

ПРОГНОЗУВАННЯ ОБСЯГІВ ВИРОБНИЦТВА СІЛЬСЬКОГОСПОДАРСЬКОЇ ПРОДУКЦІЇ ЗА ДОПОМОГОЮ МЕТОДІВ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ДАНИХ

С.М. КОНЦЕБА, Р.І. ЛІЩУК, С.Д. СКУРТОЛ, Г.Ю. РОДАЩУК,
І.П. ВАСИЛЬЧЕНКО

Анотація. Спрогнозовано майбутні значення показників виробництва зернових і зернобобових культур у сільськогосподарських підприємствах Черкаської області на основі часових рядів динаміки, рівні яких виражено в натуральних одиницях вимірювання. Під час дослідження використано одну з методик інтелектуального аналізу даних — аналіз часових рядів, що дозволяє на основі моделі ряду динаміки, побудованої на базі фактичних значень попередніх показників минулих років, спрогнозувати виробництво на наступні роки. Методика містить такі етапи: графічний аналіз (для підбору відповідного рівняння моделі), виділення і аналіз детермінованих складових ряду, згладжування і фільтрація часового ряду, дослідження випадкової компоненти, побудова і перевірка на адекватність моделі часового ряду, прогнозування поведінки часового ряду на підставі проведених досліджень.

Ключові слова: інтелектуальний аналіз даних, часовий ряд, прогнозування, тренд, сільськогосподарське підприємство, модель.

ВСТУП

Питання прогнозування економічних процесів стоять досить гостро в періоди нестійкого економічного стану виробництва сільськогосподарської продукції. Застосування методик інтелектуального аналізу даних дозволяє підвищити ефективність виробництва продукції в сільськогосподарських підприємствах за рахунок оптимального управління запасами, доходами, витратами. Зростання ролі прогнозування розвитку виробництва сільськогосподарської продукції зумовлено ускладненням управлінських функцій в умовах динамічної і нестабільної ситуації як у зовнішньому середовищі, так і на самому ринку (поведінка фермерських господарств, які суттєво впливають на ринкову ситуацію, експорт-імпорт, особливо в умовах відкритості ринку, тінізація економіки тощо). Поряд з тим виробництво і реалізація сільськогосподарської продукції в Україні займає значну частку в економіці країни, тому використання методів інтелектуального аналізу для прогнозування показників діяльності підприємств агропромислового комплексу є актуальним.

Проблеми прийняття рішень в економічних системах на основі прогнозування показників з використанням методів інтелектуального аналізу даних розглянуто у працях вітчизняних авторів Р.О. Петрова, О.Я. Кучерука [1]. Прогнозуючи терміни продажу товарів, звертають увагу на те, що для прогнозування продажів найчастіше використовуються класичні методи аналізу

часових рядів та дерева рішень. П. І. Бідюк, С.М. Савченко, А.С. Савченко [2] визначили переваги методів інтелектуального аналізу та їх комбінацій (гібридні методи інтелектуального аналізу даних) для прийняття раціональних рішень у системі управління та прогнозування конкурентоспроможності вітчизняних підприємств. Група вчених Г. Чорноус, С. Рибальченко [3], П.І. Бідюк, А.В. Федоров [4], Л.О. Коршевнюк [5], О.Ю. Берзлев, М.М. Маляр, В.В. Ніколенко [6, 7] застосовують методи інтелектуального аналізу даних для прогнозування процесів ціноутворення, біржових показників, оцінювання фінансових ризиків.

Значно менше уваги звертають учені на використання методів інтелектуального аналізу даних для прогнозування показників діяльності підприємств агропромислового комплексу. В.Д. Кишенько, М.А. Сич [8] зазначають, що інтелектуальний аналіз даних, а саме часові ряди, дозволяє визначити стан технологічного комплексу цукрового заводу, зробити висновки про якість його функціонування, дати рекомендації щодо пошуку й усунення проблемних ситуацій. І.І. Глаголева, А.Ю. Берко [9] вважають, метод інтелектуального аналізу даних на основі прогнозування часових рядів для земельного кадастру дає змогу аналізувати дані та прогнозувати майбутні значення показників використання земельних ресурсів.

Питання прогнозування виробництва продукції в сільськогосподарських підприємствах з використанням часових рядів є недостатньо вивченим, залишається дискусійним і потребує подальших досліджень.

Мета роботи — спрогнозувати майбутні значення показників виробництва зернових і зернобобових культур у сільськогосподарських підприємствах Черкащини на 2020–2021 рр. на основі часових рядів динаміки, рівні яких виражені в натуральних одиницях вимірювання.

МЕТОДИКА ДОСЛІДЖЕНЬ

Під час дослідження використовувалась одна з методик інтелектуального аналізу даних, а саме: аналіз часових рядів для того, щоб на основі моделі ряду динаміки, побудованої на базі фактичних значень показника за 2000–2019 рр., виконати прогноз виробництва на 2020–2021 рр.

Методика містить такі етапи: графічний аналіз, виділення й аналіз детермінованих складових ряду, згладжування і фільтрація часового ряду, дослідження випадкової компоненти, побудова і перевірка на адекватність моделі часового ряду, прогнозування поведінки часового ряду на підставі проведених досліджень.

Для прикладу використано значення показників виробництва зернових і зернобобових культур у сільськогосподарських підприємствах Черкащини. Аналогічну методику досліджень можна використати для будь-яких видів продукції сільськогосподарського виробництва як у галузі рослинництва, так і в галузі тваринництва.

РЕЗУЛЬТАТИ ДОСЛІДЖЕНЬ

Показники багатьох економічних явищ і процесів змінюються в часі, характерними особливостями яких є те, що рівень показників у наступному часо-

вому періоді значною мірою залежить від їх рівня в минулому. Тенденцію зміни в часі кількісної міри досліджуваного показника можна подати аналітичною формулою, а у вибраній системі координат — плавною траєкторією або трендом. Послідовність спостережень одного показника (ознаки), упорядкована залежно від послідовно зростаючих або спадних значень часу, називають часовим рядом, який подано обсягами виробництва зернових і зернобобових культур у Черкаській області за 2006–2019 рр. (табл. 1).

Таблиця 1. Обсяги виробництва зернових і зернобобових культур у Черкаській області, тис. ц

Рік	Обсяги виробництва	Рік	Обсяги виробництва
2006	21086,0	2013	40684,8
2007	18424,2	2014	36996,5
2008	29780,3	2015	37454,9
2009	31939,6	2016	40917,0
2010	25310,7	2017	29265,5
2011	37618,8	2018	46440,1
2012	33105,5	2019	45598,3

Джерело: Рослинництво (1995–2019 рр.). Головне управління статистики у Черкаській області. URL: http://www.ck.ukrstat.gov.ua/source/arch/2020/roslinnictvo_19.pdf

Для моделювання тренду потрібно перевірити однорідність та порівнянність рівнів часового ряду y_t , а також визначити можливості їх зіставлення і стійкості. Попередній аналіз даних табл. 1 показує, що обсяги виробництва зернових і зернобобових культур в області за період 2006–2009 рр. нестабільні, варіюють у великому діапазоні, і тому для побудови часового ряду візьмемо дані за 2010–2019 рр., попередньо замінивши рівні часового ряду за 2010, 2011, 2013 та 2017 роки середньоарифметичними попередніх і наступних років (табл. 2).

Таблиця 2. Скориговані обсяги виробництва зернових і зернобобових культур для побудови тренду, тис. ц

Рік	Обсяги виробництва y_t	Рік	Обсяги виробництва y_t
2010	4698,0	2015	5665,0
2011	4739,0	2016	6073,0
2012	5194,0	2017	6034,0
2013	5051,0	2018	6117,0
2014	5257,0	2019	5928,0

Джерело: Розраховано авторами з використанням джерел: Рослинництво (1995–2019 рр.). Головне управління статистики у Черкаській області. URL: http://www.ck.ukrstat.gov.ua/source/arch/2020/roslinnictvo_19.pdf

Перевірені за методом Ірвіна [11, с. 409] рівні y_t скоригованого часового ряду не є аномальними ($\lambda_p < \lambda_{0,05} = 1,3$, якщо $n = 10$).

Для перевірки гіпотез про однорідність дисперсій та наявність тренду застосуємо метод перевірки різниць середніх рівнів.

Для дослідження однорідності дисперсій початковий часовий ряд $y_1, y_2, \dots, y_t, \dots, y_n$ розбивається на дві частини приблизно однієї довжини $n_1 = n_2 = 10/2 = 5$ і для кожної частини розраховуються [11, с. 409] (табл. 3):

- середні значення $y_{\text{сеп}_1} = \frac{\sum_t^{n_1} y_{t_1}}{n_1}$ і $y_{\text{сеп}_2} = \frac{\sum_t^{n_2} y_{t_2}}{n_2}$;
- дисперсії $\delta_1^2 = \frac{\sum_t^{n_1} (y_t - y_{\text{сеп}_1})^2}{(n_1 - 1)}$ і $\delta_2^2 = \frac{\sum_t^{n_2} (y_t - y_{\text{сеп}_2})^2}{(n_2 - 1)}$.

Таблиця 3. Розрахунок середніх значень та дисперсії часового ряду

t	y_t	$y_t - y_{\text{сеп}_1}$	$(y_t - y_{\text{сеп}_1})^2$
1	34779,2	951,1	904667,3
2	29208,1	-4620,0	21344030,4
3	33105,5	-722,6	522093,0
4	35051,0	1222,9	1495582,2
5	36996,5	3168,4	10039012,0
Разом	169140,3	-	34305384,9
Середнє значення $y_{\text{сеп}_1}$	33828,1	Дисперсія δ_1^2	8576346,2
6	37454,9	-5362,9	28760481,9
7	40917,0	-1900,8	3612964,6
8	43678,6	860,8	741011,1
9	46440,1	3622,3	13121202,2
10	45598,3	2780,5	7731291,5
Разом	214088,9	-	53966951,2
Середнє значення $y_{\text{сеп}_2}$	42817,8	Дисперсія δ_2^2	13491738

Джерело. Розраховано авторами з використанням праці [11, с. 409].

Однорідність дисперсій обох частин перевіряється порівнянням розрахованого критерію Фішера [11, с. 410] $F_p = \frac{\delta_2^2}{\delta_1^2} = \frac{13491738,0}{8576346,2} = 1,573$ з теоретичним значенням $F_{(0,05;n_1-1;n_2-1)} = 6,388$. Розраховане значення F_p менше за теоретичне, тому з імовірністю 95% гіпотеза про однорідність дисперсій приймається.

Гіпотеза про наявність тренду перевіряється за допомогою t -критерію Стьюдента, який розраховується за формулою [11, с. 410]

$$t_p = \frac{-|y_{\text{сеп}_1} - y_{\text{сеп}_2}|}{\delta \sqrt{1/n_1 + 1/n_2}} = \frac{-(33808,1 - 42817,8)}{3321,8 \sqrt{1/5 + 1/5}} = 4,279,$$

де δ — середньоквадратичне відхилення різниці середніх, яке розраховується за формулою

$$\delta = \sqrt{\frac{[(n_1 - 1)\delta_1^2 + (n_2 - 1)\delta_2^2]}{n_1 + n_2 - 2}} =$$

$$= \sqrt{\frac{[(5 - 1) * 8576346,3 + (5 - 1) * 13491738,0]}{5 + 5 - 2}} = 3321,8.$$

Розраховане значення $t_p = 4,279$ більше за теоретичне $t_{(0,05; 7)} = 2,365$, тому з імовірністю 95% гіпотеза про існування тренда приймається.

Аналітичне вирівнювання часового ряду виконаємо за методом експоненціального згладжування (табл. 4), використовуючи формулу

$$x_t = \gamma y_t + (1 - \gamma)x_{t-1},$$

де $\gamma = 0,03$ — параметр згладжування;

$$x_0 = x_1 \frac{34779,2 + 29208,1 + 33105,5}{3} = 32364,3 ;$$

x_{t-1} — рівні згладженого часового ряду.

Тенденції зміни часового ряду можуть бути встановлені добиранням емпіричної функції, яка має назву кривої зростання. Вибір форми кривої виконаємо за згладженим рядом на основі графічного зображення часового ряду (на ЕОМ – *Excel – Мастер діаграм*). Візуально найкращою формою залежності буде поліноміальна крива третього порядку (див. рисунок):

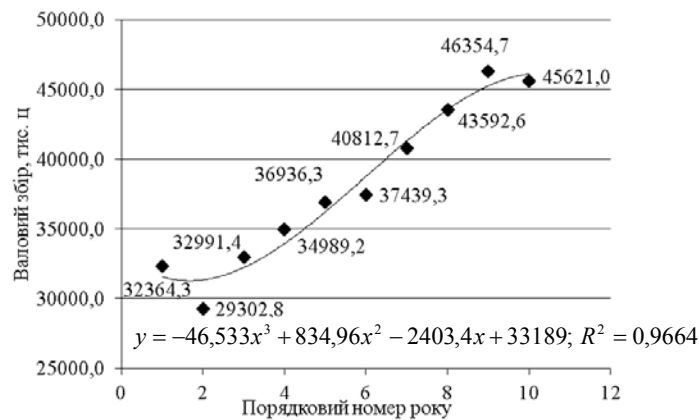
Таблиця 4. Згладжений часовий ряд

n	x_t
1	32364,3
2	29302,8
3	32991,4
4	34989,2
5	36936,3
6	37439,3
7	40812,7
8	43592,6
9	46354,7
10	45621,0

Джерело. Розраховано авторами.

$$y = 33189 - 2103,4t + 834,96t^2 - 46,533t^3.$$

Перевірка адекватності моделі тренду ґрунтується на перевірці виконання в залишках u_t таких властивостей:



Динамічний ряд виробництва зернових і зернобобових культур у Черкаській області за досліджуваний період

1. Незалежність значень рівнів залишків або відсутність між ними автокореляції досліджується за критерієм Дарбіна–Уотсона [11, с. 433] (табл. 5).

Таблиця 5. Розрахунки фактичного значення критерію Дарбіна–Уотсона

y_t	y_{tm}	u_t	u_t^2	u_{t-1}	$u_t - u_{t-1}$	$(u_t - u_{t-1})^2$
34779,2	31574,0	3205,2	10273134,0	–	–	–
29208,1	31349,8	–2141,7	4586776,1	3205,2	–5346,8	28588794,2
33105,5	32237,0	868,5	754207,1	–2141,7	3010,1	9060864,6
35051,0	33956,6	1094,4	1197606,3	868,5	225,9	51031,3
36996,5	36229,4	767,1	588480,8	1094,4	–327,2	107077,5
37454,9	38776,0	–1321,1	1745389,8	767,1	–2088,3	4360817,3
40917,0	41317,4	–400,4	160337,0	–1321,1	920,7	847708,7
43678,6	43574,3	104,3	10869,3	–400,4	504,7	254698,9
46440,1	45267,6	1172,5	1374749,2	104,3	1068,2	1141138,8
45598,3	46118,0	–519,7	270088,1	1172,5	–1692,2	2863530,7
Разом	–	–	20961637,6	–	–	47275662,0

Джерело. Розраховано авторами.

Фактичне значення критерію Дарбіна–Уотсона

$$DW_p = \frac{\sum_t^n (u_t - u_{t-1})^2}{\sum_t^n u_t^2} = \frac{47275662,0}{20961637,6} = 2,255.$$

Розраховане значення критерію Дарбіна–Уотсона $DW_p = 2,181$ міститься в межах $1,5 < DW_p < 2,5$, що свідчить про відсутність автокореляції в моделі тренду.

2. Дослідження відповідності розподілу ймовірностей рівнів залишків нормальному закону виконаємо за *RS*-методом [11, с. 429]: розраховується величина розмаху між максимальним та мінімальним рівнями ряду залишків

$$R = u_{\max} - u_{\min} = 3205,2 - (-2141,7) = 5346,9$$

та їх стандартне відхилення

$$S = \sqrt{\frac{\sum u_t^2}{n-1}} = \sqrt{\frac{20961637,6}{10-1}} = 1926,13.$$

Розраховане значення

$$RS_p = \frac{R}{S} = \frac{5346,9}{1926,13} = 2,775$$

потрапляє в інтервал між табличними межами, якщо $n = 10$, від 2,67 до 3,18, тому гіпотеза про нормальний закон випадкової складової приймається.

3. Рівність нулю математичного сподівання залишків — якщо ряд залишків підпорядкований нормальному закону розподілу ($S_n = 0$), то розраховується критерій Стюдента за формулою [11, с. 430]

$$t_p = \frac{u_{\text{сеп}} \cdot \sqrt{n}}{\delta_y} = \frac{282,9}{1618,7} \cdot 3,162 = 0,553,$$

де $u_{\text{сеп}}$ — середнє значення залишків; δ_y — середнє квадратичне відхилення; n — довжина часового ряду.

Розраховане значення $t_p = 0,553$ менше за теоретичне $t_{(0,05;7)} = 2,365$, тому гіпотеза про рівність нулю математичного сподівання залишків приймається.

4. Для дослідження випадковості коливань залишкових рівнів часового ряду застосуємо критерій піків (поворотних точок) [11, с. 428].

Крок 1. У послідовності залишків обираються поворотні точки (табл. 6).

Таблиця 6. Визначення піків або поворотних точок

t	y_t	y_{tm}	u_t	П
1	34779,2	31574,0	3205,2	–
2	29208,1	31349,8	–2141,7	1
3	33105,5	32237,0	868,5	0
4	35051,0	33956,6	1094,4	1
5	36996,5	36229,4	767,1	0
6	37454,9	38776,0	–1321,1	1
7	40917,0	41317,4	–400,4	0
8	43678,6	43574,3	104,3	0
9	46440,1	45267,6	1172,5	1
10	45598,3	46118,0	–519,7	–
Разом				4

Джерело. Розраховано авторами.

Крок 2. Розраховується загальна кількість піків (поворотних точок), яка дорівнює $\Pi = 4$.

Крок 3. Обчислюється математичне сподівання поворотних точок Π^* та їх дисперсія δ_n^2 за умови випадкової вибірки залишків:

$$\Pi^* = (2/3)(n - 2) = (2/3)(10 - 2) = 5,333$$

та

$$\delta_n^2 = \frac{16n - 29}{90} = \frac{16 \cdot 10 - 29}{90} = 1,456.$$

Крок 4. Критерієм випадковості залишків з 5%-м рівнем значущості має бути виконання нерівності

$$\Pi = 5 > [\Pi^* - 1,96\sqrt{\delta n^2}] = [5,333 - 1,96\sqrt{1,456}] = 2,969,$$

де квадратні дужки означають цілу частину числа. Оскільки $\Pi = 4 > [\Pi^* - 1,96\sqrt{\delta n^2}] = 2$, то випадковість коливань залишків з імовірністю 95% підтверджується.

Таким чином, побудована модель тренду (1) адекватна.

Точковий прогноз Y_{n+L} визначається показником, коли в рівняння моделі тренду підставляється значення часу $t = n + L$, що відповідає періоду упередження або прогнозованому періоду $t = n + 1 = 10 + 1 = 11$ (2020 р.) або

$t = n + 2 = 10 + 2 = 12$ (2021 р.) і становить:

- для 2020 р.

$$y_{11} = 33189 - 2103,4 \cdot 11 + 834,96 \cdot 11^2 - 46,533 \cdot 11^3 = 45846,3 \text{ тис. ц.}$$

- для 2021 р.

$$y_{12} = 33189 - 2103,4 \cdot 12 + 834,96 \cdot 12^2 - 46,533 \cdot 12^3 = 44173,4 \text{ тис. ц.}$$

Інтервальний прогноз для тренду, що має вид полінома третього порядку, розраховується за формулою [11, с. 437]

$$U_y = Y_{n+L} \mp t_\alpha S_y \sqrt{1 + \frac{1}{n} + \frac{t_L}{\sum t^2} + \frac{\sum t^4 - 2t_L^2 \sum t^2 + nt_L^4}{n \sum t^4 - (\sum t^2)^2}},$$

де Y_{n+L} — точковий прогноз за моделлю на $(n + L)$ -й період часу; t_α — теоретичне значення критерію Стьюдента для рівня значущості α ;

$$S_y = \sqrt{\frac{\sum u_t^2}{n - m}} = \sqrt{\frac{20961637,6}{10 - 3}} = 1730,5;$$

S_y — середнє квадратичне відхилення; t — порядковий номер рівня ряду ($t = 1, 2, \dots, n$); t_L — відповідає $(n + L)$ -му періоду часу, для якого виконується прогноз; N — довжина часового ряду; m — кількість параметрів моделі тренду.

Інтервальний прогноз виробництва зернових і зернобобових культур сільськогосподарськими підприємствами в Черкаській області становить:

- для 2020 р.

$$U_{11} = 45846,3 \mp 2,365 \cdot 1730,5 \times$$

$$\times \sqrt{1 + \frac{1}{10} + \frac{11}{385} + \frac{25333 - 2 \cdot 121 \cdot 385 + 10 \cdot 14641}{10 \cdot 25333 - 148225}} = 45846,3 \mp 5604,8;$$

- для 2021 р.

$$U_{12} = 44173,4 \mp 2,365 \cdot 1730,5 \times$$

$$\times \sqrt{1 + \frac{1}{10} + \frac{12}{385} + \frac{25333 - 2 \cdot 144 \cdot 385 + 10 \cdot 20736}{10 \cdot 25333 - 148225}} = 44173,4 \mp 6192,4.$$

Показники точності моделі тренда розраховуються за формулами [11, с. 430]:

- 1) середнє квадратичне відхилення

$$S_y = \sqrt{\frac{\sum u_t^2}{n - m}} = 1730,5;$$

- 2) середня відносна помилка апроксимації

$$\varepsilon_t = \frac{\frac{1}{n} \sum |u_t|}{y_t} \cdot 100 = \frac{32,77}{10} \cdot 100 = 3,3,$$

яка не перевищує 5%;

3) коефіцієнт збіжності

$$\varphi^2 = \frac{\sum_{t=1}^n u_t^2}{\sum_{t=1}^n (y_t - y_{cp})^2} = \frac{20961637,6}{290310000} = 0,07,$$

4) коефіцієнт детермінації

$$R^2 = 1 - \varphi^2 = 1 - 0,07 = 0,93.$$

Розраховані показники свідчать про високу точність прогнозу, оскільки середня відносна помилка дорівнює 3,3%, коефіцієнт детермінації вказує, що отримані спостереження на 93% підтверджують модель.

ВИСНОВКИ

Для дослідження прогнозованих значень показників виробництва зернових і зернобобових культур у сільськогосподарських підприємствах Черкаської області використано одну з методик інтелектуального аналізу даних: аналіз часових рядів для того, щоб на основі моделі ряду динаміки, побудованої на базі фактичних значень попередніх показників минулих років, виконати прогноз виробництва на наступні роки. Застосована методика включає в себе такі етапи: графічний аналіз (дає змогу підібрати рівняння моделі), виділення і аналіз детермінованих складових ряду, згладжування і фільтрація часового ряду, дослідження випадкової компоненти, побудова і перевірка на адекватність моделі часового ряду, прогнозування поведінки часового ряду на підставі результатів досліджень. Практична апробація моделі дозволила визначити, що зі збереженням тенденцій за попередні роки прогнозне значення обсягу виробництва зернових і зернобобових культур у Черкаській області на 2021 р. буде коливатися від 4091,4 до 5551,8 тис. ц, а на 2021 р. – від 4587,7,4 до 6317,9 тис. ц.

ЛІТЕРАТУРА

1. Р.О. Петров та О.Я. Кучерук, *Прогнозування термінів продажу товарів методами інтелектуального аналізу даних. Актуальні проблеми комп'ютерних наук*. 2019. Доступно: http://elar.khnu.km.ua/jspui/bitstream/123456789/7933/1/APKN-2019_%28v_2_0%29-156-158.pdf (дата звернення 20.10.2020).
2. П.І. Бідюк, С.М. Савченко, та А.С. Савченко, *Методи інтелектуального аналізу даних в прогнозуванні конкурентоспроможності підприємств*. Доступно: <http://www.ei-journal.in.ua/index.php/journal/article/view/61/48> (дата звернення 20.10.2020).
3. Г. Черноус та С. Рибальченко, “Оптимізація ціноутворення на основі моделей інтелектуального аналізу даних”, *Вісник Київського національного університету імені Тараса Шевченка*, №7 (172), с. 52–58, 2015.
4. П.І. Бідюк та А.В. Федоров, “Ймовірнісне прогнозування процесів ціноутворення на фондових ринках”, *Системні дослідження та інформаційні технології*, № 1, с. 65–73, 2009.
5. Л.О. Коршевнюк та Бідюк П.І., “Інформаційно-аналітична система для адаптивного прогнозування фінансових процесів та оцінювання ризиків”, *Наукові праці. Комп'ютерні технології*, вип. 201, т. 213, с. 59–62, 2013.
6. О.Ю. Берзлев, М.М. Маляр, та В.В. Ніколенко, “Адаптивні комбіновані моделі прогнозування біржових показників”, *Вісник Черкаського держ. технолог. ун-ту. Серія: технічні науки*, № 1, с. 50–54, 2011.

7. А.Ю. Берзлев, “Оценка эффективности прогнозирования и принятия решений на финансовом рынке”, *Problems of Computer Intellectualization*, V.M. Glushkov Institute of Cybernetics of NAS of Ukraine. Kyiv-Sofia: ITHEA, 2012, с. 249–257.
8. В.Д. Кишенько та М.А. Сич, “Інтелектуальний аналіз в задачах прогнозування тенденцій розвитку технологічного комплексу цукрового заводу”, *Цукор України*, № 6–7 (126–127), с. 36–40, 2016.
9. І.І. Глаголева та А.Ю. Берко, *Застосування методів інтелектуального аналізу даних для прогнозування використання земельних ресурсів*. Доступно: file:///C:/Users/user/Downloads/VNULPICM_2013_770_20.pdf (дата звернення 20.10.2020).
10. О.І. Черняк та П.В. Захарченко, *Інтелектуальний аналіз даних: підручник*. Київ, 2014, 599 с.
11. С.І. Наконечний, Т.О. Терещенко, та Т.П. Романюк, *Економетрія: підручник*, 2-ге вид., допов. та перероб. Київ: КНЕУ, 2000, 296 с.
12. О.В. Ульянченко, *Дослідження операцій в економіці : підручник для студентів вузів*. Харків: Гриф, 2002, 580 с.
13. К. Холден, Д.А. Піл, та Дж. Л. Томпсон, *Економічне прогнозування. Вступ*. Київ: Інформтехніка, 1996, 216 с.
14. С. Chatfield, *The Analysis of Time Series: An Introduction*. London, 1996, 283 p.

Надійшла 27.10.2020

INFORMATION ON THE ARTICLE

Serhii M. Kontseba, ORCID: 0000-0003-4161-5581, Uman National University of Horticulture, Ukraine, e-mail: kontseba@meta.ua

Roman I. Lishchuk, ORCID: 0000-0002-2051-5365, Uman National University of Horticulture, Ukraine, e-mail: roma0lir@gmail.com

Svitlana D. Skurtol, ORCID: 0000-0003-4517-2466, Uman National University of Horticulture, Ukraine, e-mail: skurtol@i.ua

Halyna Yu. Rodashchuk, ORCID: 0000-0002-0336-3435, Uman National University of Horticulture, Ukraine, e-mail: galya.rodashchuk@gmail.com

Ivan P. Vasylychenko, ORCID: 0000-0002-9341-539X, Uman National University of Horticulture, Ukraine, e-mail: ipvasylchenko@gmail.com

ПРОГНОЗИРОВАНИЕ ОБЪЕМОВ ПРОИЗВОДСТВА СЕЛЬСКОХОЗЯЙСТВЕННОЙ ПРОДУКЦИИ С ПОМОЩЬЮ МЕТОДОВ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ДАННЫХ / С.М. Концеба, Р.И. Лишук, С.Д. Скуртол, Г.Ю. Родащук, И.П. Васильченко

Аннотация. Спрогнозированы будущие значения показателей производства зерновых и зернобобовых культур в сельскохозяйственных предприятиях Черкасской области на основе временных рядов динамики, уровни которых выражены в натуральных единицах измерения. При исследовании использовалась одна из методик интеллектуального анализа данных — анализ временных рядов, что позволяет на основе модели ряда динамики, построенной на базе фактических значений показателей прошлых лет, спрогнозировать производство на будущие годы. Методика содержит следующие этапы: графический анализ (для подбора подходящего уравнения модели), выделение и анализ детерминированных составляющих ряда, сглаживание и фильтрация временного ряда, исследования случайной компоненты, построение и проверка на адекватность модели временного ряда, прогнозирование поведения временного ряда на основе проведенных исследований.

Ключевые слова: интеллектуальный анализ данных, временной ряд, прогнозирование, тренд, сельскохозяйственное предприятие, модель.

FORECASTING OF AGRICULTURAL PRODUCTION VOLUMES USING METHODS OF DATA MINING / S.M. Kontseba, R.I. Lishchuk, S.D. Skurtol, H.Yu. Rodashchuk, I.P. Vasylenko

Abstract. In this article, the future values of indicators were forecasted for production of grains and legumes on farms in Cherkasy region based on the time series expressed in physical units. Time series analysis as one of the data mining techniques was used during the research in order to make a forecast of production using the data (based on the model of dynamic series) from past years to predict the future production volumes. This method contains the following steps: a graphical analysis (allows you to choose the model equation in the best way), separation and analysis of deterministic components of the series, smoothing and filtering of time series, study of random components, construction and testing for the adequacy of the time series model, forecasting the behavior of the time series based on the conducted research.

Keywords: data mining, time series, forecasting, trend, agricultural enterprise, model.

REFERENCES

1. R.O. Petrov and O.Y. Kucheruk, "Forecasting the terms of sale of goods by the methods of intellectual data analysis", *Current problems of computer science*, pp. 156–158, 2019. Accessed on: October 20, 2020. [Online]. Available: http://elar.khnu.km.ua/jspui/bitstream/123456789/7933/1/APKN-2019_%28v_2_0%29-156-158.pdf
2. P. Bidyuk, A. Savchenko, and S. Savchenko, *Methods of intellectual data analysis in forecasting competitiveness of enterprises*. [Online]. Available: <http://www.ei-journal.in.ua/index.php/journal/article/view/61/48> [Accessed on: October 20, 2020].
3. H. Chornous and S. Rybalchenko, "Pricing optimization based on data mining models", *Bulletin of Taras Shevchenko National University of Kyiv*, no. 7 (172), pp. 52–58, 2015.
4. P. Bidyuk and A. Fedorov, "Probabilistic forecasting of pricing processes in stock markets", *System Research and Information Technologies*, no. 1, pp. 65–73, 2009.
5. L. Korshevnyuk and P. Bidyuk, "Information-analytical system for adaptive forecasting of financial processes and risk estimation", *Scientific works. Computer Technology*, vol. 201, no. 213, pp. 59–62, 2013.
6. O. Berzlev, M. Malyar, and V. Nikolenko, "Adaptive combined models of forecasting stock market indicators", *Bulletin of Cherkasy State Technological University*, pp. 50–54, 2011.
7. A. Berzlev, "Assessment of the effectiveness of forecasting and decision-making in the financial market", *Problems of Computer Intellectualization, V.M. Glushkov Institute of Cybernetics of NAS of Ukraine*. Kyiv-Sofia: ITHEA, pp. 249–257, 2012.
8. V. Kyshenko and M. Sych, "Intellectual analysis in the tasks of forecasting the trends in the technological complex of the sugar factory development", *Sugar of Ukraine*, no. 6-7 (126-127), pp. 36–40, 2016.
9. I. Hlaholeva and A. Berko, *Application of data mining methods for forecasting land use*. [Online]. Available: <http://science.lpnu.ua/sisn/all-volumes-and-issues/volume-770-2013/zastosuvannya-zasobiv-intelektualnogo-analizu-danih-dlya> [Accessed on: October 20, 2020].
10. O. Chernyak and P. Zakharchenko, *Data mining*. Kyiv, 2014, 599 p.
11. S. Nakonechnyy, T. Tereshchenko, and T. Romanyuk, *Econometric*, 2nd edition. Kyiv National Economic University, 2000, 296 p.
12. O. Ulyanchenko, *Research of operations in economics: a textbook for university students*. Kharkiv: Gryf, 2002, 580 p.
13. K. Holden, D.A. Peel, and J.L. Thompson, *Economic forecasting: an introduction*. Kyiv: Information technology, 1996, 216 p.
14. C. Chatfield, *The Analysis of Time Series: An Introduction*. London, 1996, 283p.