

ВДОСКОНАЛЕННЯ ПРОЦЕДУРИ ІДЕНТИФІКАЦІЇ ГІБРИДНИХ ТВАРИН ЗА УМОВ НЕВИЗНАЧЕНОЇ КІЛЬКОСТІ КЛАСІВ

Гітис І.В.

Науковий керівник – д.т.н., проф. Філатов В.О.

Харківський національний університет радіоелектроніки, каф. ШІ
м. Харків, Україна

тел.: +38(095) 822-07-84, e-mail: iryna.hitis@nure.ua

This paper considers the classification of animals obtained as a result of crossing genetically different forms within the same species, using the example of horses in the absence of a priori data. It is proposed to use methods for evaluating the quality of clustering to improve the efficiency of the algorithms. The application of these methods in the field of animal husbandry and, in particular, horse breeding will help specialists automate the process of grouping hybrid individuals in conditions where it is not known for certain how many classes the studied sample of animals should be divided into.

У природі та сільському господарстві зустрічаються організми, які були отримані шляхом схрещення генетично відмінних форм. Таких гібридних тварин досить важко класифікувати у межах одного виду з приводу відсутності приналежності до жодної з існуючих порід.

У роботі [1] описано створення бази даних, яка надає можливість зберігання та обробки даних тварин на прикладі коней. Для класифікації гібридних особин було запропоновано використання кластерного аналізу методом k-середніх. Проаналізована вибірка складалася із 40 тварин. Породи коней прийнято поділяти на три великі категорії: верхові, важковози та поні [2]. Тому алгоритмом k-середніх було виділено 3 кластери.

Слід зазначити, що через значні відмінності між гібридами не завжди може бути апріорно відомо, на скільки груп слід розділяти тварин.

Тому метою роботи було вдосконалення алгоритмів інтелектуального аналізу даних зареєстрованих тварин для підвищення ефективності розпізнавання гібридних особин.

Для визначення кількості кластерів, на які слід ділити екземпляри вибірки, використали метод оцінки силуету кластерів.

У табл. 1 наведено результати розрахунку коефіцієнтів силуету.

Таблиця 1 – Коефіцієнт силуету для кожної кількості кластерів

Кількість кластерів	Коефіцієнт силуету
2	0,57075
3	0,58974
4	0,57111
5	0,50106

Коефіцієнт силуету перебуває у діапазоні $[-1; 1]$, де більш високе значення відноситься до моделі з більш когерентними кластерами [3].

На рис. 1 представлено діаграму, що відображає оцінки силуетів при поділі на 2 та 3 кластери.

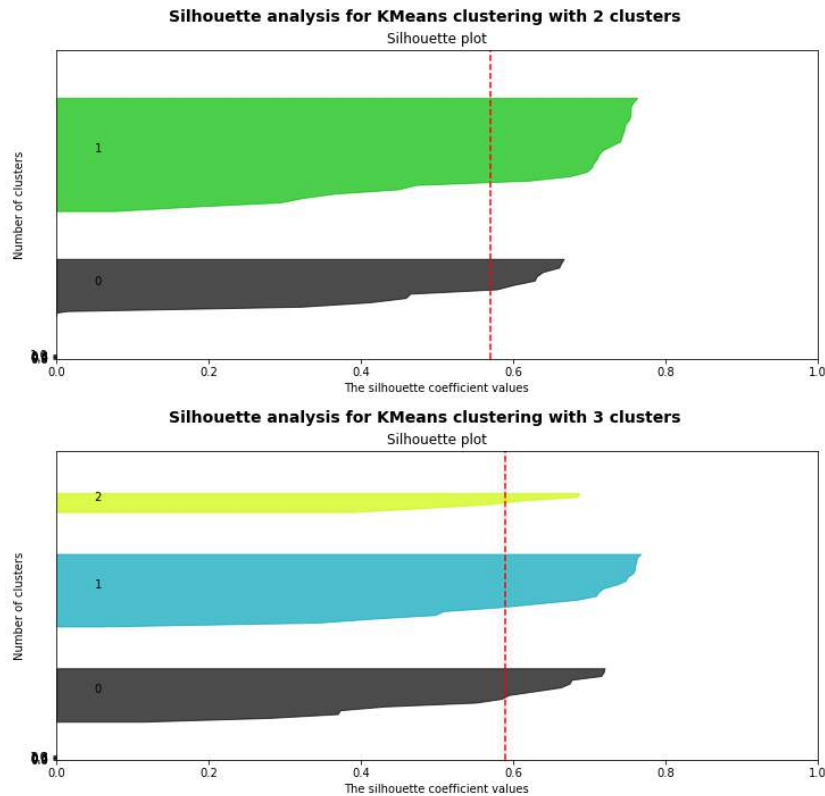


Рисунок 1 – Коефіцієнти силуету після кластеризації на 2 та 3 кластери

Аналіз отриманих результатів дозволяє зробити висновок, що при оцінці результатів кластеризації найвище значення силуету (0,58974) відповідає розбиттю на 3 кластери. Отже, ця кількість груп є оптимальною для вирішення поставленої задачі.

Таким чином, використання методів інтелектуального аналізу даних допоможе фахівцям у галузі тваринництва більш ефективно виконувати класифікацію гібридних особин. При цьому методи оцінки якості кластеризації дозволяють оптимізувати роботу з угруповання тварин за умов невизначеності.

Список використаних джерел:

1. Гітіс, І. В. (2022). Застосування методів кластерного аналізу для ідентифікації гібридних тварин. У 26-й Міжнародний молодіжний форум «Радіоелектроніка та молодь у XXI столітті» (Т. 6). ХНУРЕ.
2. Сольди, А. (2007). Лошади (Н. В. Ярош, Пер.).
3. Belyadi, H., & Haghghat, A. (2021). Machine learning guide for oil and gas using Python. Elsevier. <https://doi.org/10.1016/c2019-0-03617-5>.