

ГРИЦЮК П. М.

Національний університет водного господарства та природокористування, м. Рівне  
<https://orcid.org/0000-0002-3683-4766>  
e-mail: [p.m.hrytsiuk@nuwm.edu.ua](mailto:p.m.hrytsiuk@nuwm.edu.ua)

БАБИЧ Т. Ю.

Національний університет водного господарства та природокористування, м. Рівне  
<https://orcid.org/0000-0001-6927-7313>  
e-mail: [t.iu.babych@nuwm.edu.ua](mailto:t.iu.babych@nuwm.edu.ua)

КРАСЬКО Б. В.

Національний університет водного господарства та природокористування, м. Рівне  
e-mail: [krasko\\_ak21@nuwm.edu.ua](mailto:krasko_ak21@nuwm.edu.ua)

## КЛАСИФІКАЦІЙНІ МЕТОДИ ПРОГНОЗУВАННЯ ВРОЖАЙНОСТІ

*Аналіз врожайності пшениці за останні 20 років показує, що врожайність у всіх областях України зростає. Це є результатом зростання інвестиційної привабливості галузі зерновиробництва та значних інвестицій, які надійшли у галузь. Однак, тенденція до зростання супроводжується значними коливаннями врожайності, причиною яких здебільшого є погодно-кліматичні фактори. Найбільш помітним є вплив кліматичних факторів для степових регіонів України. В роботі запропоновано класифікаційний підхід до прогнозування врожайності пшениці з врахуванням погодно-кліматичних факторів та використанням методів машинного навчання. Для побудови комп'ютерних моделей було використано набір даних врожайності пшениці у чотирьох областях степової зони України за період 2000 – 2021 роки та відповідні значення 12 метеофакторів. Числові значення врожайності були перетворені у бінарну змінну, яка набуває лише двох значень: «0» (низька врожайність) чи «1» (висока врожайність). Для моделювання бінарних даних використано probit- та logit-моделі, результуючі значення яких можуть трактуватися як імовірність значного перевищення врожайністю трендового значення. Точність побудованих моделей перевищує 80%.*

*Ключові слова: інвестиції, модель.*

Petro HRYTSIUK, Tetiana BABYCH, Bohdan KRASKO  
The National University of Water and Environmental Engineering, Rivne

## CLASSIFICATION METHODS OF THE YIELD FORECASTING

*Analysis of wheat yields over the past 20 years shows that wheat yield are growing in all regions of Ukraine. This is a result of the growing investment attractiveness of the grain industry and the significant investments made in the industry. However, the upward trend is accompanied by significant fluctuations in wheat yields, which are mostly caused by weather and climatic factors. The most noticeable is the climatic factors influence for the steppe regions of Ukraine. The paper proposes a classification approach to forecasting wheat yield taking into account weather and climatic factors and the use of machine learning methods. A set of wheat yield data in four regions of the steppe zone of Ukraine for the period 2000-2021 and the corresponding values of 12 meteorological factors were used to build computer models. Cross-validation with division into 10 groups was used to build and test of models. Numerical values of wheat yield were converted into a binary variable, which acquires only two values: "0" (low yield) or "1" (high yield). If the value of wheat yield exceeding over the trend value is greater than the specified threshold value, the new value of binary variable is 1, otherwise the binary variable value is 0.*

*Two models are constructed to classify the excess of yield over the trend value, which are based on the regression logit and regression probit. The obtained model values can be interpreted as the probability of a significant excess of yield over the trend value. Both models provide high quality classification with a slight advantage of the logit model. The aggregate classification quality of both models exceeds 88%, which is a good result for such a complex task as modeling the dependence of yield on meteorological factors in a small sample.*

*Keywords: investing, model*

### Постановка проблеми у загальному вигляді

#### та її зв'язок із важливими науковими чи практичними завданнями

Зернове виробництво є основою агропромислового комплексу України, визначає обсяги пропозиції та вартість основних видів продовольства. Надходження значних інвестицій в аграрне виробництво забезпечує стабільне зростання врожайності зернових. Впродовж останніх років Україна увійшла до групи провідних виробників та експортерів зерна. У 2021 році зібрано рекордний врожай зернових та зернобобових в розмірі 84 млн тон. Аграрний сектор формує близько 17 % валового внутрішнього продукту нашої держави, аграрний експорт становить майже 45 % від всіх валютних надходжень. Разом з тим, зерновиробництво є високо ризиковим виробництвом, ефективність якого у значній мірі залежить від кліматичних факторів. Для забезпечення стабільності зерновиробництва необхідно враховувати залежність врожайності від кліматичних факторів та мати адекватні моделі такої залежності.

Дискусійним моментом є критерій якості прогнозової моделі. Похибка прогнозу не може виступати таким критерієм, адже більш важливими є показники зростання/спадання врожайності та перевищення/недосягнення врожайністю середнього/трендового значення. Тому часто використовується критерій точності прогнозу, який будується на основі співвідношення кількості правильно передбачених напрямків зміни врожайності (приріст – спад) до загальної кількості прогнозів.

У даній роботі запропоновано класифікаційний підхід до прогнозування врожайності пшениці в областях степової зони України з використанням комп'ютерних моделей класифікації даних, та проведено порівняння ефективності використаних моделей.

### Аналіз досліджень та публікацій

У роботі [Грицюк, Бачишина, 2016] здійснено аналіз впливу метеорологічних факторів на врожайність зернових культур в регіонах України. Відзначено, що вирішальне значення для майбутньої врожайності мають погодні умови квітня, травня і червня. Схожий висновок було отримано у дослідженні залежності врожайності пшениці від кліматичних факторів у Ірландії [Irish Journal of Agricultural Research. - Vol. 8, No. 3, Dec., 1969. pp. 349-357. - J. Lee and M. J. Connaughton. Effects of Weather on Wheat Yields ]

В аналітичному огляді [Даніель Мюллер, Анне Юнгандreas, Фрідріх Кох та ін. Вплив кліматичних змін на виробництво пшениці в Україні. Німецько-український агрополітичний діалог. – Київ, 2016. – 45с.] проаналізовано оцінили вплив кліматичних змін на врожайність озимої пшениці на території усієї України та окремо у трьох чітко визначених екологічних зонах країни. Відзначено негативний вплив змін клімату на очікувані зміни врожайності зернових і, особливо, у степовому регіоні України.

У роботі Кернасюк Ю.В. Адаптація до кліматичних змін: економіка технологій поливу. – Економічний гектар. 03.06.2020. <http://agro-business.com.ua/agro/ekonomichniyi-hektar/item/17754-adaptatsiia-do-klimatichnykh-zmin-ekonomika-tekhnologii-polyvu.html> проаналізована економічна ефективність поливу з метою підвищення врожайності сільськогосподарських культур.

У роботі [О.О. Кривошеїн, Л.П. Однолеток, Л.П. Дзюба. Оцінка впливу погодних умов та організаційно-технологічних заходів на урожайність озимої пшениці за її кліматичним потенціалом. – Наукові праці УкрНДГМІ, 2016, Вип. 269. – с.151-158] використано поняття кліматичного потенціалу врожайності як прогностичної ознаки врожайності озимої пшениці.

У роботі [Кобченко Ю.Ф., Кобченко О.Ю., Резуненко В.А. Вплив погодних факторів на формування урожаю зернових культур у Харківській області. – Вісник Харківського національного університету. – 2014, №1098. – с.86-91] показано, що в умовах зони лісостепу вирішальний вплив на формування урожаю пшениці відіграє достатня кількість опадів. Цей же висновок підтверджено у роботі [Impact of extreme weather conditions on European crop production in 2018. Damien Beillouin].

Проблемам трансформації зерновиробництва в контексті змін клімату присвячено праці вітчизняних та зарубіжних вчених серед яких: Л. Бачишина, У. Гофман, П. Грицюк, В. Данкевич, Т. Ільєнко, А. Кучер, О. Нечипоренко, А. Польовий, Л. Попитченко, Дж. Портер, Г. Робертс, А. Сохнич, Ю. Тараріко та ін.

У роботі [Kurt Heil, Anna Lehner and Urs Schmidhalter. Influence of Climate Conditions on the Temporal Development of Wheat Yields in a Long-Term Experiment in an Area with Pleistocene Loess. - Climate 2020, 8, 100, 17p. [www.mdpi.com/journal/climate](http://www.mdpi.com/journal/climate)] було запропоновано використовувати залишки від тренду врожайності як індикатори впливу кліматичних міжрічних змін.

Прогнозування врожайності є складною задачею, суть якої визначається випадковим характером багатьох впливаючих факторів. Тому для вирішення цієї задачі доцільно застосовувати методи інтелектуального аналізу даних із застосуванням сучасних комп'ютерних технологій. Останнім часом з'явилося декілька робіт такого спрямування [Beillouin D, Schauburger B, Bastos A, Ciais P, Makowski D. 2020 Impact of extreme weather conditions on European crop production in 2018. Phil. Trans. R. Soc. B 375; Konduri V., Vandal T., Ganguly S., Ganguly A. Data science for weather impacts on crop yield. Front. Sustain. Food Syst., Vol. 4, May 2020 ]

### Виклад основного матеріалу.

#### Аналіз динаміки врожайності пшениці в областях України.

Для аналізу та моделювання динаміки врожайності пшениці ми використали офіційні дані, отримані з урядового порталу Державної служби статистики України [21]. Аналіз динаміки врожайності в областях України за останні 20 років показує, що врожайність у всіх областях зростає. Зростання врожайності стало результатом зростання інвестиційної привабливості галузі зерновиробництва та значних інвестицій, які в результаті надійшли у галузь. В результаті покращилась насіннева база, зросла агротехнічна культура, триває розбудова логістичної мережі (елеватори, вагони-зерновози, порти). У 2021 році в Україні був отриманий рекордний врожай зернових та зернобобових – 84 млн тон. Однак, тенденція до зростання врожайності супроводжується значними коливаннями врожайності, причиною яких здебільшого є погоднокліматичні фактори. Ілюстрацією може слугувати динаміка врожайності озимої пшениці в Херсонській області (рис. 1). Величина відхилень від тренду прямо залежить від впливу кліматичних факторів, основним з яких є посухи (2003 та 2012 роки на рис. 1). Для оцінки залежності врожайності від кліматичних факторів ми пропонуємо використовувати критерій, який є сумою модулів відхилень врожайності  $y_i$  від трендових значень  $tr_i$

$$S = \sum_{i=2000}^{2020} |y_i - tr_i|. \quad (1.1)$$

Розподіл областей за значенням кліматичного критерію представлений на рис. 2. Найбільше відчують вплив кліматичного фактора області, які належать до степової зони – Херсонська, Одеська, Запорізька, Кіровоградська. Саме ці 4 області є об'єктом нашого наукового дослідження.

Для прогнозування врожайності використовують різні методи. Одним з найпростіших є метод екстраполяції лінійного тренду, який дозволяє отримати точковий прогноз врожайності. Більш поширеним є інтервальний прогноз, який дозволяє встановити рівень надійності прогнозування. Для побудови інтервального прогнозу врожайності необхідно перевірити гіпотезу про нормальний розподіл трендових перевищень

$$eps_t = y_t - tr_t . \tag{1.2}$$

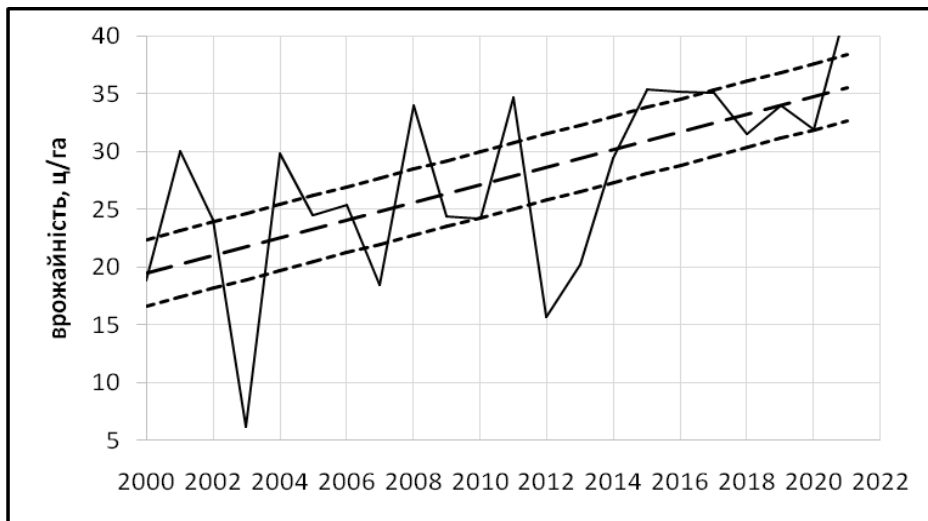


Рис. 1. Динаміка врожайності озимої пшениці у Херсонській області. Штрихова лінія – лінійний тренд. Пунктирні лінії – межі високої та низької врожайності.

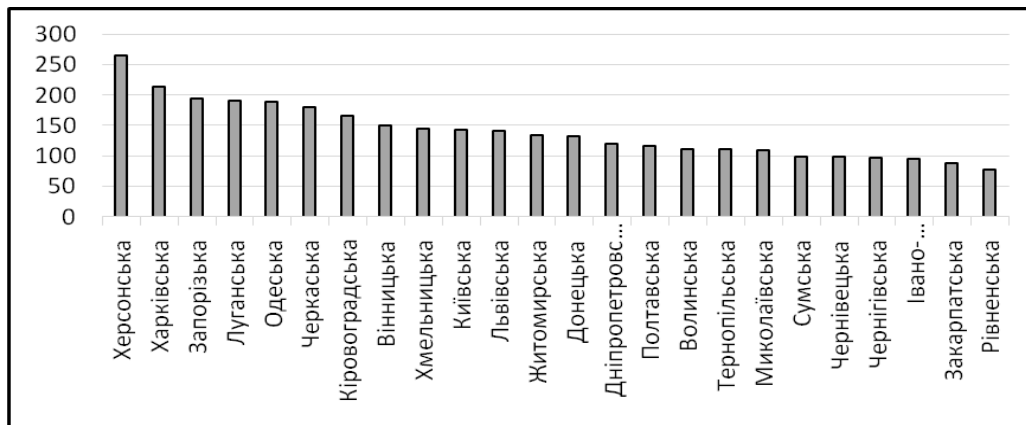


Рис. 2. Сортуння областей за значенням критерію залежності врожайності від кліматичного фактора

Для перевірки гіпотези про нормальний розподіл трендових перевищень ми використали об'єднану вибірку трендових перевищень для чотирьох областей степової зони (88 значень). Гіпотеза про нормальний розподіл перевищень була підтверджена з використанням критерію Колмогорова-Смірнова та критерію Хі-квадрат Пірсона (рис. 3).

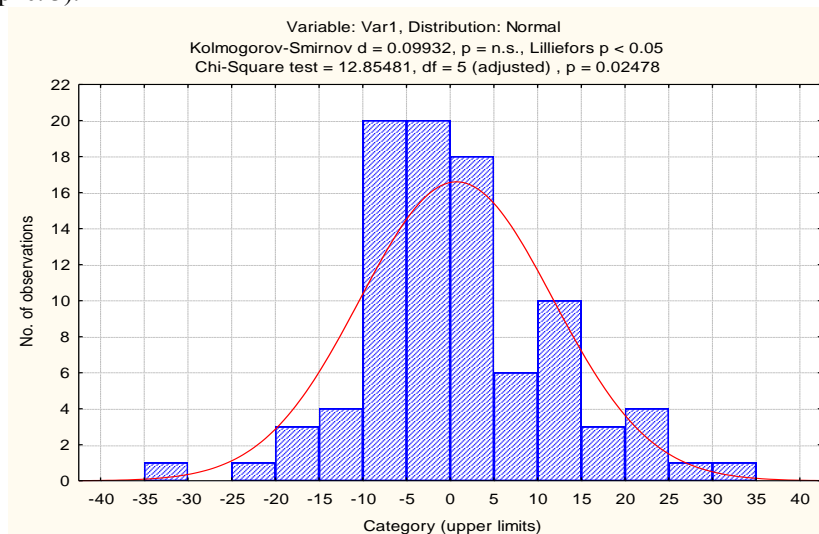


Рис. 3. Перевірка гіпотези про нормальний розподіл трендових перевищень врожайності. Об'єднана вибірка для Херсонської, Одеської, Кіровоградської та Запорізької областей. Дані за 2000 – 2021 роки.

Для вирішення багатьох проблем планування аграрного бізнесу не обов'язково мати точний прогноз

врожайності. Наприклад, для прийняття рішення щодо інвестування у конкретний проект достатньо мати оцінку майбутньої врожайності у термінах «висока врожайність» - «низька врожайність». При цьому термін «висока врожайність» означає таке значення врожайності, яке значно перевищує середній рівень врожайності. Такий підхід уможливує використання класифікаційних методів при прогнозуванні врожайності. Підтвердження нормального розподілу перевищень врожайності спрощує їх бінарну класифікацію. Ми пропонуємо віднести до групи «висока врожайність» ті значення врожайності, які з імовірністю  $p > 0.67$  розміщені на інтегральній кривій нормального розподілу трендових залишків  $F(eps)$  (рис. 4), тобто ті, для яких виконується умова

$$F(eps) > 0.67. \quad (1.3)$$

Значення врожайності, для яких не виконується умова (3), будуть віднесені до групи «низька врожайність».

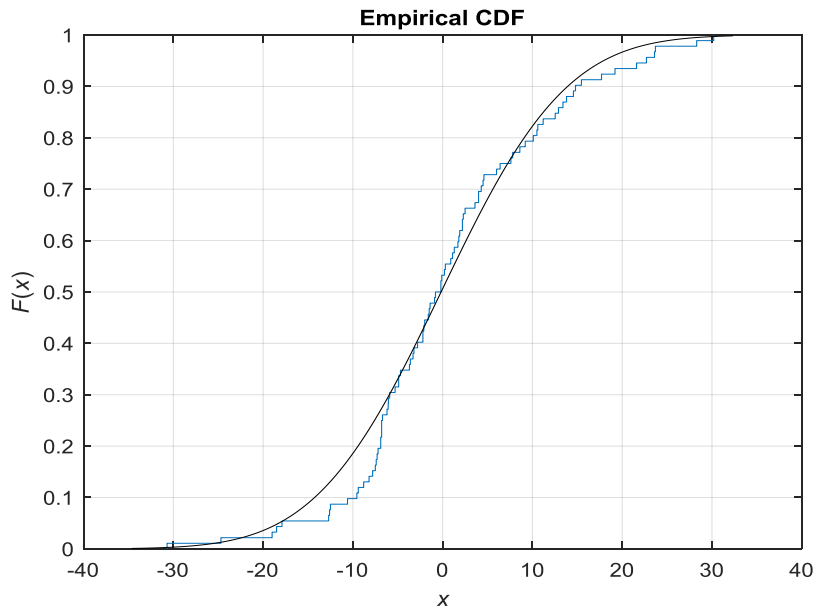


Рис. 4. Інтегральна крива нормального розподілу трендових залишків врожайності для 4-х областей степової зони України

Для реалізації класифікаційного підходу до прогнозування врожайності ми вводимо бінарну змінну  $bin$ , яка має лише два значення: 1 («висока врожайність») та 0 («низька врожайність»). Якщо головним завданням є прогнозування високої врожайності, значення змінної  $bin$  визначається за правилом

$$bin = \begin{cases} 1, & \text{if } F(eps) > 0.67; \\ 0, & \text{if } F(eps) \leq 0.67. \end{cases} \quad (1.4)$$

У деяких ситуаціях потрібно передбачити не високу врожайність, а її низьке значення (неврожай). Тоді вище описана задача класифікаційного прогнозування змінюється наступним чином. Термін «низька врожайність» означає таке значення врожайності, яке є значно нижчим від середнього рівня врожайності. До групи «низька врожайність» будуть віднесені ті значення врожайності, які з імовірністю  $p < 0.33$  розміщені на інтегральній кривій нормального розподілу трендових залишків, тобто ті, для яких виконується умова

$$F(eps) < 0.33. \quad (1.5)$$

Значення врожайності, для яких не виконується умова (1.5), будуть віднесені до групи «висока врожайність».

Бінарна змінна врожайності  $bin$  тепер буде визначатися за наступним правилом

$$bin = \begin{cases} 1, & \text{if } F(eps) < 0.33; \\ 0, & \text{if } F(eps) \geq 0.33. \end{cases} \quad (1.6)$$

#### Методика класифікаційного моделювання врожайності

Для оцінки врожайності у термінах «висока врожайність», «низька врожайність» необхідно здійснити попередню обробку даних врожайності. Для цього ми використали процедуру біннінгу, алгоритм якої полягає у розбитті набору історичних змінних на дві категорії, базуючись на оцінці перевищення врожайності над середнім (трендовим) значенням. Таким чином, побудова моделі врожайності передбачатиме проведення таких етапів [Клебан]:

- 1) збір даних для дослідження, формування тренувальної та тестової вибірок;
- 2) розбиття значень пояснюючих змінних на дві непересічні категорії;
- 3) побудова класифікаційної моделі на навчальній вибірці;
- 4) оцінка адекватності побудованих моделей на тестовій вибірці за критерієм AUROC (*area under receiver operating characteristic* [Fawcett, Tom]);

5) аналіз отриманих результатів та формування висновків щодо ефективності моделі.

Для побудови математичної моделі було використано набір даних врожайності пшениці у чотирьох областях степової зони України (Херсонська, Одеська, Запорізька, Кіровоградська) за період 2000 – 2021 роки та відповідні значення метеофакторів. Проблема вибору метеорологічних факторів, які найбільш впливають на врожайність зернових була розглянута в роботі [Грицюк, Бачишина]. Було показано, що найбільш суттєвим для майбутньої врожайності є вплив метеорологічних умов квітня, травня та червня, тобто періоду, коли відбувається основний ріст та формування рослини. У даній роботі досліджується вплив середньодекадних температур та місячних сум опадів вказаного періоду на значення врожайності пшениці. Місячні значення суми опадів були обрані нами з огляду на той факт, що в умовах степової зони багато декадних значень сум опадів є близькими до нуля. Ми використовували наступні позначення факторів (табл. 1)

Таблиця 1

Позначення впливаючих факторів

Позначення	Фактор	Період
<i>t1</i>	Середньодекадна температура	Перша декада квітня
<i>t2</i>	Середньодекадна температура	Друга декада квітня
<i>t3</i>	Середньодекадна температура	Третя декада квітня
<i>t4</i>	Середньодекадна температура	Перша декада травня
<i>t5</i>	Середньодекадна температура	Друга декада травня
<i>t6</i>	Середньодекадна температура	Третя декада травня
<i>t7</i>	Середньодекадна температура	Перша декада червня
<i>t8</i>	Середньодекадна температура	Друга декада червня
<i>t9</i>	Середньодекадна температура	Третя декада червня
<i>R10</i>	Місячна сума опадів	Квітень
<i>R20</i>	Місячна сума опадів	Травень
<i>R30</i>	Місячна сума опадів	Червень

Масив досліджуваних даних містить статистичні дані щодо врожайності пшениці [21] та 12 метеорологічних факторів [Климатический монитор] для 4-х областей степової зони за період 2000 – 2021 роки (22 роки), які мають числовий характер. Залежною змінною є бінарний показник *ers*, що вказує на факт перевищення врожайності пшениці над трендовим значенням (значення «1» - «висока врожайність», значення «0» - «низька врожайність»). Структура вхідного набору даних представлена у табл. 2. Дані навчальної та тестової вибірки розділені у пропорції 75/25, і не повторюються в межах моделі.

Таблиця 2

Початковий набір даних		
	«висока врожайність»	«низька врожайність»
Значення бінарної змінної	1	0
Кількість записів, од	31	57
Частка, %	35,2	64,8

Для побудови та тестування моделей використовувалася кросвалідація з розбиттям на 10 груп [Berrar, D]. У процесі перехресного тестування моделі склад навчальної та тестової вибірки змінювався випадковим чином, тому їх структуру вказати неможливо.

При моделюванні перевищення врожайності над трендовим значенням результатом розрахунку є дійсне число, неперервне на інтервалі [0;1], що визначає ймовірність перевищення. У машинному навчанні оцінка прогнозної здатності моделі здійснюється з використанням кривої AUC-ROC, яка дає змогу оцінити та візуалізувати ефективність класифікації [Fawcett]. Площа під ROC-кривою AUC (Area Under Curve) є агрегованою характеристикою якості класифікації. Чим вище значення AUC, тим краще модель ідентифікує зразок з характеристикою «висока врожайність». Дуже хороша модель класифікації має значення AUC, близьке до 1, погана модель має значення AUC менше від 0,5. При значеннях AUC > 0.75 класифікаційна модель вважається хорошою.

### Модель логістичної регресії

Класифікаційний підхід до прогнозування врожайності передбачає бінаризацію (біннінг) історичних даних врожайності, тобто переведення числових значень врожайності у бінарну змінну, яка набуватиме лише двох значень: «0» (низька врожайність) чи «1» (висока врожайність). Тоді регресійне моделювання врожайності полягає у моделюванні неперервної змінної, що приймає значення з інтервалу [0;1].

Для моделювання бінарних даних використовують моделі бінарного вибору: *probit*- та *logit*-моделі, результуючі значення яких можуть трактуватися як перевищення (неперевищення) врожайністю деякого порогового значення. Логістична регресія - регресія, у якій пояснювана змінна - факторна, а пояснюючі змінні можуть бути і кількісними і факторними.

Формальна постановка задачі полягає в наступному. Нехай змінна  $eps$  визначає величину перевищення врожайності над середнім (трендовим) значенням врожайності та набуває значень 0 чи 1 відповідно. Кожен об'єкт характеризується  $m$ -мірним вектором ознак, що в сукупності утворює масив інформації  $X = (x_{ij})$ , ( $j = 1, m$  – номер показника;  $i = 1, n$  – номер часового періоду).

Класичне рівняння лінійної регресії має вигляд:

$$P_1 = P(eps = 1) = X\beta \tag{2.2}$$

де  $\beta_m$  – шукані коефіцієнти регресії, змінна  $P_1 = P(eps=1)$  - ймовірність високої врожайності. Але лінійна модель ймовірності типу (2.2) не дасть потрібних значень залежної змінної на проміжку  $[0;1]$ . Для вирішення цієї проблеми використовують *probit*- та *logit*-моделі, в яких припускається, що:

$$P_1 = F(X\beta) \tag{2.4}$$

де  $F$  – функція, область значень якої належить інтервалу  $[0;1]$ .

Для реалізації функції  $F$  зазвичай використовують два види функцій розподілу:

- функція стандартного нормального розподілу (для *probit*-моделі):

$$F(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} ; \tag{2.5}$$

- функція логістичного розподілу (для *logit*-моделі):

$$F(z) = \frac{e^z}{1 + e^z} . \tag{2.6}$$

Для вибору найкращої моделі ймовірності необхідно оцінити значення коефіцієнтів  $\beta$  моделі (2.2). Найпоширенішим способом оцінки параметрів є метод максимальної правдоподібності.

З метою порівняння ефективності різних моделей для прогнозування перевищень врожайності побудуємо *probit*- та *logit*-моделі на основі навчальної вибірки з 66 випадково вибраних спостережень та 12 факторів. Для оцінки ефективності класифікації використано такі показники: матриця помилок (*Confusion matrix*), загальна точність класифікації за певного рівня розмежування (*Accuracy*), чутливість (*Sensitivity*), специфічність (*Specificity*) та площа під кривою (*AUROC*). Всі розрахунки проводилися в програмному середовищі R.

Матриця помилок будується на основі результатів класифікації моделлю та фактичної приналежності спостережень до класів. В контексті поставленої задачі в матриці наводиться чотири випадки:

- TP (*True Positives*) – модель вірно виявила перевищення врожайності;
- FP (*False Positives*) – низька врожайність помилково розпізнається як висока врожайність;
- FN (*False Negatives*) – висока врожайність помилково розпізнається як низька врожайність;
- TN (*True Negatives*) – модель вірно виявила випадок низької врожайності.

Матриця помилок має наступний вигляд:

Фактичні дані	Результати тесту	
	Низька врожайність	Висока врожайність
Низька врожайність	TN	FP
Висока врожайність	FN	TP

Побудуємо GLM-модель логістичної регресії для описаного вище набору даних, який описує 4 області степового регіону України. Змінна-відгук (перевищення врожайності над трендом) була бінаризована за правилом (1.4). Випадковим чином розіб'ємо набір на дві частини: навчальна вибірка (66 рядків) та тестова вибірка (22 рядки). У якості факторів GLM-моделі нами були вибрані змінні  $t7$  та  $R10$ , які, згідно з даними кореляційного аналізу, чинять найбільший вплив на врожайність. Характеристики моделі, побудованої на навчальній вибірці представлені у табл. 7. На основі побудованої моделі ми побудували прогноз на тестовій вибірці. Якість прогнозної класифікації визначає таблиця прогнозних оцінок табл. 8.

Таблиця 7

Характеристики GLM-моделі

	Estimate	SE	Z value	Pr(> z )
Intercept	14.3641	3.7441	3.84	0.00012
t7	-0.8398	0.2054	-4.09	4.3e-05
R10	0.0331	0.0151	2.20	0.02809
Null deviance: 84.020 on 65 degrees of freedom				
Residual deviance: 53.196 on 63 degrees of freedom				
AIC: 59.20				

Таблиця 2.8

## Характеристики якості класифікації за моделлю GLM

№	Модель	AUC	Sensitivity	Specificity	Accuracy
1	Модель <i>logit</i>	0.8890	0.8000	0.9167	0.8636
2	Модель <i>probit</i>	0.8800	0.7778	0.8462	0.8182

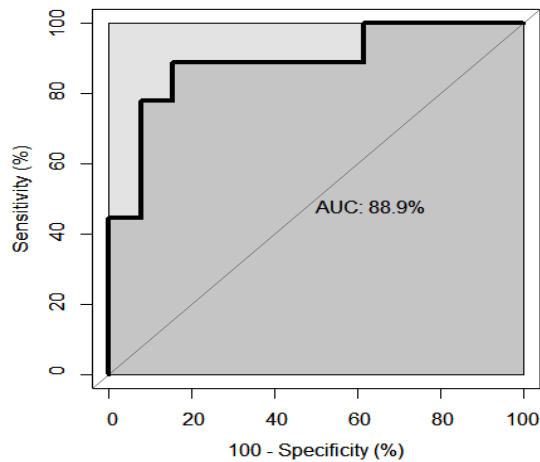


Рис. 4. ROC крива для тестової вибірки логіт моделі

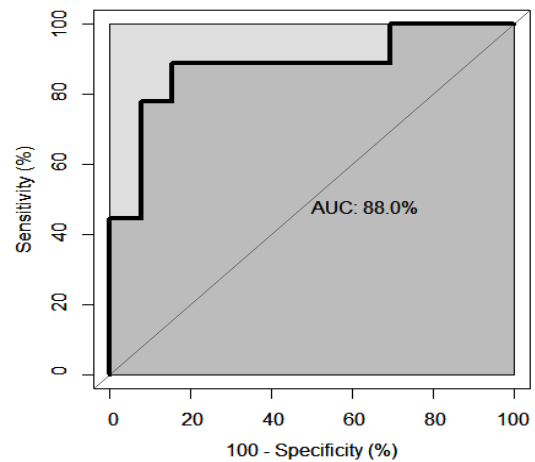


Рис. 2.5. ROC крива для тестової вибірки пробіт моделі

Якість логіт моделі є дещо вищою від якості пробіт моделі. Точність логіт моделі, яка вказує на частку правильно класифікованих значень врожайності на тестовій вибірці, становить 86.36%. Це свідчить про хорошу якість класифікаційної моделі. Для візуальної оцінки коректності моделі бінарної класифікації на рис 4 представлена ROC-крива помилок. Як видно з рисунка, агрегована характеристика якості класифікації логіт моделі становить AUC = 88.9%. Агрегована якість класифікація пробіт моделі становить 88.0%.

#### Висновки з даного дослідження і перспективи подальших розвідок у даному напрямі

Прогнозування врожайності є складною задачею, суть якої визначається випадковим характером багатьох впливаючих факторів. Тому для вирішення цієї задачі доцільно застосовувати методи інтелектуального аналізу даних із застосуванням сучасних комп'ютерних технологій. Дискусійним моментом є критерій якості прогнозу моделі. Похибка прогнозу не може виступати таким критерієм, адже більш важливими є показники правильності передбачення зростання/спадання врожайності та перевищення/недосягнення врожайності середнього/трендового значення.

Для вирішення багатьох задач при плануванні аграрного бізнесу не обов'язково мати точний прогноз врожайності. Для прийняття рішення щодо інвестування у конкретний проект достатньо мати оцінку майбутньої врожайності у термінах «висока врожайність» - «низька врожайність». При цьому термін «висока врожайність» означає таке значення врожайності, яке значно перевищує середнє/трендове значення врожайності. У даній роботі запропоновано класифікаційний підхід до прогнозування врожайності пшениці в областях степової зони України з використанням комп'ютерних моделей класифікації даних.

Нами побудовані дві моделі для класифікації перевищення врожайності над трендовим значенням, які ґрунтуються на логіт регресії та пробіт регресії. Обидві моделі забезпечують високу якість класифікації з незначною перевагою логіт моделі. Агрегована якість класифікації обох моделей перевищує 88%, що є дуже хорошим результатом для такої складної задачі як моделювання залежності врожайності від метеорологічних факторів в умовах невеликої вибірки даних. Запропонований нами підхід може бути використаний для класифікації значень врожайності інших сільськогосподарських культур з використанням інших методів машинного навчання. Проведені дослідження та запропонована методика можуть бути використані при прийнятті рішень щодо інвестування у проекти аграрної економіки.

#### Література

1. Грицюк П.М., Бачишина Л.Д. Вплив зміни кліматичних умов на динаміку врожайності зернових в Україні // Економіка України. – 2016, № 6 (655), с.68 – 75
2. J. Lee and M. J. Connaughton. Effects of Weather on Wheat Yields. - Irish Journal of Agricultural Research. - Vol. 8, No. 3, Dec., 1969. pp. 349-357.
3. Даніель Мюллер, Анне Юнгандreas, Фрідріх Кох та ін. Вплив кліматичних змін на виробництво пшениці в Україні. Німецько-український агрополітичний діалог. – Київ, 2016. – 45с.
4. Кернасюк Ю.В. Адаптація до кліматичних змін: економіка технологій поливу. – Економічний гектар. 03.06.2020. <http://agro-business.com.ua/agro/ekonomichnyi-hektar/item/17754-adaptatsiia-do-klimatychnykh-zmin-ekonomika-tekhnologii-polyvu.html>
5. О.О. Кривошеїн, Л.П. Однолеток, Л.П. Дзюба. Оцінка впливу погодних умов та організаційно-



технологічних заходів на урожайність озимої пшениці за її кліматичним потенціалом. – Наукові праці УкрНДГМІ, 2016, Вип. 269. – с.151-158

6. Кобченко Ю.Ф., Кобченко О.Ю., Резуненко В.А. Вплив погодних факторів на формування урожаю зернових культур у Харківській області. – Вісник Харківського національного університету. – 2014, №1098. – с.86-91

7. Kurt Heil, Anna Lehner and Urs Schmidhalter. Influence of Climate Conditions on the Temporal Development of Wheat Yields in a Long-Term Experiment in an Area with Pleistocene Loess. - *Climate* 2020, 8, 100, 17p. [www.mdpi.com/journal/climate](http://www.mdpi.com/journal/climate)

8. Beillouin D, Schauburger B, Bastos A, Ciaï P, Makowski D. 2020 Impact of extreme weather conditions on European crop production in 2018. - *Phil. Trans. R. Soc. B* 375. <https://doi.org/10.1098/rstb.2019.0510>

9. Konduri V., Vandal T., Ganguly S., Ganguly A. Data science for weather impacts on crop yield. *Front. Sustain. Food Syst.*, Vol. 4, May 2020

10. Державна служба статистики України. <http://www.ukrstat.gov.ua/>

11. Клебан Ю.В. Дослідження способів трансформації даних в контексті підвищення ефективності моделей кредитного скорингу. Нейронечіткі технології моделювання в економіці. Науково-аналітичний журнал. Київ, 2019. №8. С.94-123.

12. Fawcett, Tom. An Introduction to ROC Analysis. *Pattern Recognition Letters*. 27 (8) (2011). P. 861–874

13. Климатический монитор <http://www.pogodaiklimat.ru/>

14. Berrar, D. Cross-Validation. *Encycl. Bioinform. Comput. Biol.* 2019, 1, 542–545. URL: [https://www.researchgate.net/publication/324701535\\_Cross-Validation](https://www.researchgate.net/publication/324701535_Cross-Validation)

### References

1. Hrytsiuk P.M., Bachyshyna L.D. Vplyv zminy klimatychnyh umov na dynamiku vrozhaynosty zernovyh v Ukraini // *Economicka Ukrainy*. - 2016, № 6 (655), p.68 – 75

2. J. Lee and M. J. Connaughton. Effects of Weather on Wheat Yields. - *Irish Journal of Agricultural Research*. - Vol. 8, No. 3, Dec., 1969. pp. 349-357.

3. Daniel Muller, Anne Yunhandreas, Fridrih Koch et al. Vplyv klimatychnyh zmin na vyrobnyctvo pshenyци v Ukraini. *Nimecko-ukrainsky ahropolitychny dialoh*. – Kyiv, 2016. – 45p.

4. Kernasiuk Yu.V. Adaptaciya do klimatychnyh zmin: ekonomika tehnolohiy polyvu. – *Ekonomichny hektar*. 03.06.2020. <http://agro-business.com.ua/agro/ekonomichni-hektar/item/17754-adaptatsiia-do-klimatychnykh-zmin-ekonomika-tehnolohii-polyvu.html>

5. O.O. Kryvoshein, L.P. Odnoletok, L.P. Dzyuba. Ocinka vplyvu pohodnyh umov ta orhanizaciyno-tehnolohichnyh zahodiv na urozhaynist ozymoi pshenyци za ii klimatychnym potencialom. – *Naukovi pracі UkrNDHMI*, 2016, Vyp. 269. – с.151-158.

6. Kobchenko Yu.F., Kobchenko O.Yu., Rezunenکو V.A. Vplyv pohodnyh factoriv na formuvannya urozhayu zernovyh kultur u Harkivskiy oblasti. – *Visnyk Harkivskoho nacionalnoho universytetu*. – 2014, №1098. – p.86-91

7. Kurt Heil, Anna Lehner and Urs Schmidhalter. Influence of Climate Conditions on the Temporal Development of Wheat Yields in a Long-Term Experiment in an Area with Pleistocene Loess. - *Climate* 2020, 8, 100, 17p. [www.mdpi.com/journal/climate](http://www.mdpi.com/journal/climate)

8. Beillouin D, Schauburger B, Bastos A, Ciaï P, Makowski D. 2020 Impact of extreme weather conditions on European crop production in 2018. - *Phil. Trans. R. Soc. B* 375. <https://doi.org/10.1098/rstb.2019.0510>

9. Konduri V., Vandal T., Ganguly S., Ganguly A. Data science for weather impacts on crop yield. *Front. Sustain. Food Syst.*, Vol. 4, May 2020

10. Derzhavna sluzhba statystyky Ukrainy. <http://www.ukrstat.gov.ua/>

11. Kleban Yu.V. Doslidzhennia sposobiv transformacii danyh v kontexti pidvyshchennia efektyvnosti modeley kredytnoho skorynhu. *Nejronechitki tehnolohii modeluvannya v ekonomici*. Naukovo-analitychny zhurnal. Kyiv, 2019. №8. p.94-123.

12. Fawcett, Tom. An Introduction to ROC Analysis. *Pattern Recognition Letters*. 27 (8) (2011). P. 861–874

13. Klimaticheskiy monitor <http://www.pogodaiklimat.ru/>

14. Berrar, D. Cross-Validation. *Encycl. Bioinform. Comput. Biol.* 2019, 1, 542–545. URL: [https://www.researchgate.net/publication/324701535\\_Cross-Validation](https://www.researchgate.net/publication/324701535_Cross-Validation)