозування урожайності зернових культур в Україні з використанням сучасних методів машинного навчання.

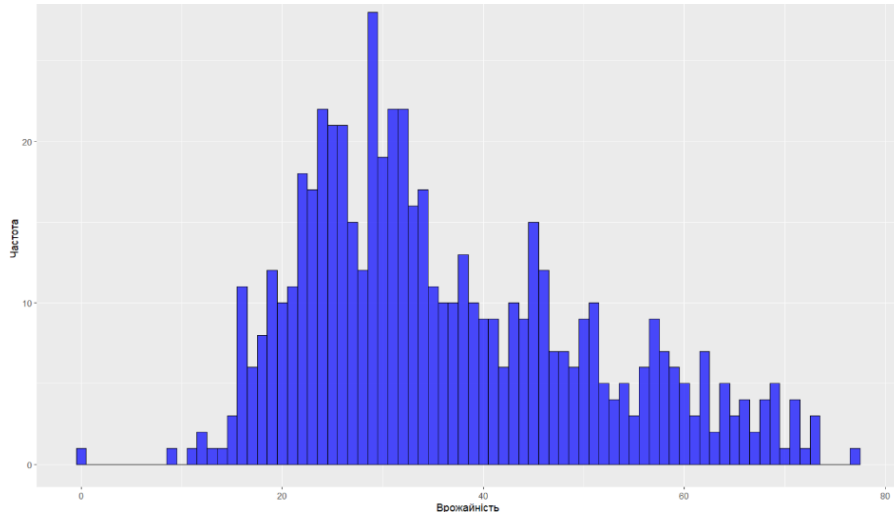
***Виклад основного матеріалу дослідження.***

У дослідженні використано дані з відкритого джерела AgroStats [4], що охоплюють 24 області України з 2000 по 2023 роки, а набір даних містить 576 спостережень. Основна змінна — урожайність зернових культур (ц/га). Для аналізу нормальності розподілу даних було застосовано тести Шапіро-Вілка та Колмогорова-Смірнова, результати яких наведені у таблиці 1, а також побудовано гістограму розподілу врожайності, як показано на рисунку 1.

**Таблиця 1. Результати тестової перевірки**

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

*Джерело: розроблено автором*



**Рис.1. Розподіл урожайності зернових культур**

*Джерело: розроблено автором*

З даних, що представлені на рисунку 1, стає очевидним асиметричний розподіл з правим «хвостом». Це означає, що у вибірці є значна частка високих значень врожайності. Крім того, спостерігаються декілька піків, що потенційно свідчить про мультимодальність - наявність декількох кластерів з різними середніми показниками врожайності. 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). 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:  $\approx 31.6$  одиниць (урожайність значно збільшилася у 2023 році). t-статистика: -8.4984. Ступені свободи: 24.962. p-value:  $7.766e-09$ . 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 аналіз ANOVA.

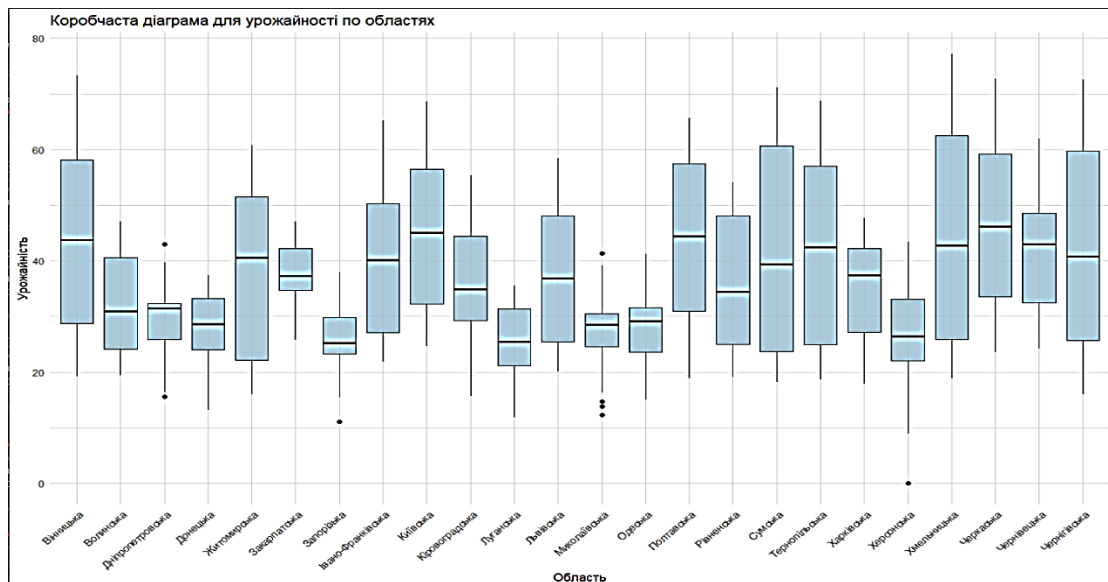
**Таблиця 2. Результат ANOVA**

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

*Джерело: розроблено автором*

Результати наведені у таблиці 2, презентують наступне: Sum Sq (Сума квадратів) дає нам інформацію про те, як розподіляється варіація: між різними групами та в межах кожної з них. Для фактору «Область» дане значення 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 лишилися, – 162. F-значення дорівнює 7.612. Значне F-значення свідчить про те, що різниця між групами набагато більша, ніж різниця всередині груп. Pr(>F) являє собою p-значення для перевірки гіпотези, що всі групи мають однакове середнє. P-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. Для глибшого аналізу даних було побудовано коробчасту діаграму (рис. 2) між змінною Урожайність та Область, графік розсіювання для кожної області, графік трендів та графік розсіяння з кольоровими точками для різних областей.

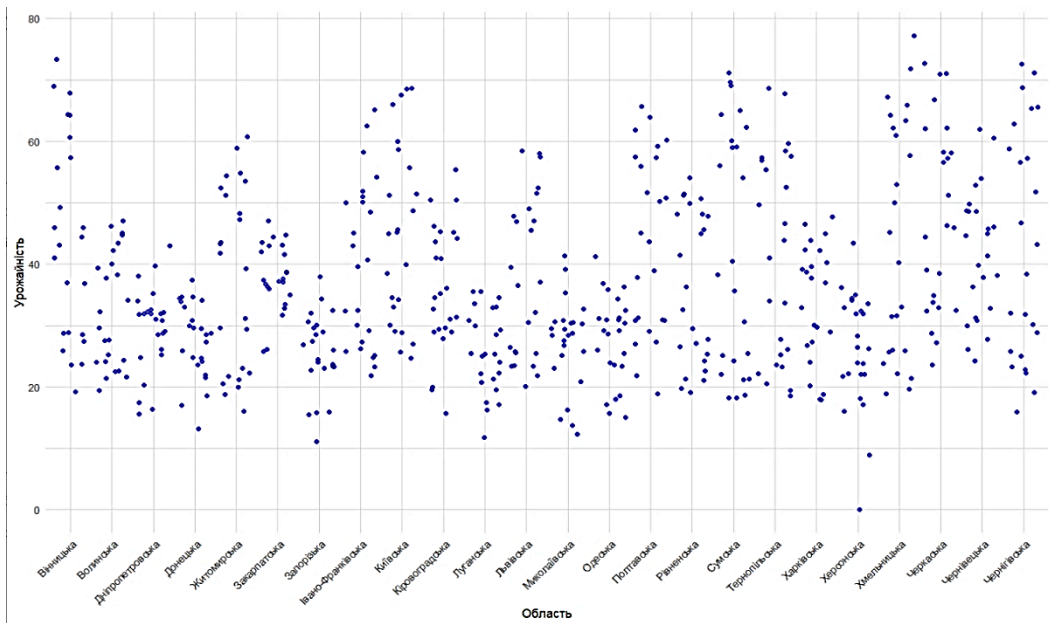


**Рис. 2. Результати коробчастої діаграми**

*Джерело: розроблено автором*

З графік показує, що Черкащина, Хмельниччина, Сумщина, Тернопільщина, Київщина вирізняються високим медіанним значенням, що вказує на стабільно значний рівень врожайності. Житомирщина, Закарпаття теж демонструють досить широкий розкид показників врожайності. Івано-Франківщина, Миколаївщина, Херсонщина, Одещина показують медіану нижчу, що, ймовірно, пов'язано з кліматичними умовами або іншими особливостями регіонів. У деяких областях помітні випадки, зокрема, в Івано-Франківській, Херсонській, Миколаївській, Львівській областях зустрічаються кілька аномально низьких показників урожайності, що може бути наслідком неврожайних років або інших факторів впливу. Отож, центральна та північна частини України в цілому демонструють вищі показники. Південні та західні регіони характеризуються меншими значеннями або більшою варіативністю.

Графіки розсіювання, побудовані для кожної області наведені нижче на рисунку 3.



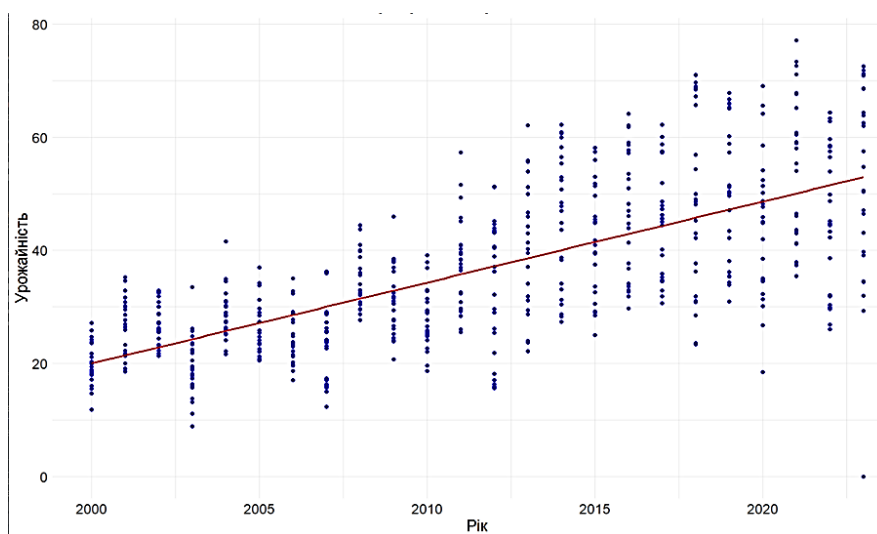
**Рис. 3. Результат графіку розсіювання для різних областей України**

*Джерело: розроблено автором*

На рисунку 3 зображено розподіл урожайності по регіонах України через точкову діаграму. З графіка стає очевидним, що найбільші показники урожайності (вище 70) зафіксовані в таких областях: Черкаська, Хмельницька, Сумська, Тернопільська, Чернігівська та Вінницька. Це 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 агропромисловості 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.

Дослідження показало, що існує загальна тенденція до збільшення урожайності, 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 демонструють суттєво більшу середню урожайність порівняно з іншими, припускається, що це можуть бути західні або центральні області. 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. У результаті моделювання було підтверджено позитивний тренд урожайності (приріст 1.43 ц/га на рік) (див. рис. 4).



**Рис.4. Графік лінійної регресії**  
*Джерело: розроблено автором*

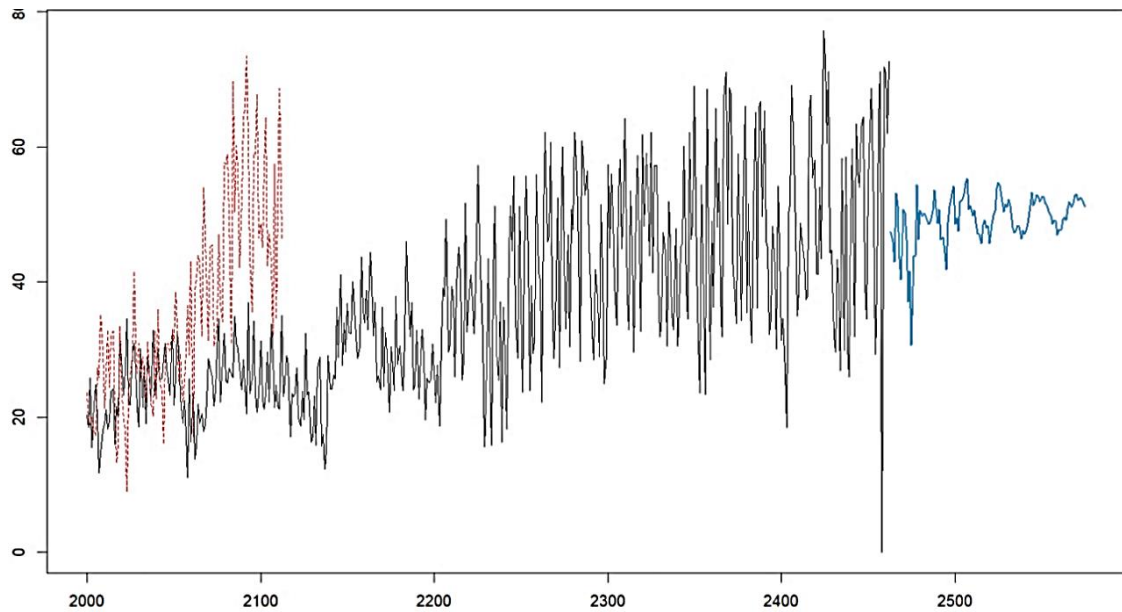
Для глибшого аналізу даних урожайності зернових культур по областях України побудовані моделі із застосуванням сучасних методів машинного навчання, зокрема: дерева рішень, Random Forest, XGBoost, SVM, нейронні мережі, ETS, ARIMA. Моделювання проводилось у середовищі R з використанням пакетів forecast, nnfor, randomForest, xgboost. Вибір цих інструментів зумовлено їх широким використанням у міжнародних дослідженнях. RMSE було обрано як основний критерій оцінки точності моделей. Порівняння результатів дозволило визначити найбільш точні алгоритми для прогнозування аграрної врожайності.

**Таблиця 3. Значення RMSE кожної моделі**

Модель	RMSE
Лінійна регресія	7.837
Дерева рішень	9.403
Випадкові ліси	6.338
SVM	7.267
XGBoost	5.218
Нейронні мережі однокрокова (nnetar)	1.730
Нейронні мережі багат шарова (nnfor)	9.429
ETS	25.482
ARIMA	23.669

*Джерело: розроблено автором*

З даних таблиці 3, очевидно, що найкраща модель – це одноступенева нейронна мережа. Вона продемонструвала найнижчий рівень RMSE. Як підтвердження результатів, було побудовано графічне зображення прогнозу, сформованого за допомогою одноступеневої нейронної мережі (див. рис. 5). На графіку показано фактичні значення урожайності зернових культур у навчальному та тестовому періодах, а також прогнозовані значення, отримані на основі цієї моделі.



**Рис. 5. Графік прогнозу урожайності з використанням нейронної мережі авторегресії (NNAR)**

*Джерело: розроблено автором*

На рисунку 5 зображено графік передбачення врожайності, сформований за допомогою нейронної мережі авторегресії. 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.

***Висновки та перспективи подальших розвідок у даному напрямі.*** У результаті проведеного дослідження здійснено комплексний аналіз урожайності

зернових культур в Україні за період 2000–2023 років. За допомогою описової статистики та тестів на нормальність виявлено значну варіативність урожайності між роками та регіонами, що свідчить про нерівномірність аграрного розвитку в країні. Результати дисперсійного аналізу (ANOVA) підтвердили наявність статистично значущих відмінностей між областями, а візуалізація даних допомогла окреслити просторово-часові тренди розвитку сільського господарства.

Ключовим елементом дослідження стало прогнозування урожайності з використанням як класичних методів (лінійна регресія, дерева рішень), так і сучасних підходів машинного навчання (ARIMA та ETS, Random Forest, SVM, XGBoost, штучні нейронні мережі). Оцінювання моделей за критерієм RMSE засвідчило, що найбільш точну модель забезпечує одноступенева нейронна мережа ( $RMSE \approx 1.73$ ), яка продемонструвала найкращі прогностичні характеристики серед усіх розглянутих моделей. Водночас моделі ARIMA та ETS, показали найменшу точність прогнозу. Графічна інтерпретація прогнозу підтвердила ефективність нейронної мережі у відтворенні складних динамічних патернів урожайності, зокрема довгострокових трендів та короткотермінових коливань. Прогнозні значення на найближчі роки залишаються в діапазоні 40–50 ц/га, що свідчить про стабільне зростання та зменшення ризику продовольчої нестабільності.

Отримані результати мають важливе прикладне значення для аграрної політики, зокрема в контексті стратегічного планування, формування державного резерву зерна та управління ризиками. Запропонована модель може бути адаптована для інших сільськогосподарських культур і регіонів, а також інтегрована в цифрові платформи управління аграрним виробництвом. Подальші дослідження можуть зосереджуватися на включенні кліматичних змінних, супутникових даних та соціально-економічних факторів задля підвищення точності та адаптивності моделей прогнозування.

## Література

1. Lobell D. B., Cassman K. G. and Field C. B. Crop yield gaps: Their importance, magnitudes, and causes. *Annual Review of Environment and Resources*. 2009. № 34. P. 179-204. URL: <https://doi.org/10.1146/annurev.enviro.041008.093740>
2. You J., Li X., Low M., Lobell D. and Ermon S. Deep Gaussian process for crop yield prediction based on remote sensing data. *AAAI Conference on Artificial Intelligence*. 2017. № 31(1). URL: <https://ojs.aaai.org/index.php/AAAI/article/view/11369>
3. Месель-Веселяк В. Я. Виробництво зернових культур в Україні: потенційні можливості. *Економіка АПК*. 2018. №5. С. 5-14. URL: <https://doi.org/10.1609/aaai.v31i1.11172>
4. Статистика сільського господарства України. URL: <https://agrostats.uhmi.org.ua/> (дата звернення: 15.05.2025).
5. Kmytiuk T.L. and Majore G.M. Time series forecasting of agricultural product prices using Elman and Jordan recurrent neural networks. *Neuro-Fuzzy Modeling Techniques in Economics*. 2021. № 10, P. 67-85. URL: <http://doi.org/10.33111/nfmte.2021.067>
6. Kmytiuk T.L., Majore G.M. and Bilyk T.O. Time series forecasting of price of the agricultural products using data science. *Agricultural and Resource Economics: International Scientific E-Journal*. 2024. № 10 (3), P. 5-33. URL: <https://doi.org/10.51599/are.2024.10.03.01>
7. Shin S., Ahn J. and Kim N. Predicting crop yields with long short-term memory neural networks and remote sensing data. *Sustainability*. 2020. № 12 (4). URL: <https://doi.org/10.3390/su12041441>

## References

1. Lobell, D.B. Cassman, K.G. and Field, C.B. (2009), "Crop yield gaps: Their importance, magnitudes, and causes", *Annual Review of Environment and*

*Resources*, vol. 34, pp. 179-204. DOI: <https://doi.org/10.1146/annurev.enviro.041008.093740>

2. You, J. Li, X. Low, M. Lobell, D. and Ermon, S. (2017), “Deep Gaussian process for crop yield prediction based on remote sensing data”, *AAAI Conference on Artificial Intelligence*, pp. 4559-4565. DOI: <https://doi.org/10.1609/aaai.v31i1.11172>

3. Mesel-Veselyak V.Ya. (2018), “Grain production in Ukraine: potential opportunities”, *Ekonomika APK*, vol. 5, pp. 5-14. available at: [https://eapk.com.ua/web/uploads/pdf/eapk\\_2018\\_05\\_p\\_5\\_122-5-14.pdf](https://eapk.com.ua/web/uploads/pdf/eapk_2018_05_p_5_122-5-14.pdf) (Accessed 25 Sept 2025).

4. AgroStats (2025), “Statistics of agriculture in Ukraine”, available at: <https://agrostats.uhmi.org.ua/> (Accessed 15 May 2025).

5. Kmytiuk, T.L. and Majore G.M. (2021), “Time series forecasting of agricultural product prices using Elman and Jordan recurrent neural networks”, *Neuro-Fuzzy Modeling Techniques in Economics*, vol. 10, pp. 67-85. DOI: <http://doi.org/10.33111/nfmte.2021.067>

6. Kmytiuk, T.L. Majore, G.M. and Bilyk T.O. (2024), “Time series forecasting of price of the agricultural products using data science”, *Agricultural and Resource Economics: International Scientific E-Journal*, vol. 10(3), pp. 5-33. DOI: <https://doi.org/10.51599/are.2024.10.03.01>

7. Shin, S. Ahn, J. and Kim, N. (2020), “Predicting crop yields with long short-term memory neural networks and remote sensing data”, *Sustainability*, vol. 12(4). DOI: <https://doi.org/10.3390/su12041441>

*Стаття надійшла до редакції 07.10.2025 р.*