

IMPLEMENTASI *K-MEANS* *CLUSTERING* DAN *COSINE SIMILARITY* PADA *CASE BASE REASONING* UNTUK DIAGNOSIS PENYAKIT SAPI (STUDI KASUS: PENYAKIT SAPI DI KOTA BENGKULU)

Zahira Salsabila Barly¹, Funny Farady Coastera², Mochammad Yusa³

^{1,2,3}Program Studi Informatika, Fakultas Teknik, Universitas Bengkulu
Jl. WR. Supratman Kandang Limun Bengkulu 38371A Indonesia
(telp: 0736-341022; fax:0736-341022)

¹zsalsabilabarly@gmail.com

²ffaradyc@unib.ac.id

³mochammad.yusa@unib.ac.id

Abstrak: Dari berbagai macam jenis hewan ternak yang paling banyak dipelihara atau ditenakkan oleh peternak adalah sapi. Sebagai negara yang memiliki iklim tropis, berbagai macam keadaan cuaca dapat menyebabkan kesehatan sapi perlu diperhatikan karena dapat menimbulkan penyakit. Dengan dilakukan penelitian ini dapat mempermudah peternak sapi dalam mendiagnosa penyakit pada sapi dengan gejala-gejala yang diderita menggunakan sistem yang mengimplementasikan algoritma *K-Means Clustering* dan *Cosine Similarity* pada *Case Base Reasoning* untuk diagnosa penyakit sapi. Sistem pakar yang dirancang dapat mendeteksi 7 penyakit dengan total 34 gejala serta memberikan solusinya. Algoritma *K-Means Clustering* digunakan untuk mengelompokkan basis pengetahuan. Terdapat 3 kelompok yang ditentukan menggunakan metode *elbow*. *Cosine similarity* digunakan untuk mendapatkan nilai kemiripan kasus dengan kasus lama untuk menentukan penyakit yang diderita oleh sapi. Proses pengujian fungsionalitas sistem pakar ini berjalan dengan baik menggunakan *black box testing* dan menghasilkan keberhasilan fungsional sebesar 100%. Penelitian ini dilakukan dengan menguji 40 data uji dan menghasilkan tingkat akurasi sebesar 92,5%.

Kata Kunci: Sistem Pakar, Penyakit Sapi, *K-Means Clustering*, *Cosine Similarity*, *Case Base Reasoning*.

Abstract: *the various types of livestock and the most widely kept or bred by farmers are cows. As a country that has a tropical climate, various kinds of weather conditions can cause the health of cows to be considered because they can cause disease. By doing this research, it can make it easier for cow farmers to diagnose diseases in cow with symptoms suffered using a system that implements the K-Means Clustering and Cosine Similarity algorithm in Case Base Reasoning for diagnosing cow disease. The expert system designed can detect 7 diseases with a total of 34 symptoms and provide solutions. K-Means Clustering algorithm is used to group the knowledge base. There are 3 groups determined using the elbow method. Cosine similarity is used to obtain the similarity value of the case to the old case to determine the disease suffered by the cow.*

The process of testing the functionality of this expert system went well using black box testing and resulted in 100% functional success. This research was conducted by testing 40 test data and resulted in an accuracy rate of 92.5%.

Keywords: *Expert System, Cow Diseases, K-Means Clustering, Cosine Similarity, Case Base Reasoning*

I. PENDAHULUAN

Peternakan adalah salah satu sub sektor penting dari sektor pertanian. Sektor peternakan penting karena selain berkontribusi terhadap tekanan ekonomi, sektor peternakan juga dapat memenuhi kebutuhan gizi masyarakat[1]. Dari berbagai macam jenis hewan ternak yang paling banyak

dipelihara atau ditenakkan oleh peternak adalah sapi. Sapi menduduki peringkat pertama sebagai komoditas unggulan, diikuti oleh ayam dan kambing[2].

Sebagai negara yang memiliki iklim tropis, berbagai macam keadaan cuaca dapat menyebabkan kesehatan sapi perlu diperhatikan karena dapat menimbulkan penyakit. Jumlah sapi yang terdapat di Kota Bengkulu sesuai dengan data yang didapatkan dari [3], pada tahun 2012 terdapat 4.877 ekor sapi, pada tahun 2013 terdapat 5.915 ekor sapi, pada tahun 2014 terdapat 6.050 ekor sapi. Untuk tahun 2017, dengan sumber dari [4], terdapat 6.080 ekor sapi, tahun 2018 terdapat 7.960 ekor sapi, tahun 2019 terdapat 8.360 ekor sapi, dan di tahun 2020 terdapat 8.310 ekor sapi. Penyebab penyakit pada sapi berbagai macam seperti virus, bakteri, parasit, cacing, dan sebagainya[5].

Menurut narasumber yaitu Bapak Ekwan Sinto, S.Pt, penyakit yang sering ditemukan pada sapi-sapi di Kota Bengkulu adalah *Scabies*, Hipokalsemia, Demam 3 Hari / BEF (*Bovine Ephemeral Fever*), Jembrana, *Pink Eye*, MCF (*Malignant Catarrhal Fever*) dan Kecacingan. Untuk dokter hewan yang dimiliki Kota Bengkulu saat ini diperkirakan berjumlah 3 orang dan saat ini peternak sapi di Kota Bengkulu diperkirakan berjumlah 500-an orang. Dengan jumlah hewan ternak yang banyak dan dokter hewan yang sedikit maka peternak seringkali kesulitan jika ingin berkonsultasi terkait permasalahan tentang hewan ternak sapi.

Case Base Reasoning (CBR) merupakan salah satu metode pemecahan masalah yaitu dengan mencari solusi dari suatu kasus yang baru, lalu sistem akan melakukan pencarian terhadap solusi dari kasus lama yang memiliki permasalahan yang

sama dan sudah pernah terjadi sebelumnya[6]. Penyajian pengetahuan (*knowledge representation*) dibuat dalam bentuk kasus-kasus (*cases*). Cara kerja CBR adalah dengan membandingkan kasus baru dengan kasus lama, jika kasus baru tersebut mempunyai kemiripan dengan kasus lama maka CBR akan melakukan adaptasi dengan memasukkan kasus baru tersebut ke dalam basis data penyimpanan kasus (*case base*), sehingga secara tidak langsung pengetahuan CBR akan bertambah[7].

Karena berbasis kasus, maka seiring waktu kasus yang ada akan bertambah banyak. Jika kasus lama yang telah ada pada basis kasus (*case base*) memiliki jumlah yang banyak, maka akan muncul kendala dalam lamanya waktu yang diperlukan untuk memperoleh hasil dari proses pencarian kasus yang mirip atau sama pada data karena sistem harus menghitung nilai kemiripan kasus baru terhadap semua kasus lama yang telah ada. Oleh karena itu, diperlukan proses *clustering*.

Proses *clustering* merupakan proses pengelompokan kasus yang ada berdasarkan fitur yang ditentukan[7]. Metode ini berusaha mengelompokkan data yang ada ke dalam beberapa kelompok, sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu *cluster* yang sama dan data yang memiliki karakteristik yang berbeda dikelompokkan ke dalam *cluster* yang lain[8].

Di dalam CBR untuk menyelesaikan permasalahan baru yaitu dengan cara mengadaptasi solusi yang terdapat pada kasus sebelumnya dengan cara menghitung tingkat kemiripan (*similaritas*). Pada penelitian ini, penulis menggunakan metode *cosine similarity*. *Cosine similarity* adalah metode pengukuran kemiripan

antara dua objek yang berbeda. Nilai *cosine similarity* terletak antara 0 dan 1. Untuk nilai 1 menyatakan kemiripan hingga 100%, sedangkan jika nilai 0 maka menyatakan 100% tidak ada kemiripan antara dua objek[9].

Untuk menunjang penelitian yang akan dibuat, maka peneliti membutuhkan referensi mengenai penelitian terkait. terdapat penelitian yang dilakukan oleh referensi[10]. Tujuan dilakukannya penelitian memudahkan mengidentifikasi penyakit pada sapi Bali dengan metode *Certainty Factor* (CF) dan menghasilkan nilai tingkat kepastian dari suatu penyakit yang menyerang sapi bali dan dapat membantu dokter hewan dalam mengidentifikasi jenis-jenis penyakit pada sapi bali[10].

Penelitian selanjutnya terkait algoritma *clustering K-Means* pernah dilakukan oleh [11] memiliki tujuan untuk mengelompokkan tingkat penyebaran Covid-19 di daerah-daerah di Indonesia menggunakan algoritma *cluster K-Means* yang menghasilkan 3 *cluster* yaitu *cluster* pertama merupakan daerah penyebaran Covid-19 terbesar di Indonesia dengan kasus positif dan kasus meninggal paling besar yaitu Provinsi DKI Jakarta, *cluster* kedua dan ketiga merupakan daerah penyebaran kasus Covid-19 yang cukup potensial yaitu berada pada provinsi Jawa Barat dan Banten[11].

Lalu penelitian yang dilakukan oleh [12] bertujuan untuk membuktikan bahwa metode *K-Means* dapat mengurangi kesalahan pada saat *centroid* awal yang dipilih dengan hasil akurasi *clustering* dalam *centroid* dengan hanya menggunakan metode *K-Means* sebesar 86.67% sedangkan untuk penggunaan *K-Means* dan *cosine similarity* dihasilkan akurasi sebesar 89.7%[12].

Melihat dari kebutuhan penanganan yang mengharuskan cepat tanggap untuk menghadapi penyakit yang dapat menyerang sapi di Kota Bengkulu, peneliti mengusulkan sebuah sistem informasi berbasis web dengan judul “Implementasi *K-Means Clustering* dan *Cosine Similarity* pada *Case Base Reasoning* untuk Diagnosa Penyakit Sapi” dengan studi kasus penyakit sapi di Kota Bengkulu.

II. METODOLOGI

A. Jenis Penelitian

Pada penelitian ini akan dibangun sebuah aplikasi yang akan digunakan untuk mendiagnosa penyakit pada hewan ternak sapi dengan menggunakan *K-Means Clustering* dan algoritma *Cosine Similarity* serta metode *Case Base Reasoning*. Penelitian yang dilakukan merupakan jenis penelitian terapan (*applied research*), dimana penelitian terapan bertujuan untuk dapat memecahkan permasalahan yang dihadapi dan menemukan solusi bagi masyarakat atau organisasi tertentu.

B. Metode Pengumpulan Data

1. Metode Wawancara

Data yang dikumpulkan dan digunakan merupakan data yang diperoleh dari hasil wawancara oleh beberapa pakar yaitu Bapak Ekwan Sinarto, S.Pt. sebagai inseminator sapi dan Ibu drh. Henny Kusumadewi sebagai dokter hewan di Dinas Pangan dan Pertanian Kota Bengkulu bagian kesehatan hewan, yang mana pakar-pakar tersebut merupakan sumber utama informasi data penyakit dan gejala yang dialami oleh sapi-sapi yang ada di Kota Bengkulu serta solusi dari penyakit yang ada.

2. Sumber Data

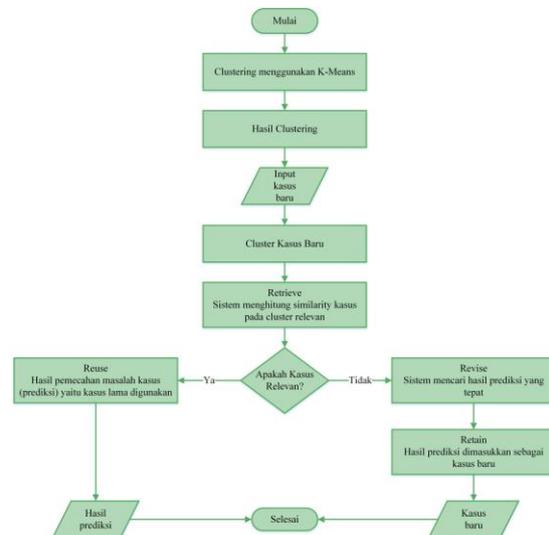
Data yang dikumpulkan dan digunakan merupakan data yang diperoleh dari hasil wawancara oleh beberapa pakar yaitu Bapak Ekwan Sinarto, S.Pt. sebagai inseminator sapi dan Ibu drh. Henny Kusumadewi sebagai dokter hewan di Dinas Pangan dan Pertanian Kota Bengkulu bagian kesehatan hewan, yang mana pakar-pakar tersebut merupakan sumber utama informasi data penyakit dan gejala yang dialami oleh sapi-sapi yang ada di Kota Bengkulu serta solusi dari penyakit yang ada.

3. Metode Observasi

Metode observasi dilakukan untuk mengumpulkan data dengan cara pengamatan langsung dengan hal-hal yang berkaitan dengan penelitian, sehingga peneliti melakukan pengamatan langsung ke peternakan sapi di daerah Kota Bengkulu, yaitu di daerah Semarang yang dikelola oleh Bapak Yanto dan Bapak Bayu. Peneliti juga mendokumentasikan beberapa sapi beserta penyakitnya dan kegiatan di peternakan bersama pakar dan pengelola peternakan.

III. HASIL DAN PEMBAHASAN

A. Diagram Alir Sistem



Gambar 1. Diagram Alir Sistem

Diagram alir sistem yang ditampilkan pada Gambar 1 digambarkan bahwa langkah awal yang dilakukan yaitu melakukan *clustering* menggunakan *K-Means* dan akan menghasilkan hasil *clustering*. Dari hasil *clustering* akan diinput kasus baru dengan perhitungan jarak paling dekat terhadap *centroid*, lalu akan terbentuk *cluster* baru dengan kasus baru. Kasus baru yang digunakan didapatkan dari masukan berupa data berbentuk kata-kata yaitu gejala. Kemudian sistem akan melakukan proses *Retrieve* (menghitung kesamaan gejala menggunakan algoritma *Cosine Similarity*). Apabila kasusnya relevan atau mirip maka sistem akan melakukan proses *Reuse* data kasus lama untuk digunakan kembali sehingga langsung didapatkan hasil diagnosanya. Sedangkan apabila kasusnya tidak relevan atau tidak mirip, maka sistem akan melakukan proses *Revise* dan *Retain* untuk menjadikan gejala dari penyakit yang ada sebagai kasus baru.

Dengan menggunakan kedua metode yaitu *K-Means Clustering* dan *Cosine Similarity*, penelitian yang dilakukan dapat menghasilkan diagnosa

dengan ketelitian yang cukup tinggi karena algoritma *clustering* yang digunakan cukup efisien untuk pengolahan objek. Selain itu, terdapat perbedaan antara penelitian sebelumnya yang hanya menggunakan algoritma *similarity*, pada penelitian ini tidak perlu mencari nilai kesamaan pada seluruh kasus yang ada pada sistem, hanya mencari nilai kesamaan pada seluruh kasus yang memiliki *cluster* yang sama dengan kasus baru.

1) Pengujian Algoritma K-Means Clustering

a. Menentukan jumlah *cluster*

Untuk menentukan jumlah *cluster* atau jumlah kelompok digunakan metode *elbow* dan didapatkan bahwa nilai k atau jumlah *cluster* (jumlah kelompok) terbaik adalah sama dengan 3.

b. Menentukan pusat *cluster* atau *centroid* awal

c. Melakukan perhitungan jarak pada setiap kasus terhadap *centroid* awal

$$d(p, q) = \sqrt{(X_{1p} - X_{1q})^2 + \dots + (X_{np} - X_{nq})^2}$$

Jarak data Kasus 91 ke Pusat Cluster 1

$$= \sqrt{\frac{(0.25 - 0.75)^2 + (0.25 - 0.25)^2 + (0 - 0)^2 + (0.25 - 0)^2 + (0 - 0)^2 + (0 - 0)^2 + (0 - 0)^2 + (0.25 - 0)^2 + (0 - 0)^2}{(0.25 - 0)^2 + (0 - 0.25)^2 + (0 - 0.25)^2}} = \sqrt{0.375} = \underline{\underline{0.612}}$$

Jarak data Kasus 91 ke Pusat Cluster 2

$$= \sqrt{\frac{(0.25 - 0)^2 + (0.25 - 0)^2 + (0 - 0)^2 + (0.25 - 0.25)^2 + (0 - 0.25)^2 + (0 - 0)^2 + (0 - 0.25)^2 + (0 - 0)^2 + (0.25 - 0)^2 + (0 - 0.25)^2}{(0.25 - 0)^2 + (0 - 0.25)^2}} = \sqrt{0.375} = \underline{\underline{0.612}}$$

Jarak data Kasus 91 ke Pusat Cluster 3

$$= \sqrt{\frac{(0.25 - 0)^2 + (0.25 - 0)^2 + (0 - 0.32)^2 + (0.25 - 0)^2 + (0 - 0.16)^2 + (0 - 0)^2 + (0 - 0.16)^2 + (0 - 0.16)^2 + (0.25 - 0)^2 + (0 - 0.16)^2}{(0 - 0.16)^2 + (0.25 - 0)^2 + (0 - 0.16)^2}} = \sqrt{0.454} = \underline{\underline{0.674}}$$

d. Melakukan perhitungan jarak terhadap *centroid* baru

Perhitungan yang dilakukan tetap menggunakan rumus *euclidean distance* dan tetap

menentukan jarak paling dekat (nilai paling kecil) dari setiap kasus terhadap *centroid* baru. Perhitungan terus dilakukan hingga nilai *centroid* sudah tidak berubah lagi. Pada perhitungan Kasus 91 masuk ke dalam *cluster* 1 dengan 6 kali iterasi.

2) Pengujian Cosine Similarity

Pengujian dilakukan untuk mengetahui kesesuaian penyakit yang didiagnosa dari hasil perhitungan manual dan dari hasil perhitungan sistem dengan nilai kepercayaan yang sudah ditentukan.

$$\cos(\emptyset) = \frac{A \cdot B}{\|A\| \cdot \|B\|} = \frac{\sum A \cdot B}{\sqrt{\sum A^2 \cdot \sum B^2}}$$

Kasus 91	= A	=	0.4,	0.6,	0.9,	0.5
Kasus 9	= B	=	0,	1,	1,	1

$$A \cdot B = \sum A \cdot B_1 = (0.4 \times 0) + (0.6 \times 1) + (0.9 \times 1) + (0.5 \times 1) = 2$$

$$\|A\| = \sqrt{\sum A^2} = \sqrt{0.4^2 + 0.6^2 + 0.9^2 + 0.5^2} = \sqrt{2.06} = 1.435$$

$$\|B\| = \sqrt{\sum B^2} = \sqrt{0^2 + 1^2 + 1^2 + 1^2} = \sqrt{3} = 1.732$$

$$\cos(\emptyset) = \frac{\sum A \cdot B}{\sqrt{\sum A^2 \cdot \sum B^2}} = \frac{2}{1.435 \times 1.732} = \frac{0.80}{80\%}$$

Dari hasil yang didapatkan, nilai kemiripan Kasus 91 memiliki nilai kemiripan paling besar dengan Kasus 9 yaitu sebanyak 80% dan didiagnosa dengan penyakit Scabies.

3) Pengujian Akurasi

Pengujian akurasi dilakukan untuk mengetahui akurasi dari hasil diagnosa penyakit pada sapi yang diperoleh dari analisa pakar dan perhitungan pada sistem. Berdasarkan hasil pengujian yang

berjumlah 40 data, diperoleh nilai akurasi menggunakan persamaan yang mengacu pada (5) seperti di bawah berikut:

$$\text{Nilai Akurasi} = \frac{\text{Jumlah data akurat}}{\text{Jumlah seluruh data}} \times 100\% \quad \dots(5)$$

$$\text{Nilai Akurasi} = \frac{37}{40} \times 100\% = 92.5\%$$

Maka dapat disimpulkan bahwa akurasi sistem berdasarkan 40 data uji adalah sebesar 92.5 % yang menunjukkan bahwa sistem berfungsi dengan cukup baik sesuai dengan diagnosa pakar.

IV. KESIMPULAN DAN SARAN

A. Kesimpulan

Berdasarkan hasil dan pembahasan yang telah dijabarkan sebelumnya, maka kesimpulan yang dapat diambil adalah sebagai berikut:

1. Penelitian ini telah menghasilkan Sistem Informasi Penyakit Sapi berbasis web dengan mengimplementasikan metode *K-Means Clustering* dan *Cosine Similarity*. Pengujian fungsional sistem dengan menggunakan *black box testing* menghasilkan keberhasilan fungsional sebesar 100%.
2. Sistem informasi penyakit sapi menghasilkan nilai akurasi sebesar 92,5% berdasarkan pengujian sebanyak 40 data.

B. Saran

Dalam membangun Sistem Informasi Penyakit Sapi masih belum sempurna dan masih banyak kekurangannya. Oleh karena itu perlu dikembangkan dan penyempurnaan lebih lanjut. Adapun saran yang dapat diberikan untuk pengembangan penelitian selanjutnya adalah sebagai berikut:

1. Sistem Informasi Penyakit Sapi dengan metode *K-Means Clustering* dan *Cosine Similarity* ini dapat dikembangkan lebih lanjut dalam bentuk berbasis android.
2. Sistem Informasi Penyakit Sapi dapat dikembangkan dengan menggunakan metode atau algoritma lainnya.

REFERENSI

- [1] Sugihandono, A., Kusri, & Al Fatta, H., "Case Base Reasoning Sebagai Alat Bantu Diagnosis Penyakit Ternak Sapi menggunakan Sorenson Coefficient (Studi Kasus: Di Kabupaten Pati)," *Jurnal Teknologi Informasi*, hal. 31-42, 2019.

- [2] Muzakkir, I., & Botutihe, M. H., "Case Based Reasoning Method untuk Sistem Pakar Diagnosa," *ILKOM Jurnal Ilmiah*, hal. 25-31, 2020.
- [3] Badan Pusat Statistik, 'Populasi Ternak menurut Jenisnya di Kota Bengkulu 2012-2014', 2014. [Online]. Tersedia:
<https://bengkulukota.bps.go.id/indicator/24/66/1/populasi-ternak-menurut-jenisnya-di-kota-bengkulu.html>. (Diakses: 6 April 2021).
- [4] Badan Pusat Statistik, 'Populasi Ternak Menurut Kabupaten/Kota dan Jenis Ternak di Provinsi Bengkulu (Ribuan ekor)', 2020. [Online]. Tersedia:
https://www.bps.go.id/indikator/indikator/view_data_public/1700/api_pub/S2ViU1dwVTlpSXRwU1MvendHN05Cdz09/da_05/1. (Diakses: 6 April 2021).
- [5] Faransyah, S. P., Mola, S. A., & Nabuasa, Y. Y., "Implementasi Case Base Reasoning Menggunakan Metode Cosine Similarity Untuk Mendiagnosa Penyakit Pada Sapi," *J-ICON*, hal. 47-52, 2018.
- [6] Putri, T. E., Andreswari, D., & Efendi, R., "Implementasi Metode CBR (Case Based Reasoning) Dalam Pemilihan Pesticida Terhadap Hama Padi Sawah Menggunakan Algoritma K-Nearest Neighbor (KNN) (Studi Kasus Kabupaten Seluma)," *Jurnal Rekursif*, vol. 4, 2016.
- [7] Rismawan, T., & Hartati, S., "Case-Based Reasoning untuk Diagnosa Penyakit THT (Telinga Hidung dan Tenggorokan)," *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, hal. 67-78, 2012.
- [8] S.Y., H., Rismayani, & Syam, A., "Data Mining Menggunakan Algoritma K-Means Pengelompokan Penyebaran Diare Di Kota Makassar," *Prosiding Seminar Ilmiah Sistem Informasi dan Teknologi Informasi*, hal. 73-82, 2019.
- [9] Fauzi, M. A., & dkk., "Automatic Essay Scoring System Using N-Gram and Cosine Similarity for Gamification Based E-Learning," *ICAIP (International Conference on Advances in Image Processing)*, hal. 151-155, 2017.
- [10] Rasyid, M., & Sumijan, "Sistem Pakar dalam Mengidentifikasi Penyakit pada Sapi Bali Menggunakan Metode Certainty Factor," *Jurnal Informasi dan Teknologi*, hal. 174-180, 2021.
- [11] Dwitri, N., & dkk., "Penerapan Algoritma K-Means Dalam Menentukan Tingkat Penyebaran Pandemi Covid-19 di Indonesia," *JurTI (Jurnal Teknologi Informasi)*, vol. 4, 2020.
- [12] Azmi, F., Utama, K., Gurning, O. T., & Ndraha, S., "Optimalisasi Centroid Awal Algoritma K-Means dengan Cosine Similarity," *JITE (Journal of Informatics and Telecommunication Engineering)*, vol. 3, hal. 224-231, 2020.