

Evaluasi Kinerja Naïve Bayes, Decision Tree, Dan Random Forest Serta Voting Ensemble Pada Klasifikasi Multi-Kelas Penyakit Sapi Berbasis Gejala

Nazwa Diajeng Istika Rahmadhani*, Nurgiyatna

Fakultas Komunikasi dan Informatika, Prodi Teknik Informatika, Universitas Muhammadiyah Surakarta, Surakarta, Indonesia

Email: ¹*l200220251@student.ums.ac.id, ²nurgiyatna@ums.ac.id

Email Penulis Korespondensi: l200220251@student.ums.ac.id

Submitted: 18/04/2026; Accepted: 02/06/2026; Published: 05/06/2026

Abstrak—Sapi merupakan komoditas ternak penting, namun proses diagnosis penyakit masih menjadi kendala bagi peternak akibat kemiripan gejala klinis dan keterbatasan akses tenaga ahli. Penelitian ini bertujuan untuk membandingkan kinerja algoritma *Naïve Bayes*, *Decision Tree*, dan *Random Forest* serta mengevaluasi metode *Voting Ensemble* dalam sistem diagnosis penyakit sapi. Metodologi yang digunakan adalah CRISP-DM dengan tahapan pengolahan data, pemodelan, dan evaluasi menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score*. Hasil pengujian menunjukkan bahwa *Naïve Bayes* memberikan performa terbaik dengan *accuracy* 0.951 dan *F1-score* 0.920. *Random Forest* memperoleh *accuracy* 0.799, sedangkan *Decision Tree* menunjukkan performa terendah dengan *accuracy* 0.265. Metode *ensemble* seperti *Voting NB+RF*, *Voting Weighted*, *Voting Soft*, dan *Voting Hard* menghasilkan performa yang bervariasi dengan akurasi masing-masing sebesar 0.912, 0.900, 0.853, dan 0.792. Hasil ini menunjukkan bahwa *Naïve Bayes* lebih sesuai untuk karakteristik data gejala yang bersifat *high-dimensional* dan *sparse*, sehingga memberikan performa paling stabil dibandingkan model lainnya. Sistem kemudian diimplementasikan sebagai sistem pakar berbasis web, dengan hasil pengujian *usability* menggunakan *System Usability Scale* (SUS) sebesar 77 (kategori *Good*). Penelitian ini menunjukkan bahwa *machine learning* dapat digunakan sebagai sistem pendukung keputusan dalam diagnosis penyakit sapi.

Kata Kunci: Sistem Pakar; Diagnosa Penyakit Sapi; Machine Learning; Voting Ensemble; Komparasi Algoritma

Abstract—Cattle are an important livestock commodity; however, farmers often face difficulties in disease diagnosis due to the similarity of clinical symptoms and limited access to veterinary experts. This study aims to compare the performance of three machine learning classification algorithms, namely *Naïve Bayes*, *Decision Tree*, and *Random Forest*, and to evaluate the effectiveness of an ensemble approach using a *Voting Ensemble* method for cattle disease diagnosis. The study adopts the CRISP-DM methodology, consisting of data preprocessing, modeling, and evaluation stages. Model performance is assessed using *accuracy*, *precision*, *recall*, and *F1-score* metrics. The experimental results show that *Naïve Bayes* achieves the best performance with an *accuracy* of 0.951 and an *F1-score* of 0.920. *Random Forest* obtains an *accuracy* of 0.799, while *Decision Tree* performs the lowest with an *accuracy* of 0.265. Ensemble methods, including *Voting NB+RF*, *Voting Weighted*, *Voting Soft*, and *Voting Hard*, achieve *accuracies* of 0.912, 0.900, 0.853, and 0.792, respectively. These findings indicate that *Naïve Bayes* is more suitable for high-dimensional and sparse symptom-based data, providing the most stable performance among the evaluated models. The developed system is implemented as a web-based expert system. Usability evaluation using the *System Usability Scale* (SUS) yields a score of 77, categorized as “Good.” This study demonstrates that machine learning can support decision-making in cattle disease diagnosis.

Keywords: Expert System; Cattle Disease Diagnosis; Machine Learning; Voting Ensemble; Algorithm Comparison Expert

1. PENDAHULUAN

Hewan ternak yang memberikan banyak manfaat bagi manusia di antaranya adalah sapi, sebagai penghasil daging dan susu, serta mendukung kegiatan pertanian dan penggunaan tenaga kerja di berbagai bidang. Bahkan kotorannya dapat dimanfaatkan sebagai pupuk maupun bahan bakar alternatif, sehingga keberadaannya sangat penting bagi kehidupan masyarakat [1]. Hal ini menjadikan sapi memiliki peran strategis dalam mendukung ketahanan pangan serta perekonomian masyarakat. Peningkatan kebutuhan protein hewani, khususnya daging sapi, turut mendorong meningkatnya minat masyarakat dalam beternak sapi. Pada tahun 2023, konsumsi daging sapi diperkirakan mencapai sekitar 2,5 kg per kapita per tahun. Selain itu, pemenuhan kebutuhan susu nasional masih belum optimal karena sekitar 80% kebutuhan susu dalam negeri masih bergantung pada impor [2]. Namun, peningkatan produktivitas sapi sering terkendala oleh berbagai penyakit yang disebabkan oleh bakteri, virus, parasit, maupun faktor lingkungan. Laporan situasi 11 penyakit prioritas dari iSIKHNAS menunjukkan bahwa 6 di antaranya menyerang sapi, sehingga mengindikasikan bahwa penyakit hewan pada sapi masih menjadi tantangan utama di Indonesia. Sejak wabah Penyakit Mulut dan Kuku (PMK) pada tahun 2022, Indonesia mencatat ratusan ribu kasus secara kumulatif yang kemudian menurun signifikan pada 2023 berkat vaksinasi dan pengendalian, namun hingga 2025–2026 masih ditemukan sekitar 7.349 kasus secara sporadis, sementara *Lumpy Skin Disease* (LSD) dilaporkan sekitar 1.587 kasus pada periode yang sama. Selain itu, penyakit lain seperti *antraks*, *brucellosis*, dan penyakit jembrana juga masih ditemukan secara sporadis dan termasuk dalam penyakit prioritas pengawasan kesehatan hewan di Indonesia.

Kondisi tersebut terjadi dikarenakan keterbatasan biaya serta penanganan yang masih bersifat tradisional, dan diagnosis penyakit yang kurang akurat. Peternak mengalami kesulitan dalam membedakan gejala awal penyakit, sebab tanda klinis seperti demam dan penurunan nafsu makan sering muncul pada berbagai jenis penyakit, sehingga kerap diabaikan atau terlambat ditangani. Kondisi ini membuat proses identifikasi manual menjadi kurang efektif dan

berpotensi menimbulkan kesalahan diagnosis. Penanganan yang diberikan tidak tepat sasaran, sehingga penyakit dapat menyebarluas, meningkatkan kematian ternak, dan menimbulkan kerugian ekonomi yang signifikan bagi peternak [3].

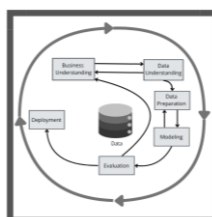
Untuk mengatasi permasalahan tersebut, pemanfaatan teknologi sistem pakar dapat digunakan guna mendukung proses pengambilan keputusan secara otomatis. Pendekatan ini memungkinkan sistem komputer meniru cara berpikir seorang ahli dalam memberikan analisis serta rekomendasi terhadap suatu permasalahan yang kompleks [4]. Sistem yang dikembangkan memanfaatkan metode klasifikasi berbasis *machine learning* dengan algoritma *Decision Tree*, *Naïve Bayes*, dan *Random Forest*, yang digabungkan menggunakan *Voting Ensemble*. *Decision Tree* adalah metode analisis berbentuk struktur pohon yang lebih kecil dan homogen berdasarkan karakteristik atau atribut yang dimilikinya [5]. Klasifikasi *Naïve Bayes* merupakan salah satu algoritma yang memiliki berbagai keunggulan, di antaranya mudah untuk diimplementasikan karena tingkat kompleksitasnya rendah serta tidak memerlukan data *training* dalam jumlah besar [6]. Di sisi lain, *Random Forest* merupakan metode *ensemble learning* yang membangun sejumlah *decision tree* pada tahap pelatihan, kemudian menggabungkan hasil prediksi untuk menghasilkan keputusan akhir, di mana pada kasus klasifikasi digunakan mekanisme voting (mode) dan pada regresi digunakan nilai rata-rata dari seluruh pohon, sehingga pendekatan ini membuat model lebih stabil serta memiliki akurasi yang lebih baik dibandingkan satu *decision tree* saja [7]. *Ensemble Learning* adalah teknik yang menggabungkan beberapa algoritma *Machine Learning* untuk mencapai akurasi prediksi tinggi dibandingkan dengan menggunakan satu algoritma saja [8].

Penelitian yang sudah dilakukan menunjukkan bahwa algoritma *Decision Tree*, *Naïve Bayes*, dan *Random Forest* memiliki karakteristik performa yang berbeda pada berbagai kasus klasifikasi. *Decision Tree* dikenal memiliki *interpretabilitas* yang baik dan proses komputasi yang relatif sederhana, namun rentan mengalami *overfitting*, terutama pada data yang tidak seimbang. *Naïve Bayes* memiliki keunggulan dari sisi efisiensi dan kecepatan proses, tetapi performanya dapat menurun ketika asumsi *independensi* antar atribut tidak terpenuhi [9]. Sementara itu, *Random Forest* sebagai metode berbasis *ensemble* cenderung memberikan performa yang lebih stabil dan akurasi yang lebih tinggi melalui kombinasi banyak pohon keputusan, meskipun hasilnya tetap dipengaruhi oleh karakteristik *dataset* yang digunakan [10]. Beberapa penelitian juga menunjukkan bahwa pendekatan *ensemble learning* mampu meningkatkan akurasi dan stabilitas prediksi dibandingkan penggunaan model tunggal, namun peningkatan tersebut tidak selalu konsisten pada *dataset* dengan distribusi fitur yang kompleks atau tidak seimbang. Pada bidang kesehatan hewan, sistem diagnosis penyakit sapi yang telah dikembangkan masih memiliki keterbatasan, seperti cakupan jenis penyakit yang terbatas, ketergantungan pada sumber data tertentu, serta antarmuka yang kurang praktis bagi pengguna non-teknis [11]-[14]. Namun demikian, kajian komparatif yang secara khusus membandingkan performa *Decision Tree*, *Naïve Bayes*, dan *Random Forest* pada *dataset* penyakit sapi berbasis teks dengan karakteristik *high-dimensional* dan *multi-kelas* masih relatif terbatas. Oleh karena itu, penelitian ini difokuskan untuk mengevaluasi performa ketiga algoritma tersebut serta menilai efektivitas pendekatan *ensemble* dalam konteks diagnosis penyakit sapi.

Berdasarkan studi terdahulu, algoritma *Decision Tree*, *Naïve Bayes*, dan *Random Forest* telah banyak digunakan dalam berbagai kasus klasifikasi, termasuk pada bidang kesehatan dan peternakan. Namun, kajian komparatif ketiga algoritma tersebut pada *dataset* penyakit sapi dengan karakteristik *high-dimensional* dan *multi-kelas* masih relatif terbatas, khususnya pada data penyakit terbaru yang berbentuk *input* teks deskriptif hasil observasi gejala dan divalidasi oleh tenaga ahli (dokter hewan). Selain itu, penerapan *ensemble learning* pada konteks tersebut belum secara konsisten menunjukkan peningkatan performa yang signifikan dibandingkan model tunggal, karena sangat bergantung pada karakteristik distribusi data yang digunakan. Di sisi lain, implementasi sistem sebelumnya masih menghadapi keterbatasan dari segi efisiensi akses serta kompleksitas antarmuka yang kurang ramah pengguna. Oleh karena itu, penelitian ini difokuskan pada analisis komparatif performa algoritma *Naïve Bayes*, *Decision Tree*, dan *Random Forest* serta evaluasi efektivitas pendekatan *Voting Ensemble* pada *dataset* gejala penyakit sapi berbasis teks yang kompleks dan berdimensi tinggi. Model yang telah dievaluasi kemudian diimplementasikan ke dalam sistem pakar berbasis web dengan antarmuka *input* teks deskriptif untuk mendukung proses identifikasi awal penyakit secara lebih praktis, fleksibel, dan dapat digunakan sebagai alat bantu keputusan pengguna di bawah pengawasan tenaga ahli.

2. METODOLOGI PENELITIAN

Metode yang digunakan dalam penelitian ini adalah CRISP-DM (*Cross Industry Standard Process for Data Mining*), yaitu sebuah metode standar yang digunakan dalam proses *data mining*. Metode ini bertujuan untuk membantu peneliti maupun praktisi dalam melakukan analisis data secara sistematis dan terstruktur, sehingga dapat menemukan pola serta merumuskan solusi terhadap permasalahan yang sedang dihadapi secara lebih efektif [15].



Gambar 1. Metode CRISP-DM

Metode CRISP-DM terdiri dari enam tahapan utama yang saling terhubung dan bersifat iteratif [16], yaitu:

2.1. Business Understanding

Tahap *Business Understanding* merupakan langkah awal dalam penelitian yang berfokus pada pemahaman secara menyeluruh terhadap konteks dan karakteristik permasalahan yang diselesaikan [17]. Pada tahapan ini dilakukan identifikasi terhadap kondisi nyata di lapangan, khususnya yang berkaitan dengan proses penentuan diagnosis penyakit sapi yang masih dilakukan secara manual oleh peternak. Keterbatasan pengetahuan, kurangnya akses terhadap tenaga ahli seperti dokter hewan, serta keterlambatan dalam mengenali gejala penyakit menjadi faktor utama yang menyebabkan proses diagnosis sering kali tidak dilakukan secara cepat dan akurat.

2.2. Data Understanding

Tahap *Data Understanding* dilakukan untuk mendapatkan pemahaman secara menyeluruh terhadap struktur, karakteristik, dan kondisi data yang digunakan dalam penelitian. Pada tahapan ini dilakukan pengumpulan data yang relevan dengan permasalahan, yaitu data gejala dan jenis penyakit sapi bersumber dari *dataset* publik di Kaggle. Selanjutnya, data tersebut dikonsultasikan dengan dokter hewan untuk memastikan validitas, kesesuaian, serta keakuratan dengan kondisi nyata di lapangan.

2.3. Data Preparation

Data Preparation merupakan proses *multi*-tahap di mana data diidentifikasi, diekstrak, dibersihkan, dan diintegrasikan sebelum digunakan untuk analisis [18]. Alur *Data Preparation* dijabarkan pada Gambar 2 dibawah.



Gambar 2. Alur Tahapan *Data Preparation*

Tahap *cleaning* merupakan proses penting dalam analisis teks yang bertujuan membersihkan dan menyiapkan data dengan menghapus elemen tidak relevan serta memastikan konsistensi agar hasil analisis lebih akurat dan dapat diandalkan [10]. Pada tahap ini dilakukan penanganan *missing value* dengan menghapus data yang kosong guna menjaga konsistensi *dataset*, serta penghapusan data duplikat untuk menghindari terjadinya bias dalam analisis. Selain itu, dilakukan pengecekan konsistensi atribut untuk memastikan keseragaman format dan tipe data, serta melakukan transformasi data, seperti mengubah variabel kategori ke dalam bentuk numerik agar dapat diproses oleh algoritma klasifikasi. Label *encoding* merupakan teknik yang digunakan untuk mengonversi data kategorikal yang berbentuk teks menjadi nilai numerik, Proses ini dilakukan dengan memberikan nilai numerik unik pada setiap kategori dalam variabel tersebut, sehingga data dapat dipahami dan diolah algoritma *machine learning* yang hanya mengenali *input* berupa angka [19]. Selanjutnya data terlebih dahulu *distandarisasi* untuk menyelaraskan format dan skala sehingga konsisten, memudahkan model dalam memprosesnya. *Preprocessed* data adalah data diproses dengan membersihkan duplikasi, menormalisasi teks, dan mengubahnya menjadi format numerik (*feature encoding* atau *vectorization*) agar siap digunakan dalam pemodelan *machine learning*.

Pada penelitian ini, tahapan yang dilakukan meliputi *preprocessing* data, modeling, dan evaluasi. *Dataset* memiliki karakteristik *high-dimensional sparse* dengan 679 sampel, 411 fitur gejala biner, dan 101 kelas penyakit. Kondisi ini menyebabkan distribusi data sangat tidak seimbang dengan rata-rata sampel per kelas yang rendah (± 6 sampel), sehingga meningkatkan risiko *class imbalance* dan *curse of dimensionality* dalam proses pembelajaran model. Sebelum tahap modeling, dilakukan penyaringan data dengan mempertahankan kelas yang memiliki minimal dua sampel untuk mengurangi ketidakseimbangan ekstrem pada *dataset*. Proses evaluasi model dilakukan menggunakan *Stratified 2-Fold Cross Validation* untuk menjaga distribusi kelas tetap seimbang pada setiap *fold*, sehingga setiap kelas tetap terwakili secara proporsional pada data latih maupun data uji. Untuk mengatasi keterbatasan jumlah data per kelas, penelitian ini menggunakan *Stratified K-Fold Cross Validation* ($K=2$) tanpa pembagian *hold-out* terpisah, sehingga seluruh data digunakan secara bergantian sebagai data latih dan data uji. Pendekatan ini dipilih untuk memastikan setiap kelas tetap terwakili pada setiap *fold* mengingat distribusi data yang terbatas.

2.4. Modeling

Tahap *Modeling* dalam metodologi CRISP-DM merupakan proses penerapan berbagai teknik pemodelan pada data yang telah melalui tahap *preparation*, dengan tujuan membangun model yang mampu menghasilkan prediksi atau keputusan dengan sasaran penelitian [20]. Pada penelitian ini digunakan empat model, yaitu *Naïve Bayes*, *Decision Tree*, *Random Forest*, dan *Voting Ensemble* untuk membandingkan performa dalam klasifikasi penyakit sapi. Perbandingan dilakukan melalui pengujian model tunggal (*Naïve Bayes*, *Decision Tree*, *Random Forest*) serta model

ensemble menggunakan *Voting Ensemble* untuk meningkatkan stabilitas prediksi. Implementasi seluruh model dilakukan menggunakan pustaka *scikit-learn* dengan konfigurasi standar untuk klasifikasi multi-kelas.

$$Entropy(S) = \sum_{i=1}^n -p_i \log_2(p_i) \quad (1)$$

Decision Tree membangun model klasifikasi dengan cara membagi data secara rekursif berdasarkan atribut yang menghasilkan penurunan *impurity* terbesar. *Entropy* digunakan untuk mengukur tingkat ketidakmurnian data, di mana semakin kecil nilai *entropy*, semakin homogen suatu *subset* data. Pemilihan atribut terbaik dilakukan berdasarkan *information gain*, sehingga struktur pohon yang terbentuk merepresentasikan pola keputusan dari data gejala penyakit sapi.

$$P(C|X) = \frac{P(C|X).P(C)}{P(X)} \quad (2)$$

Naïve Bayes mengasumsikan bahwa setiap fitur bersifat independen terhadap fitur lainnya, sehingga probabilitas gabungan dapat dihitung sebagai hasil perkalian probabilitas masing-masing fitur terhadap kelas tertentu. Meskipun asumsi ini tidak selalu terpenuhi pada data nyata, *Naïve Bayes* tetap efektif digunakan pada data berdimensi tinggi dan *sparse* seperti data gejala penyakit sapi karena kesederhanaan dan kestabilannya [21]. Selanjutnya, penelitian ini menggunakan metode *Random Forest*, yaitu *ensemble learning* yang membangun beberapa pohon keputusan dari *subset* data secara acak sebagaimana ditunjukkan pada Persamaan (3).

$$Dt \sim Bootstrap(D) \quad (3)$$

Random Forest merupakan metode *ensemble* berbasis pohon keputusan yang membangun banyak pohon dari *subset* data yang diambil secara acak menggunakan *bootstrap* sampling. Setiap pohon dilatih dengan *subset* data dan *subset* fitur yang berbeda untuk meningkatkan variasi model. Hasil akhir diperoleh melalui *majority* voting dari seluruh pohon, sehingga dapat mengurangi *overfitting* dan meningkatkan stabilitas prediksi dibandingkan *Decision Tree* tunggal. Metode *ensemble* pada penelitian ini menggabungkan hasil prediksi dari beberapa model klasifikasi untuk meningkatkan stabilitas dan akurasi sistem diagnosis. Pada metode *Soft Voting*, probabilitas tersebut digabungkan secara rata-rata, sedangkan pada *Hard Voting* digunakan mekanisme *majority* voting berdasarkan kelas dengan jumlah suara terbanyak. Sementara itu, *Weighted Voting* memberikan bobot berbeda pada setiap model sesuai dengan kontribusi dan performanya. Selain itu, pendekatan *Voting* NB+RF digunakan dengan mengombinasikan dua model utama yang memiliki karakteristik berbeda, yaitu *Naïve Bayes* yang stabil pada data berdimensi tinggi dan *Random Forest* yang mampu menangkap pola kompleks. Hasil akhir diperoleh dengan memilih kelas yang memiliki nilai probabilitas tertinggi berdasarkan fungsi *argmax*, sehingga keputusan akhir merupakan hasil konsensus dari beberapa model, bukan hanya satu model tunggal.

2.5. Evaluation

Proses evaluasi dilakukan secara bertahap untuk menjawab tujuan perbandingan [22]. Pada tahap pertama, dilakukan pengujian terhadap algoritma *Decision Tree*, *Naïve Bayes*, dan *Random Forest* secara terpisah menggunakan data uji yang telah disiapkan. Pada tahap kedua, dilakukan pengujian terhadap hasil penggabungan model menggunakan *Voting Ensemble*. Untuk mengukur performa model, digunakan beberapa metrik evaluasi yaitu *confusion matrix*, *accuracy*, *precision*, *recall*, dan *F1-score*. *Confusion matrix* digunakan untuk melihat kesesuaian antara hasil prediksi dan data aktual pada setiap kelas. *Accuracy* digunakan untuk mengukur tingkat ketepatan prediksi secara keseluruhan, sedangkan *precision* dan *recall* digunakan untuk melihat kemampuan model dalam melakukan klasifikasi pada masing-masing kelas. *F1-score* digunakan sebagai ukuran keseimbangan antara *precision* dan *recall*. Hasil evaluasi dari kedua tahap kemudian dibandingkan untuk menganalisis efektivitas masing-masing algoritma serta melihat pengaruh penggunaan metode *ensemble* dalam meningkatkan stabilitas dan performa klasifikasi penyakit sapi.

2.6. Deployment

Tahapan *Deployment* merupakan tahap akhir dalam metodologi CRISP-DM yang berfungsi untuk mengaplikasikan model yang telah dikembangkan dan dievaluasi ke dalam sistem nyata agar dapat digunakan secara langsung oleh pengguna. Pada fase ini, model sistem pakar diagnosis penyakit sapi telah melalui tahapan *Modeling* dan *Evaluation* akan diimplementasikan melalui *dashboard website* yang dirancang agar mudah diakses oleh pengguna utama, yaitu peternak. Fokus utama tahap *Deployment* adalah memastikan sistem mampu memberikan prediksi diagnosis secara cepat, *real-time*, dan akurat berdasarkan gejala yang dimasukkan, serta menghasilkan rekomendasi pengobatan yang tepat. Berikut merupakan rancangan antarmuka (*wireframe*) dari sistem pakar diagnosis penyakit sapi yang akan diimplementasikan pada tahap *deployment*.

3. HASIL DAN PEMBAHASAN

Bagian ini merupakan inti dari penelitian yang berisi hasil pengujian serta pembahasan mengenai pengembangan sistem pakar diagnosa penyakit sapi. Pembahasan disusun secara sistematis berdasarkan tahapan CRISP-DM, mulai

dari pemahaman masalah, pengolahan data, pemodelan, hingga evaluasi sistem. Fokus utama penelitian ini adalah membandingkan performa beberapa algoritma klasifikasi serta melihat pengaruh metode *ensemble* dalam menghasilkan diagnosa yang lebih stabil dan akurat untuk membantu peternak di lapangan.

3.1 Business Understanding

Permasalahan utama peternakan sapi skala rakyat di Indonesia adalah tingginya risiko kerugian akibat keterlambatan penanganan penyakit, yang disebabkan oleh anggapan remeh terhadap gejala awal seperti penurunan nafsu makan atau demam serta keterbatasan tenaga medis hewan, sehingga diperlukan sistem untuk deteksi dini dan rekomendasi penanganan awal. Oleh karena itu, diperlukan sistem yang dapat membantu deteksi dini serta memberikan rekomendasi penanganan. Pada penelitian ini dikembangkan sistem diagnosis penyakit sapi berbasis kecerdasan buatan dengan memanfaatkan data gejala yang dimasukkan pengguna. Penelitian ini menggunakan tiga algoritma klasifikasi, yaitu Naïve Bayes, Decision Tree, dan Random Forest, serta metode Voting Ensemble untuk meningkatkan stabilitas dan konsistensi prediksi.

3.2 Data Understanding

Keberhasilan model *machine learning* sangat dipengaruhi oleh kualitas data yang digunakan. Pada penelitian ini, *dataset* terdiri dari 679 data rekam medis penyakit sapi yang mencakup 101 jenis penyakit. Data diperoleh dari sumber publik *Kaggle* dan telah divalidasi oleh dokter hewan untuk memastikan kesesuaian antara gejala dan kondisi klinis di lapangan. Secara teknis, *dataset* ini memiliki 411 fitur gejala unik, sehingga tergolong data berdimensi tinggi (*high-dimensional* data). Kondisi ini menuntut proses pemodelan yang tepat agar model mampu menangkap pola secara efektif tanpa mengalami penurunan performa. Selain itu, terdapat ketidakseimbangan jumlah data pada beberapa kelas penyakit yang perlu diperhatikan pada tahap pelatihan model.

Validasi oleh dokter hewan menjadi tahap penting dalam memastikan hubungan antara gejala dan jenis penyakit sesuai dengan kondisi klinis aktual. Sebagai contoh, pada Tabel 1, setiap baris menunjukkan hubungan antara diagnosis, gejala, penyakit, penanganan, dan risiko. Pada kasus Bovine Ephemeral Fever, meskipun memiliki gejala umum seperti demam dan anoreksia, penyakit ini dapat dibedakan melalui karakteristik spesifik berupa durasi gejala yang lebih khas dibandingkan penyakit lainnya. Dengan demikian, validasi ahli memastikan bahwa representasi data yang digunakan lebih akurat dan relevan untuk proses pemodelan.

Tabel 1. Daftar Penyakit Sapi

	Diagnosa	Gejala	Penyakit	Penanganan	Resiko
0	Abses	Abses, Benjolan yang membesar dan nyeri di...	Abses	1.000unit dengan ...	NaN
1	Avitaminosis	Anemia, Anorexia, Bulu Kusam, Demam, Diare, Ga...	Avitaminosis	1.000 bolus dengan PYRONIL...	Kekurangan vitamin ...
2	Babesiosis	Bulu kusam,Gatal,Kelainan...	Babesiosis	8000gram dengan TRYPONIL...	NaN
3	Bovine Ephemeral Fever	Anorexia, Demam, Demam, Diare, Durasi gejala...	Bovine Ephemeral Fever	10.000 ml dengan B-SANPLEX dan...	
4	Corpus Luteum	Anestrus, Siklus estrus tidak teratur	Corpus Luteum Persisten	5000 ml dengan CAPRIGLADIN INJ...	NaN

Hasil eksplorasi data pada Tabel 1 juga menunjukkan adanya kemiripan gejala antar beberapa penyakit. Gejala demam tercatat paling dominan, yaitu sebanyak 147 kemunculan atau sekitar 21,65% dari total data. Hal ini mengindikasikan bahwa proses klasifikasi tidak dapat hanya bergantung pada satu gejala tunggal, melainkan harus mempertimbangkan kombinasi beberapa gejala agar hasil diagnosis lebih akurat.

3.3 Data Preparation

Tahap persiapan data merupakan proses yang cukup penting karena data gejala yang masih berbentuk teks harus diubah ke dalam format yang dapat diproses oleh mesin tanpa menghilangkan informasi utamanya.

3.3.1 Pembersihan Teks dan Normalisasi Sinonim

Teks gejala yang dimasukkan oleh peternak sering kali masih tidak baku dan menggunakan istilah yang berbeda-beda. Oleh karena itu, tahap awal dilakukan *text cleaning* yang meliputi perubahan huruf menjadi *lowercase*, penghapusan karakter yang tidak diperlukan, serta perapihan spasi. Tantangan utama pada tahap ini adalah variasi istilah dengan makna yang sama, seperti penggunaan kata panas, badan hangat, atau suhu naik untuk menyebut demam. Untuk mengatasi hal tersebut, penelitian ini menerapkan kamus sinonim kustom agar istilah-istilah tersebut diseragamkan.

Langkah ini penting karena tanpa normalisasi, sistem akan menganggap setiap istilah sebagai fitur yang berbeda, sehingga dapat menurunkan ketepatan model dalam mengenali gejala yang sama.

3.3.2 Vektorisasi dengan MultiLabelBinarizer

Setiap data gejala diubah ke dalam bentuk numerik agar dapat diproses oleh model. Pada penelitian ini digunakan metode MultiLabelBinarizer untuk mengonversi data gejala menjadi vektor biner berdasarkan seluruh fitur gejala yang tersedia. Tabel 2 menunjukkan setiap baris direpresentasikan ke dalam sejumlah kolom sesuai dengan jumlah gejala unik, di mana nilai 1 menunjukkan keberadaan gejala dan 0 menunjukkan ketiadaannya. Meskipun metode ini mendukung representasi multi-gejala dalam satu kasus, pada dataset yang digunakan setiap data umumnya hanya memiliki satu gejala dominan. Oleh karena itu, hasil encoding cenderung menghasilkan satu nilai aktif (1) pada setiap baris. Hasil proses ini membentuk matriks data biner yang digunakan sebagai input pada tahap pemodelan.

Tabel 2. Gejala yang sudah di *encoding*

	abses	ada butiran kristal di bulu kemaluan	agresif jika disentuh	air kencing berwarna merah tua coklat seperti kopi	air kencing bisa berwarna merah coklat
0	1	0	0	0	0
1	0	0	0	0	0
2	0	0	0	0	0

3.3.3 Stratified K-Fold Cross Validation

Dataset memiliki 101 kelas dengan distribusi yang tidak seimbang, Data tidak seimbang karena jumlah sampel pada setiap kelas penyakit berbeda-beda, sehingga beberapa kelas memiliki data jauh lebih sedikit dibanding kelas lainnya. Oleh karena itu penerapan *Stratified K-Fold Cross Validation* memberikan keuntungan dalam menjaga proporsi setiap kelas tetap konsisten pada setiap *fold*. Dalam penelitian ini digunakan *Stratified 2-Fold Cross Validation*, di mana seluruh data digunakan secara bergantian sebagai data latih dan data uji. Pendekatan ini memungkinkan setiap sampel berkontribusi dalam proses pelatihan maupun pengujian, sehingga hasil evaluasi lebih representatif terhadap distribusi seluruh kelas. Pemilihan nilai $K = 2$ didasarkan pada kondisi *dataset* yang memiliki jumlah sampel relatif sedikit pada setiap kelas (rata-rata ± 6 data per kelas). Penggunaan nilai K yang lebih besar berpotensi menghasilkan pembagian data yang tidak stabil dan kurang merepresentasikan distribusi kelas secara optimal.

3.4 Modeling

Pada tahap ini dilakukan analisis terhadap kinerja masing-masing algoritma dalam memproses data gejala sapi yang memiliki jumlah fitur cukup banyak.

3.4.1 Naïve Bayes

Naïve Bayes tetap efektif digunakan pada data teks maupun data biner. *MultinomialNB* digunakan untuk menghitung kemungkinan suatu penyakit berdasarkan kombinasi gejala yang terdeteksi. Keunggulan metode ini terletak pada proses komputasi yang cepat, efisien, serta cukup baik dalam menangani fitur yang kurang relevan. *Naïve Bayes* memperoleh performa tertinggi dengan *mean accuracy* sebesar 0.951 dan *mean F1-macro* sebesar 0.920, serta standar deviasi yang sangat kecil (0.001) yang menunjukkan kestabilan hasil antar *fold*.

3.4.2 Decision Tree

Decision Tree model ini menyerupai alur pemeriksaan, misalnya jika sapi mengalami demam maka sistem akan memeriksa gejala lain sebagai penentu keputusan berikutnya. Keunggulan utama metode ini adalah hasil prediksi mudah dipahami karena aturan keputusan dapat dilihat secara jelas. Namun, dengan jumlah fitur gejala yang banyak, pohon yang terbentuk dapat menjadi terlalu kompleks dan berisiko mengalami *overfitting* terhadap data latih. *Decision Tree* menghasilkan performa terendah dengan *mean accuracy* 0.265 dan *mean F1-macro* 0.139.

3.4.3 Random Forest

Sebagai pengembangan dari *Decision Tree*, *Random Forest* membangun 200 pohon keputusan yang dibentuk secara acak. Setiap pohon dilatih menggunakan sebagian data dan sebagian fitur, sehingga model tidak bergantung pada satu pola saja. Hasil prediksi akhir ditentukan berdasarkan gabungan keputusan dari seluruh pohon. Pada penelitian ini, *Random Forest* menunjukkan peningkatan dibanding *Decision Tree* dengan *mean accuracy* 0.799 dan *mean F1-macro* 0.650.

3.4.4 Mekanisme Voting Ensemble

Pada metode *ensemble* digunakan beberapa pendekatan, yaitu *Voting Soft*, *Voting Hard*, *Voting Weighted*, dan *Voting NB+RF*. Berdasarkan hasil pengujian, *Voting NB+RF* memperoleh nilai *mean accuracy* sebesar 0.912 dan *mean F1-macro* sebesar 0.849. model *Naïve Bayes* memperoleh *accuracy* sebesar 0.951, *macro precision* 0.915, *macro recall* 0.929, dan *macro F1-score* 0.920. *Decision Tree* memperoleh *accuracy* sebesar 0.265, *macro precision* 0.139, *macro*

recall 0.149, dan macro F1-score 0.139. Random Forest memperoleh accuracy sebesar 0.799, macro precision 0.644, macro recall 0.667, dan macro F1-score 0.650. Pada metode ensemble, Voting Soft memperoleh accuracy sebesar 0.853, macro precision 0.748, macro recall 0.767, dan macro F1-score 0.754. Voting NB+RF memperoleh accuracy sebesar 0.912, macro precision 0.843, macro recall 0.863, dan macro F1-score 0.849. Voting Weighted memperoleh accuracy sebesar 0.900, macro precision 0.830, macro recall 0.847, dan macro F1-score 0.836. Voting Hard memperoleh accuracy sebesar 0.792, macro precision 0.646, macro recall 0.663, dan macro F1-score 0.649.

3.5 Evaluation

Kinerja model dinilai menggunakan metrik accuracy, precision, recall, dan F1-score agar performa dapat dianalisis secara komprehensif. Berdasarkan hasil pengujian komparatif pada Tabel 3, model Naive Bayes menunjukkan performa terbaik dengan akurasi sebesar 0.951, diikuti oleh Voting NB+RF (0.912) dan Voting Weighted (0.900). Sementara itu, Decision Tree menunjukkan performa terendah dengan akurasi sebesar 0.265, yang mengindikasikan keterbatasan model dalam menangkap pola pada data berdimensi tinggi.

Tabel 3. Hasil Komparatif Performa Model

Model	Accuracy	Macro Precision	Macro Recall	Macro F1-Score	Weighted Precision	Weighted Recal	Weighted F1-Score
Naive Bayes	0.951	0.915	0.929	0.920	0.926	0.951	0.935
Decision Tree	0.265	0.139	0.149	0.139	0.249	0.265	0.249
Random Forest	0.799	0.644	0.667	0.650	0.760	0.799	0.770
Voting Soft	0.853	0.748	0.767	0.754	0.819	0.853	0.829
Ensemble							
Voting NB+RF	0.912	0.843	0.863	0.849	0.875	0.912	0.887
Voting Weighted	0.900	0.830	0.847	0.836	0.869	0.900	0.880
Voting Hard	0.792	0.646	0.663	0.649	0.762	0.792	0.768
Ensemble							

Perbedaan performa antar model menunjukkan bahwa pendekatan probabilistik seperti Naive Bayes mampu menangkap pola distribusi gejala secara lebih efektif dibandingkan model berbasis pohon keputusan. Hal ini diperkuat oleh nilai F1-macro yang tinggi, yang menunjukkan kemampuan model dalam menangani ketidakseimbangan kelas pada dataset.

3.5.1 Perbandingan Performa Model (Baseline vs Noise)

Dilakukan simulasi dengan menambahkan gangguan pada data input berupa flipping nilai fitur secara acak dengan tingkat tertentu untuk mensimulasikan ketidaksempurnaan data di dunia nyata. Hal ini bertujuan untuk menguji ketahanan model terhadap data yang tidak bersih (noisy data).

Tabel 4. Perbandingan Performa Model (Baseline vs Noise)

Model	Mean Accuracy	Std Accuracy	Mean F1 Macro	Std F1 Macro	Noise Mean Accuracy	Noise Std Accuracy	Noise Mean F1 Macro	Noise Std F1 Macro
Naive Bayes	0.951	0.001	0.920	0.001	0.922	0.007	0.882	0.014
Decision Tree	0.265	0.000	0.139	0.000	0.214	0.019	0.109	0.011
Random Forest	0.799	0.000	0.650	0.002	0.779	0.012	0.596	0.025
Voting Soft	0.853	0.012	0.754	0.022	0.839	0.019	0.737	0.023
Voting Hard	0.792	0.007	0.649	0.002	0.768	0.010	0.612	0.025
Voting Weighted	0.900	0.018	0.836	0.029	0.888	0.015	0.821	0.032
Voting NB+RF	0.912	0.009	0.849	0.016	0.897	0.018	0.829	0.031

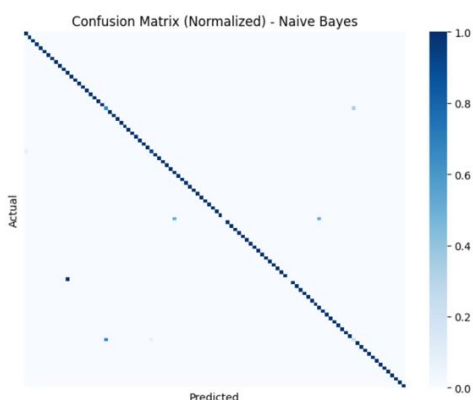
Berdasarkan Tabel 4, seluruh model mengalami penurunan performa pada kondisi data dengan noise, baik pada mean accuracy maupun mean F1-macro. Naive Bayes mengalami penurunan dari 0.951 menjadi 0.922, Random Forest dari 0.799 menjadi 0.779, dan Decision Tree dari 0.265 menjadi 0.214. Pada metode ensemble, Voting Soft, Voting Hard, Voting Weighted, dan Voting NB+RF juga menunjukkan penurunan performa dengan selisih yang relatif kecil. Selain itu, nilai standar deviasi pada kondisi noise cenderung meningkat, yang menunjukkan adanya variasi performa yang lebih besar dibandingkan kondisi baseline.

3.5.2 Analisis Stabilitas dan Robustness

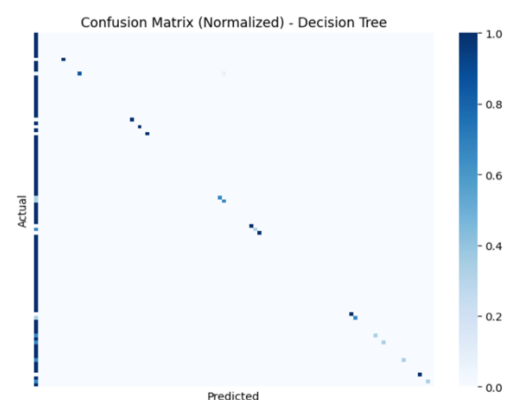
Analisis stabilitas model dilakukan berdasarkan nilai standar deviasi hasil *cross-validation* pada kondisi *baseline* dan data *noise*. Stabilitas diukur melalui konsistensi performa antar-*fold*, di mana nilai standar deviasi yang lebih kecil menunjukkan performa yang lebih konsisten. Hasil menunjukkan bahwa *Naïve Bayes* memiliki variansi paling rendah pada kondisi *baseline* (± 0.001) dan tetap relatif rendah pada kondisi *noise*. *Random Forest* menunjukkan variansi yang lebih besar ($\pm 0.002-0.012$), sedangkan metode *ensemble* berada pada rentang ($\pm 0.009-0.018$). *Decision Tree* memiliki variansi yang sangat kecil pada *baseline*, namun meningkat pada kondisi *noise*. Pola yang sama juga terlihat pada pengujian dengan data *noise*, di mana seluruh model mengalami peningkatan variansi. Meskipun demikian, *Naïve Bayes* tetap menunjukkan konsistensi performa yang relatif lebih stabil dibandingkan model lainnya. Hal ini berkaitan dengan karakteristik *dataset* yang bersifat *high-dimensional* dan *sparse*, sehingga pendekatan *probabilistik* lebih sesuai dibandingkan model berbasis pohon.

3.5.3 Analisis Kesalahan Prediksi (*Error Analysis*)

Hasil *confusion matrix* divisualisasikan secara agregat menggunakan *class-grouping / top-N classes* untuk meningkatkan interpretabilitas.

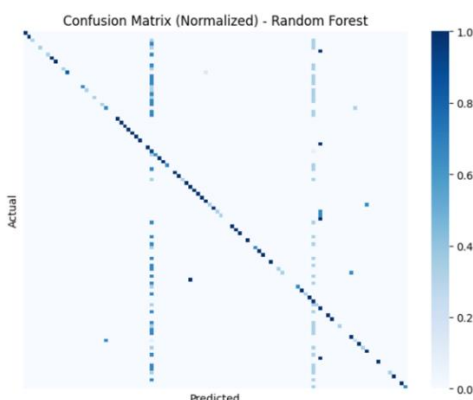


Gambar 3. *Confusion Matrix Model Naïve Bayes*

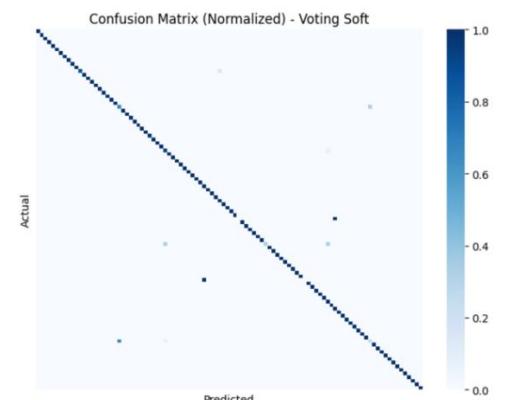


Gambar 4. *Confusion Matrix Model Decision Tree*

Visualisasi *confusion matrix* pada model *Naïve Bayes* (Gambar 3) menunjukkan dominasi warna gelap pada garis diagonal utama, yang mengindikasikan bahwa model mampu melakukan klasifikasi dengan akurasi tinggi dan tingkat kesalahan yang minimal. Namun, adanya ketidakseimbangan jumlah data pada beberapa kelas (*class imbalance*) berpotensi menyebabkan bias dan meningkatkan risiko *overfitting*, sehingga perlu diperhatikan dalam interpretasi hasil. Sebaliknya, visualisasi pada model *Decision Tree* (Gambar 4) menunjukkan penumpukan prediksi pada satu kolom vertikal, yang mengindikasikan bahwa model cenderung bias terhadap satu kelas dan kurang mampu membedakan antar kategori. Hal ini menunjukkan bahwa kemampuan generalisasi model *Decision Tree* relatif rendah, yang kemungkinan dipengaruhi oleh ketidakseimbangan data.

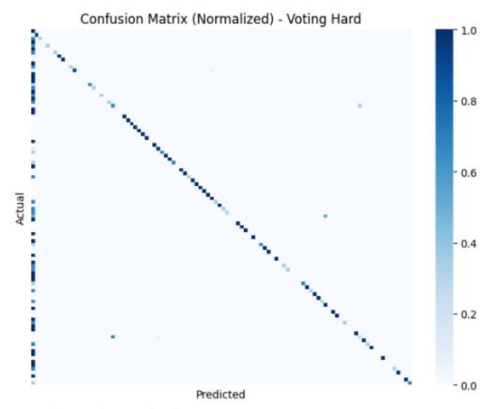


Gambar 5. *Confusion Matrix Random Forest*

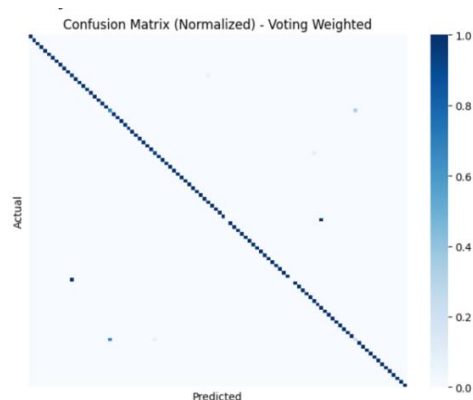


Gambar 6. *Confusion Matrix Model Soft Voting*

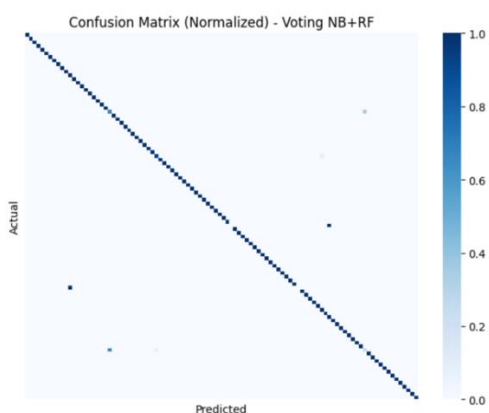
Visualisasi *confusion matrix* pada model *Random Forest* (Gambar 5) menampilkan pola diagonal yang cukup jelas, yang menunjukkan bahwa model memiliki kemampuan klasifikasi yang baik pada sebagian besar kelas, meskipun masih terdapat penyebaran kesalahan secara vertikal yang mengindikasikan adanya *false positive* pada beberapa kategori. Visualisasi *confusion matrix* pada model *Voting Soft* (Gambar 6) menunjukkan garis diagonal yang sangat solid dengan tingkat kesalahan yang minimal, yang mengindikasikan performa klasifikasi yang sangat tinggi dan stabil. Namun, kompleksitas model yang lebih tinggi berpotensi meningkatkan risiko *overfitting* apabila tidak divalidasi menggunakan data yang cukup beragam.



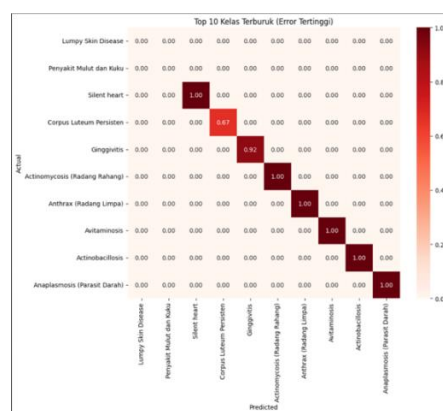
Gambar 7. Confusion Matrix Voting Hard



Gambar 8. Confusion Matrix Voting Weighted



Gambar 9. Confusion Matrix Voting NB+RF



Gambar 10. Evaluasi Top 10 Kelas Terburuk

Visualisasi *confusion matrix* pada model *Voting Hard* (Gambar 7) menunjukkan pola prediksi yang kurang merata, ditandai dengan adanya penumpukan pada satu kolom vertikal yang mengindikasikan bias terhadap kelas tertentu serta kemampuan klasifikasi yang kurang optimal. Sementara itu, model *Voting Weighted* (Gambar 8) dan *Voting Naive Bayes + Random Forest* (Gambar 9) menampilkan garis diagonal yang sangat dominan dan minim kesalahan, yang menunjukkan performa klasifikasi yang sangat tinggi dan stabil. Evaluasi Top 10 kelas dengan *error* tertinggi (Gambar 10) menunjukkan adanya ketidakseimbangan performa antar kelas, di mana beberapa penyakit seperti *Lumpy Skin Disease* dan *Penyakit Mulut dan Kuku* tidak berhasil dikenali dengan baik akibat keterbatasan representasi data dan kemiripan gejala antar kelas. Kondisi ini mengindikasikan bahwa distribusi data yang tidak seimbang masih menjadi faktor utama yang memengaruhi kesalahan klasifikasi dan diperlukan peningkatan kualitas serta jumlah data pada kelas minoritas.

Analisis *confusion matrix* menunjukkan bahwa sebagian besar prediksi berada pada kelas yang benar dengan tingkat *true positive* yang tinggi. Kesalahan klasifikasi terutama terjadi pada kelas dengan jumlah data kecil serta memiliki kemiripan gejala (*class overlap*), yang meningkatkan *false positive* dan *false negative* pada beberapa kategori. Secara kuantitatif, hasil evaluasi *confusion matrix* menunjukkan bahwa tingkat akurasi per kelas pada model terbaik berada pada kisaran $>85\%$ untuk sebagian besar kelas dominan, dengan rata-rata kesalahan klasifikasi sekitar $\pm 2,2\%$. Rata-rata tingkat kesalahan klasifikasi berada pada kisaran $\pm 2,2\%$, yang terutama berasal dari kelas minoritas. Kesalahan tersebut terutama terjadi pada kelas dengan jumlah sampel kecil serta memiliki kemiripan gejala antar penyakit (*class overlap*), yang menyebabkan model kesulitan membedakan pola secara tegas. Hal ini mengindikasikan bahwa pendekatan *ensemble* berbasis pembobotan maupun kombinasi model terbaik mampu meningkatkan akurasi dan mengurangi kesalahan prediksi secara signifikan dibandingkan model individual. Stabilitas model dievaluasi berdasarkan standar deviasi hasil *cross-validation*. *Naive Bayes* menunjukkan variansi paling rendah (± 0.001), diikuti *Random Forest* ($\pm 0.002-0.012$) dan metode *ensemble* ($\pm 0.007-0.018$). Hasil ini menunjukkan bahwa meskipun beberapa metode *ensemble* meningkatkan akurasi, stabilitasnya tidak selalu lebih baik dibanding *Naive Bayes* tunggal. Namun, kompleksitas model yang lebih tinggi pada pendekatan tersebut berpotensi meningkatkan risiko *overfitting* apabila tidak didukung oleh data yang cukup beragam dan proses validasi yang tepat.

3.5.4 Feature Importance

Hasil analisis *feature importance* menunjukkan bahwa pada model *Decision Tree*, fitur dengan kontribusi tertinggi adalah sempoyongan (0.0365), diikuti oleh gejala seperti *mencret*, *anemia*, lemah, dan pincang dengan nilai yang

relatif berdekatan. Selain itu, gejala lain seperti liur berdarah dan gangguan pada mata juga termasuk dalam fitur penting dengan kontribusi lebih rendah. Pada model *Random Forest*, fitur dengan nilai *importance* tertinggi adalah liur berlebihan (0.0327), diikuti oleh sempoyongan, kekurusan, *abses*, dan *anestrus*. Fitur lain seperti rahim bernanah, kelainan kulit, *anorexia*, dan *prolaps* anus juga berkontribusi dalam proses klasifikasi. Beberapa fitur seperti sempoyongan dan kekurusan muncul pada kedua model, yang menunjukkan perannya dalam membedakan kelas penyakit. Perbedaan nilai dan urutan kontribusi antar model menunjukkan adanya variasi dalam pemanfaatan fitur selama proses klasifikasi. Secara umum, fitur dengan kontribusi lebih tinggi cenderung berasal dari gejala yang lebih spesifik, dibandingkan gejala umum yang muncul pada banyak kelas.

3.5.5 Pengujian Diagnosa dan Justifikasi Stabilitas Model

Untuk menguji kinerja model pada kondisi nyata, dilakukan simulasi dengan memasukkan sejumlah gejala klinis, antara lain gelisah, kesulitan dalam proses persalinan, keterlambatan waktu persalinan, posisi atau presentasi janin yang tidak normal, keluarnya cairan atau darah dari vulva, serta tanda-tanda kesakitan seperti sapi merintih dan tampak tidak nyaman. *Input* tersebut kemudian diproses melalui tahap *preprocessing*, ekstraksi fitur, dan klasifikasi menggunakan model yang telah dibangun. Hasil diagnosis ditampilkan pada Tabel 5.

Tabel 5. Hasil Uji Coba Diagnosa Penyakit Sapi

	Hasil Diagnosa Penyakit Tingkat Keyakinan (%)		Penanganan
0	<i>Distokia</i>	98.2%	1.000 unit dengan Traksi dan 10.000 ml dengan PENSTREP-400 dan 8.000 ml dengan BIODIN, 3.000 ml dengan BIODIN dan 1.000 unit dengan 2.000 ml dengan VET-OXY LA dan 1.000 ml dengan OXYTOCIN
1	<i>Konjunktivitas</i>	0.32 %	0.300 ml dengan Dexatozoom Injeksi; @0.200 ml dengan XYLA; @0.200 ml dengan KETAMINE 10% INJ; @0.500 ml dengan PENSTREP-400, 0.500 ml dengan VET-OXY SB; @0.300 ml dengan Dexatozoom Injeksi
2	<i>Kecelakaan</i>	0.23 %	0.500 ml dengan Dexatozoom Injeksi; @0.300 ml dengan KETAMINE 10% INJ; @0.300 ml dengan XYLA; @0.500 ml dengan Hematodin, 0.500 ml dengan PENSTREP-400; @0.300 ml

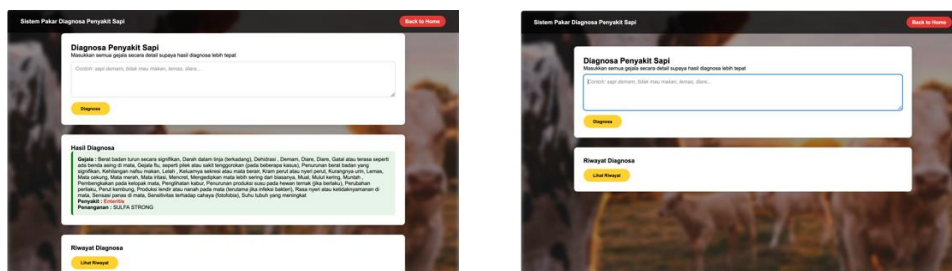
Berdasarkan hasil pengujian, sistem memprediksi *Distokia* sebagai penyakit utama dengan tingkat keyakinan sebesar 98,2%, diikuti oleh kemungkinan lain seperti *Konjunktivitas* (0,32%) dan *Kecelakaan* (0,23%). Perbedaan probabilitas yang signifikan menunjukkan bahwa model memiliki tingkat kepercayaan tinggi terhadap kelas utama karena kombinasi gejala yang bersifat spesifik. Penyajian beberapa kemungkinan penyakit beserta nilai probabilitas menunjukkan bahwa model tidak hanya memberikan satu keputusan, tetapi juga tingkat keyakinan untuk setiap kelas, sehingga lebih informatif dalam mendukung pengambilan keputusan. Model *Naïve Bayes* dipilih sebagai model akhir karena menunjukkan performa terbaik dengan akurasi tinggi, stabilitas yang baik, serta kompleksitas yang rendah dibandingkan metode lain. Meskipun pendekatan *ensemble* mampu meningkatkan performa pada beberapa skenario, peningkatannya relatif kecil ($\Delta \approx 0,01-0,02$) dan disertai kompleksitas yang lebih tinggi. Selain itu, *Naïve Bayes* menunjukkan konsistensi performa yang tinggi dengan variasi yang sangat kecil, sehingga lebih stabil pada berbagai kondisi pengujian. Karakteristik *dataset* yang bersifat *high-dimensional sparse* juga mendukung efektivitas model ini, karena asumsi *independensi* antar fitur memungkinkan pemanfaatan fitur gejala secara efisien. Dengan demikian, *Naïve Bayes* dinilai optimal sebagai model akhir karena mampu memberikan keseimbangan antara akurasi, stabilitas, dan efisiensi komputasi dalam sistem diagnosis awal penyakit sapi.

3.6 Deployment

Tahapan *Deployment* merupakan fase paling krusial untuk mengonversi hasil riset *algoritma* ke dalam bentuk sistem yang dapat dioperasikan secara nyata oleh pengguna akhir. Model *Naïve Bayes* yang telah dilatih kemudian diekstraksi dan diimplementasikan ke dalam kerangka aplikasi berbasis web menggunakan server Flask dengan basis data terintegrasi SQLite.

3.6.1 Integrasi Antarmuka dan Algoritma Pencocokan *Input*

Halaman utama sistem diagnosis dirancang interaktif dan mudah digunakan untuk mengidentifikasi penyakit sapi berdasarkan gejala. Sistem dilengkapi fitur *fuzzy matching* yang mampu mengenali variasi penulisan, kesalahan ketik, atau istilah berbeda dari pengguna, kemudian mencocokkannya ke dalam 411 fitur gejala baku yang digunakan model. Dengan fitur ini, proses diagnosis menjadi lebih fleksibel dan akurat. Tampilan halaman utama ditunjukkan pada Gambar 11.



Gambar 11. Halaman Diagnosa Penyakit Sapi

Setelah vektor fitur terbentuk, data diproses oleh model *Naïve Bayes* untuk menghitung probabilitas klasifikasi. Hasilnya ditampilkan melalui antarmuka yang memuat gejala terdeteksi, prediksi penyakit, penanganan, serta riwayat diagnosis pada *database*, sehingga sistem menjadi lebih informatif dan mudah digunakan.

3.6.2 Pengujian Penerimaan Sistem (*System Usability Scale*)

Sebagai bentuk validasi akhir dan pengukuran penerimaan teknologi, penelitian ini melakukan pengujian kegunaan (*usability testing*) menggunakan instrumen *System Usability Scale* (SUS). Survei evaluasi ini melibatkan 37 responden dengan latar belakang yang beragam, meliputi elemen mahasiswa, pengelola sistem (admin), dan peternak sapi. Kuesioner SUS dirancang menggunakan 10 item pernyataan tertutup yang secara objektif mengukur dimensi kemudahan navigasi, kompleksitas fitur, hingga tingkat kepercayaan pengguna.

Tabel 6. Hasil pengujian SUS

Skor Hasil Hitung										Jumlah	Nilai (Jumlah x 2.5)
Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10		
3	3	2	3	3	3	3	3	3	3	29	73
3	3	3	3	3	3	3	3	3	3	30	75
3	1	4	1	4	0	2	0	2	0	17	43
4	4	4	4	4	4	4	4	4	4	40	100
4	0	0	2	2	0	3	0	2	1	14	35
3	3	3	1	4	3	3	3	3	1	27	68
4	0	4	0	4	0	4	0	4	0	20	50
3	4	2	3	3	3	3	3	4	2	30	75
4	3	3	4	2	3	3	3	4	2	31	78
4	3	2	3	3	4	4	3	3	2	31	78
2	4	4	3	4	3	3	4	4	3	34	85
3	3	4	4	3	3	3	4	4	4	35	88
4	3	3	4	4	3	4	4	3	3	35	88
3	3	3	3	3	3	4	4	3	2	31	78
4	3	3	3	2	4	4	3	3	3	32	80
4	3	3	4	2	3	3	3	4	3	32	80
3	3	3	4	3	3	4	3	3	3	32	80
3	3	3	3	3	4	3	3	4	3	32	80
2	3	4	3	2	3	4	3	4	3	31	78
3	3	3	4	2	3	4	3	3	3	31	78
3	3	3	3	4	3	3	4	4	2	32	80
2	3	3	3	4	4	3	3	4	2	31	78
4	3	3	3	2	4	3	3	4	2	31	78
3	3	4	4	3	3	4	4	4	3	35	88
3	3	4	4	3	3	4	4	4	3	35	88
3	3	3	3	2	4	3	3	4	3	31	78
3	3	4	4	4	3	3	3	4	3	34	85
4	3	3	4	4	3	3	4	4	3	35	88
3	4	4	3	3	4	4	3	4	2	34	85
3	3	4	4	4	3	3	3	4	3	34	85
Skor Rata-rata (Hasil Akhir)											77

Berdasarkan Tabel 6, perhitungan standar SUS terhadap seluruh respons kuantitatif, sistem memperoleh skor rata-rata sebesar 77. Merujuk pada panduan pemeringkatan SUS, nilai yang diperoleh termasuk dalam kategori baik (*Good*) serta dapat diterima (*Acceptable*). Temuan ini memberikan konfirmasi akhir bahwa sistem pakar diagnosis penyakit sapi yang dikembangkan tidak hanya kokoh pada aspek akurasi algoritmik, tetapi juga telah

memenuhi standar fungsionalitas dan aspek kegunaan (*usability*) yang esensial sebagai instrumen pendamping medis di lapangan.

4. KESIMPULAN

Sistem pakar diagnosis penyakit sapi berbasis kecerdasan buatan berhasil dikembangkan dengan mengimplementasikan algoritma *Naïve Bayes*, *Decision Tree*, dan *Random Forest* yang dikombinasikan melalui metode *Voting Ensemble*. Hasil penelitian menunjukkan bahwa *Naïve Bayes* memberikan performa terbaik dibandingkan model lainnya, sementara metode *ensemble* berperan dalam meningkatkan stabilitas dan konsistensi hasil prediksi meskipun peningkatannya relatif tidak signifikan. Hal ini mengindikasikan bahwa karakteristik *dataset* memiliki pola fitur yang cukup jelas dan dapat ditangani secara efektif oleh model probabilistik sederhana, sehingga perbedaan performa antar model tidak terlalu jauh. Selain itu, sistem yang dikembangkan memperoleh skor *System Usability Scale* (SUS) sebesar 77 yang termasuk dalam kategori *Good* dan *Acceptable*, yang menunjukkan bahwa sistem dapat diterima dengan baik oleh pengguna serta mampu membantu proses identifikasi awal penyakit sapi secara lebih mudah dan efisien. Penelitian ini masih memiliki keterbatasan terutama pada jumlah *dataset* yang terbatas serta adanya ketidakseimbangan distribusi kelas yang memengaruhi performa pada beberapa jenis penyakit. Oleh karena itu, penelitian selanjutnya disarankan untuk meningkatkan kualitas dan variasi *dataset* melalui penambahan data yang lebih representatif dari lapangan agar model dapat lebih *general* dalam melakukan prediksi. Selain itu, pengembangan sistem dapat diarahkan pada integrasi data pendukung yang lebih beragam seperti rekam medis digital atau data observasi klinis untuk meningkatkan ketepatan diagnosis. Pengembangan antarmuka ke dalam bentuk aplikasi *mobile* juga dapat menjadi alternatif untuk memperluas aksesibilitas sistem bagi pengguna di lapangan.

REFERENCES

- [1] H. A. Febriani, D. P. Wijaya, A. Pramuntadi, and W. D. Prastowo, "Sistem Pakar Diagnosa Penyakit Lambung Menggunakan Metode Certainty Factor Berbasis Web," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 4, no. 4, pp. 1290–1300, 2024, doi: 10.57152/malcom.v4i4.1402.
- [2] Malikul Fajar Lubis, Wahyu Hadinata, Gideon Syahputra, VarelSimanjuntak, and Kiagus Muhammad Zain, "Analisis Perkembangan Populasi Dan Produktivitas Ternak Sapi Di Indonesia," *Botani : Publikasi Ilmu Tanaman dan Agribisnis*, vol. 2, no. 1, pp. 172–181, 2024, doi: 10.62951/botani.v2i1.171.
- [3] M. Alfian, W. Aprianti, H. Rhomadhona, and J. Permadi, "Sistem Pakar Penyakit Sapi Menggunakan Rule Based Reasoning dengan Forward Chaining," *Teknomatika: Jurnal Informatika dan Komputer*, vol. 17, no. 2, pp. 22–29, 2024, doi: 10.30989/teknomatika.v17i2.1541.
- [4] G. Setiawan and G. Setia Budi, "Implementasi Metode Forward Chaining pada Sistem Pakar untuk Penyakit DBD," *Dike: Jurnal Ilmu Multidisiplin*, vol. 1, no. 2, pp. 44–48, Aug. 2023. [Online]. Available: <https://garuda.kemdiktisaintek.go.id/documents/detail/3647798>
- [5] H. Imaduddin, B. A. Hermansyah, and F. A. Salsabilla B, "Comparison of Support Vector Machine and Decision Tree Methods in the Classification of Breast Cancer," *Cyberspace: Jurnal Pendidikan Teknologi Informasi*, vol. 5, no. 1, p. 22, 2021, doi: 10.22373/cj.v5i1.8805.
- [6] Dirgantara Yudhanata and E. Sudarmilah, "Analisis Sentimen Terhadap Isu Islamofobia pada Platform Twitter Menggunakan Metode Klasifikasi Naïve Bayes," Universitas Muhammadiyah Surakarta, 2023. Available: <https://eprints.ums.ac.id/123003/>
- [7] H. Aisyah and Y. S. Nugroho, Exploring Algorithmic Approaches for Academic Interest Classification in Application UMS Study Case, in Proceeding of the 4th International Conference on Education for All ICEDUALL 2024, e ISSN 3025 9517, 2024. Available <https://proceedings.alptkptm.org/index.php/iceduall/article/download/48/47/208>
- [8] Y. I. Fajarendra and S. Uyun, "Penerapan Metode Ensemble Learning dalam Klasifikasi Risiko Abrasi Menggunakan Citra Satelit Google Earth Engine," *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK)*, vol. 13, no. 1, pp. 131–140, Feb. 2026. [Online]. Available: <https://garuda.kemdiktisaintek.go.id/documents/detail/6072242>
- [9] N. T. Rahman, "Analisa Algoritma Decision Tree Dan Naïve Bayes Pada Pasien Penyakit Liver," *Jurnal Fasilkom*, vol. 10, no. 2, pp. 144–151, 2020, doi: 10.37859/jf.v10i2.2087.
- [10] C. V. Angkoso, N. M. A. Thrisna, B. D. Satoto, and A. Kusumaningsih, "Optimasi Klasifikasi Sentimen Menggunakan Random Forest dengan Preprocessing K-Means Clustering dan SMOTE," *JEPIN (Jurnal Edukasi dan Penelitian Informatika)*, vol. 10, no. 3, pp. 389–400, 2024. doi: 10.26418/jp.v10i3.84514. Available: <https://jurnal.untan.ac.id/index.php/jepin/article/view/84514>
- [11] K. Dineva and T. Atanasova, "Health Status Classification for Cows Using Machine Learning and Data Management on AWS Cloud," *Animals*, vol. 13, no. 20, Oct. 2023, doi: 10.3390/ani13203254.
- [12] E. G. Mounq, C. C. Wooi, M. M. Sufian, C. K. On, and J. A. Dargham, "Ensemble-based face expression recognition approach for image sentiment analysis," *International Journal of Electrical and Computer Engineering*, vol. 12, no. 3, pp. 2588–2600, Jun. 2022, doi: 10.11591/ijece.v12i3.pp2588-2600.
- [13] D. Sasya, F. Khanun, and R. Ummah, "Penerapan Sistem Ensemble Learning untuk Analisis Sentiment Pembaruan One UI 7 pada Perangkat Samsung Menggunakan Algoritma Naïve Bayes, SVM, LSTM, Bi-LSTM, dan IndoBERT," Universitas Muhammadiyah Surakarta, 2024. Available: <https://eprints.ums.ac.id/143388/1/NASKAH%20PUBLIKASI.pdf>
- [14] "M. R. Zamroni, Q. C. K. N., and A. Wahyudi, "Sistem Pakar Diagnosa Penyakit Sapi sebagai Upaya Pencegahan Penyebaran Wabah PMK di Lamongan," *Jurnal Ilmiah Informatika*, vol. 10, no. 2, 2022, doi: 10.33884/jif.v10i02.6373.
- [15] M. Fitriani, G. F. Nama, and M. Mardiana, "Implementasi Association Rule Dengan Algoritma Apriori Pada Data Peminjaman Buku UPT Perpustakaan Universitas Lampung Menggunakan Metodologi CRISP-DM," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 10, no. 1, pp. 41–49, 2022, doi: 10.23960/jitet.v10i1.2263.



- [16] F. Rahmadayanti, S. Muntari, and R. Putriani, “Pengelompokan Tingkat Resiko Penyakit Jantung Berdasarkan Usia Menggunakan Algoritma K-Means,” *JUTIM (Jurnal Teknik Informatika Musirawas)*, vol. 8, no. 2, pp. 124–130, 2024, doi: 10.32767/jutim.v8i2.2108.
- [17] Y. A. Singgalen, “Analisis Sentimen dan Sistem Pendukung Keputusan Menginap di Hotel Menggunakan Metode CRISP-DM dan SAW,” *Journal of Information System Research (JOSH)*, vol. 4, no. 4, pp. 1343–1353, 2023, doi: 10.47065/josh.v4i4.3917.
- [18] A. A. A. Fernandes, M. Koehler, N. Konstantinou, P. Pankin, N. W. Paton, and R. Sakellariou, “Data Preparation: A Technological Perspective and Review,” *SN Comput. Sci.*, vol. 4, no. 4, pp. 1–20, 2023, doi: 10.1007/s42979-023-01828-8.
- [19] V. R. Maulani, M. A. Barata, and P. E. Yuwita, “A Improving House Price Clustering Results with K-means through the Implementation of One-hot Encoding Pre-processing Technique,” *Journal of Applied Informatics and Computing*, vol. 9, no. 3, pp. 741–748, 2025, doi: 10.30871/jaic.v9i3.9481.
- [20] Sugiyono and S. S. Hartinah, “Pemodelan Data Mining Transaksi Penjualan Menggunakan Algoritma Apriori (Studi Kasus: Kedai Ngodeng & Smoothies),” *Jurnal JMIMIK*, vol. 5, no. 3, 2024. doi: 10.35870/jimik.v5i3.992.
- [21] N. F. Imam, “Komparasi Naïve Bayes dan Random Forest untuk Prediksi Tingkat Stres Berdasarkan Pola Penggunaan Medsos,” 2026. [Online]. Available: <https://jurnal.bsi.ac.id/index.php/ijec/>
- [22] M. Elkalawy, A. Al-Sakkaf, E. Mohammed Abdelkader, and G. Alfalah, “CRISP-DM-Based Data-Driven Approach for Building Energy Prediction Utilizing Indoor and Environmental Factors,” *Sustainability (Switzerland)*, vol. 16, no. 17, 2024, doi: 10.3390/su16177249.
- [23] M. A. Kosim, S. R. Aji, and M. Darwis, “Pengujian Usability Aplikasi PeduliLindungi dengan Metode System Usability Scale (SUS),” *Jurnal Sistem Informasi dan Sains Teknologi*, vol. 4, no. 2, 2022. doi: 10.31326/sistek.v4i2.1326.