

ANALISIS CLUSTER LOVEBIRD BERDASARKAN CIRI FISIK DAN JENISNYA MENGGUNAKAN ALGORITMA K-MEANS DENGAN PENGHITUNGAN JARAK MANHATTAN

Mohammad Rofii¹⁾, Nilam Ramadhani²⁾

^{1,2}Teknik Informatika, Fakultas Teknik, Universitas Madura

^{1,2}Jalan Raya Panglegur KM 3.5 Pamekasan, Jawa Timur, Indonesia

¹opekhunter@gmail.com, ²nilam_ramadhani@yahoo.com

ABSTRAK

Data penelitian dapat dikelompokkan berdasarkan kemiripan ciri, objek maupun metode penelitian. Hasil pengelompokan data penelitian dapat memperlihatkan bagaimana pola kemiripan penelitian dan variasi ciri karakteristik penelitian dari waktu ke waktu. Hasil pengelompokan juga dapat memperlihatkan karakteristik yang banyak diambil data ciri burung lovebird. Informasi tersebut diharapkan dapat membantu Paguyuban Sendang Lovebird Farm dalam mengevaluasi pengelompokan data yang telah dilakukan. Penelitian ini mengelompokkan data burung lovebird di Paguyuban Sendang Lovebird Farm Sumenep karena jumlah data penelitiannya cukup banyak. Pengelompokan data penelitian yang umumnya berbentuk teks dapat dilakukan dengan text mining dengan metode clustering. Metode clustering yang digunakan pada penelitian ini adalah kombinasi antara metode K-Means Clustering dan penghitungan jarak Manhattan. Clustering data burung menghasilkan maksimal 5 cluster. Hasil cluster dianalisa keterkaitan ciri fisik dan jenisnya.

Kata Kunci : Clustering, K-Means Clustering, Data Mining, Lovebird

ABSTRACT

Research data can be grouped based on similarity of features, objects and research methods. The results of grouping research data can show how the patterns of similarity of research and variations in the characteristics of research characteristics over time. The results of the grouping can also show the characteristics that many lovebird bird characteristic data taken. The information is expected to help Sendang Lovebird Farm Association in evaluating the grouping of data that has been done. This study grouped lovebird birds data in the Sumenep Farm Lovebird Sendang Society because the amount of research data was quite large. Grouping research data which is generally in the form of text can be done by text mining with clustering methods. The clustering method used in this study is a combination of the K-Means Clustering method and Manhattan distance calculation. Clustering bird data produces a maximum of five clusters. The results of the cluster were analyzed the interrelationship of physical characteristics and types.

Keywords: Clustering, K-Means Clustering, Data Mining, Lovebird

PENDAHULUAN

Salah satu hewan peliharaan yang populer di Indonesia adalah burung. Jenis burung yang banyak dipelihara diantaranya burung kicauan. Jenis burung kicauan ini sering digunakan dalam berbagai kontes dan perlombaan burung. Hobi memelihara burung di Indonesia memiliki banyak peminat dengan alasan yang bermacam-macam.

Burung memiliki keunikan, keindahan suara dan kecantikan warna warni bulunya yang mampu memberikan kepuasan tersendiri bagi pemiliknya. Suara dari beberapa jenis burung sangat merdu yang tentunya memukau telinga para pecinta burung sehingga sering diikutsertakan dalam berbagai kontes burung kicauan.

Salah satu jenis burung yang jadi

primadona sebagai peliharaan adalah lovebird. Lovebird merupakan salah satu dari Sembilan spesies dari genus Agapornis, asal mula kata berasal dari negara Yunani Agape yang berarti cinta dan Ornis yang berarti burung. Hewan ini adalah tipe burung yang sosial atau berkelompok dan dekat dengan keluarga bayan. Delapan spesies Lovebird berasal dari benua Afrika, sementara Grey-headed Lovebird berasal dari Madagaskar. Lovebird merupakan tipe burung yang monogamy atau setia pada pasangan dalam jangka waktu yang lama. Beberapa spesies yang dibiakkan sebagai hewan peliharaan dengan berbagai warna yang cantik merupakan hasil persilangan yang selektif di peternakan burung. Lovebird pada umumnya dapat berumur 10 sampai 15 tahun. Lovebird termasuk burung kecil dengan ukuran maksimalnya 13-17 cm dan berat badan mencapai 40-60 gram. Lovebird adalah burung terkecil dari keluarga betet. Lovebird memiliki tubuh kompak, ekor pendek berujung tumpul, paruh tajam. Lovebird liar di dominasi warna hijau dengan berbagai warna pada tubuh bagian atas, tergantung spesies. Lovebird Fischer's, Lovebird Black-cheeked, dan Masked Lovebird memiliki cincin putih terkemuka di sekitar mata mereka. Di Indonesia tipe ini disebut dengan Lovebird kaca mata atau juga dapat disebut Lovebird Clep. Populasi liar Lovebird spesies Fischeri Masked berada di kota-kota di Afrika Timur. Lovebird Hibrida memiliki warna coklat kemerahan pada kepala dan oranye di dada bagian atas dan sedikit mirip dengan jenis Masked Lovebird. Lovebird memiliki kecenderungan untuk menjalin ikatan baik dengan sesama burung Lovebird atau berinteraksi dengan manusia.

Terdapat penelitian tentang budidaya lovebird oleh Nopry Adi Winata, menurut penelitian tersebut peminat burung lovebird di kota solo berdasarkan jumlah member komunitas yaitu 35883 member semakin banyak

dan berkembang. Seiring banyaknya peminat pembeli atau peminat hobi lovebird yang baru cenderung awam dalam pengetahuan dari lovebird. Hal tersebut menyebabkan dalam pembelian lovebird tidak mengetahui jenis-jenis sedangkan jenis burung lovebird. Sehingga perlu adanya penelitian yang dapat bermanfaat untuk masyarakat awam dalam mengetahui jenis jenis lovebird.

Dalam upaya mengetahui dan membedakan jenis-jenis lovebird dapat dilakukan dengan cara dapat dilakukan K-Means Clustering. Clustering merupakan proses pengelompokan data. Secara detail teknik ini menggunakan ukuran ketidakmiripan untuk mengelompokkan obyek. Ketidakmiripan dapat diterjemahkan dalam konsep jarak. Dua obyek dikatakan mirip jika jarak dua objek tersebut dekat. Semakin tinggi nilai jarak, semakin tinggi nilai ketidakmiripannya.

Penelitian ini diharapkan dapat memberikan solusi terhadap permasalahan yang dihadapi masyarakat yaitu permasalahan pengelompokan jenis burung lovebird berdasarkan cirinya. Sehingga dengan mengimplementasikan aplikasi ini dapat membantu dalam menentukan jenis burung lovebird.

Clustering

Clustering adalah metode penganalisaan data, yang sering dimasukkan sebagai salah satu metode Data Mining, yang tujuannya adalah untuk mengelompokkan data dengan karakteristik yang sama ke suatu kelompok yang sama dan data dengan karakteristik yang berbeda ke kelompok yang lain.

Ada beberapa pendekatan yang digunakan dalam mengembangkan metode *Clustering*. Dua pendekatan utama adalah *Clustering* dengan pendekatan partisi dan *Clustering* dengan pendekatan hirarki. *Clustering* dengan pendekatan partisi atau sering disebut dengan *partition-based*

Clustering mengelompokkan data dengan memilah-milah data yang dianalisa ke dalam *cluster-cluster* yang ada. *Clustering* dengan pendekatan hirarki atau sering disebut dengan *hierarchical Clustering* mengelompokkan data dengan membuat suatu hirarki berupa *dendogram* dimana data yang mirip akan ditempatkan pada hirarki yang berdekatan dan yang tidak pada hirarki yang berjauhan. Di samping kedua pendekatan tersebut, ada juga *Clustering* dengan pendekatan *automatic mapping (Self-Organising Map/SOM)*.

Clustering adalah metode data mining yang *Unsupervised* yaitu data yang tak terawasi, karena tidak ada satu atributpun yang digunakan untuk memandu proses pembelajaran, jadi seluruh atribut input diperlakukan sama dan pada *cluster* ini tidak memiliki target atribut. Kebanyakan Algoritma *Clustering* membangun sebuah model melalui serangkaian pengulangan dan berhenti ketika model tersebut telah memusat atau berkumpul (batasan dari segmentasi ini telah stabil).

Proses pengelompokan sekumpulan obyek kedalam kelas obyek yang sama disebut *clustering*. Tujuan pengelompokan adalah untuk mengidentifikasi pengelompokan alami data dari satu set data yang besar untuk menghasilkan representasi ringkas dari perilaku sistem. Pengelompokan merupakan satu dari sekian banyak fungsi proses data mining untuk menemukan kelompok atau identifikasi kelompok obyek yang hampir sama. *Clustering* berbeda dengan group, group memiliki kelompok yang sama dimana kondisinya jika tidak ya maka sudah pasti bukan kelompoknya. Namun tidak demikian dengan *cluster*, *cluster* tidak harus sama akan tetapi pengelompokannya berdasarkan pada kedekatan dari suatu karakteristik sample data yang ada. Analisis *cluster* merupakan usaha untuk mengidentifikasi kelompok obyek yang mirip dan membantu menemukan pola penyebaran

dan pola hubungan dalam sekumpulan data yang besar. Hal penting dalam proses pengelompokan adalah menyatakan sekumpulan pola ke kelompok yang sesuai yang berguna untuk menemukan kesamaan dan perbedaan sehingga dapat menghasilkan kesimpulan yang berharga

K-Means

Metode *k-means* pertama kali diperkenalkan oleh MacQueen JB pada tahun 1976. Metode ini adalah salah satu metode *non hierarchi* yang umum digunakan. Metode ini termasuk dalam teknik penyekatan (*partition*) yang membagi atau memisahkan objek ke *k* daerah bagian yang terpisah.

Metode ini mempartisi data ke dalam *cluster* 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 lain.

Digunakan algoritma *k-means* yang di dalamnya memuat aturan sebagai berikut:

- a. Jumlah *cluster* yang diinginkan.
- b. Hanya memiliki atribut bertipe numerik.

Kelebihan metode *k-means* diantaranya adalah mampu mengelompokkan objek besar dengan sangat sehingga mempercepat proses pengelompokan. Adapun kekurangan yang dimiliki oleh *k-means* diantaranya:

1. Sangat sensitif pada pembangkitan titik pusat awal secara acak.
2. Memungkinkan suatu gerombol tidak mempunyai anggota.
3. Hasil pengelompokan bersifat tidak unik (selalu berubah-ubah) terkadang bagus terkadang tidak.
4. Sangat sulit mencapai global optimum.
5. Algoritma *k-means Clustering* walaupun proses pengerjaannya cepat tetapi keakuratannya tidak dijamin.

Adapun tujuan dari data *Clustering* ini adalah untuk meminimalisasikan *objective function* yang diset dalam proses *Clustering*, yang pada umumnya berusaha meminimalisasikan variasi di dalam suatu *cluster* dan memaksimalkan variasi antar *cluster*. Data *Clustering* menggunakan metode *K-Means* ini secara umum dilakukan dengan algoritma dasar sebagai berikut:

- a. menentukan jumlah *cluster*.
- b. Mengalokasikan data ke dalam *cluster* secara random.
- c. Menghitung *centroid*/rata-rata dari data yang ada di masing-masing *cluster*.
- d. Mengalokasikan masing-masing data ke *centroid*/rata-rata terdekat.
- e. Kembali ke Step c, apabila masih ada data yang berpindah *cluster* atau apabila perubahan nilai *centroid* ada yang di atas nilai *threshold* yang ditentukan atau apabila perubahan nilai pada *objective function* yang digunakan di atas nilai *threshold* yang ditentukan.

Formula (1) merupakan formula yang digunakan untuk mengukur jarak antar 2 data atau lebih menggunakan *Distances Manhattan* :

$$d(x,y)=|x_1-y_1|+|x_2-y_2|+ \dots |x_n-y_n|.....(1)$$

METODE PENELITIAN

Dalam pengumpulan data ini terdapat teknik-teknik, antara lain sebagai berikut :

- a. Wawancara

Berikut beberapa komponen yang sudah dilakukan wawancara:

1. Peternak burung *Lovebird*

Dari hasil wawancara diketahui bahwa, klasifikasi burung *lovebird* dilakukan didasarkan pada tanda-tanda fisik yang ada pada burung *Lovebird*. Namun tidak semua peternak, terutama pemula yang dapat melakukan klasifikasi dengan tepat. Terkadang mereka harus bertanya ke peternak yang sudah ahli. Ketika bertanya kepada

ahlinya mereka memerlukan waktu dan tenaga yang tidak sedikit.

- b. Observasi

Observasi yang dilakukan penulis adalah untuk mengamati dan mengetahui secara langsung jalannya sistem yang sedang berjalan atau proses yang dilakukan pada sistem saat itu. Dalam hal ini ada beberapa objek sasaran yang akan dijadikan data mentah pembuatan sistem diantaranya data burung *Lovebird* dan cirinya.

- c. Dokumentasi

Dokumentasi merupakan teknik pengumpulan data yang dilakukan dengan cara meminta data sistem dan data klasifikasi burung *Lovebird* dari tempat penelitian.

- d. Studi Literatur

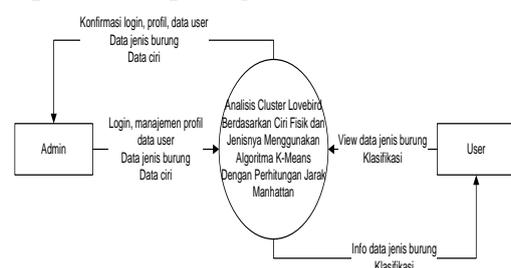
Selain itu dalam pengumpulan data ini juga didapat dari literatur yang berhubungan dengan pengolahan aplikasi, literatur yang membahas *database* dan aplikasi komputer lainnya. Dengan peran itu maka sangat butuh sekali untuk dilakukan wawancara pada pihak-pihak yang terkait ini tujuan supaya tahu keinginan, saran, harapan, keluhan dan lain sebagainya yang mengenai aplikasi sistem informasi yang akan dibuat

Data Flow Diagram (DFD)

Adapun beberapa data flow diagram sistemnya seperti dibawah ini :

- a. DFD Level 0

Data Flow Diagram Level 0 mencakup satu simbol proses yang mewakili seluruh aplikasi Analisis *Cluster Lovebird* Berdasarkan Ciri Fisik Dan Jenisnya Menggunakan Algoritma *K-Means* Dengan Penghitungan Jarak *Manhattan*. Context diagram aplikasi dapat dilihat pada gambar 1 :



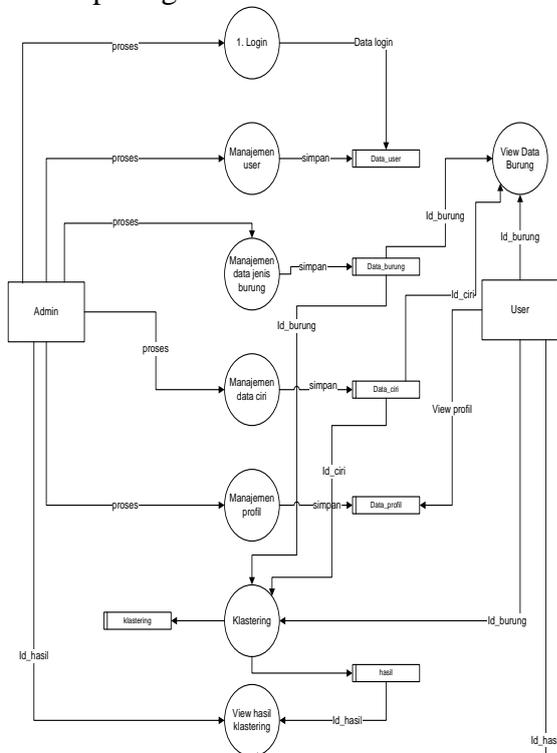
Gambar 1. DFD Level 0

Dari gambar DFD di atas dapat disimpulkan bahwa alur sistem informasi ini terdapat 2 user yaitu admin dan user. Admin adalah pihak yang akan mengelola sistem. Sebelum masuk ke dalam sistem, admin harus melakukan login terlebih dahulu. Admin melakukan manajemen data user, data profil, data jenis burung dan data ciri.

User hanya dapat membaca informasi yang diberikan oleh admin. Informasi yang dapat diakses oleh user adalah informasi data jenis burung dan melakukan klasifikasi jenis burung dengan menggunakan metode K-Means.

b. Data Flow Diagram Level 1

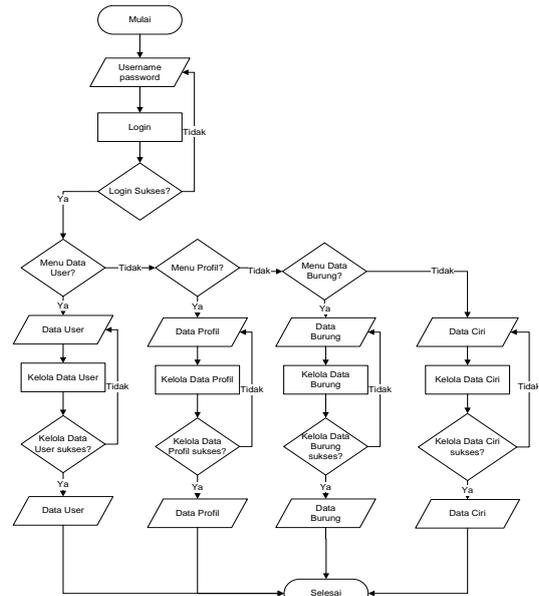
Data Flow Diagram level 1 merupakan bentuk penjabaran dengan lebih jelas dari DFD level 0. DFD level 1 untuk sistem pakar klasifikasi burung *lovebird* berdasarkan cirinya menggunakan metode certainty dapat dilihat pada gambar 2



Gambar 2. DFD Level 2

Sehingga pada gambar 2, dapat dilihat proses yang dapat dilakukan oleh setiap user, dan hak aksesnya. Beserta hubungan proses dan tabel pada seluruh sistem.

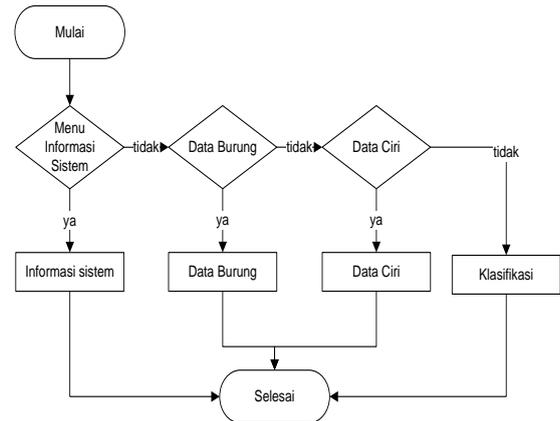
Flowchart Admin



Gambar 3. Flowchart Admin

Seperti yang terlihat pada gambar 3 flowchart atau digarm alir pada halaman admin, Admin melakukan login ke dalam sistem. Jika login berhasil, maka admin akan memasuki menu admin. Menu admin terdiri dari menu data user, data profil, data burung, dan data ciri. Admin dapat melakukan pengolahan (input, edit dan delete data) tiap menu yang ada.

Flowchart User



Gambar 4. Flowchart User

Selain diagram alir pada halaman admin seperti yang ditunjukkan pada gambar 3, flowchart user akan menjelaskan aliran diagram pada halaman user seperti yang dijelaskan pada gambar 4. User dapat masuk ke dalam sistem tanpa harus melakukan

login. User dapat membaca informasi sistem, data burung, data ciri dan melakukan klasifikasi dengan menggunakan *Clustering* dengan perhitungan jarak menggunakan Manhattan

HASIL DAN PEMBAHASAN

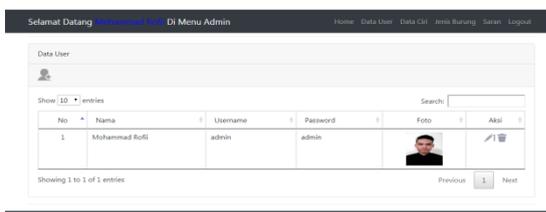
Sesuai dengan permasalahan yang telah dibahas dalam dalam bab sebelumnya, maka *sistem* aplikasi yang dikembangkan diharapkan mampu menjawab segala permasalahan yang telah diungkapkan. Berikut ini merupakan implementasi dari hasil analisis dan perancangan yang telah dilakukan.

Dari rancangan antarmuka yang telah dipaparkan pada bab sebelumnya, berikut ini adalah hasil rancangan antarmuka setelah diimplementasikan dengan menggunakan bahasa pemrograman php dan database mysql. Secara umum, menu aplikasi terdiri dari dua menu yaitu menu admin dan menu user. Gambar 5 merupakan gambar halaman login admin

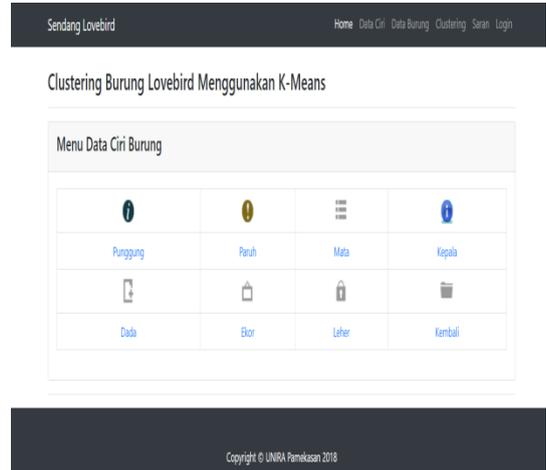


Gambar 5. Login Admin

Admin adalah pihak yang akan mengelola konten sistem. untuk masuk ke menu admin, admin harus melakukan login. Jika login berhasil, maka sistem akan menampilkan menu admin, setelah berhasil mengisi form login pada gambar 5, maka user yang berhasil login akan masuk ke halaman data user pada gambar 6

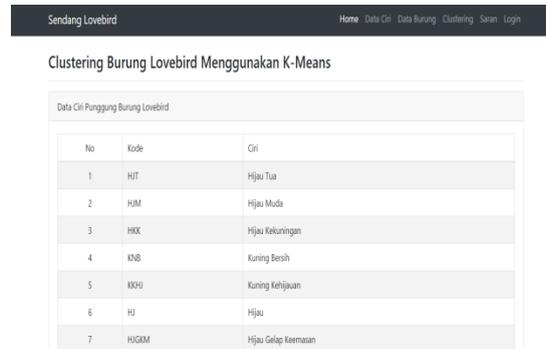


Gambar 6. Menu Data User
Menu ini akan menampilkan data user sistem. data yang telah dimasukkan dapat diedit dan dihapus dengan cara menekan tombol Edit dan Hapus. Admin dapat menambah data dengan cara menekan tombol Tambah Data

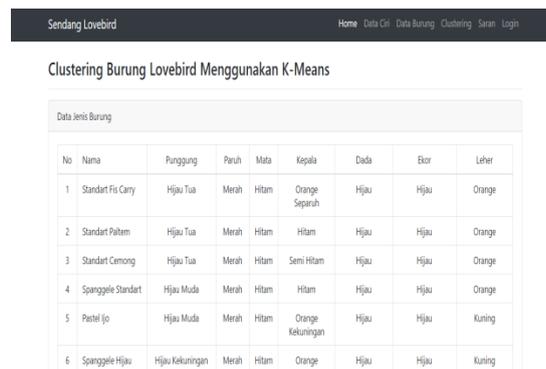


Gambar 7. Menu Ciri

Gambar 7 merupakan Menu yang akan menampilkan data ciri burung yang terdiri dari menu ciri punggung pada gambar 8, ciri paruh, ciri mata, ciri kepala, ciri dada, ciri ekor dan ciri leher.



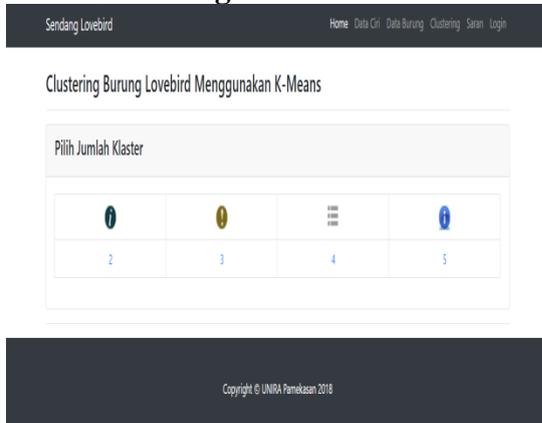
Gambar 8. Menu Ciri Punggung



Gambar 9. Menu Data Burung
Gambar 9, merupakan Menu yang

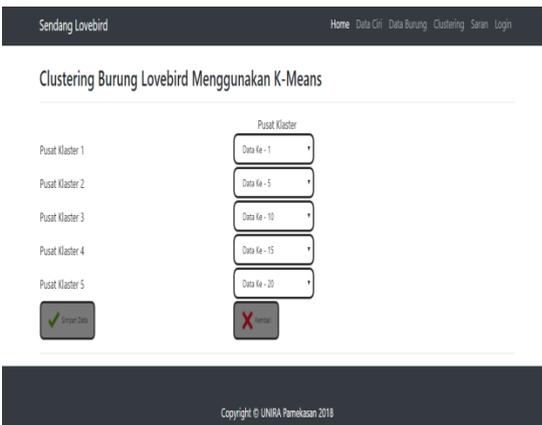
akan menampilkan data burung lovebird beserta data identitas setiap burung, seperti warna paruh, mata, kepala, dada, ekor dan leher

Menu Clustering



Gambar 10. Menu Clustering

Gambar 10 merupakan tampilan Menu yang akan menampilkan form proses clustering. User memilih jumlah Cluster sesuai dengan tombol yang telah ditentukan yaitu cluster dengan jumlah dua, tiga, empat dan lima. Tombol pilihan cluster yang telah dipilih user, merupakan jumlah cluster yang akan digunakan dalam proses selanjutnya, yaitu proses Clustering.



Gambar 11. Pusat Data Cluster

Setelah melakukan pemilihan jumlah cluster pada halaman sebelumnya, selanjutnya yang harus dilakukan user adalah memilih pusat data cluster pada gambar 11. Menu ini akan menampilkan proses pemilihan pusat Cluster. Tombol Simpan Data untuk menyimpan data pusat Cluster dan melanjutkan ke proses selanjutnya.

Selamat Datang Di Menu User Home Data Ciri Data Burung Clustering Saran Login

Iterasi ke - 2

Data Ciri Burung

No	Punggung	Jarak Data					Kelompok Data				
		C1	C2	C3	C4	C5	C1	C2	C3	C4	C5
1	HJT	10,67	25,08	33	28,8	38	1	0	0	0	0
2	HIM	10	16,75	31,5	24,8	38	1	0	0	0	0
3	HKK	18	8,08	23,5	14	30	0	1	0	0	0
4	KNB	28,67	22,75	37	31,6	42	0	1	0	0	0
5	KDH	22	6,42	23,5	14	30	0	1	0	0	0
6	HJ	10,67	23,08	36,5	30,8	38	1	0	0	0	0
7	HGKM	18	13,08	27	19,8	32	0	1	0	0	0
8	HGPD	20,67	7,08	23,5	14	30	0	1	0	0	0
9	HGKK	20,67	12,58	23,5	17,4	30	0	1	0	0	0
10	BSEM	33,33	22,75	8	20	24	0	0	1	0	0

Gambar 12. Proses Clustering

Form ini akan menampilkan proses clustering. Hal ini perlu diperhatikan karena data belum konvergen sampai system memberikan validasi bahwa iterasi sudah konvergen, seperti yang ditampilkan pada gambar 12.

Selamat Datang Di Menu User localhost menyatakan Data sudah konvergen Data Burung Clustering Saran Login

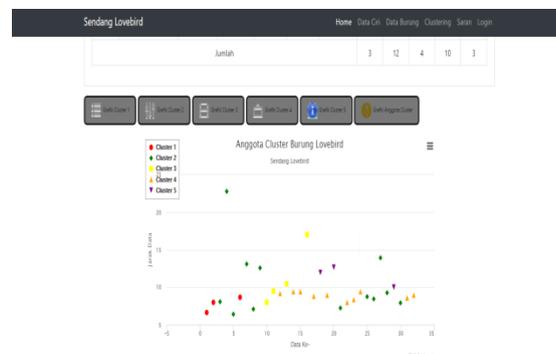
Iterasi ke - 2

Data Ciri Burung

No	Punggung	Jarak Data					Kelompok Data				
		C1	C2	C3	C4	C5	C1	C2	C3	C4	C5
1	HJT	10,67	25,08	33	28,8	38	1	0	0	0	0
2	HIM	10	16,75	31,5	24,8	38	1	0	0	0	0
3	HKK	18	8,08	23,5	14	30	0	1	0	0	0
4	KNB	28,67	22,75	37	31,6	42	0	1	0	0	0
5	KDH	22	6,42	23,5	14	30	0	1	0	0	0
6	HJ	10,67	23,08	36,5	30,8	38	1	0	0	0	0
7	HGKM	18	13,08	27	19,8	32	0	1	0	0	0
8	HGPD	20,67	7,08	23,5	14	30	0	1	0	0	0

Gambar 13. Data Konvergen

Data konvergen terjadi jika pengelompokan data tidak mengalami perubahan. Hal ini terjadi karena pusat Cluster baru tidak mengalami perubahan, seperti yang ditampilkan pada gambar 13 yaitu tampilan data-data konvergen.



Gambar 14. Grafik Anggota Cluster Tampilan pada gambar 14, merupakan tampilan Grafik yang menampilkan jarak data anggota Cluster yang telah konvergen.

