

# BAB 1

## PENDAHULUAN

### 1.1 Latar Belakang Masalah

Abalone merupakan salah satu jenis moluska laut yang memiliki nilai ekonomi tinggi dan seringkali menjadi objek penelitian dalam bidang biologi, akuakultur, dan ilmu kelautan [1]. Dikenal karena keindahan cangkangnya dan nilai gizinya yang tinggi, abalone sering kali menjadi target perburuan manusia untuk diambil dagingnya dan juga memiliki nilai komersial dalam industri perhiasan. Selain itu, kemampuan abalone untuk hidup di habitat laut yang beragam menjadikannya sebagai objek penelitian yang menarik dalam bidang ekologi [2]. Salah satu aspek penting dalam pemahaman tentang abalone adalah memahami variasi fitur fisiknya. Fitur fisik ini meliputi Jenis kelamin abalone, Panjang, Diameter, Tinggi, Berat keseluruhan, Berat daging, Berat organ dalam, Berat cangkang, dan Jumlah cincin pada cangkang. Jumlah cincin ini secara khusus memberikan indikasi usia abalone [3]. Variasi ini tidak hanya mempengaruhi morfologi abalone, tetapi juga dapat memberikan petunjuk tentang kondisi lingkungan dan kesehatan populasi abalone [4]. Variabel variabel tersebut perlu dilakukan klusterisasi karena terdapat variasi karakteristik fisik yang sangat bervariasi. Dengan melakukan klusterisasi, maka fitur fisik abalone dapat dikelompokkan ke dalam kluster yang sama sehingga dapat memberikan informasi yang berharga untuk memahami variasi dan pola dalam populasi abalone, serta membantu dalam pengelolaan dan konservasi sumber daya abalone secara lebih efisien [5].

Klusterisasi adalah salah satu metode dalam pembelajaran mesin (*machine learning*) yang bertujuan untuk mengelompokkan data berdasarkan kesamaan tertentu tanpa memerlukan label [6]. Metode klusterisasi berbasis *centroid*, seperti *K-Means*, sering digunakan karena sederhana dan efektif dalam memisahkan data

ke dalam kelompok (*cluster*) yang saling berbeda [7]. Namun, metode ini rentan terhadap hasil yang tidak optimal karena pemilihan initial *centroid* yang dilakukan secara acak, yang dapat memengaruhi kualitas akhir klasterisasi [8].

Untuk mengatasi kelemahan ini, penelitian ini mengusulkan penggunaan *Cuckoo Search Algorithm* (CSA), sebuah metode optimasi *metaheuristik* yang terinspirasi dari perilaku reproduksi burung cuckoo. CSA bekerja dengan memanfaatkan prinsip *Lévy Flight* untuk pencarian lokal dan mekanisme seleksi sarang yang memastikan eksplorasi global ruang pencarian [9]. CSA dikenal memiliki keunggulan dalam menemukan solusi optimal pada ruang pencarian yang kompleks dan berdimensi tinggi [10].

Hasil dari penelitian ini diharapkan dapat meningkatkan akurasi dan efisiensi dalam pengelompokan fitur fisik abalone, sekaligus memberikan kontribusi dalam penerapan metode optimasi *metaheuristik* pada analisis data biologis. Dengan pendekatan ini, penelitian tidak hanya memberikan solusi yang lebih baik untuk klasterisasi, tetapi juga membantu pemahaman tentang pola dan struktur dalam populasi abalone, yang penting untuk mendukung konservasi dan pengelolaan abalone secara berkelanjutan.

## **1.2 Rumusan Masalah**

Abalone merupakan jenis kerang laut yang memiliki nilai ekonomis tinggi. Mengetahui variasi dan pola pada populasi abalone menjadi hal yang penting dalam memahami karakteristik serta informasi terkait abalone. *Clustering* menjadi alternatif cara dalam mengelompokkan karakteristik abalone. Penelitian ini menggunakan algoritma *Cuckoo Search* untuk menemukan *centroid* optimum yang dapat digunakan sebagai acuan dalam mengklasifikasikan abalone berdasarkan fitur fisiknya seperti jenis kelamin, panjang, diameter, tinggi, berat total, berat daging, berat bagian dalam (organ), berat cangkang, dan jumlah cincin. Dengan menentukan *centroid* optimum, diharapkan dapat memberikan hasil *cluster* yang optimal.

### 1.3 Pertanyaan Penelitian

Pada bagian ini, ada beberapa pertanyaan penelitian yang menjadi fokus utama dengan tujuan untuk memberikan arah penelitian dan memberikan panduan dalam menganalisis data serta mengevaluasi hasil yang diperoleh. Pertanyaan-pertanyaan penelitian tersebut adalah sebagai berikut:

1. Bagaimana cara menentukan *centroid optimum* menggunakan *Cuckoo Search Algorithm*?
2. Apakah penggunaan CSA untuk menentukan *initial centroid* dapat meningkatkan performa metode *K-Means*?
3. Bagaimana performa *CSA K-Means* dalam hal akurasi waktu dan *Silhouette score* dibandingkan dengan *K-Means* dan *cluster manual*?

### 1.4 Batasan Masalah

Untuk memastikan penelitian ini tetap fokus dan terarah, terdapat beberapa batasan yang harus diperhatikan. Adapun batasan masalah dalam penelitian ini adalah sebagai berikut:

1. Data yang digunakan adalah data karakteristik fisik meliputi jenis kelamin, panjang, diameter, tinggi, berat total, berat daging, berat bagian dalam (organ), berat cangkang, dan jumlah cincin Abalone
2. Performa komputasi dipengaruhi oleh device yang digunakan yaitu RAM 8 gb SSD 256 gb dengan processornya intel core i3 Gen 11 2 core 4 thread
3. Sumber data yang digunakan adalah hanya menggunakan satu website resmi tentang abalone yaitu UCI Machine Learning Repository
4. Penelitian hanya berfokus untuk menentukan klaster terbaik dari dataset abalone tanpa ada fokus berkelanjutan

### 1.5 Tujuan Penelitian

Penelitian ini memiliki beberapa tujuan utama yang ingin dicapai. Tujuan-tujuan ini diharapkan dapat memberikan kontribusi yang signifikan dalam bidang analisis data dan optimasi *clustering*. Tujuan-tujuan penelitian ini adalah sebagai berikut:

1. Menentukan *centroid* optimum pada klusterisasi menggunakan *Cuckoo Search Algorithm (CSA)*.
2. Mengevaluasi performa *CSA K-Means*.
3. Membandingkan *CSA K-Means* dalam hal akurasi dan waktu komputasi dibandingkan dengan *K-Means* dan *cluster* manual dalam klusterisasi data.

## **1.6 Manfaat Penelitian**

Penelitian ini diharapkan dapat memberikan manfaat yang berarti, baik dari segi teori maupun aplikasi praktis. Manfaat-manfaat yang diharapkan dari penelitian ini antara lain adalah sebagai berikut:

1. Memberikan pemahaman yang lebih dalam mengenai algoritma *clustering*, khususnya *K-Means*, dan bagaimana optimasi dengan *Cuckoo Search* dapat meningkatkan kinerja algoritma tersebut.
2. Penelitian ini diharapkan dapat meningkatkan akurasi dan efektivitas *clustering* pada dataset yang kompleks.
3. Penelitian ini dapat menjadi dasar bagi penelitian lanjutan yang mengeksplorasi lebih jauh tentang penggabungan algoritma optimasi lainnya dengan teknik *clustering*.
4. Penelitian ini dapat digunakan sebagai bahan ajar dalam mata kuliah yang berkaitan dengan data mining, machine learning, dan analisis statistik, sehingga dapat membantu mahasiswa dan praktisi dalam memahami konsep-konsep tersebut secara lebih mendalam.