

СИСТЕМА УПРАВЛІННЯ ПРОЦЕСОМ ПРИГОТУВАННЯ ОПАРИ НА ОСНОВІ МОДЕЛІ РЕГРЕСІЙНОГО ДЕРЕВА

Н. А. Заєць, доктор технічних наук, професор

Національний університет біоресурсів і природокористування України

Н. М. Луцька, доктор технічних наук, професор

Національний університет харчових технологій

Л. О. Власенко, кандидат технічних наук, доцент

Національний університет біоресурсів і природокористування України

E-mail: z-n@ukr.net, vlasenko.lidia1@gmail.com, lutkanm2017@gmail.com

Анотація. Розроблена модель машинного навчання для системи управління процесом приготування опари на хлібозаводі, яка включає моніторинг, контроль та прогнозування якості опари. Система управління дозволить забезпечити постійний аналіз параметрів опари, таких як температура, вологість, підйомна сила та титранова кислотність, що забезпечить вчасне виявлення відхилень в якості опари та уникнення виробничих проблем. Крім того, вона сприятиме автоматизації процесу контролю якості та забезпечить стабільність виробничих процесів, що позитивно позначиться на репутації підприємства та задоволенні потреб споживачів. У рамках дослідження було розглянуто сім моделей машинного навчання, включаючи три лінійні та чотири різновиди бінарних дерев рішень. Проведений аналіз найкращої моделі – регресійного дерева рішень, підтверджена її ефективність та обґрунтованість отриманої ієрархічної структури, зокрема щодо передбачення важливих показників опари з високим ступенем достовірності, що свідчить про її потенціал у практичних застосуваннях у виробництві.

Ключові слова: *опара, хлібозавод, машинне навчання, модель, прогнозування, дерево рішень*

Актуальність. Початок 21-го століття характеризується зміною парадигми в підходах до харчового виробництва, що характеризується швидким розвитком технологій, підвищенням автоматизації та зростанням уваги до сталого розвитку. Зазвичай стратегія сталого розвитку хлібобулочного виробництва базується на принципі оцінки життєвого циклу хлібобулочних виробів, що включає комплексування традиційних виробничих [1-4], економічних, соціальних, політичних, безпекових, екологічних аспектів та новітніх підходів. Виробництво все

більше спрямовується на мінімізацію відходів та використання екологічно чистої упаковки. Підтримка здоров'я споживачів, прозорість виробництва та використання місцевих ресурсів стають ключовими пріоритетами, а активна комунікація зі споживачами допомагає пояснити переваги сталого розвитку.

Для досягнення сталого розвитку в хлібобулочному виробництві необхідно постійно проводити моніторинг технологічних процесів з подальшим прогнозуванням розвитку виробничих ситуацій. Це включає контроль за використанням енергії та ресурсів [5], оптимізацію виробничих процесів для зменшення відходів, впровадження екологічно чистих технологій та виробничих методів, а також постійний аналіз і вдосконалення виробничих процесів, пов'язаних з роботою технологічного обладнання [6-7] та розумних сенсорних систем [8] з метою покращення ефективності виробництва.

У сучасному харчовому виробництві, особливо на заводах з виробництва хлібобулочних виробів, якісні показники виробництва є одним із ключових факторів успішності. Заводи, що спеціалізуються на виготовленні хліба та інших хлібобулочних виробів, здійснюють складний технологічний процес, в якому кожний етап має велике значення для кінцевого результату. У цьому контексті важливість якісних показників виробничих процесів стає очевидною, оскільки вони впливають не лише на якість кінцевої продукції, але й на витрати сировини, робочої сили та енергії. Забезпечення стабільної якості продукції на заводі є важливою складовою конкурентоспроможності та впевненості споживачів у продукті.

У зв'язку з цим, методи автоматизації та впровадження сучасних технологій стають невід'ємною частиною управління виробництвом на хлібозаводах. Автоматизація дозволяє не лише підвищити ефективність та точність виробничих процесів, але й зменшити ризики людського фактору та впливу випадкових чинників на якість продукції. Одним із перспективних напрямків автоматизації виробництва на хлібозаводах є використання моделей технологічних процесів. Ці моделі, побудовані на базі моделей машинного навчання, можуть ефективно аналізувати великі обсяги даних, враховувати множину факторів та виконувати прогнози щодо параметрів виробництва. Розробка та впровадження таких моделей

може значно покращити якість та стабільність виробництва на хлібо заводах, сприяючи підвищенню ефективності та конкурентоспроможності підприємства.

Створення математичної моделі процесу приготування опари з прогнозуванням якісних показників може бути корисним з різних причин. Зокрема, така модель може допомогти виробникам оптимізувати процес виробництва опари, роблячи прогнози щодо оптимальних значень якісних змінних, які впливають на кінцеву продукцію. Також шляхом аналізу та прогнозування параметрів опари можна вдосконалити якість виробленого хліба або інших хлібобулочних виробів. Наприклад, оптимізація температури та вологості може покращити текстуру і смак хліба. Знання прогнозованих значень показників опари дозволить уникнути зайвих витрат матеріалів та енергії, що може призвести до економічних вигод для підприємства. Модель може бути використана для моніторингу та контролю якості продукції в реальному часі, що дозволить операторам швидко виявляти будь-які відхилення в показниках процесу та вчасно коригувати процес. Також шляхом аналізу змін у змінних процесу виробництва опари можна виявити передвісники можливих виробничих аварій чи відхилень у якості продукції, що дозволить вчасно приймати заходи для їх уникнення.

Отже, створення моделі процесу виробництва опари з прогнозуванням важливих якісних показників є невід'ємною частиною системи управління відділенням, що підвищить ефективність та якість виробництва в цілому на хлібо заводі.

Об'єктом дослідження є процеси при виготовленні опари як ключовий етап виробництва хлібобулочних виробів, що впливає на якість кінцевої продукції.

Предметом дослідження є моделі машинного навчання для визначення та прогнозування якості опари, що будуть використані в системі управління відділенням для моніторингу та контролю якості продукції в реальному часі. Це дозволить операторам швидко виявляти будь-які відхилення в показниках процесу виробництва опари та вчасно коригувати процес.

Аналіз останніх досліджень та публікацій. Машинне навчання, як напрямок штучного інтелекту, займається розробкою алгоритмів, які можуть навчатися з

даних [9], виявляє закономірності та патерни в процесі аналізу великих масивів даних, що дає змогу проводити моніторинг, прогнозування [10] процесів. Це допомагає аналізувати поведінку, виявляти причинно-наслідкові зв'язки в неочевидних залежностях. Широкого поширення алгоритми машинного навчання набули [11, 12] при: прогнозуванні на основі історичних даних; розпізнаванні образів і зображень; створенні інструментів і систем для аналізу і прийняття рішень на основі великих обсягів даних. Існує багато різних алгоритмів машинного навчання, кожен з яких має свої сильні та слабкі сторони. Деякі з найпоширеніших типів алгоритмів машинного навчання включають [13]: дерева рішень, методи Байєса, опорних векторів, нейронні мережі.

Дерева рішень реалізують спосіб організації даних у вигляді ієрархічної структури, де кожен "вузол" – це окреме рішення, отримане на основі обробки значень певних ознак [14]. Під час побудови дерева дані рекурсивно розбиваються на менші групи на основі конкретних правил, таких як порівняння значень ознак [15]. Цей процес триває до досягнення критерію зупинки, наприклад, заданої глибини дерева чи мінімальної кількості прикладів у кожному "листку". До переваг використання дерев рішень відноситься: легка інтерпретованість, можливість роботи з «сирими» (необробленими) та неоднорідним даними, ефективна робота з багатоозначковими наборами даних, доволі висока швидкість прогнозування. Недоліки дерев рішень зумовлені в переважній більшості обмеженнями й умовами їх застосування. Зокрема до них відносяться: схильність до перенавчання; нестійкість результату до навіть незначних змін в навчальних вибірках; неефективність при роботі з великими даними через значні часові затрати; недостатня точність прогнозування в порівнянні з деякими іншими методами, наприклад, нейронні мережі чи метод опорних векторів тощо.

Варіацією машинного навчання є ансамблевий метод випадкових лісів [16, 17], що створений для покращення точності та стійкості до перенавчання на основі обробки множини дерев рішень. Кожне дерево з множини тренується на підмножині навчальних даних, які включають випадкові атрибути. Після навчання всіх дерев

новий набір даних проходить кожне дерево в лісі, а потім визначається клас, який найчастіше зустрічається.

Таким чином, запропонована в роботі модель для прогнозування процесу виробництва опари допоможе покращити ефективність процесу приготування опари, зменшити відхилення якості продукції та забезпечити стабільність виробничих процесів.

Мета дослідження – розробка моделі щодо якісних показників опари для системи моніторингу та контролю якості опари в реальному часі, що надасть змогу підвищити якість кінцевої продукції.

Матеріали та методи дослідження. Якість опари має значний вплив на подальший процес виготовлення тіста та випікання хлібобулочних виробів. Опара є ключовим етапом виробництва, і її якість визначає такі аспекти, як текстура, смак, аромат та характеристики зовнішнього вигляду кінцевого продукту. Зокрема, якість опари впливає на здатність тіста до розширення під час ферментації. Якщо опара має високу підйомну силу, тісто легше розширюється, що призводить до більшого об'єму та пухкості кінцевого продукту. Якість опари впливає на структуру тіста, опара допомагає утворити стабільну мережу глютену, яка в свою чергу забезпечує еластичність та жорсткість тіста. Якщо опара недостатньо розвинена або має високий рівень титранової кислотності, це може призвести до слабкої структури тіста та низької підйомної сили. Якість опари впливає на смак та аромат кінцевого продукту, добра опара додає до тіста приємний аромат та багатий смак, що робить хліб більш смачним та привабливим для споживачів. Опара також може впливати на кольоровий вигляд кінцевого продукту. Наприклад, опара з надмірною температурою або недостатньо розвинена може призвести до неправильного забарвлення хліба.

Шляхом опитування досвідчених технологів, які працюють на хлібозаводах України довгий час, було встановлено 11 вхідних змінних та 4 вихідні змінні для моделі процесу приготування опари для приготування хліба українського, що включає дозування, заміс та бродіння. Ці змінні наведені в табл. 1.

1. Основні змінні процесу для моделі

Позначення		Опис
Вхідні змінні		
x_1	u_1	Витрата дріжджевого молока
x_2	z_1	Кількість і якість клейковини у борошні
x_3	z_2	Водопоглинаюча здатність борошна
x_4	z_3	Газоутворююча здатність борошна
x_5	z_4	Автолітична активність борошна
x_6	z_5	Кислотність борошна
x_7	z_6	Підйомна сила дріжджів
x_8	z_7	Кислотність дріжджів
x_9	u_2	Температура води
x_{10}	u_3	Час замісу опари
x_{11}	u_4	Час бродіння опари
Вихідні змінні		
y_1		Температура опари
y_2		Вологість опари
y_3		Підйомна сила опари
y_4		Титранова кислотність опари



Рис. 1. Залежність забарвленості та чистоти попередньо дефекованого соку від лужності

Для управління дозуванням продуктів використовується автоматизована система дозування, яка контролює подачу інгредієнтів, забезпечуючи правильну

пропорцію відповідно до рецептури. Це допомагає зберегти потрібну консистенцію та якість опари. Кількість дріжджового молока та температура води визначають основні умови для бродіння дріжджів, тому температура води та опари є критичними для активації дріжджів. Висока або низька температура може впливати на швидкість і рівномірність бродіння. Використання нагрівальних або охолоджувальних елементів дозволяє підтримувати стабільну температуру на кожному етапі процесу. Також час замісу впливає на структуру тіста та рівномірність розподілу інгредієнтів. Занадто короткий або довгий заміс може негативно позначитися на текстурі. Час бродіння регулюється в залежності від підйомної сили дріжджів, дозволяючи дріжджам виробляти газ, що створюють пухку та повітряну структуру опари.

Для моніторингу процесу необхідно контролювати якість сировини, а саме: показники якості борошна (клейковина, водопоглинання) та дріжджів (підйомна сила, кислотність), які перевіряються перед початком процесу, щоб забезпечити належну якість продукту. Це дозволяє виявити відхилення ще на ранньому етапі. Крім того, необхідно відстежувати показники опари, тобто під час процесу регулярно вимірюються температура, вологість та кислотність опари, що допомагає уникати відхилень. Ці параметри вказують на активність дріжджів та ступінь готовності опари до наступних етапів обробки. Датчики для відстеження температури, вологості та кислотності підключені до системи автоматичного контролю, яка може оперативно реагувати на відхилення та самостійно коригувати параметри, підтримуючи оптимальні умови для бродіння.

Дані моніторингу використовуються на моделі для прогнозування вихідних показників, що дозволяє адаптувати параметри управління. Система зворотного зв'язку допомагає оцінювати ефективність корекційних дій та вдосконалювати технологічний процес, зберігаючи стабільну якість та мінімізуючи витрати.

Результати досліджень та їх обговорення. Математична модель для оцінки та прогнозування якісних показників приготування опари описується такою залежністю:

$$\bar{y} = h(\bar{x}),$$

$$\bar{y} = [y_1, y_2, y_3, y_4], \bar{x} = [x_1, x_2, \dots, x_{11}], \quad (1)$$
$$\bar{x} = [z_1, \dots, z_7, u_1, \dots, u_4],$$

h – нелінійна регресійна модель (гіпотеза); y_i , x_i – відповідно входи та виходи моделі, причому вектор входів складається з вектору збурень та вектору управлінь (табл. 1).

Під час збору виробничих даних було виявлено декілька обмежень, що призвело до малої та недостатньо репрезентативної вибірки. По-перше, обмежений доступ до обладнання для автоматичного збору даних ускладнює процес, оскільки давачі та системи моніторингу можуть бути дорогими або важкодоступними. Додаткові труднощі створює недостатній персонал або обмежений час для збору даних, оскільки навантаження на заводі часто високе, що ускладнює виділення часу для цього процесу. Питання конфіденційності також обмежують доступ до даних для досліджень. Крім того, збір даних у рамках експериментальних проектів може мати низький пріоритет, що призводить до їх недостатнього обсягу для аналізу. Через обмежену кількість даних виникає потреба в ресемплінгу для покращення доступності та репрезентативності. Техніка ресемплінгу [18-19] допомагає збалансувати вибірку, включаючи рідкісні або недостатньо представлені випадки, що сприяє покращенню точності та надійності аналізу і моделювання.

Аналіз даних та побудова множини моделей машинного навчання виконувався в середовищі Google Colaboratory мовою Python (модуль sklearn). На рис. 2 та рис. 3 зображені діаграми розкидів вхідних і вихідних даних. Дані поділені на навчальну та тестову вибірки. Через обмежену кількість даних і можливість потрапляння важливих точок у тестову вибірку, для тестування були відібрані дані, найбільш схожі за матрицею евклідової відстані. Навчальна вибірка була збільшена за допомогою техніки бутстреп-ресемплінгу з поверненням.

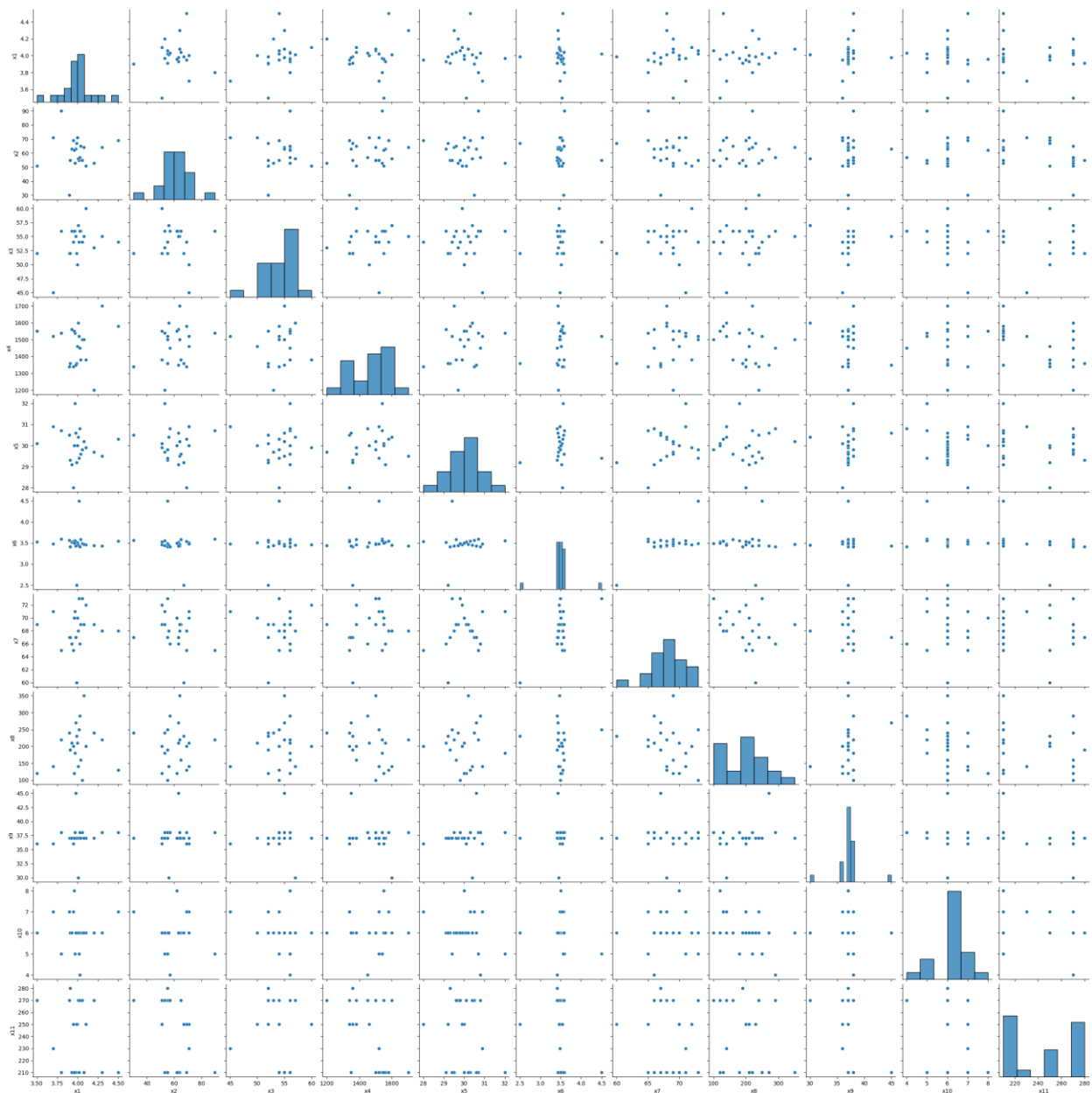


Рис. 2. Діаграма розкиду вхідних ознак

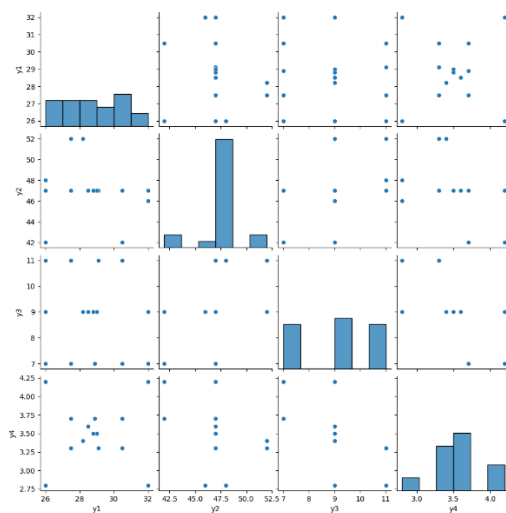


Рис. 3. Діаграма розкиду вихідних змінних

На рис. 4 відображено кореляційну матрицю (використовується теплова карта heatmap), що демонструє ступінь кореляції між різними змінними у наборі даних (Табл. 1). Цей вид візуалізації дозволяє швидко виявити та оцінити зв'язки між різними змінними. Зокрема, можна побачити, що всі обрані вхідні ознаки в різній мірі корелюються в вихідними ознаками.

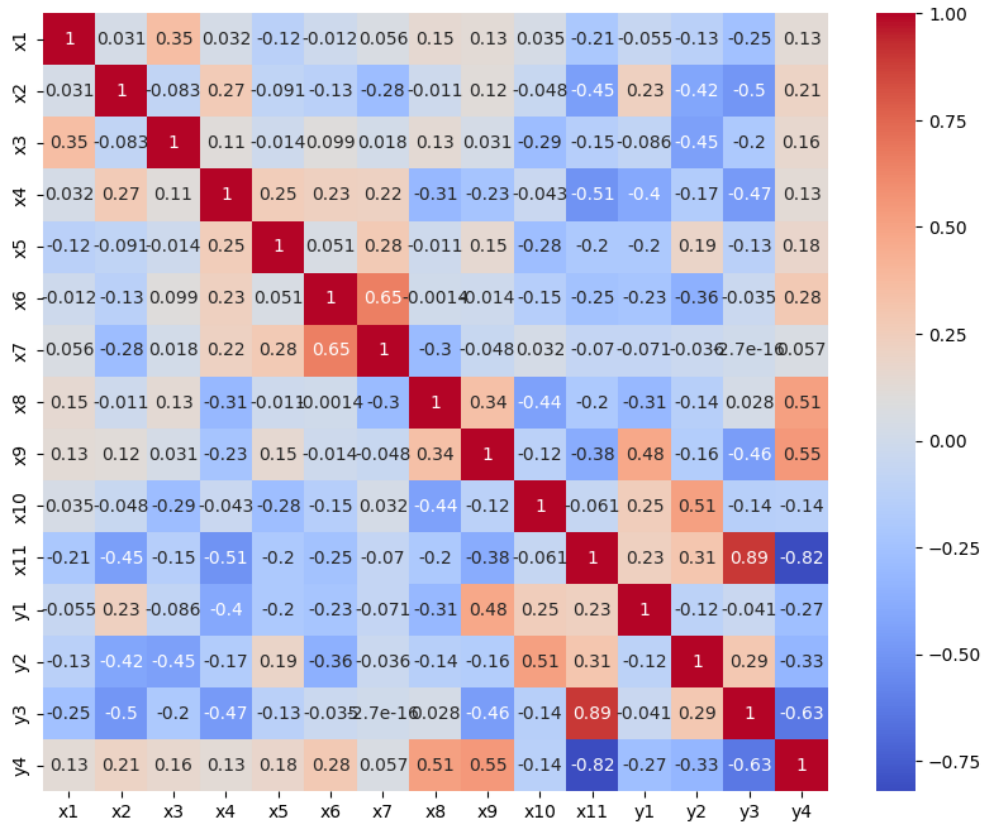


Рис. 4. Кореляційна матриця

Для побудови моделі обрано такі методи навчання під наглядом: лінійна модель – цей метод відмінно підходить для початкового аналізу даних, оскільки дозволяє оцінити лінійні відносини між вхідними та вихідними змінними; поліноміальна модель – оскільки вибірка мала, використання цього методу може допомогти захопити нелінійні залежності між вхідними та вихідними змінними, оскільки поліноми дозволяють апроксимувати складні залежності, що можуть бути присутні в даних, зберігаючи при цьому простоту моделі; лінійна модель за головними компонентами - з урахуванням обмеженого обсягу даних, PCA може допомогти зменшити розмірність простору ознак, зберігаючи при цьому важливу інформацію,

тому це зробить модель більш ефективною, зменшуючи кількість вимірювань та уникнення перенавчання; модель випадкового лісу – може бути корисною для невеликих наборів даних, оскільки вона використовує багато дерев для аналізу даних та зменшення ризику перенавчання; Gradient Boosting model – алгоритмом, який може покращити прогностичну здатність моделі шляхом ітеративного навчання послідовних моделей, враховуючи обмеженість даних, цей метод може допомогти максимізувати прогностичну здатність; AdaBoost regressor – метод, який може покращити точність моделі, особливо в умовах обмежених даних, він використовує ваги для вирішення проблеми невдалих прогнозів попередніх моделей; регресійне дерево рішень - простий і зрозумілий метод, який може бути корисним для вивчення залежностей у невеликих наборах даних. Враховуючи обмежену кількість точок даних, цей метод може швидко побудувати модель, яка може давати корисні прогнози.

Як критерій порівняння було обрано коефіцієнт множинної детермінації R^2 , що є метрикою, яка використовується в регресійному аналізі для вимірювання того, наскільки добре модель пояснює варіацію залежних змінних:

$$R^2 = 1 - \frac{\sum_{m=1}^M \sum_{n=1}^{N=4} [y_{mn} - h(\bar{x}_m)]^2}{\sum_{m=1}^M \sum_{n=1}^{N=4} [y_{mn} - \text{mean}(y_n)]^2}, \quad (2)$$

де M – кількість даних у вибірці; $\text{mean}(y_n)$ – вектор середніх значень n -го виміру в векторі \bar{y} .

Побудовано множину моделей машинного навчання під наглядом для оцінки та прогнозування якісних показників приготування опари, в процесі навчання було здійснено оптимізацію гіперпараметрів. Найкращі моделі та їх гіперпараметри наведені в табл. 2.

Найкраща модель – це модель бінарного дерева, значення множинного коефіцієнта детермінації на тестовій вибірці становить 0,96. Розглянемо причини, при яких можна було отримати такий результат. Порівняно з лінійними моделями дерево рішень має однозначну перевагу за рахунок вловлювання нелінійних залежностей. Крім того, імовірно дані містять малу частку шуму або викидів, тому це може негативно вплинути на результати лінійних моделей, таких як лінійна,

поліноміальна, лінійна PLS. Модель регресійного дерева більш стійка до шуму та викидів. Також модель регресійного дерева схильна до перенавчання меншою мірою завдяки своїй простій структурі в порівнянні з більш складними моделями, такими як RandomForestRegressor, GradientBoosting, AdaBoost.

1. Оцінки точності моделей

№	Модель	R^2	
		Навчальна вибірка	Тестова вибірка
1	Лінійна модель з масштабуванням: Ridge(alpha=0.1)	0.89	0.23
2	Поліноміальна модель з масштабуванням: make_pipeline(PolynomialFeatures(degree=2), Ridge(alpha=500))	0.68	0.54
3	Лінійна модель за головними компонентами: PLSRegression(n_components=5)	0.83	0.57
4	Регресійне дерево рішень: DecisionTreeRegressor(max_features=11, max_leaf_nodes=15, splitter='random')	0.99	0.96
5	Модель випадкового лісу (без ресімплінгу на даних): RandomForestRegressor(n_estimators=100, max_depth=12)	0.87	0.51
6	Gradient Boosting model (без ресімплінгу на даних): MultiOutputRegressor(GradientBoostingRegressor(n_estimators=20, max_features=11, validation_fraction=0.1, n_iter_no_change=10, learning_rate=0.1))	0.82	0.71
7	AdaBoost regressor (без ресімплінгу на даних): MultiOutputRegressor(AdaBoostRegressor(n_estimators=10))	0.97	0.69

Зважаючи на складність даних, їх малий обсяг та варіативність, множину розглянутих гіперпараметрів, ймовірність перенавчання та долю ймовірності в процесі навчання, для системи моніторингу обирається модель регресійного дерева. Крім того, така модель може бути більш зрозумілою та інтерпретованою порівняно з

іншими моделями. Це може бути корисно для розуміння того, як вхідні ознаки впливають на вихідні.

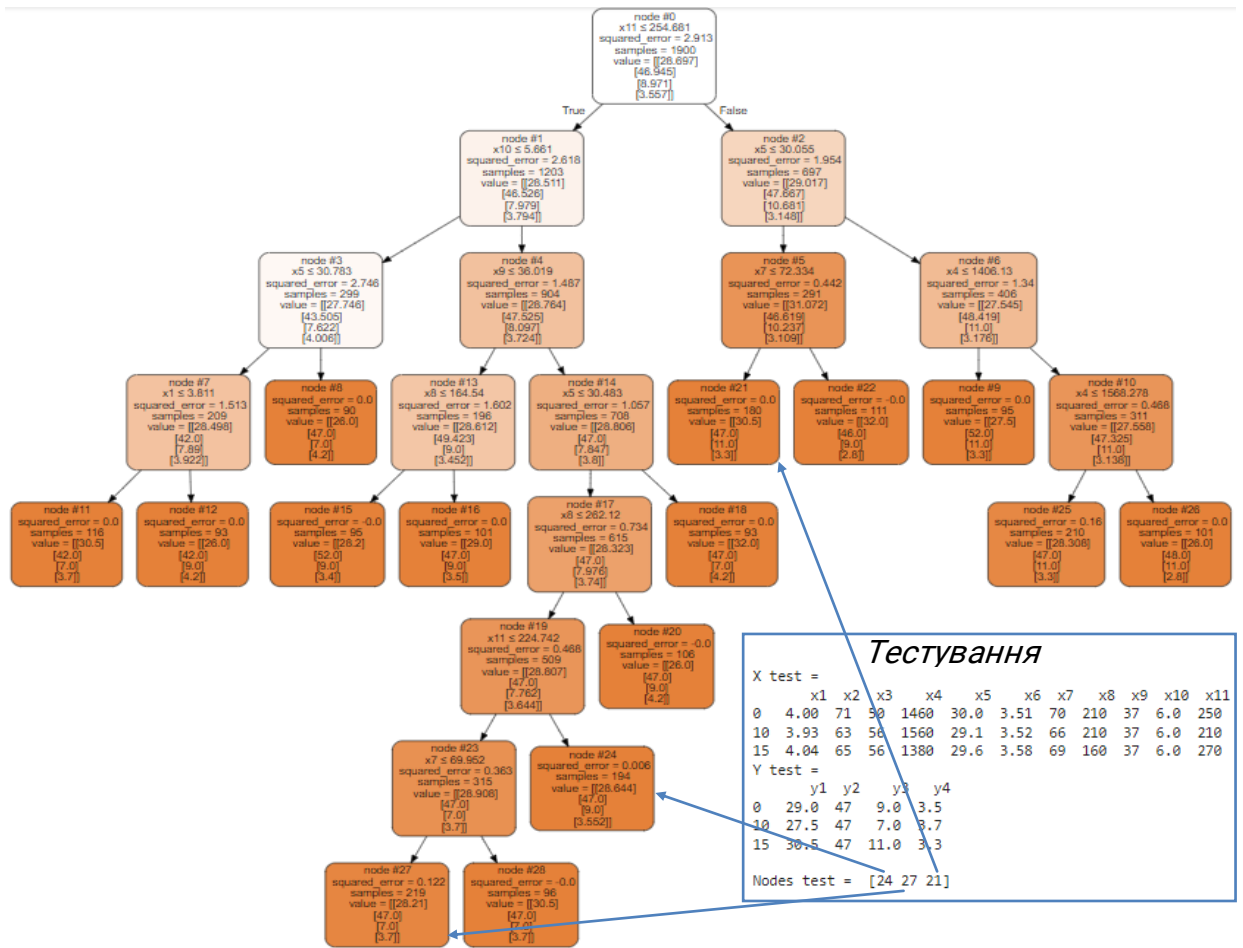


Рис. 5. Модель двійкового регресійного дерева та її тестування

На рис. 5 зображено обрану модель, а також результат навчання на тестових даних. На рис.6 відображено важливість ознак у дереві рішень. Найважливішими ознаками є автолітична активність борошна, час бродіння та час замісу опари. Ці ж ознаки формують корінь дерева – nodes #0 – #3. Вища автолітична активність призводить до більш інтенсивного розкладання клейковини та інших складових борошна під час процесу замісу опари, що впливає на якість та консистенцію опари. У моделі машинного навчання автолітична активність борошна може вважатися важливою ознакою через її потенційний вплив на характеристики опари, такі як підйомна сила та титранова кислотність опари, які ми спостерігаємо на вузлі #2, #3, #14. Умова для часу бродіння опари (node #0) з'являється в корені дерева, що відображає її як головну ознаку, де різні величини часу бродіння відокремлюють

різні шляхи розвитку моделі. Також ця ознака з'являється на node #19. Час замісу з'являється лише на nodes #1, тобто ця ознака є вирішальним для визначення властивостей опари чи характеристик кінцевого продукту лише в певних умовах, які визначаються іншими ознаками на шляху лівої гілки. На правій гілці спостерігаємо, що інформація, яку представляв час замісу, може бути узагальнена або представлена через інші ознаки, які були більш важливими для моделі на цій гілці.

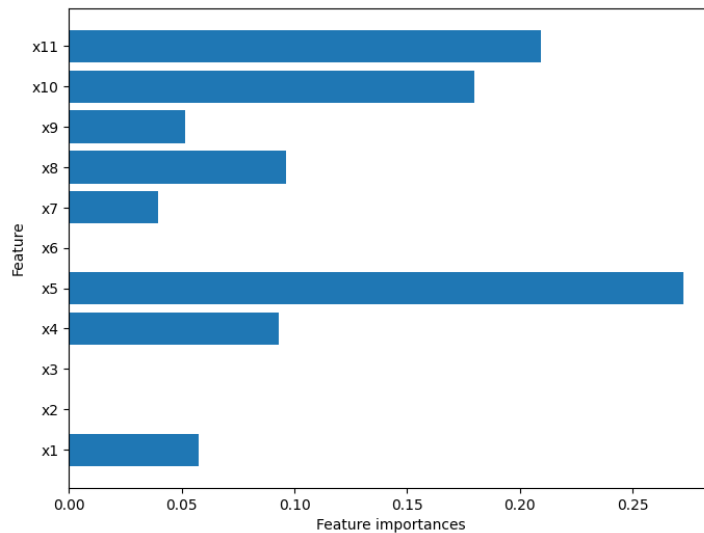


Рис. 6. Важливість ознак регресійного дерева

На рис. 7 зображено результати прогнозування поряд з реальними даними для двох важливих ознак, за якими можна зробити висновок про хорошу підгонку моделі.

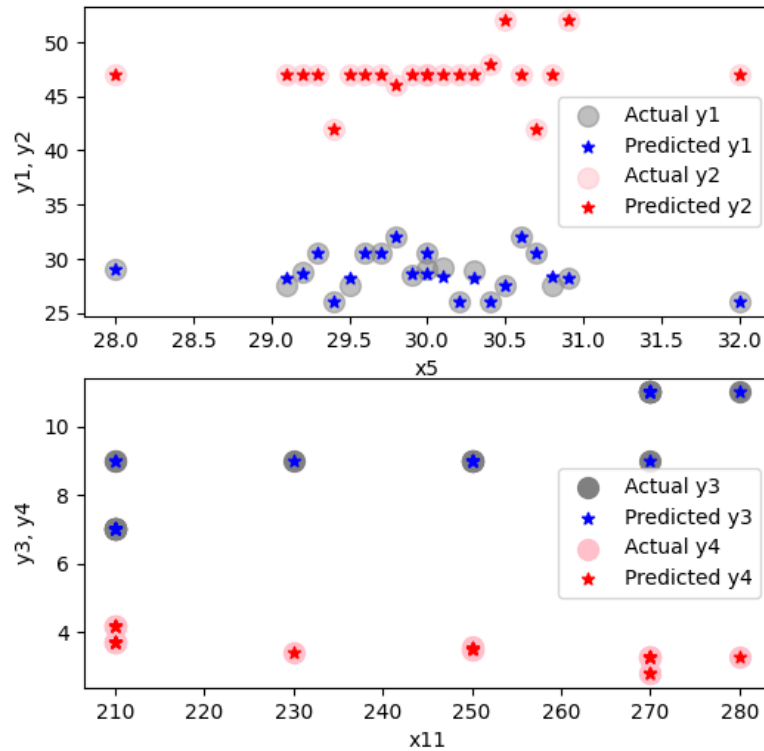


Рис. 7. Реальні дані та результати прогнозування

Висновки і перспективи. У роботі розроблена модель машинного навчання для системи моніторингу та контролю якості опари на хлібозаводі. Ця система забезпечує безперервний моніторинг параметрів опари, таких як температура, вологість, підйомна сила та титранова кислотність, що дозволяє вчасно виявляти відхилення в якості та запобігати виробничим проблемам, підвищуючи якість кінцевої продукції. Автоматизація контролю якості також сприяє стабільності процесів і позитивно впливає на репутацію підприємства та задоволення потреб споживачів. У дослідженні було побудовано сім моделей машинного навчання для оцінки якості опари: три лінійні моделі та чотири різновиди бінарних дерев рішень. Найкращою моделлю виявилось регресійне дерево рішень, яке дозволило визначити найважливіші ознаки, що впливають на якість опари, та забезпечити точне передбачення її параметрів.

У майбутньому планується розробка подібних моделей для інших виробничих етапів хлібозаводу і створення системи моніторингу всієї виробничої лінії.

References

1. R. Suchintita Das, B.K. Tiwari and M. Garcia-Vaquero (2023). "The Fundamentals of Bread Making: The Science of Bread." *Traditional European Breads: An Illustrative Compendium of Ancestral Knowledge and Cultural Heritage*. Cham: Springer International Publishing, 1-40, doi: [10.1007/978-3-031-23352-4_1](https://doi.org/10.1007/978-3-031-23352-4_1).
2. L. N. Meda, T. N. Sipiwe, and D. Ignatius (2023). "Optimization of baking parameters in producing bread through development of a generic control system." in *American Institute of Physics Conference Series*, 2581(1),. 030007, doi: [10.1063/5.0126294](https://doi.org/10.1063/5.0126294).
3. Dymchenko, G. Milan and G. Tomáš Gregor (2023). "Trends in bread waste utilisation," in *Trends in Food Science & Technology*, 132, 93-102, doi: [10.1016/j.tifs.2023.01.004](https://doi.org/10.1016/j.tifs.2023.01.004) .
4. K. Świąder and M. Marczevska (2021). "Trends of using sensory evaluation in new product development in the food industry in countries that belong to the eit regional innovation scheme", *Foods*, 10, 446, doi:[10.3390/foods10020446](https://doi.org/10.3390/foods10020446).
5. N. Zaiets, N. Lutska and L. Vlasenko (2022). "Improving the Efficiency of a Multistage Evaporator Station for Sugar Production Using Neural Networks," *2022 IEEE 4th International Conference on Modern Electrical and Energy System (MEES)*, Kremenchuk, Ukraine, 2022, 1-5, doi: [10.1109/MEES58014.2022.10005745](https://doi.org/10.1109/MEES58014.2022.10005745).
6. N. Lutska, N. Zaiets, L. Vlasenko (2024). "Development of Diagnostic System for the State of Electric Drives of Food Enterprise," in *Przegląd Elektrotechniczny*, 2024(2), 164, doi: [10.15199/48.2024.02.33](https://doi.org/10.15199/48.2024.02.33).
7. N. Zaiets, N. Lutska, L. Vlasenko and A. Zhylytsov (2023). "Forecasting Breakdowns of Electric Motors of a Sugar Factory Using Machine Learning Methods," *2023 IEEE 4th KhPI Week on Advanced Technology (KhPIWeek)*, Kharkiv, Ukraine, 2023, 1-6, doi: [10.1109/KhPIWeek61412.2023.10312974](https://doi.org/10.1109/KhPIWeek61412.2023.10312974).
8. B. di Diodoro, C. Bax, N. Dellarosa, F. Corazza, G. Langfelder and L. Capelli (2022). "Bread baking monitoring by smart sensory system: a feasibility study," *2022 IEEE International Symposium on Olfaction and Electronic Nose (ISOEN)*, Aveiro, Portugal, 2022, 1-4, doi: [10.1109/ISOEN54820.2022.9789664](https://doi.org/10.1109/ISOEN54820.2022.9789664).
9. Kumar, J. Rawat, N. Mohd, and S. Husain (2021). "Opportunities of artificial intelligence and machine learning in the food industry," *Journal of Food Quality*, 2021, 1-10, doi: [10.1155/2021/4535567](https://doi.org/10.1155/2021/4535567).
10. J.P. Usuga Cadavid, S. Lamouri, B. Grabot, R. Pellerin and A. Fortin (2020). "Machine learning applied in production planning and control: A state-of-the-art in the era of industry 4.0", *Journal of Intelligent Manufacturing*, 31, 1531-1558, doi: [10.1007/s10845-019-01531-7](https://doi.org/10.1007/s10845-019-01531-7).
11. J. Lee, Y. Kim and S. Kim (2023). "The Study of an Adaptive Bread Maker Using Machine Learning," *Foods*, 12(22), 4160, doi: [10.3390/foods12224160](https://doi.org/10.3390/foods12224160).
12. C. Gonzalez Viejo, N. M. Harris, and S. Fuentes (2022). "Quality Traits of Sourdough Bread Obtained by Novel Digital Technologies and Machine Learning Modelling," *Fermentation*, 8(10), 516, doi: [10.3390/fermentation8100516](https://doi.org/10.3390/fermentation8100516).
13. B. Charbuty, and A. Abdulazeez (2021). "Classification based on decision tree algorithm for machine learning," *Journal of Applied Science and Technology Trends*, 2(01), 20-28, doi: [10.38094/jastt20165](https://doi.org/10.38094/jastt20165).

14. ZH Zhou, Decision Trees (2021). *Machine Learning*. Springer, Singapore, 79–102, doi: 10.1007/978-981-15-1967-3_4

15. H. H. Patel, and P. Prajapati (2018). "Study and analysis of decision tree based classification algorithms," *International Journal of Computer Sciences and Engineering*, 6(10), 74-78, 2018.

16. S. Kaparthy, and D. Bumblauskas (2020). "Designing predictive maintenance systems using decision tree-based machine learning techniques," *International Journal of Quality & Reliability Management*, 37(4), 659-686, doi: [10.1108/IJQRM-04-2019-0131](https://doi.org/10.1108/IJQRM-04-2019-0131).

17. Antoniadis, S. Lambert-Lacroix, and J. M. Poggi (2021). "Random forests for global sensitivity analysis: A selective review," *Reliability Engineering & System Safety*, 206, 107312, doi: 10.1016/j.ress.2020.107312.

18. G. James, D. Witten, T. Hastie, R. Tibshirani and J. Taylor (2023). "Resampling methods," in *An Introduction to Statistical Learning: with Applications in Python*. Cham: Springer International Publishing, 201-228, doi: 10.1007/978-3-031-38747-0_5.

19. Yu Chong Ho (2019). "Resampling methods: concepts, applications, and justification," in *Practical Assessment, Research, and Evaluation*, 8 (19), 19, doi: [10.7275/9cms-my97](https://doi.org/10.7275/9cms-my97).

CONTROL SYSTEM FOR THE PROCESS OF PREPARATION OF DOUGH BASED ON THE REGRESSION TREE MODEL

N. Zaiets, N. Lutska, L. Vlasenko

Abstract. *The paper develops a machine learning model for the dough preparation process control system at a bakery, which includes dough quality monitoring, control and forecasting. The control system will ensure continuous analysis of dough parameters, such as temperature, humidity, lifting force and titran acidity, which will ensure timely detection of deviations in dough quality and avoidance of production problems. In addition, it will contribute to the automation of the quality control process and ensure the stability of production processes, which will have a positive impact on the company's reputation and customer satisfaction. The study considered seven machine learning models, including three linear and four types of binary decision trees. The analysis of the best model - a regression decision tree - confirmed its effectiveness and the validity of the resulting hierarchical structure, in particular, the prediction of important dough indicators with a high degree of reliability, which indicates its potential in practical applications in production.*

Key words: *dough, baking production, machine learning, monitoring, forecasting, decision tree*