

УДК 004.45

ЗАСТОСУВАННЯ КЛАСТЕРИЗАЦІЇ НА ОСНОВІ МАШИННОГО НАВЧАННЯ ДЛЯ ВДОСКОНАЛЕННЯ УПРАВЛІННЯ ПОПУЛЯЦІЯМИ БЕЗДОМНИХ ТВАРИН

Бегунова В.Д.

Науковий керівник – к.т.н., асист. Кобилін І.О.

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

e-mail: veronika.biehunova@nure.ua

This work is devoted to developing methods for managing populations of stray animals, particularly through analyzing the duration of their stay in shelters. The primary focus is on creating a technological solution that adapts to the unique characteristics of each animal and provides flexibility in response to changing conditions and data limitations. The main goal is to identify the specific needs and behavioral characteristics of animals based on their time spent in shelters, in order to optimize care processes, reduce the time it takes to find new homes for animals, and improve resource allocation between shelters. The application of clustering algorithms, such as *k*-means, will allow grouping animals with similar behavioral characteristics and needs based on their shelter stay time, thereby facilitating the development of targeted adaptation programs and accelerating the process of finding new owners.

Завдання подання вивчення, регуляції популяцій бездомних тварин або тварин, що залишилися без нагляду, догляду та опіки, є значною проблемою, яка досі залишається недостатньо дослідженою і вимагає технологічного рішення. Вирішення цієї задачі стикається з низкою важливих труднощів.

По-перше, кожна тварина має яскраво виражені поведінкові характеристики, які можуть бути об'єднані в підгрупи при загальному аналізі. Однак, при детальному вивченні кожної окремої тварини стає очевидним, що унікальні особливості історії, стану здоров'я та оточення можуть сильно впливати на їх поведінку та потреби.

По-друге, обмежений обсяг доступних спостережень у вибірці створює перешкоди застосування класичних статистичних методів. Нерівномірні часові інтервали між спостереженнями ускладнюють тимчасову впорядкованість даних і унеможливають використання традиційних метрик, що застосовуються в кластерному аналізі. Таким чином, стандартні методи, такі як евклідова метрика, можуть виявитися непридатними в даному контексті [3].

По-третє, у більшості випадків інформація про бездомних тварин розрізнена і не піддається систематичному збору та аналізу. Відсутність централізованих та стандартизованих систем обліку та моніторингу ство-

рює труднощі в аналізі та прогнозуванні динаміки популяцій, а також ускладнює розробку ефективних стратегій регулювання та допомоги [5].

Для подолання цих труднощів необхідно розробити технологічне рішення, яке враховує унікальні характеристики кожної тварини, при цьому забезпечуючи адаптивність до умов і обмежень даних, що змінюються.

Для вирішення цієї проблеми можна застосувати метод кластеризації на основі часу перебування тварини у притулку [1]. Мета цієї кластеризації полягатиме у групуванні тварин відповідно до тривалості їх перебування у притулку, що допоможе виявити патерни поведінки та особливості догляду за ними залежно від часу перебування. Такий підхід дозволить ефективно управляти популяціями бездомних тварин і розробляти персоналізовані підходи до їхнього догляду та регуляції.

Кластерний аналіз є потужним інструментом для групування подібних об'єктів у наборі даних [2]. Одним з найпоширеніших методів кластеризації є алгоритм k -середніх (k -means), який знаходить оптимальні центри кластерів шляхом мінімізації середньоквадратичної відстані між кожним об'єктом і центром його кластера. Як працює цей алгоритм:

1. Ініціалізація: вибирається кількість кластерів (k), та випадковим чином вибираються початкові центри кластерів.

2. Приналежність до кластера: кожен об'єкт призначається до кластера, який має найближчий до нього центр.

3. Перерахунок центрів: для кожного кластера обчислюється новий центр як середнє арифметичне всіх об'єктів, що до нього відносяться.

4. Повторення: кроки 2 і 3 повторюються до збіжності, коли приналежність об'єктів до кластерів не змінюється або зміни відбуваються дуже малі.

Для виконання поставленої задачі було обрано саме алгоритм k -means. Для цього спочатку необхідно визначити часові інтервали, наприклад, тимчасове перебування (до 1 місяця), короткострокове перебування (1–6 місяців), середньострокове перебування (6–12 місяців), тривале перебування (12–24 місяців) та довгострокове перебування (понад 24 місяців). Потім можна використовувати алгоритм кластеризації, такий як k -середніх, щоб розділити тварин на групи на основі часу перебування в притулку [4].

Результати кластеризації (рис. 1) дозволять виявити, наприклад, які категорії тварин більш схильні до тривалого перебування в притулку, що може вказувати на потребу особливої турботи та уваги. Також це допоможе керувати ресурсами притулку більш ефективно, передбачати завантаженість та вживати заходів для зменшення часу перебування тварин у притулку, наприклад, шляхом активізації програм з пошуку нових будинків та збільшення зусиль щодо залучення потенційних власників.

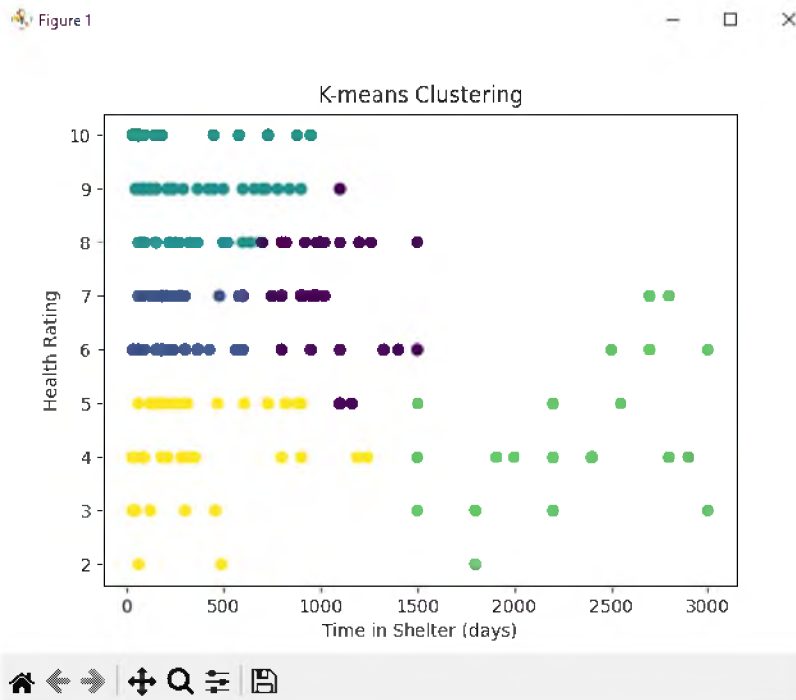


Рисунок 1 – Результат виконання кластерного аналізу алгоритмом *k*-means

Список використаних джерел:

1. Bodyanskiy Y., Kobylin I., Rashkevych Y., Vynokurova O., Peleshko D. Hybrid fuzzy-clustering algorithm of unevenly and asynchronously spaced time series in computer engineering. 2018 14th International Conference on Advanced Trends in Radioelectronics, Telecommunications and Computer Engineering (TCSET). IEEE, 2018. P. 930–935.
2. Bodyanskiy Y., Vynokurova O., Kobylin I., Kobylin O. Adaptive fuzzy clustering of short time series with unevenly distributed observations in Data Stream Mining tasks. Information Technology and Management Science. 2016. Vol. 19, No. 1. P. 23–28.
3. Ermshaus A., Schäfer P., Leser U. Window Size Selection in Unsupervised Time Series Analytics: A Review and Benchmark. AALTD, 2023. P. 83–101.
4. Keogh E., Taposh D. R., Naik U., Agrawal A. Multidataset timeseries anomaly detection competition. ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2021.
5. Lu Y., Srinivas TVA., Nakamura T., Imamura M., Keogh E. Matrix Profile XXX: MADRID: A Hyper-Anytime and Parameter-Free Algorithm to Find Time Series Anomalies of all Lengths. 2023 IEEE International Conference on Data Mining (ICDM). IEEE, 2023. P. 1199–1204.