

KLASIFIKASI DAERAH RAWAN PENYAKIT HEWAN MENULAR STRATEGIS MENGGUNAKAN ALGORITMA *DECISION TREE* C4.5

Nadya Susanti Hasibuan¹, Baenil Huda², Shofa Shofiah Hilabi³, April Lia Hananto⁴

1,2,3,4) Sistem Informasi, Fakultas Ilmu Komputer, Universitas Buana Perjuangan Karawang, Indonesia

Article Info

Article history:

Received: 09 April 2026

Revised: 14 April 2026

Accepted: 17 April 2026

ABSTRACT

Abstrak

Penelitian ini memiliki tujuan untuk mengklasifikasikan tingkat kerawanan daerah berdasarkan kasus penyakit hewan menular strategis di wilayah kabupaten atau kota Provinsi Nusa Tenggara Barat dengan memanfaatkan teknik data mining. Dalam studi ini, metode yang diterapkan adalah algoritma *Decision Tree* C4.5 dengan menggunakan teknik evaluasi *5-fold cross validation* sebagai penilaian efektivitas model. Dataset yang digunakan merupakan data sekunder periode 2020–2024 yang melalui tahapan *preprocessing* meliputi *data cleaning*, agregasi, *feature engineering*, dan pelabelan menjadi tiga kategori kerawanan: rendah, sedang, dan tinggi. Temuan dari penelitian ini menunjukkan bahwa variabel rata-rata kasus menjadi faktor utama dalam menentukan tingkat kerawanan wilayah. Model yang dibangun mampu menghasilkan klasifikasi dengan performa yang baik, ditunjukkan oleh akurasi rata-rata sebesar 94% dan nilai tinggi untuk *precision*, *recall*, serta *F1-score* di setiap kategori. Selain itu, model memiliki kemampuan interpretasi yang jelas melalui struktur pohon keputusan yang dihasilkan. Dengan demikian, pendekatan yang digunakan terbukti efektif dalam mengidentifikasi wilayah berisiko dan dapat mendukung pengambilan keputusan dalam upaya pengendalian penyakit hewan menular strategis secara lebih tepat dan berbasis data.

Kata Kunci: Klasifikasi, Penyakit Hewan Menular, *Decision Tree* C4.5, Data Mining, Kerawanan Wilayah.

Abstract

This study aims to classify regional vulnerability levels based on strategic infectious animal disease cases in districts or cities in West Nusa Tenggara Province using data mining techniques. The study employed the Decision Tree C4.5 algorithm, employing a 5-fold cross-validation evaluation technique to assess model effectiveness. The dataset used was secondary data from 2020–2024, which underwent preprocessing stages including data cleaning, aggregation, feature engineering, and labeling into three vulnerability categories: low, medium, and high. The findings of this study indicate that the average case rate is the primary factor in determining regional vulnerability levels. The developed model was able to produce classifications with good performance, demonstrated by an average accuracy of 94% and high values for precision, recall, and F1-score in each category. Furthermore, the model has clear interpretability through the resulting decision tree structure. Thus, the approach used has proven effective in identifying at-risk areas and can support decision-making in efforts to control strategic infectious animal diseases more precisely and based on data.

Keywords: Classification, Infectious Animal Diseases, *Decision Tree* C4.5, Data Mining, Regional Vulnerability.

Djtechno: Jurnal Teknologi Informasi oleh Universitas Dharmawangsa Artikel ini bersifat open access yang didistribusikan di bawah syarat dan ketentuan dengan Lisensi Internasional Creative Commons Attribution NonCommercial ShareAlike 4.0 ([CC-BY-NC-SA](https://creativecommons.org/licenses/by-nc-sa/4.0/)).



Corresponding Author:

E-mail : nadyasusantihisibuan@gmail.com

1. PENDAHULUAN

Bidang peternakan memiliki kontribusi yang signifikan dalam mendukung ketahanan pangan, perekonomian daerah, serta kesejahteraan masyarakat di Indonesia. Namun, keberlanjutan sektor ini masih menghadapi berbagai tantangan, salah satunya adalah penyebaran penyakit hewan menular strategis (PHMS) yang dapat menurunkan produktivitas ternak serta menimbulkan kerugian ekonomi yang signifikan. Beberapa penelitian menunjukkan bahwa penyebaran penyakit hewan menular tidak hanya memengaruhi kesehatan hewan, tetapi juga berhubungan dengan kesejahteraan hewan, stabilitas ekonomi peternakan, serta kebijakan kesehatan Masyarakat [1], [2]. Selain itu, peningkatan mobilitas ternak, perubahan iklim, serta keterbatasan sistem deteksi dini turut memperbesar potensi penyebaran penyakit hewan di wilayah tropis seperti Indonesia [3].

Di Provinsi Nusa Tenggara Barat (NTB), berbagai jenis penyakit hewan menular seperti *Anthrax*, *Rabies*, *Scabies*, *Theleziasis*, dan *Coccidiosis* masih ditemukan setiap tahunnya. Informasi dari dinas peternakan & kesehatan hewan Provinsi NTB menunjukkan bahwa penyebaran penyakit tersebut masih menjadi masalah yang perlu mendapatkan perhatian serius [4]. Penelitian sebelumnya juga menyebutkan bahwa kasus rabies masih ditemukan secara rutin di wilayah NTB [5]. Selain itu, hasil survei dari balai besar veteriner Denpasar tahun 2024 menunjukkan penyakit *Septicaemia Epizootica* (SE) masih berpotensi menjadi ancaman dengan tingkat seropositif sebesar 23,01% pada sampel sapi, yang mengindikasikan rendahnya tingkat kekebalan kelompok ternak [6].

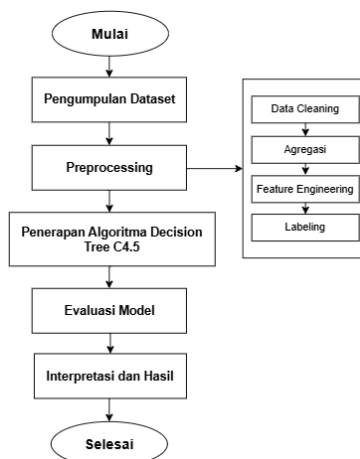
Pemanfaatan teknik data mining adalah cara pendekatan yang bisa dilakukan guna menganalisis pola penyebaran penyakit serta memprediksi tingkat kerawanan suatu wilayah [7], [8]. Metode klasifikasi dengan pendekatan pohon keputusan telah ramai digunakan pada berbagai sistem prediksi penyakit karena dapat menghasilkan model yang sederhana dan memiliki tingkat akurasi yang tinggi [9], [10]. Penelitian sebelumnya memperlihatkan bahwa algoritma *Decision Tree* C4.5 menunjukkan kinerja yang sangat baik pada proses klasifikasi data, yang memperoleh tingkat akurasi sebesar 96,80% dalam klasifikasi data operasional [11]. Pada penelitian lain menunjukkan penggunaan algoritma C4.5 cukup efektif dalam menganalisis hubungan antar variabel untuk mengidentifikasi faktor risiko berbasis wilayah [12]. Pada penelitian lain menunjukkan bahwa metode 5-fold cross validation efektif digunakan sebagai teknik evaluasi akurasi dalam klasifikasi data [13].

Penelitian yang secara khusus memetakan tingkat kerawanan penyakit hewan menular berbasis wilayah masih terbatas. Maka dari itu, tujuan dari penelitian ini adalah untuk mengembangkan model pengklasifikasian daerah yang rawan terhadap penyakit *zoonosis* di Provinsi NTB menggunakan algoritma *Decision Tree* C4.5. Metode pengujian dalam penelitian ini dengan memakai metode evaluasi *5-fold cross validation* untuk menilai tingkat akurasi dari model klasifikasi yang dibangun. Diharapkan hasil penelitian dapat menyajikan data serta informasi yang tepat serta mudah dipahami guna membantu

pengambilan keputusan di berbagai daerah provinsi NTB dalam upaya mitigasi risiko, penentuan prioritas vaksinasi, serta perencanaan tindakan pencegahan penyakit hewan berbasis data.

2. METODE PENELITIAN

Metode penelitian ini dirancang secara sistematis untuk menguraikan langkah-langkah yang dilakukan dalam menganalisis dan mengklasifikasikan daerah dengan tingkat kerawanan penyakit hewan menular strategis Provinsi NTB. Berikut Adalah tahapan alur penelitian yang dilakukan:



Gambar 1. Metode Penelitian

Sumber Gambar: Dokumentasi Pribadi

Gambar 1 menunjukkan alur metode penelitian, diawali dengan proses pengumpulan data penyakit hewan menular yang diambil dari portal resmi Provinsi NTB. Lalu dilakukan tahap *preprocessing* data, tahap berikutnya proses klasifikasi dengan menerapkan algoritma Decision Tree C4.5 dan tahap terakhir Adalah evaluasi model untuk mengetahui kinerja model dalam menghasilkan Tingkat kerawanan penyakit hewan menular.

2.1 Pengumpulan Dataset

Data yang dipakai pada penelitian ini ialah kumpulan informasi pendukung mengenai kasus penyakit hewan menular strategis di kabupaten atau kota Provinsi Nusa Tenggara Barat untuk periode antara tahun 2020 sampai 2024. Data diperoleh dari laman resmi dataNTB yang terdiri dari 999 data Excel (xlsx.) berisi informasi mengenai jenis penyakit, jumlah kasus, wilayah, dan tahun kejadian.

2.2 Preprocessing

Dilakukan tahap pra-pemrosesan guna menyiapkan data sebelum digunakan dalam langkah pemodelan. Maka dilakukan beberapa tahapan yang meliputi:

Data Cleaning

Data dibersihkan untuk memastikan kualitas data yang baik dengan melakukan pengecekan terhadap nilai yang hilang (*missing value*), duplikasi data, serta penyesuaian format data numerik agar dapat diproses dengan baik oleh sistem [14].

Transformasi Data (Agregasi)

Data yang semula berbentuk per jenis penyakit ditransformasikan melalui proses agregasi berdasarkan kabupaten/kota dan tahun. Proses ini bertujuan untuk mengubah unit analisis menjadi wilayah per tahun. Hasil dari proses ini menghasilkan dua variabel utama, yaitu total kasus dan jumlah jenis penyakit pada masing-masing wilayah.

Feature Engineering

Feature engineering dilakukan dengan membentuk variabel baru berupa rata-rata kasus, yang diperoleh dari pembagian total kasus dengan jumlah jenis penyakit. Variabel ini digunakan untuk merepresentasikan intensitas penyebaran penyakit pada setiap wilayah.

Labeling

Label kerawanan dibentuk berdasarkan total kasus menggunakan metode *discretization* dengan pendekatan kuantil (*quantile-based discretization*)[15]. Dibagi menjadi tiga kategori, yaitu rendah, sedang, dan tinggi, sehingga dapat digunakan sebagai target dalam proses klasifikasi.

2.3 Penerapan Algoritma *Decision Tree* C4.5

Model klasifikasi dibentuk dengan menerapkan algoritma *Decision Tree* C4.5 melalui pendekatan entropi sebagai kriteria pemilihan atribut terbaik [16]. Variabel yang berperan sebagai input pada penelitian ini mencakup jumlah jenis penyakit, rata-rata kasus, dan tahun, sedangkan variabel target adalah kategori tingkat kerawanan.

2.4 Evaluasi Model

Evaluasi model diterapkan dengan pendekatan *5-fold cross validation* guna mendapatkan hasil yang lebih konsisten serta mengurangi potensi bias yang diakibatkan oleh pemisahan data. Dataset dibagi menjadi lima segmen, setiap segmen secara bergantian digunakan sebagai data pengujian, sementara segmen-segmen lainnya digunakan sebagai data pelatihan [17]. Selanjutnya dilakukan evaluasi kinerja model dengan menerapkan *confusion matrix* serta metrik evaluasi lain seperti *accuracy*, *precision*, *recall*, dan *F1-score* [18]. Metode ini digunakan untuk menilai sejauh mana model dapat mengklasifikasikan tingkat kerawanan secara tepat pada setiap kategori[19].

2.5 Interpretasi Hasil

Tahap interpretasi dilakukan dengan menganalisis struktur pohon keputusan yang dihasilkan oleh algoritma *Decision Tree* [20]. Model divisualisasikan dalam bentuk diagram pohon untuk memahami proses pengambilan keputusan dalam klasifikasi [21]. Selain itu, dilakukan analisis terhadap aturan-aturan klasifikasi (*if-then rules*) serta variabel yang paling berpengaruh dalam menentukan tingkat kerawanan[22]. Hasil interpretasi ini digunakan untuk memberikan pemahaman yang lebih mendalam terhadap pola penyebaran penyakit serta sebagai dasar dalam perolehan keputusan.

3. HASIL DAN PEMBAHASAN

3.1 Preprocessing

Setelah informasi terkumpul, langkah berikutnya adalah melakukan *preprocessing*. Di tahap ini akan dilakukan pengecekan terhadap struktur data sebelum masuk ke tahap penerapan model.

Data Cleaning

Data Cleaning dilakukan guna menghapus data duplikat yang berpotensi memengaruhi hasil klasifikasi, dan juga menyeragamkan penulisan atribut. Hasil pembersihan dataset yang terdiri dari 999 data tersebut menunjukkan bahwa tidak ditemukan *missing value* maupun data yang terduplikasi. Hal ini menunjukkan bahwa dataset siap dan layak digunakan pada tahap selanjutnya

Transformasi Data (Agregasi)

Agregasi dilakukan dengan menjumlahkan seluruh kasus penyakit pada setiap wilayah untuk memperoleh variabel total kasus, serta menghitung jumlah jenis penyakit yang muncul pada masing-masing wilayah. Hasil dari proses ini menghasilkan dataset baru dengan jumlah 50 baris data baru yang merepresentasikan kondisi penyebaran penyakit pada setiap kabupaten/kota per tahun.

Tabel 1. Hasil Agregasi Data

Kabupaten/Kota	Tahun	Total Kasus	Jumlah Jenis Penyakit
Kabupaten Bima	2020	1654	20
Kabupaten Bima	2021	1559	20
Kabupaten Bima	2022	10378	20
Kabupaten Bima	2023	2997	20
Kabupaten Bima	2024	24117	20

Sumber Tabel: Dokumentasi Pribadi

Berdasarkan Tabel 1, hasil agregasi menunjukkan bahwa setiap baris data merepresentasikan satu wilayah pada satu tahun tertentu dengan nilai total kasus dan jumlah jenis penyakit yang berbeda-beda. Data ini kemudian digunakan sebagai dasar dalam proses pembentukan fitur dan klasifikasi tingkat kerawanan.

Feature Engineering

Pada tahap feature engineering dilakukan pembentukan variabel baru untuk memperkaya informasi dalam dataset. Variabel yang dibentuk adalah rata-rata kasus, yang diperoleh dari pembagian total kasus dengan jumlah jenis penyakit pada setiap wilayah. Variabel ini digunakan untuk merepresentasikan intensitas penyebaran penyakit dan merupakan salah satu elemen krusial dalam proses pengklasifikasian tingkat kerawanan daerah.

Labeling

Selain itu, dilakukan proses pelabelan data berdasarkan total kasus menggunakan metode *discretization* yang membagi ke dalam tiga kategori: "rendah", "sedang", dan "tinggi". Hasil akhir preprocessing menghasilkan sebanyak 50 data yang siap digunakan dalam proses klasifikasi.

Tabel 2. Distribusi Kategori Data

Kategori	Jumlah Data
Rendah	17
Sedang	16
Tinggi	17

Sumber Tabel: Dokumentasi Pribadi

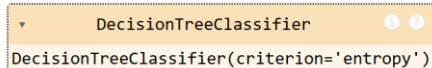
Dapat dilihat pada Tabel 2, bahwa distribusi data pada setiap kategori relatif seimbang, sehingga tidak terdapat ketimpangan kelas yang signifikan dalam dataset.

3.2 Penerapan *Decision Tree* C4.5

Pada tahap penerapan algoritma, model klasifikasi dibentuk dengan menggunakan algoritma pohon keputusan (*Decision Tree*) yang memakai kriteria entropi sebagai dasar dalam pemilihan atribut terbaik. Penggunaan kriteria entropy bertujuan untuk mengukur tingkat ketidakpastian data dalam proses pembentukan pohon keputusan. Berikut Adalah kode program yang digunakan:

```
from sklearn.tree import DecisionTreeClassifier

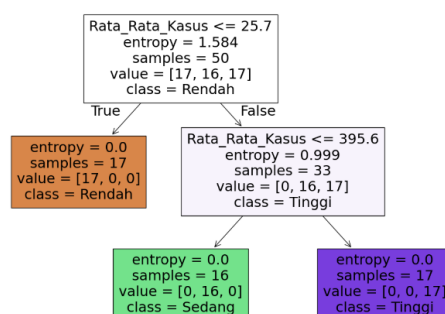
DecisionTreeClassifier(criterion="entropy")
```


Gambar 2. Kode Program *Decision Tree* C4.5

Sumber Gambar: Dokumentasi Pribadi

Implementasi model dijalankan dengan menggunakan *library scikit-learn* pada bahasa pemrograman Python, dengan memanfaatkan kelas *DecisionTreeClassifier*. Parameter yang diterapkan pada penelitian ini adalah *criterion = entropy*, yang sesuai dengan konsep dasar algoritma C4.5 dalam menciptakan struktur pohon keputusan.

Model kemudian dilatih dengan memanfaatkan dataset yang telah melalui tahap *preprocessing* guna menghasilkan aturan klasifikasi dalam bentuk pohon keputusan. Model klasifikasi dirancang dengan menerapkan algoritma *Decision Tree* C4.5 dengan pendekatan *entropy*. Variabel input yang digunakan meliputi jumlah jenis penyakit, rata-rata kasus, dan tahun, sedangkan variabel target adalah kategori tingkat kerawanan. Tahapan pelatihan model diproses dengan menerapkan metode *5-fold cross validation* guna memastikan bahwa model menunjukkan kinerja yang konsisten terhadap variasi pembagian data.

Gambar 3. Pohon Keputusan *Decision Tree* C4.5

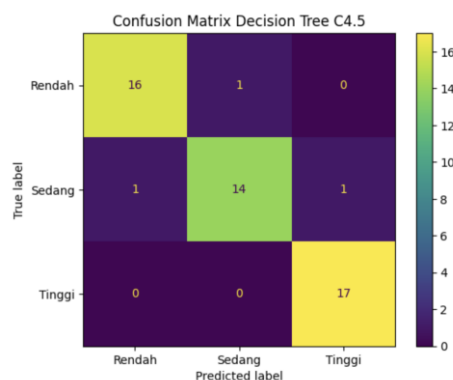
Sumber Gambar: Dokumentasi Pribadi

Berdasarkan hasil visualisasi pohon keputusan pada gambar 4, diketahui bahwa variabel Rata-Rata Kasus merupakan faktor utama (*root node*) dalam menentukan tingkat kerawanan daerah. Hal ini ditunjukkan dengan posisinya sebagai *root node* pada pohon keputusan. Model membagi data berdasarkan nilai ambang sebesar 395.6, di mana daerah dengan nilai rata-rata kasus di atas ambang tersebut cenderung diklasifikasikan sebagai daerah dengan tingkat kerawanan tinggi.

Sebaliknya, daerah dengan nilai rata-rata kasus di bawah ambang tersebut cenderung masuk ke dalam kategori rendah atau sedang. Hasil ini menunjukkan bahwa intensitas penyebaran penyakit (yang direpresentasikan oleh rata-rata kasus) lebih berpengaruh dibandingkan hanya jumlah jenis penyakit dalam menentukan tingkat kerawanan daerah.

3.3 Evaluasi Model

Proses evaluasi model dijalankan dengan menerapkan pendekatan *5-fold cross validation method*. Berdasarkan hasil pengujian, diperoleh nilai akurasi pada masing-masing fold sebesar 0.90, 0.80, 1.00, 1.00, dan 1.00 dengan rata-rata akurasi sebesar 0.94 atau 94%. Hasil penelitian ini mengindikasikan bahwa model menghasilkan performa yang bagus dan cukup stabil dalam mengklasifikasikan tingkat kerawanan daerah. Gambar 5 menunjukkan hasil dari *confusion matrix*.



Gambar 4. *Confusion Matrix*

Sumber Gambar: Dokumentasi Pribadi

Berdasarkan gambar *confusion matrix* diatas, terlihat bahwa hampir seluruh data sukses diklasifikasikan dengan tepat. Pada kelas "rendah", sebanyak 16 data diklasifikasikan dengan benar dan 1 data mengalami kesalahan klasifikasi. Pada kelas "sedang", terdapat 14 data yang diklasifikasikan dengan benar, sedangkan 2 data mengalami kesalahan. Sementara itu, pada kelas "tinggi" seluruh data berhasil diklasifikasikan dengan tepat tanpa adanya kesalahan.

Selain itu, hasil evaluasi yang dilakukan menggunakan *precision*, *recall*, serta *F1-score* ditampilkan pada Gambar 6.

Classification Report:				
	precision	recall	f1-score	support
Rendah	0.94	0.94	0.94	17
Sedang	0.93	0.88	0.90	16
Tinggi	0.94	1.00	0.97	17
accuracy			0.94	50
macro avg	0.94	0.94	0.94	50
weighted avg	0.94	0.94	0.94	50

Gambar 5. Hasil Evaluasi Model

Sumber Gambar: Dokumentasi Pribadi

Terlihat bahwa model menunjukkan performa yang memuaskan pada setiap kelas dengan nilai evaluasi di atas 0.90. Kelas "tinggi" memperoleh skor *recall* sebesar 1.00, yang menandakan bahwa seluruh data dengan tingkat kerawanan tinggi berhasil terdeteksi oleh model.

3.4 Interpretasi

Berlandaskan hasil pemodelan menggunakan algoritma *Decision Tree* C4.5, diperoleh struktur pohon keputusan untuk mengklasifikasikan tingkat kerawanan daerah. Variabel rata-rata kasus menjadi faktor utama dalam klasifikasi, yang ditunjukkan sebagai *root node*, sehingga menunjukkan bahwa intensitas kasus sangat mempengaruhi tingkat kerawanan. Model membagi data berdasarkan nilai ambang pada rata-rata kasus, di mana nilai yang tinggi cenderung diklasifikasikan sebagai kerawanan tinggi, sedangkan nilai yang lebih rendah masuk ke kategori sedang atau rendah. Variabel jumlah jenis penyakit juga berpengaruh, namun tidak sebesar rata-rata kasus.

Hasil evaluasi menunjukkan akurasi sebesar 94% dengan nilai *recall* 1.00 pada kelas tinggi, yang berarti seluruh daerah dengan tingkat kerawanan tinggi berhasil teridentifikasi. Kesalahan klasifikasi sebagian besar terjadi pada kategori sedang karena memiliki karakteristik yang berada di antara dua kelas lainnya. Secara keseluruhan, model memiliki performa yang baik dan mampu memberikan interpretasi yang jelas, sehingga bisa berfungsi sebagai sarana dalam membuat pilihan mengenai pengelolaan maupun pengendalian penyakit hewan menular strategis.

4. SIMPULAN

Penelitian ini menghasilkan kesimpulan bahwa penggunaan algoritma *Decision Tree* C4.5 mampu mengklasifikasikan tingkat kerawanan penyakit hewan menular strategis pada setiap kabupaten/kota secara efektif. Proses transformasi data melalui agregasi serta pembentukan fitur seperti rata-rata kasus berhasil merepresentasikan kondisi penyebaran penyakit sehingga mendukung kinerja model dalam proses klasifikasi. Hasil pemodelan menunjukkan bahwa intensitas kasus menjadi faktor utama dalam menentukan tingkat kerawanan, yang sejalan dengan tujuan penelitian untuk mengidentifikasi wilayah berisiko berdasarkan data kasus penyakit. Selain itu, model yang dibuat tidak hanya menunjukkan kinerja yang memuaskan, tetapi juga dapat memberikan dan menciptakan penjelasan yang jelas melalui format pohon keputusan. Oleh karena itu, studi ini membuktikan bahwa teknik yang diterapkan bisa menjadi cara

yang relevan dan praktis dalam membantu proses pengambilan keputusan mengenai pengendalian penyakit hewan menular yang strategis di suatu area.

UCAPAN TERIMAKASIH

Penulis mengucapkan terima kasih yang sebesar-besarnya kepada Bapak Agustia Hananto, M.Kom. selaku dosen yang telah memberikan dukungan dan kontribusinya sebagai pemberi dana dalam pelaksanaan penelitian ini. Bantuan yang diberikan sangat berarti dalam menunjang kelancaran serta keberhasilan proses penelitian, sehingga penelitian ini dapat diselesaikan dengan baik. Semoga segala kebaikan yang telah diberikan mendapatkan balasan yang setimpal.

References

- [1] B. Bose And S. Siva Kumar, "Economic Burden Of Zoonotic And Infectious Diseases On Livestock Farmers: A Narrative Review," *J. Health Popul. Nutr.*, Vol. 44, No. 1, P. 158, May 2025, Doi: 10.1186/S41043-025-00913-3.
- [2] A. Kappes *Et Al.*, "Livestock Health And Disease Economics: A Scoping Review Of Selected Literature," *Front. Vet. Sci.*, Vol. 10, Sep. 2023, Doi: 10.3389/Fvets.2023.1168649.
- [3] S. A. Muhammad, A. Rauf, H. Ali, And M. Awais, "Impact Of Climate Change On Animal Health," *International Journal Of Agriculture And Biosciences*, Vol. 1, No. 1, Pp. 8–13, 2023, Doi: 10.47278/Book.Oht/2023.02.
- [4] Balai Besar Penelitian Veteriner, *Informasi Dan Diskripsi Singkat Penyakit Phms (Penyakit Hewan Menular Strategis)*. Bogor: Badan Penelitian Dan Pengembangan Pertanian, Kementerian Pertanian, 2023.
- [5] R. Rahmad *Et Al.*, "Rabies Outbreak In Dompu, West Nusa Tenggara, Indonesia: A Case Study," *Journal Of Community Health And Preventive Medicine*, Vol. 1, No. 1, Pp. 1–8, Jan. 2021, Doi: 10.21776/Ub.Jochapm.2021.001.01.1.
- [6] Balai Besar Veteriner Denpasar, *Laporan Surveilans Dan Monitoring Balai Besar Veteriner Denpasar*. Bali: Kementerian Pertanian Direktorat Jendral Peternakan Dan Kesehatan Hewan, 2024.
- [7] Z. I. Alfianti, "Pengelompokan Wilayah Penyebaran Covid-19 Di Kabupaten Karawang Menggunakan Algoritma K-Means," *Jurnal Ilmiah Informatika Komputer*, Vol. 26, No. 2, Pp. 111–122, Aug. 2021, Doi: 10.35760/Ik.2021.V26i2.4155.
- [8] S. Anggraini, M. Akbar, A. Wijaya, H. Syaputra, And M. Sobri, "Klasifikasi Gejala Penyakit Coronavirus Disease 19 (Covid-19) Menggunakan Machine Learning," *Journal Of Software Engineering Ampera*, Vol. 2, No. 1, Pp. 57–68, Feb. 2021, Doi: 10.51519/Journalsea.V2i1.105.
- [9] M. Bin Q. Al-Asiri And A. A. Al-Asmari, "Disease Prediction System Using Data Mining Techniques Based On Classification Mechanism: Survey Study," *Journal Of Pioneering Medical Science*, Vol. 13, No. 4, Pp. 25–31, Jul. 2024, Doi: 10.61091/Jpms202413404.
- [10] D. Safitri, S. S. Hilabi, And F. Nurapriani, "Analisis Penggunaan Algoritma Klasifikasi Dalam Prediksi Kelulusan Menggunakan Orange Data Mining," *Rabit : Jurnal Teknologi Dan Sistem Informasi Univrab*, Vol. 8, No. 1, Pp. 75–81, Jan. 2023, Doi: 10.36341/Rabit.V8i1.3009.
- [11] R. Pratama, B. Huda, E. Novalia, And H. Kabir, "Perbandingan Algoritma C4.5 Dan Naïve Bayes Dalam Menentukan Persediaan Stok," *Metik Jurnal*, Vol. 6, No. 2, Pp. 115–122, Dec. 2022, Doi: 10.47002/Metik.V6i2.379.

-
- [12] E. Indra, K. Ho, Arlinanda, R. Hakim, D. Sitanggang, And O. Sihombing, "Application Of C4.5 Algorithm For Cattle Disease Classification," *J. Phys. Conf. Ser.*, Vol. 1230, No. 1, P. 012070, Jul. 2019, Doi: 10.1088/1742-6596/1230/1/012070.
- [13] K. Pal And Biraj. V. Patel, "Data Classification With K-Fold Cross Validation And Holdout Accuracy Estimation Methods With 5 Different Machine Learning Techniques," In *2020 Fourth International Conference On Computing Methodologies And Communication (Iccmc)*, Ieee, Mar. 2020, Pp. 83–87. Doi: 10.1109/Iccmc48092.2020.Iccmc-00016.
- [14] M. A. Fadli, T. Tukino, E. Novalia, And A. L. Hananto, "Klasifikasi Ulasan Aplikasi Kopi Kenangan Pada Google Playstore Menggunakan Algoritma Naive Bayes," *Djtechno: Jurnal Teknologi Informasi*, Vol. 6, No. 2, Pp. 593–605, Aug. 2025, Doi: 10.46576/Djtechno.V6i2.7037.
- [15] R. Mazumder, X. Meng, And H. Wang, "Quant-Bnb: A Scalable Branch-And-Bound Method For Optimal Decision Trees With Continuous Features," In *International Conference On Machine Learning*, Plmr, 2022, Pp. 15255–15277.
- [16] P. Chen, "The Application Of An Improved C4.5 Decision Tree," In *2021 7th Annual International Conference On Network And Information Systems For Computers (Icnisc)*, Ieee, Jul. 2021, Pp. 392–396. Doi: 10.1109/Icnisc54316.2021.00078.
- [17] X. Ouyang *Et Al.*, "Dual-Sampling Attention Network For Diagnosis Of Covid-19 From Community Acquired Pneumonia," *Ieee Trans. Med. Imaging*, Vol. 39, No. 8, Pp. 2595–2605, Aug. 2020, Doi: 10.1109/Tmi.2020.2995508.
- [18] R. Yacouby And D. Axman, "Probabilistic Extension Of Precision, Recall, And F1 Score For More Thorough Evaluation Of Classification Models," In *Proceedings Of The First Workshop On Evaluation And Comparison Of Nlp Systems*, Stroudsburg, Pa, Usa: Association For Computational Linguistics, 2020, Pp. 79–91. Doi: 10.18653/V1/2020.Eval4nlp-1.9.
- [19] F. M. Delta Maharani, A. L. Hananto, S. S. Hilabi, F. N. Apriani, A. Hananto, And B. Huda, "Perbandingan Metode Klasifikasi Sentimen Analisis Penggunaan E-Wallet Menggunakan Algoritma Naive Bayes Dan K-Nearest Neighbor," *Metik Jurnal*, Vol. 6, No. 2, Pp. 97–103, Dec. 2022, Doi: 10.47002/Metik.V6i2.372.
- [20] R. U. Ginting, M. Y. Harefa, Rianto Sitanggang, And Immanuel H G Manurung, "Klasifikasi Penyebaran Covid-19 Menggunakan Algoritma C4.5 Di Kecamatan Medan Helvetia," *Jurnal Mahajana Informasi*, Vol. 8, No. 2, Pp. 99–107, Dec. 2023, Doi: 10.51544/Jurnalmi.V8i2.4650.
- [21] B. Huda *Et Al.*, "Analisis Sentimen E-Learning X Terhadap Antarmuka Pengguna Menggunakan Kombinasi Multinomial Naive Bayes Dan Pendekatan Design Thinking," *Jurnal Teknologi Informasi Dan Ilmu Komputer*, Vol. 11, No. 4, Pp. 895–902, Aug. 2024, Doi: 10.25126/Jtiik.1147686.
- [22] F. M. Hana, "Klasifikasi Penderita Penyakit Diabetes Menggunakan Algoritma Decision Tree C4.5," *Jurnal Siskom-Kb (Sistem Komputer Dan Kecerdasan Buatan)*, Vol. 4, No. 1, Pp. 32–39, Oct. 2020, Doi: 10.47970/Siskom-Kb.V4i1.173.