

DETEKSI BESAR HEWAN MAMALIA BERDASARKAN LUAS DAN KELILING MENGGUNAKAN METODE K-MEANS

Mudhi Ulfani¹, Nurul Fadillah²
Informatika, Teknik, Universitas Samudra
Informatika, Teknik, Universitas Samudra
¹mudhiulfanii@gmail.com, ²nurulfadillah@unsam.ac.id

Abstract

Mammals are a class of vertebrate animals with features such as hair and mammary glands. Mammals are spread almost all over the world and occupy different types of habitats, from the arctic to the equator, from the sea to the land. In this study, a program will be built using the K-means clustering method. The program will classify 60 images from 3 types of mammal images, namely 20 images of bats, 20 images of fish and 20 images of frogs. The cluster results are presented in diagrammatic form. After conducting research on the number of centroids, it can be concluded that the more the number of centroids in each clustering process, the more specific the resulting cluster groups will be. Thus making conclusions on similarities in cluster groups is easier.

Keywords: mammals; K-means; clustering;

Abstrak

Mamalia merupakan salah satu kelas dari hewan vertebrata dengan ciri seperti adanya rambut dan kelenjar susu. Hewan mamalia tersebar hampir diseluruh dunia dan menempati tipe habitat yang berbeda-beda, mulai dari daerah kutub sampai khatulistiwa, mulai dari laut hingga daratan. Dalam penelitian ini akan dibangun program dengan menerapkan metode K-means clustering. Program akan mengelompokkan 60 citra dari 3 jenis citra hewan mamalia yaitu 20 citra kelelawar, 20 citra ikan dan 20 citra katak. Hasil cluster disajikan dalam bentuk diagram. Setelah dilakukan penelitian terhadap jumlah centroidnya, dapat ditarik kesimpulan bahwa semakin banyak jumlah centroid dalam setiap proses clustering, maka makin spesifik kelompok cluster yang dihasilkan. Dengan demikian pengambilan kesimpulan kesamaan dalam kelompok cluster makin mudah.

Kata kunci: Mamalia, K-means, Clustering;

1. PENDAHULUAN

Mamalia merupakan salah satu kelas dari hewan vertebrata dengan ciri seperti adanya rambut dan kelenjar susu. Hewan mamalia tersebar hampir di seluruh dunia dan menempati tipe habitat yang berbeda-beda, mulai dari daerah kutub sampai khatulistiwa, mulai dari laut hingga daratan (Nurul, I. 2016). Di dunia, hewan mamalia terdiri dari 19 ordo, 122 famili, 1017 genus dengan jumlah jenis kurang lebih 12.000. Hal ini menunjukkan bahwa dalam hal kekayaan

mamalia, je isnya cukup beragam. Di Indonesia terdapat 515 jenis mamalia (12% dari jenis mamalia yang ada di dunia) (Departemen Kehutanan, 2005). Mamalia kecil merupakan kelompok heterogen dari sudut pandang taksonomi, mereka termasuk ke dalam ordo Insektivora dan Rodentia. Berdasarkan ukuran tubuh mamalia ini dibagi ke dalam mamalia besar dan mamalia kecil.

Teknik perhitungan luas dan keliling hewan mamalia yaitu dengan menggunakan

rumus. Namun, hal ini tidak efisien, karena kemungkinan kekeliruan masih bisa terjadi dan datanya sudah pasti tidak akurat. Untuk mendapatkan cara yang praktis, bidang Teknologi Informasi dan Komunikasi dapat digunakan deteksi besar hewan mamalia yaitu melalui image processing. Image processing dilakukan dengan proses clustering pada citra dengan menggunakan metode K-means.

Clustering merupakan proses pengelompokan data dalam kelas-kelas atau cluster-cluster sehingga data-data dalam suatu cluster memiliki tingkat persamaan yang tinggi satu dengan yang lainnya tetapi sangat berbeda dengan data pada cluster lain. Metode K-means adalah salah satu teknik yang paling sering digunakan. Algoritma K-means termasuk partitioning clustering yang memisahkan data ke daerah bagian yang terpisah. Algoritma K-means sangat terkenal karena kemudahan dan kemampuannya untuk mengcluster data besar dan data outlier dengan sangat cepat. Namun kelemahan dari metode K-means clustering adalah hasil clustering dengan K-means bersifat selalu berubah-ubah.

Pada penelitian kali ini akan dibahas contoh yang berkaitan dengan deteksi besar hewan mamalia berdasarkan luas dan keliling menggunakan metode K-means. Alasan kuat untuk membahas judul ini karena ingin membentuk variasi baru dengan menerapkan metode K-means.

2. LANDASAN TEORI

2.1 Hewan Mamalia

Mamalia merupakan salah satu kelas dari hewan vertebrata dengan ciri seperti adanya rambut dan kelenjar susu. Hewan mamalia tersebar hampir diseluruh dunia dan menempati tipe habitat yang berbeda-beda, mulai dari daerah kutub sampai khatulistiwa, mulai dari laut hingga daratan. Di dunia, hewan mamalia terdiri dari 19 ordo, 122 famili, 1017 genus dengan jumlah jenis kurang lebih 12.000. Hal ini menunjukkan bahwa dalam hal kekayaan mamalia, jenisnya cukup beragam. (Nurul, I. 2016).

2.2 Deteksi Besar Hewan Mamalia berdasarkan Luas dan Keliling

Teknik perhitungan luas dan keliling hewan mamalia yaitu dengan menggunakan rumus. Namun, hal ini tidak efisien, karena kemungkinan kekeliruan masih bisa terjadi dan datanya sudah pasti tidak akurat. Untuk mendapatkan cara yang praktis, bidang Teknologi Informasi dan Komunikasi dapat digunakan deteksi besar hewan mamalia yaitu melalui image processing. Image processing dilakukan dengan proses clustering pada citra dengan menggunakan metode K-means.

2.3 Metode

Metode adalah cara untuk menyampaikan materi pembelajaran dalam upaya mencapai tujuan, definisi tersebut menegaskan bahwa metode pembelajaran ialah 1) cara, 2) untuk menyampaikan, 3) materi, 4) sebagai upaya mencapai tujuan. (Ahwan Fanani, 2014).

2.4 Metode K-means

K-means merupakan salah satu metode data clustering non hierarki yang berusaha mempartisi data yang ada ke dalam bentuk satu atau lebih cluster atau kelompok sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu cluster yang sama dan data yang mempunyai karakteristik yang berbeda dikelompokkan ke dalam kelompok yang lainnya.

K-means adalah metode clustering berbasis jarak yang membagi data ke dalam sejumlah cluster dan algoritma ini hanya bekerja pada atribut numeric. Algoritma K-means termasuk partitioning clustering yang memisahkan data ke k daerah bagian yang terpisah. Algoritma K-means sangat terkenal karena kemudahan dan kemampuannya untuk mengcluster data yang besar dan data outlier dengan sangat cepat. Dalam algoritma K-means, setiap data harus termasuk ke cluster tertentu dan bisa dimungkinkan bagi setiap data yang termasuk cluster tertentu pada suatu tahapan proses, pada tahapan berikutnya berpindah cluster lainnya.

Algoritma K-means merupakan metode non hierarki yang pada awalnya mengambil sebagian banyaknya komponen populasi untuk dijadikan pusat cluster awal.

Pada tahap ini pusat cluster dipilih secara acak dari sekumpulan populasi data. Berikutnya K-means menguji masing-masing komponen di dalam populasi data dan menandai komponen tersebut ke salah satu pusat cluster yang telah didefinisikan tergantung dari jarak minimum antar komponen dengan tiap-tiap cluster. Posisi pusat cluster akan dihitung kembali sampai semua komponen data digolongkan ke dalam tiap-tiap pusat cluster dan terakhir akan terbentuk posisi pusat cluster yang baru. (Benri Melpa dan Herlina Latipa Sari, 2015).

2.5 Clustering

Clustering adalah proses mengelompokkan atau penggolongan objek berdasarkan informasi yang diperoleh dari data yang menjelaskan hubungan antar objek dengan prinsip untuk memaksimalkan kesamaan antar anggota satu kelas dan meminimumkan kesamaan antar kelas/cluster. Clustering membagi data ke dalam grup-grup yang mempunyai obyek yang karakteristiknya sama. (Lynda Rahmawati, dkk. 2014)

2.6 K-means Clustering

Metode ini berusaha untuk meminimalkan variasi antar data yang ada di dalam suatu cluster dan memaksimalkan variasi dengan data yang ada di cluster lainnya.

Berikut adalah langkah – langkah dari algoritma K-means :

1. Menentukan banyak k-cluster yang ingin dibentuk.
2. Membangkitkan nilai random untuk pusat cluster awal (centroid) sebanyak k-cluster.
3. Menghitung jarak setiap data input terhadap masing-masing centroid menggunakan rumus jarak Euclidean (Euclidean Distance) hingga ditemukan jarak yang paling dekat dari setiap data dengan centroid. Berikut adalah persamaan Euclidean Distance :

$$d(x_i, \mu_i) = \sqrt{(x_i - \mu_i)^2}$$

Dengan $d(x_i, \mu_i)$ adalah jarak antara cluster x dengan pusat cluster μ pada kata ke i , dimana x_i adalah bobot kata ke i pada cluster yang ingin dicari jaraknya, μ_i bobot kata ke i pada pusat cluster.

4. Mengklasifikasikan setiap data berdasarkan kedekatannya dengan centroid (jarak terkecil).
5. Mengupdate nilai centroid. Baru diperoleh dari rata-rata cluster yang bersangkutan dengan menggunakan rumus :

$$C_k = \frac{1}{n_k} \sum d_i$$

Dimana :

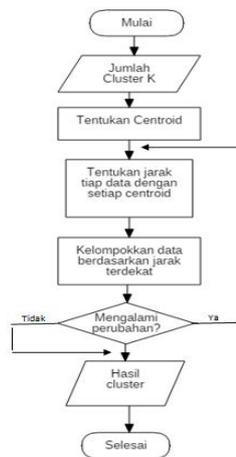
n_k = jumlah data dalam cluster.

d_i = jumlah dari nilai jarak yang masuk dalam masing-masing cluster.

6. Melakukan perulangan dari langkah 2 hingga 5 hingga anggota tiap cluster tidak ada berubah.
7. Jika langkah 6 telah dipenuhi, maka nilai rata-rata pusat cluster pada iterasi terakhir akan digunakan sebagai parameter untuk menentukan klasifikasi data. (Lynda Rahmawati, dkk. 2014)

3 Hasil dan Pembahasan

Ada beberapa proses yang akan dilakukan dalam penelitian ini. Pertama melakukan penentuan data (citra) lalu menentukan jumlah cluster. Setelah itu tentukan nilai centroid awal secara acak, selanjutnya hitung jarak tiap data terhadap masing-masing centroid, kelompokkan data-data tersebut ke kluster berdasarkan jarak paling dekat (minimum) terhadap sebuah kluster. Lalu hitung ulang nilai centroid sampai tidak lagi mengalami perubahan. Perhitungan jarak antara centroid dengan data dilakukan menggunakan persamaan Euclidean Distance. Pada contoh ini algoritma K-means digunakan untuk mengkluster 60 citra yang terdiri dari 20 citra kelelawar, 20 citra ikan, dan 20 citra katak. Pada program ini terdapat kondisi yaitu dimana apakah nilai centroid mengalami perubahan, jika mengalami perubahan ulangi perhitungan jarak tiap data terhadap masing-masing centroid hingga nilai centroid tidak lagi mengalami perubahan. Dan jika nilai centroid tidak lagi mengalami perubahan maka program selesai.



Gambar 3.1 Flowchart Clustering K-means

3.1 Memulai Penentuan Data

Proses pertama yaitu melakukan penentuan data. Datayang dimaksud adalah beberapa hewan mamalia berupa image.



Gambar 3.2 Beberapa Kelompok Hewan Mamalia

3.2 Menentukan Jumlah Cluster K

Pada proses ini, menentukan banyaknya jumlah cluster K artinya berapa banyak citra yang akan di cluster. Pada contoh ini K-means digunakan untuk mengkluster 60 citra yang terdiri dari 20 citra kelelawar, 20 citra ikan, dan 20 citra katak berupa image biner.



Gambar 3.3 60 Citra Hewan Mamalia

3.3 Menentukan Centroid

Dalam menentukan nilai centroid untuk awal iterasi, nilai awal centroid dilakukan secara acak, sedangkan jika menentukan nilai centroid yang merupakan tahap dari iterasi, maka menggunakan rumus.

3.4 Menentukan Jarak Tiap Data dengan Setiap Centroid

Perhitungan jarak antara centroid dengan data dapat dilakukan menggunakan persamaan euclidean distance, yaitu :

$$D_e = \sqrt{(x_i - s_i)^2 + (y_i - t_i)^2}$$

Dimana :

D_e adalah Euclidean Distance

i adalah banyaknya objek

(x,y) adalah koordinat object dan

(s,t) adalah koordinat centroid.

3.5 Kelompokkan Data Berdasarkan Jarak Terdekat

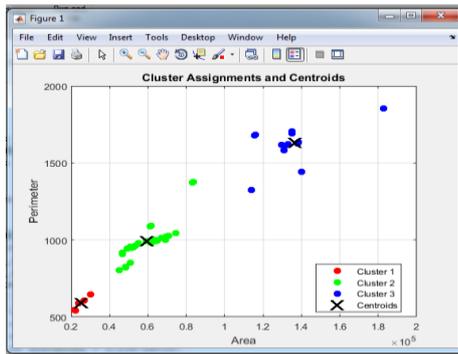
Untuk menentukan anggota cluster adalah dengan memperhitungkan jarak minimum objek. Nilai yang diperoleh dalam keanggotaan data pada distance matriks adalah 0 dan 1, dimana nilai 1 untuk data yang dialokasikan ke cluster dan nilai 0 untuk data yang dialokasikan ke cluster lain.

3.6 Kondisi

Kondisi disini menunjukkan apakah nilai centroid mengalami perubahan, jika mengalami perubahan ulang perhitungan jarak tiap data terhadap masing-masing centroid hingga nilai centroid tidak lagi mengalami perubahan.

3.7 Menampilkan Hasil Cluster

Hasil cluster ini menggunakan algoritma K-means. Selain itu nilai centroid juga sudah ditemukan melalui beberapa iterasi.



Gambar 3.4 Hasil Cluster dan Nilai Centroid

3.8 Citra Hewan Mamalia

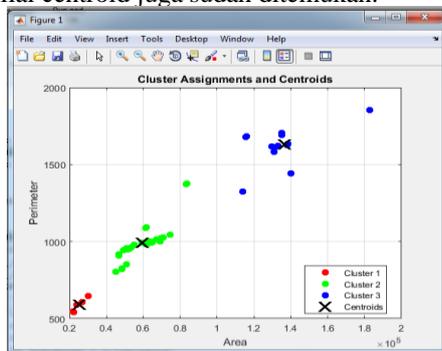
Proses penentuan data. Data yang dimaksud adalah beberapa kelompok hewan mamalia berupa image biner. Pada penelitian ini menentukan 60 citra yang akan di cluster yaitu 20 citra kelelawar, 20 citra ikan dan 20 citra katak.



Gambar 4.1 60 Citra Hewan Mamalia

3.9 Hasil Cluster 60 Citra Hewan Mamalia Menggunakan Algoritma K-means

Data citra hewan mamalia yang sudah dicluster sebanyak 60 citra. Hasil cluster ini menggunakan algoritma K-means. Selain itu nilai centroid juga sudah ditemukan.



Gambar 4.2 Hasil cluster dan nilai centroid

4 Kesimpulan

1. Algoritma K-means dapat digunakan untuk mengelompokkan beberapa hewan mamalia yang ada di dunia.
2. Algoritma K-means merupakan algoritma yang sederhana.
3. K-means clustering mampu menyelesaikan permasalahan yang ada.
4. K-means digunakan untuk mengelompokkan data sebelum menerapkan teknik analisis data. Kemudian teknik analisis data hanya perlu diterapkan pada centroid dari setiap cluster sehingga lebih efisien dalam penggunaan waktu.

REFERENSI

- Asroni dan Adrian, Ronald. 2015. *Penerapan Metode K-means untuk Clustering Mahasiswa Berdasarkan Nilai Akademik dengan Weka Interface Studi Kasus pada Jurusan Teknik Informatika UMM Magelang*. Magelang : Universitas Muhammadiyah Magelang.
- Fanani, Ahwan. 2014. *Mengurai Kerancuan Istilah Strategi dan Metode Pembelajaran*. Semarang : IAIN Walisongo Semarang.
- Metisen, Benri Melpa dan Sari, Herlina Latipa. 2015. *Analisis Clustering Menggunakan Metode K-means dalam Pengelompokan Penjualan Produk pada Swalayan Fadhila*. Bengkulu : Universitas Dehasen Bengkulu.
- Rahmawati, Lynda. 2014. *Analisa Clustering Menggunakan Metode K-means dan Hierarchical Clustering*. Surakarta : Universitas Sebelas Maret.
- Wardhani, Anindya Khrisna. 2016. *Implementasi Algoritma K-means untuk Pengelompokan Penyakit Pasien pada Puskesmas Kajen Pekalongan*. Universitas Diponegoro.
- Nurul, I. (2016). *JENIS-JENIS MAMALIA YANG MENGUNJUNGI*

KUBANGAN BABI HUTAN DI
HUTAN KONSERVASI PT TIDAR
KERINCI AGUNG DAN PT
KENCANA SAWIT INDONESIA,

SOLOK SELATAN, SUMATERA
BARAT (Doctoral dissertation,
Universitas Andalas).