

*Солдатенко Єлизавета,  
здобувачка першого (бакалаврського) рівня вищої освіти  
Навчально-науковий інститут біотехнологій та аквакультури  
Науковий керівник: **Наконечна Оксана,**  
кандидат технічних наук,  
доцент кафедри інформаційних технологій,  
Одеський державний аграрний університет,  
м. Одеса, Україна*

## **АНАЛІЗ ПРОДУКТИВНОСТІ ТВАРИН ЗА ДОПОМОГОЮ АЛГОРИТМІВ КЛАСИФІКАЦІЇ ТА КЛАСТЕРИЗАЦІЇ**

**Вступ.** Аналіз продуктивності тварин є критично важливим для підвищення ефективності тваринницької галузі. У сучасних умовах, коли аграрний сектор зосереджується на оптимізації виробничих процесів та зниженні витрат, застосування алгоритмів класифікації та кластеризації стає інноваційним рішенням для моніторингу і покращення продуктивності тварин. Використання цих методів дозволяє швидко й точно оцінювати стан кожної тварини та розподіляти їх за групами залежно від продуктивних показників. Це, в свою чергу, сприяє покращенню якості продукції, зниженню витрат на утримання та лікування, а також загальному підвищенню економічної ефективності господарств.

**Метою цього дослідження** є визначення ключових факторів, що впливають на продуктивність тварин, і застосування алгоритмів класифікації та кластеризації для їх аналізу. Це допоможе створити моделі, здатні передбачати продуктивність тварин та розподіляти їх на групи залежно від їхніх фізіологічних і поведінкових характеристик, що дозволить підвищити ефективність управління тваринницькими процесами.

Класифікація та кластеризація є ключовими поняттями в області аналізу даних, і кожен з цих методів виконує різні функції. Класифікація – це метод, що передбачає наявність заздалегідь визначених класів, до яких об'єкти можуть бути віднесені на основі їхніх характеристик. Наприклад, у тваринництві можна класифікувати тварин за їх продуктивністю, здоров'ям або породою. Цей метод базується на навчанні моделі на наявних даних, що дозволяє потім автоматично розподіляти нові об'єкти до відомих класів. Натомість, кластеризація передбачає поділ об'єктів на групи без попереднього визначення категорій. Цей підхід спирається на подібність між об'єктами, що дозволяє виявляти природні структури в даних. Розуміння цих основних відмінностей є важливим для вибору відповідного методу в залежності від задачі, яку необхідно вирішити [1].

У аналізі продуктивності тварин використовується кілька популярних алгоритмів класифікації. Один із них – дерева рішень (Decision Trees), які представляють собою інтуїтивний метод, що базується на послідовному поділі даних за допомогою критеріїв, що максимізують точність розподілу. Цей метод легко інтерпретується і дозволяє зрозуміти, як саме приймаються рішення. Іншим широко використовуваним алгоритмом є підтримуючі вектори (Support

#### Секція 4. Технології розробки інформаційних систем

Vector Machines, SVM), які створюють гіперплощину для розподілу об'єктів у просторі за класами, при цьому максимізуючи відстань між класами. Нейронні мережі, що імітують роботу людського мозку, є ще одним потужним інструментом, здатним виявляти складні взаємозв'язки в даних і ефективно працювати з великими обсягами інформації. Також, наївний байєсівський класифікатор, що базується на статистичних методах, є швидким у обробці великих обсягів даних, хоча його точність може бути помірною [2].

Для розподілу тварин за групами продуктивності можуть використовуватися різні алгоритми кластеризації. Наприклад, метод K-means ґрунтується на визначенні центроїдів груп і мінімізації відстані між точками та центроїдом кластера. Цей метод простий у використанні і швидкий, але чутливий до початкового вибору центроїдів. Ієрархічна кластеризація, у свою чергу, будує дерево кластерів, об'єднуючи схожі об'єкти на кожному рівні ієрархії. Цей метод може бути агломеративним (коли малі кластери об'єднуються) або дивізивним (коли більші кластери поділяються), що дозволяє глибше аналізувати структуру даних. Також слід згадати DBSCAN (Density-Based Spatial Clustering of Applications with Noise), алгоритм, що визначає кластери на основі щільності об'єктів у просторі. Цей метод є особливо корисним для кластеризації даних, що містять шум або аномальні значення, що робить його ефективним для аналізу в складних умовах [3].

Аналіз продуктивності тварин потребує різноманітних джерел інформації, що допомагають отримати об'єктивну картину їх стану та продуктивності. Одним із основних джерел є фізіологічні параметри, до яких належать температура тіла, частота серцебиття, рівень активності, вагові показники та інші показники, які відображають загальний стан здоров'я тварини. Ці дані є важливими, оскільки можуть вказувати на потенційні проблеми зі здоров'ям або ж на високий рівень продуктивності. Іншим важливим аспектом є дані про харчування, які включають інформацію про раціон, кількість та якість споживаної їжі. Ці показники безпосередньо впливають на ріст, розвиток і продуктивність тварин, що робить їх аналіз критично важливим. Також не слід забувати про умови утримання тварин, такі як температура, вологість, освітлення та якість повітря в приміщенні, адже всі ці фактори мають значний вплив на добробут і, відповідно, продуктивність тварин [4].

Для забезпечення точності та зменшення людського фактора у зборі даних про продуктивність тварин використовуються різноманітні технології автоматизованого збору. Серед них важливу роль відіграють датчики – біосенсори, які дозволяють безперервно відстежувати фізіологічні показники тварин, такі як температура та серцевий ритм. Це дозволяє отримувати актуальну інформацію про стан здоров'я тварин у режимі реального часу. Камери відеоспостереження також є потужним інструментом для аналізу поведінки та активності тварин. Використання технологій комп'ютерного зору для аналізу зображень та відео дозволяє виявляти аномалії у поведінці або зміни фізичного стану, що є важливим для моніторингу продуктивності. Попри переваги автоматизованих систем, традиційний ручний збір даних залишається

#### Секція 4. Технології розробки інформаційних систем

актуальним. Він включає фіксацію параметрів тварин працівниками ферми, що дозволяє підтверджувати результати автоматизованих систем і забезпечувати їх надійність [5].

Після збору даних їх підготовка до подальшого аналізу є важливим етапом, що включає кілька ключових процесів. По-перше, необхідно виявити та видалити шуми в даних, що можуть виникати через збої в роботі обладнання або інші технічні причини. Це важливо для забезпечення коректності та точності аналізу. По-друге, заповнення пропущених значень є критично важливим, оскільки відсутні дані можуть спотворити результати аналізу. Для цього використовуються різні методи коригування, такі як інтерполяція або заповнення середніми значеннями. Крім того, нормалізація та стандартизація даних дозволяє привести їх до одного формату, що підвищує точність алгоритмів класифікації та кластеризації. Цей процес забезпечує однорідність даних, що сприяє їх правильному аналізу і, як наслідок, отриманню коректних результатів у дослідженні продуктивності тварин [6].

Процес побудови класифікаційних моделей для виділення груп тварин з різною продуктивністю є важливим етапом у тваринництві, що дозволяє оптимізувати виробництво. Основним завданням є моделювання високопродуктивних груп, для чого розробляються моделі, що базуються на фізіологічних і поведінкових показниках тварин. Це передбачає збір та аналіз даних, таких як вага, вік, стан здоров'я, а також поведінка під час годування і взаємодії з іншими тваринами. Використання алгоритмів для класифікації тварин за рівнями продуктивності, таких як дерева рішень, нейронні мережі або метод опорних векторів (SVM), дозволяє ефективно ідентифікувати особин з високим потенціалом. Ці моделі допомагають передбачити продуктивність тварин у майбутньому, забезпечуючи основу для прийняття рішень щодо їхньої селекції та управління [7].

Визначення ключових факторів, що впливають на продуктивність тварин, є важливим для створення точних класифікаційних моделей. У цьому контексті проводиться аналіз значущості змінних, що дозволяє виявити, які фактори, такі як фізіологічні параметри, умови утримання та харчування, мають найбільший вплив на продуктивність. Використання статистичних методів та алгоритмів машинного навчання дає змогу не лише виявити ці фактори, але й визначити їхні вагові коефіцієнти в моделі. Це важливо для пріоритетизації тих змінних, які найбільше впливають на результати, що, в свою чергу, дозволяє краще налаштувати умови утримання та годівлі тварин, сприяючи підвищенню продуктивності [8].

Практичне застосування класифікаційних моделей для аналізу і прогнозування продуктивності тварин забезпечує можливість підвищення ефективності тваринництва. Використання різноманітних алгоритмів класифікації, таких як k-ближчих сусідів (k-NN), логістична регресія та дерева рішень, дозволяє розподілити тварин за класами продуктивності, виявляючи високопродуктивні групи. Інтеграція цих класифікаційних моделей у системи управління фермою може значно оптимізувати догляд за тваринами. Зокрема,

#### Секція 4. Технології розробки інформаційних систем

фермери зможуть вчасно реагувати на зміни в продуктивності та коригувати умови утримання, що призведе до підвищення загальної продуктивності та зменшення витрат. Це, у свою чергу, сприятиме досягненню більшої ефективності виробництва в тваринництві [9].

Кластеризація є ефективним методом для поділу тварин на групи на основі подібних фізіологічних та поведінкових ознак. Це дозволяє глибше аналізувати варіативність серед тварин, що має вирішальне значення для оптимізації умов утримання та підвищення продуктивності. Використовуючи алгоритми, такі як K-means та ієрархічна кластеризація [10], можна згрупувати тварин за показниками здоров'я, фізичними параметрами, рівнем активності та іншими продуктивними характеристиками. Завдяки цьому методологічному підходу можна виділити природні групи тварин у наявних даних, що дає змогу краще зрозуміти відмінності в продуктивності та стані здоров'я у кожній групі. Наприклад, можна розділити тварин на групи з високою, середньою та низькою продуктивністю, вивчаючи їхні фізіологічні та поведінкові патерни, що дозволяє створювати спеціалізовані стратегії для кожної групи.

Після виконання кластеризації необхідно провести детальний аналіз особливостей і характеристик кожного кластера. Цей крок дозволяє виділити типові показники для кожної групи, наприклад, продуктивність, умови утримання та реакцію на різні зовнішні фактори. Вивчення таких характеристик дає змогу визначити ключові ознаки, які найбільше впливають на продуктивність, а також зрозуміти, чому деякі групи демонструють вищу продуктивність, ніж інші. Ідентифікація цих ключових ознак допомагає у вдосконаленні управлінських рішень, оскільки стає зрозумілим, на які параметри потрібно звертати більше уваги, щоб забезпечити максимальну продуктивність і добробут тварин.

На основі отриманих кластерів та їхніх характеристик можна розробити конкретні рекомендації щодо поліпшення умов утримання для кожної групи. Ці рекомендації дозволяють адаптувати умови догляду, харчування та моніторингу відповідно до особливостей тварин у кожному кластері. Для кожної групи, виходячи з її типових ознак, можна створити спеціалізовані умови, які сприятимуть кращій продуктивності та зниженню стресу у тварин. Наприклад, групи з високою продуктивністю можуть потребувати додаткового спостереження та покращеного харчування для підтримки цього рівня, тоді як для груп з нижчою продуктивністю можуть бути корисні інші заходи, спрямовані на покращення їхніх умов утримання. Таким чином, завдяки кластеризації, умови утримання тварин можна зробити максимально ефективними, що в підсумку призведе до покращення загальної продуктивності тваринницького господарства.

Аналіз продуктивності тварин вимагає вибору оптимальних алгоритмів класифікації та кластеризації для досягнення максимальної точності та надійності результатів. Для оцінки точності класифікаційних моделей використовуються різні метрики, зокрема, точність (accuracy), F-мір, а також значення ROC-AUC. Точність є базовою мірою успіху, що визначає відсоток

#### Секція 4. Технології розробки інформаційних систем

правильно передбачених випадків у загальній кількості спостережень. F-мір є важливою метрикою для моделей з дисбалансом класів, оскільки дозволяє врахувати як точність, так і повноту класифікації, що критично важливо в ситуаціях, коли неправильно класифіковані випадки можуть призвести до втрат або ризиків для здоров'я тварин. ROC-AUC є інтегральною характеристикою ефективності моделі, що показує якість розділення класів на основі різних порогів, дозволяючи оцінити модель у випадках з декількома класами або критичними рішеннями [10].

Для оцінки якості кластеризації в завданнях групування тварин застосовуються спеціальні метрики, такі як силка (silhouette score) та інерція (inertia). Силка є індикатором компактності кластерів і дозволяє оцінити, наскільки тварини всередині одного кластера схожі між собою та водночас віддалені від тварин з інших кластерів. Це особливо корисно для визначення, чи є кластери чітко вираженими або ж спостерігається їхнє змішання. Інерція вимірює сукупну відстань від кожного об'єкта до його центроїда, і її мінімізація сприяє формуванню більш компактних та однорідних кластерів, що важливо для виявлення чітких закономірностей у поведінці чи фізіологічних показниках тварин.

Після порівняння результатів застосування різних алгоритмів класифікації та кластеризації на основі зазначених метрик можна сформулювати рекомендації щодо найбільш підходящих методів для аналізу продуктивності у тваринництві. Зокрема, для задачі класифікації доцільно використовувати дерева рішень або нейронні мережі, які добре справляються з великими обсягами даних та складними зв'язками між характеристиками. Для кластеризації ефективними є алгоритми K-means і ієрархічна кластеризація, які дозволяють формувати чіткі групи на основі фізіологічних та поведінкових показників тварин. Вибір конкретного алгоритму залежить від доступних ресурсів, обсягу та якості даних, а також від мети дослідження. Використання оптимальних методів дозволяє підвищити точність аналізу та приймати обґрунтовані рішення, спрямовані на поліпшення продуктивності та умов утримання тварин.

Одним з основних напрямів розвитку аналізу продуктивності тварин є використання більш складних алгоритмів, зокрема глибокого навчання та гібридних моделей. Сучасні методи машинного навчання, такі як нейронні мережі, дозволяють створювати моделі, здатні ефективно обробляти великі обсяги даних та знаходити складні закономірності в продуктивності тварин. Завдяки глибокому навчання можна підвищити точність прогнозів і розширити можливості аналізу, враховуючи різноманітні фактори, такі як фізіологічні, поведінкові показники, умови навколишнього середовища. Гібридні моделі, що поєднують різні методи класифікації та кластеризації, також відкривають перспективи для більш детального та точного аналізу, дозволяючи ефективніше враховувати різні показники.

Ще одним важливим напрямом є розширення набору даних. З появою нових технологій для збору інформації, таких як датчики, камери і системи комп'ютерного зору, з'являються нові джерела даних і показники, які можуть

#### Секція 4. Технології розробки інформаційних систем

поліпшити точність моделей. Наприклад, інформація про харчування, температуру тіла, фізичну активність, стресові показники та інші фізіологічні параметри може допомогти більш повно оцінити продуктивність тварин та вплив різних факторів на їхнє здоров'я. Збагачення наборів даних новими показниками сприятиме побудові більш комплексних і точних моделей, що забезпечуватимуть більш обґрунтовані рекомендації для управління продуктивністю тварин.

Окрім цього, перспективним є інтегрування аналітичних систем з іншими системами управління фермою, зокрема IoT-технологіями та великими даними. Завдяки такій інтеграції створюється комплексна система моніторингу, яка дозволяє фермеру отримувати дані в режимі реального часу з різних джерел. Система, яка поєднує дані з IoT-сенсорів, дані з аналізу продуктивності та прогнозні моделі, допомагає автоматизувати процеси управління і прийняття рішень на фермі. Це не тільки підвищує продуктивність, а й знижує ризики, пов'язані з можливими захворюваннями або відхиленнями у здоров'ї тварин. У перспективі така інтеграція сприятиме більшій ефективності аграрного сектору та оптимізації ресурсів.

**Висновки.** У ході дослідження було доведено ефективність застосування алгоритмів класифікації та кластеризації для аналізу та оцінки продуктивності тварин у сучасному тваринництві. Використання класифікаційних алгоритмів дозволяє будувати точні моделі, які здатні прогнозувати продуктивність тварин і виділяти групи з високим потенціалом. Це забезпечує можливість визначення та управління тваринами, що мають найвищий рівень продуктивності, завдяки чому можна оптимізувати витрати ресурсів та досягти підвищення загальної ефективності господарства. Класифікація також допомагає оцінити вплив різних факторів, таких як умови утримання, харчування та здоров'я, що дозволяє зосередитися на найбільш важливих змінних для подальшого підвищення продуктивності.

Алгоритми кластеризації, у свою чергу, дозволяють групувати тварин на основі фізіологічних та поведінкових показників, що сприяє глибшому розумінню внутрішньої структури груп і особливостей кожного кластера. Кластерний аналіз дає можливість виділити тварин із схожими характеристиками та адаптувати під них відповідні умови утримання та харчування. Це також допомагає розробити індивідуалізовані стратегії догляду, що сприяє зменшенню стресу у тварин та підвищенню їхнього благополуччя. Таким чином, кластеризація є корисним інструментом для побудови персоналізованого підходу до управління поголів'ям та досягнення стабільно високої продуктивності.

#### Список використаних джерел та літератури

1. Harman, Mark. The current state and future of search based software engineering // 2007 Future of Software Engineering. IEEE Computer Society, 2007. pp. 342 - 357.
2. Shtern, Mark, and Vassilios Tzerpos. Clustering methodologies for software engineering // Advances in Software Engineering 2012. pp. 1-18.
3. Mancoridis, Spiros, et al. Using Automatic Clustering to Produce High-Level

System Organizations of Source Code // IWP, 1998. pp. 45-52.

4. S. Mancoridis, B. Mitchell, Y. Chen, and E. Gansner. Bunch: A clustering tool for the recovery and maintenance of software system structures // Proceedings of the 15th International Conference on Software Maintenance. Oxford, England, 1999. pp. 50–59.

5. M. Cohen, S. B. Kooi, and W. Srisa-an. Clustering the heap in multi-threaded applications for improved garbage collection // GECCO 2006: Proceedings of the 8th annual conference on Genetic and evolutionary computation, vol. 2, Seattle, Washington, 8-12 July 2006, ACM Press. pp. 1901–1908.

6. Bazilevich R.P. Dekompozicionnye i topologicheskie metody avtomatizirovannogo konstruirovaniya jelektronnyh ustrojstv.- L'vov: Vishha shkola, 1981. 168 s.

7. Jeet, Kawal, and Renu Dhir. Software Architecture Recovery using Genetic Black Hole Algorithm // ACM SIGSOFT Software Engineering Notes 40.1, 2015. pp. 1-5.

8. Kata Praditwong, Mark Harman, Xin Yao. Software Module Clustering as a Multi-Objective Search Problem // IEEE Transactions on Software Engineering, 2011, Vol. 37, № 2. pp. 264-282.

9. R. Bazylevych, I. Podolskyi and L. Bazylevych. Partitioning optimization by recursive moves of hierarchically built clusters // Proc. of 2007 IEEE Workshop on Design and Diagnostics of Electronic Circuits and Systems. April, 2007, Krakow, Poland. pp. 235 –238.

10. Roman Bazylevych, Dmytro Yanush. Partitioning optimization by iterative reassignment of the hierarchically built clusters with border elements // 2nd Mediterian Conference on Embedded Computing, MESO 2013, 15-20 June, 2013, Budva. pp. 219 - 221.