

Peningkatan Kinerja Klasifikasi *Scabies Sapi* Menggunakan *Edited Nearest Neighbours* (ENN) pada Model *Random Forest* dan *XGBoost*

Improving the Classification Performance of Bovine Scabies Using Edited Nearest Neighbours (ENN) on Random Forest and XGBoost Models

M. Khaerul Ihsan*, Muhammad Maulana, Tanwir, Abi Mas'ud,

Naufal Hanif, Dading Oktaviadi Resmiranta

Universitas Bumigora, Mataram, Indonesia

Informasi Artikel:

Diterima: 23 Desember 2025, Direvisi: 30 Desember 2025, Disetujui: 31 Desember 2025

Abstrak-

Latar Belakang: Penyakit scabies pada sapi menyebabkan kerugian ekonomi yang signifikan bagi peternak akibat penurunan kondisi fisik dan produktivitas ternak.

Tujuan: Penelitian ini bertujuan untuk mengevaluasi efektivitas metode *Edited Nearest Neighbours* (ENN) dalam meningkatkan performa klasifikasi penyakit scabies pada sapi.

Metode: Penelitian ini menggunakan pendekatan *machine learning* dengan metode *Random Forest* dan *XGBoost*. Data gejala klinis sebanyak 600 sampel ditransformasi menjadi data numerik dan dibersihkan dari *noise* menggunakan teknik ENN.

Hasil: Hasil penelitian menunjukkan bahwa penerapan ENN secara signifikan meningkatkan akurasi model *Random Forest* dan *XGBoost* dari kisaran 0.60 menjadi 0.91. Selain itu, kedua model mencapai nilai *recall* sempurna sebesar 1.00, yang mengindikasikan kemampuan maksimal dalam mendekripsi kasus positif.

Kesimpulan: Kesimpulan penelitian ini menunjukkan bahwa reduksi *noise* dengan ENN mampu menghasilkan sistem diagnosis yang lebih akurat dan andal. Penggunaan metode ini sangat direkomendasikan untuk mengoptimalkan performa algoritma klasifikasi pada data klinis hewan yang memiliki tingkat inkonsistensi tinggi.

Kata Kunci: *Edited Nearest Neighbours; Klasifikasi; Machine Learning; Random Forest; Scabies Sapi; XGBoost.*

Abstract-

Background: *Scabies disease in cattle causes significant economic losses for farmers due to declines in the animals' physical condition and productivity.*

Objective: *This study aims to evaluate the effectiveness of the Edited Nearest Neighbours (ENN) method in improving classification performance for scabies in cattle.*

Methods: *This research employs machine learning methods, including Random Forest and XGBoost. A dataset of 600 clinical symptom samples was converted to numerical data and cleaned of noise using the ENN technique.*

Result: *Applying ENN significantly improved the accuracy of both the Random Forest and XGBoost models, increasing it from around 0.60 to 0.91. In addition, both models achieved a perfect recall of 1.00, indicating maximum capability to detect positive cases.*

Conclusion: *This study concludes that noise reduction using ENN can produce a more accurate and reliable diagnostic system. This method is highly recommended to optimize the performance of classification algorithms on animal clinical data with high levels of inconsistency.*

Keywords: *Classification; Cattle Scabies; Edited Nearest Neighbours; Machine Learning; Random Forest; XGBoost.*

How to Cite: M. K. Ihsan, M. Maulana, T. Tanwir, A. Mas'ud, N. Hanif, & D. O. Resmiranta, "Peningkatan Kinerja Klasifikasi *Scabies Sapi* Menggunakan *Edited Nearest Neighbours* (ENN) pada Model *Random Forest* dan *XGBoost*," *Jurnal Bumigora Information Technology (BITe)*, vol. 7, no. 2, pp 141–150, Des 2025. DOI: 10.30812/bite.v7i2.6055.

This is an open access article under the CC BY-SA license (<https://creativecommons.org/licenses/by-sa/4.0/>)

Penulis Korespondensi:

M. Khaerul Ihsan,
Program Studi Rekayasa Perangkat Lunak, Universitas Bumigora,
Email: ihsan@universitasbumigora.ac.id

1. PENDAHULUAN

Berdasarkan laporan *World Health Organization* (WHO), *scabies* merupakan penyakit yang memiliki persebaran global dengan estimasi sekitar 300 juta kasus terjadi setiap tahunnya di seluruh dunia, terutama pada wilayah negara-negara berkembang [1]. *Scabies* pada sapi merupakan salah satu penyakit yang disebabkan oleh infestasi parasit berupa serangga kecil pada kulit yang berasal dari tungau [2]. Penyakit ini menimbulkan rasa gatal yang hebat, munculnya luka atau kerusakan pada kulit, serta penurunan kondisi fisik secara menyeluruh. Kondisi kesehatan sapi merupakan faktor krusial yang berperan penting dalam keberhasilan usaha peternakan [3]. Penyakit pada sapi dapat menimbulkan dampak merugikan bagi peternak, antara lain kerugian akibat kematian ternak, meningkatnya biaya perawatan, menurunnya tingkat produksi, serta kurangnya efisiensi dalam pemanfaatan pakan [4]. Oleh karena itu, penting dilakukan diagnosis dini terhadap penyakit *scabies* pada sapi agar dapat dilakukan penanganan secara cepat dan tepat. Upaya ini diharapkan dapat membantu meningkatkan kesehatan ternak, menjaga produktivitas peternakan, serta mengurangi kerugian ekonomi yang ditimbulkan akibat penyakit tersebut.

Kemajuan teknologi *machine learning* memberikan potensi untuk mengotomatiskan proses klasifikasi penyakit berdasarkan pola gejala klinis, sehingga proses diagnosis dapat dilakukan dengan lebih efisien dan memiliki tingkat konsistensi yang lebih tinggi [5], [6]. Namun, data klinis yang diperoleh sering mengandung *noise* dan ketidakseimbangan kelas, yang dapat menurunkan akurasi model klasifikasi. Pembersihan data dari *noise* sebelum proses pelatihan model merupakan tahap krusial untuk meningkatkan tingkat akurasi dan kemampuan generalisasi model [7]. *Edited Nearest Neighbours* (ENN) merupakan metode pembersihan data berbasis algoritma *nearest neighbor* yang berfungsi mengeliminasi sampel yang terindikasi sebagai *noise* atau memiliki label yang keliru, sehingga data yang digunakan untuk pelatihan model menjadi lebih representatif dan berkualitas [8]. Penerapan metode ENN dengan algoritma klasifikasi seperti *Random Forest* (RF) dan *Extreme Gradient Boosting* (XGBoost) diharapkan mampu meningkatkan kinerja model dalam mengklasifikasikan penyakit scabies pada sapi berdasarkan gejala klinis.

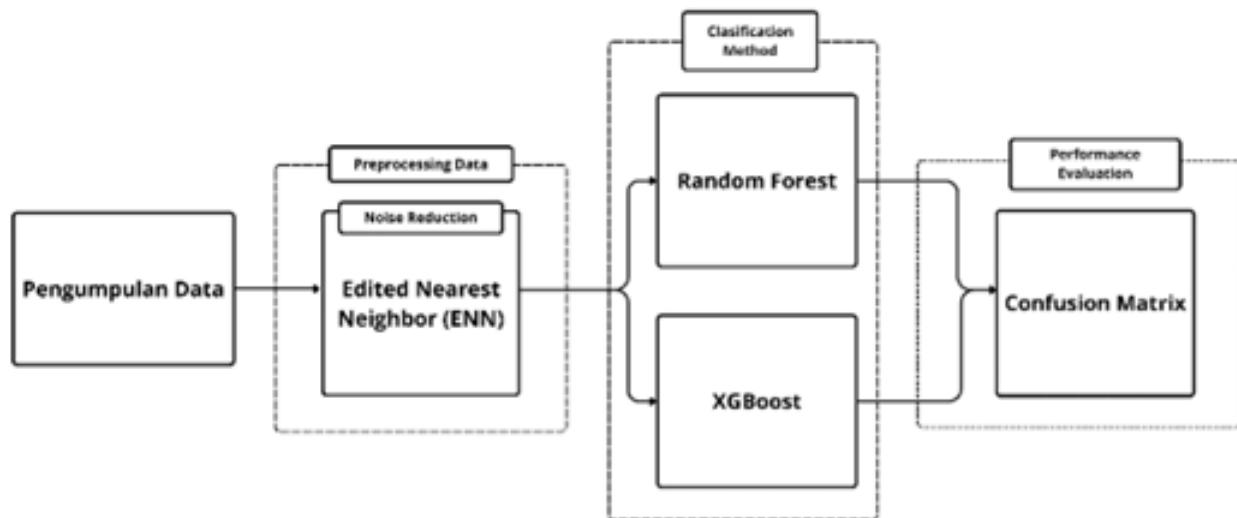
Beberapa penelitian terdahulu telah dilakukan terkait penerapan metode *machine learning* dalam mendiagnosa penyakit pada hewan. Penelitian oleh [9] menggunakan metode *Uniform Local Binary Pattern* (ULBP) sebagai teknik ekstraksi ciri dan *Random Forest* sebagai algoritma klasifikasi. Hasil penelitian menunjukkan bahwa penerapan metode ULBP mampu mengoptimalkan proses ekstraksi ciri sekaligus mempercepat waktu pemrosesan data, sehingga sistem yang dibangun menghasilkan performa sebesar 52%. Penelitian lain oleh [10] menerapkan beberapa metode *machine learning* seperti *Multiple Linear Regression* (MLR), *Linear Discriminant Analysis* (LDA), *Random Forest*, dan *Artificial Neural Network* (ANN) untuk memprediksi tingkat resistensi dan kerentanan domba terhadap infeksi *gastrointestinal nematodes*. Hasil penelitian tersebut menunjukkan bahwa model *machine learning* mampu memberikan prediksi yang cukup akurat dengan memanfaatkan data fenotip sederhana, sehingga berpotensi membantu pengambilan keputusan dalam manajemen peternakan. Sementara itu, penelitian oleh [11] menerapkan Jaringan Syaraf Tiruan Konvolusional (*Convolutional Neural Network*, CNN) untuk klasifikasi penyakit sapi seperti Demam Efemeral Sapi, Mastitis, dan Kudis. Dengan dataset sebanyak 864 data latih dan 216 data validasi, model CNN mencapai akurasi pelatihan sebesar 1.0000 dan akurasi pengujian sebesar 0.9306, yang menunjukkan bahwa metode ini efektif dalam mendeteksi penyakit sapi secara akurat.

Meskipun berbagai penelitian telah menunjukkan efektivitas metode *machine learning* dalam mendiagnosa penyakit hewan, penelitian yang secara khusus menggabungkan metode reduksi *noise* *Edited Nearest Neighbours*

(ENN) dengan algoritma *Random Forest* dan *XGBoost* dalam klasifikasi penyakit scabies pada sapi berdasarkan gejala klinis masih jarang ditemukan. Perbedaan penelitian ini dengan penelitian sebelumnya terletak pada penerapan ENN sebagai tahap praproses data untuk meningkatkan kualitas dan keseimbangan dataset, serta pemanfaatan keunggulan *Random Forest* dalam mengurangi *overfitting* dan *XGBoost* dalam meningkatkan akurasi model. Oleh karena itu, penelitian ini bertujuan untuk mengevaluasi efektivitas metode ENN dalam meningkatkan performa model klasifikasi penyakit scabies pada sapi. Dengan demikian, penelitian ini berkontribusi pada peningkatan kualitas data pelatihan serta optimalisasi performa algoritma klasifikasi sehingga diharapkan dapat menghasilkan sistem diagnosis yang lebih akurat dan andal bagi peternak maupun praktisi kesehatan hewan.

2. METODE PENELITIAN

Pada penelitian ini dilakukan beberapa tahapan sistematis untuk menghasilkan model klasifikasi penyakit scabies pada sapi yang lebih akurat dengan memanfaatkan teknik reduksi *noise* dan algoritma *machine learning*. Secara garis besar, tahapan-tahapan dalam penelitian ini dimulai dari pengumpulan data, *preprocessing* data, penerapan metode klasifikasi, hingga evaluasi model. Tahapan penelitian ini secara keseluruhan dapat dilihat pada Gambar 1.



Gambar 1. Tahapan Penelitian

2.1. Pengumpulan Data

Pengumpulan data pada penelitian ini menggunakan dataset scabies pada sapi berdasarkan gejala klinis yang diperoleh dari platform *Kaggle.com*. Dataset tersebut disediakan dalam format *.csv* dan berjumlah 600 sampel. Setiap sampel memiliki enam atribut yang merepresentasikan kondisi klinis sapi, yaitu usia, jenis kelamin, gatal, kerontokan bulu, kerak pada kulit, serta hasil tes laboratorium yang digunakan sebagai label kelas. Atribut-atribut tersebut dipilih karena secara klinis dianggap relevan dalam mengidentifikasi keberadaan scabies pada sapi. Dataset ini terdiri dari dua kelas, yakni kategori positif scabies dan negatif scabies. Dataset dapat dilihat pada Tabel 1. Data yang telah diperoleh selanjutnya menjadi input utama untuk tahap *preprocessing* dan pengembangan model *machine learning* pada penelitian ini.

Tabel 1. Data *Scabies* Sapi

No	usia	jenis_kelamin	gatal	kerontokan_bulu	kerak_pada_kulit	hasil_tes_laboratorium
1	1	Female	YA	YA	YA	POSITIF
2	3	Male	YA	YA	YA	POSITIF
3	3	Female	YA	TIDAK	YA	POSITIF

No	usia	jenis_kelamin	gatal	kerontokan_bulu	kerak_pada_kulit	hasil_tes_laboratorium
4	2	Female	TIDAK	YA	YA	NEGATIF
5	2	Female	TIDAK	YA	YA	POSITIF
6	3	Female	YA	YA	YA	POSITIF
7	2	Female	YA	TIDAK	YA	POSITIF
8	1	Female	YA	YA	TIDAK	NEGATIF
9	2	Female	YA	YA	TIDAK	NEGATIF
...
591	1	Male	TIDAK	TIDAK	TIDAK	POSITIF
592	2	Female	TIDAK	TIDAK	TIDAK	NEGATIF
593	3	Female	YA	TIDAK	TIDAK	NEGATIF
594	1	Male	YA	YA	TIDAK	POSITIF
595	1	Female	YA	TIDAK	TIDAK	POSITIF
596	3	Female	TIDAK	YA	YA	POSITIF
597	2	Male	TIDAK	TIDAK	YA	POSITIF
598	1	Female	YA	TIDAK	YA	NEGATIF
599	3	Male	TIDAK	TIDAK	YA	POSITIF
600	2	Female	YA	YA	YA	NEGATIF

2.2. Preprocessing Data

Sebelum menerapkan teknik *Edited Nearest Neighbor* (ENN), tahap pertama yang dilakukan dalam penelitian ini adalah transformasi data untuk mengubah atribut ke dalam format yang dapat dipahami oleh program atau algoritma. Nilai dari masing-masing atribut yang berupa teks, seperti jenis kelamin, gatal, kerontokan bulu, kerak pada kulit, serta atribut target berupa hasil tes laboratorium, diubah menjadi data numerik. Selanjutnya, *Edited Nearest Neighbor* (ENN) digunakan sebagai salah satu teknik reduksi *noise* atau metode pembersihan data pada tahap *preprocessing* untuk meningkatkan kualitas dataset sebelum diproses oleh algoritma klasifikasi.

Data yang tidak konsisten kemudian dihapus dari dataset sehingga menghasilkan himpunan data yang lebih bersih dan terstruktur. Dengan menghilangkan sampel yang mengganggu pola sebenarnya, ENN membantu mengurangi potensi *overfitting* serta meningkatkan kemampuan model dalam mengenali pola yang lebih relevan. Proses ini sangat bermanfaat, terutama pada dataset yang memiliki distribusi kelas yang sensitif atau rentan terhadap kesalahan pelabelan data. Penerapan ENN pada tahap awal pengolahan data dapat memberikan dampak signifikan terhadap peningkatan akurasi, stabilitas, dan kemampuan generalisasi model klasifikasi yang akan digunakan pada tahap selanjutnya.

2.3. Penerapan Metode

Random Forest merupakan algoritma *machine learning* yang bekerja dengan menggabungkan banyak pohon keputusan pada tugas regresi maupun klasifikasi. Algoritma ini membangun pohon melalui proses pemisahan biner secara rekursif hingga mencapai node akhir [12]. *Random Forest* merupakan algoritma *ensemble* yang dibangun dengan menggabungkan sejumlah pohon keputusan, di mana setiap pohon dilatih menggunakan subset data pelatihan yang dipilih secara acak melalui teknik *bootstrap*. Pada setiap proses pemisahan node, hanya sebagian fitur acak yang digunakan, sehingga meningkatkan keragaman antar pohon dan membantu menurunkan risiko *overfitting*. Setiap pohon menghasilkan prediksi secara mandiri, kemudian hasil akhir ditentukan melalui mekanisme *voting* mayoritas dalam tugas klasifikasi. Dibandingkan dengan satu pohon keputusan, metode ini lebih stabil dan akurat karena mampu mengurangi varians, meningkatkan kemampuan generalisasi, serta lebih tangguh terhadap data yang kompleks atau mengandung *noise*. Selain itu, *Random Forest* efektif dalam menangani data berdimensi tinggi dan menyediakan estimasi tingkat kepentingan fitur, yang sangat berguna untuk proses seleksi fitur dan interpretasi model [13].

XGBoost merupakan pengembangan dari metode *Gradient Boosting* yang dilengkapi dengan model

terstruktur untuk mengendalikan risiko *overfitting*. Algoritma ini menggunakan sebuah fungsi objektif yang berperan dalam menilai kualitas model yang dihasilkan selama proses pelatihan. Fungsi objektif tersebut terdiri atas dua komponen utama, yaitu nilai *loss* pada data pelatihan dan komponen regularisasi yang bertujuan untuk menjaga kompleksitas model [14], [15].

2.4. Evaluasi Model

Tahap evaluasi dilakukan untuk menilai performa model dengan memanfaatkan *confusion matrix* sebagai dasar perhitungan. Melalui matriks ini, berbagai metrik evaluasi, seperti *accuracy*, *precision*, *recall*, dan *f1-score*, dapat diperoleh untuk menggambarkan kualitas prediksi model secara menyeluruh. *Accuracy* digunakan untuk mengetahui tingkat ketepatan total hasil prediksi model dan dapat dihitung dengan persamaan (1). *Recall* merupakan metrik evaluasi yang digunakan untuk mengukur kemampuan model dalam menemukan atau mendeteksi seluruh sampel yang benar-benar termasuk dalam suatu kelas tertentu dan dapat dihitung dengan persamaan (2). *Precision* digunakan untuk mengukur tingkat ketepatan prediksi positif model yang dihitung dengan persamaan (3), sedangkan *f1-score* digunakan untuk mengukur keseimbangan antara *precision* dan *recall* yang dapat dihitung dengan persamaan (4), seperti yang ditunjukkan pada Tabel 2.

Tabel 2. *Confusion Matrix*

Aktual	Prediksi	
	Positif	Negatif
Positif	TP	FN
Negatif	FP	TN

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN} \quad (1)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (2)$$

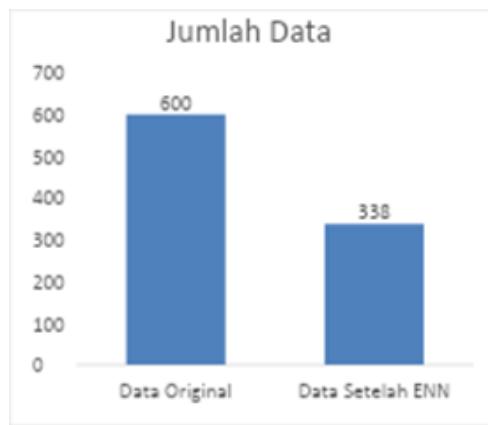
$$\text{Precision} = \frac{TP}{TP + FP} \quad (3)$$

$$F1 - Score = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

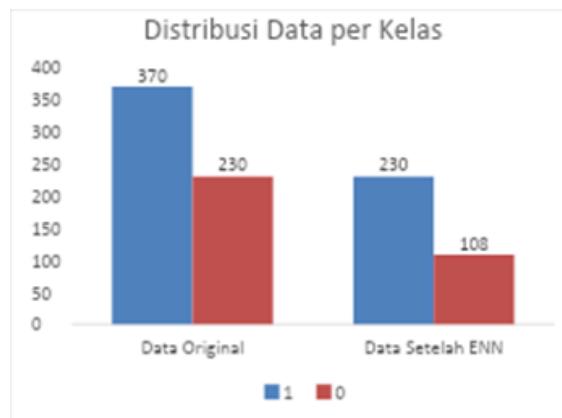
3. HASIL DAN PEMBAHASAN

Bagian ini menguraikan hasil serta temuan yang diperoleh dari penelitian ini. Data yang digunakan dalam penelitian ini berkaitan dengan penyakit scabies pada sapi berdasarkan data klinis yang diperoleh dari platform *Kaggle.com*, dengan total 600 data yang terdiri atas lima atribut dan satu atribut sebagai kelas. Tahapan berikutnya adalah *preprocessing* data yang bertujuan untuk meningkatkan kualitas dan konsistensi dataset. Proses ini diawali dengan transformasi data, yaitu mengubah masing-masing atribut yang berupa data teks menjadi data numerik. Untuk label hasil tes laboratorium, digunakan nilai satu (1) untuk kategori positif dan nilai nol (0) untuk kategori negatif. Selanjutnya, metode *Edited Nearest Neighbor* (ENN) diterapkan untuk mengidentifikasi dan menghapus data yang dianggap tidak konsisten atau mengandung *noise*.

Berdasarkan Gambar 2, dari total 600 data awal, proses ENN menghasilkan 338 data bersih, sehingga sekitar 43% dari data awal dinyatakan layak untuk tahap analisis selanjutnya. Gambar 3 menunjukkan bahwa ENN juga memengaruhi distribusi kelas, di mana kelas 1 berkurang dari 370 menjadi 230 data, sedangkan kelas 0 berkurang dari 230 menjadi 108 data. Perubahan ini menunjukkan bahwa data yang tidak konsisten lebih banyak ditemukan pada kelas 0.

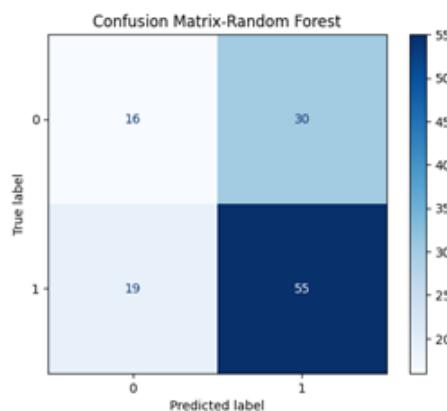


Gambar 2. Jumlah Data Sebelum dan Sesudah ENN



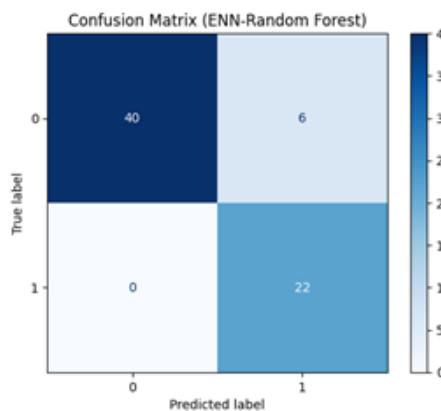
Gambar 3. Distribusi Data Perkelas Sebelum dan Sesudah ENN

Setelah tahap *preprocessing* data, selanjutnya dilakukan implementasi metode klasifikasi *Random Forest* dan *XGBoost*, baik tanpa penerapan ENN maupun dengan penerapan ENN. Tahap berikutnya adalah membandingkan kinerja kedua metode berdasarkan metrik *accuracy*, *precision*, *recall*, dan *f1-score*. Sebelum pengujian model dilakukan, data terlebih dahulu dibagi menjadi dua bagian, yaitu data latih dan data uji, dengan proporsi 80% untuk data latih dan 20% untuk data uji. Hasil klasifikasi dari kedua metode, baik tanpa ENN maupun dengan ENN, dapat dilihat pada Gambar 4–7.



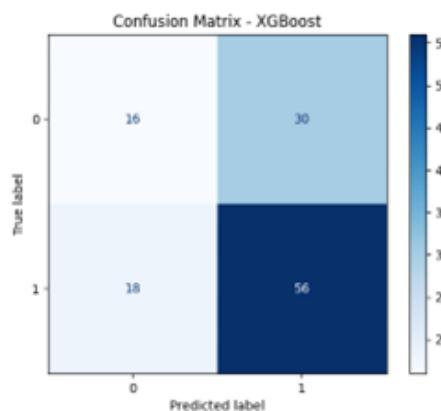
Gambar 4. Hasil Model Random Forest Tanpa ENN

Berdasarkan *confusion matrix* yang ditunjukkan pada Gambar 4, diketahui bahwa sebanyak 16 data negatif berhasil diprediksi dengan benar sebagai negatif. Namun, terdapat 30 data negatif yang keliru diklasifikasikan sebagai positif. Pada kelas positif, sebanyak 55 data berhasil diprediksi dengan tepat sebagai positif, sedangkan 19 data lainnya salah diprediksi sebagai negatif. Secara keseluruhan, *confusion matrix* ini menunjukkan bahwa model *Random Forest* memiliki kemampuan yang lebih baik dalam mengklasifikasikan data pada kelas positif dibandingkan kelas negatif, meskipun masih ditemukan kesalahan klasifikasi yang cukup signifikan, khususnya pada data negatif yang diprediksi sebagai positif.



Gambar 5. Hasil Model *Random Forest* dengan ENN

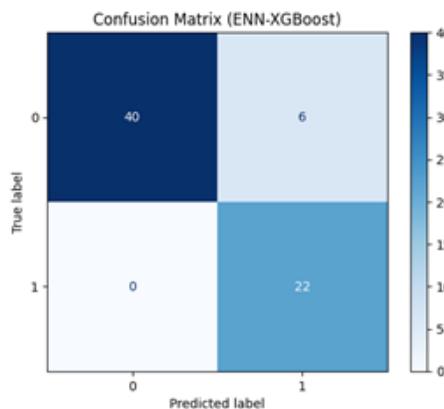
Selanjutnya, Gambar 5 menunjukkan model ENN dengan *Random Forest*. Berdasarkan hasil tersebut, sebanyak 40 data negatif berhasil diprediksi dengan benar sebagai negatif, sementara 6 data negatif keliru diprediksi sebagai positif. Pada kelas positif, seluruh data sebanyak 22 berhasil diklasifikasikan dengan tepat sebagai positif dan tidak terdapat data positif yang salah diprediksi sebagai negatif. Secara keseluruhan, hasil ini menunjukkan bahwa penerapan metode *Edited Nearest Neighbor* (ENN) sebelum algoritma *Random Forest* mampu meningkatkan kinerja klasifikasi, khususnya dalam mengurangi kesalahan prediksi pada kelas positif, sehingga model menjadi lebih akurat dan konsisten.



Gambar 6. Hasil Model *XGBoost* Tanpa ENN

Gambar 6 menunjukkan *confusion matrix* hasil klasifikasi menggunakan algoritma *XGBoost*. Diketahui bahwa sebanyak 16 data negatif berhasil diprediksi dengan benar sebagai negatif. Namun, masih terdapat 30 data negatif yang salah diklasifikasikan sebagai positif. Pada kelas positif, sebanyak 56 data berhasil diprediksi dengan tepat sebagai positif, sedangkan 18 data positif lainnya keliru diprediksi sebagai negatif. Secara keseluruhan, hasil ini menunjukkan bahwa algoritma *XGBoost* memiliki kemampuan yang cukup baik dalam mengenali kelas

positif, tetapi masih mengalami kesalahan klasifikasi yang relatif tinggi pada data kelas negatif yang diprediksi sebagai positif.



Gambar 7. Hasil Model *XGBoost* Dengan ENN

Sementara itu, Gambar 7 merupakan *confusion matrix* hasil klasifikasi metode ENN dengan *XGBoost*. Diketahui bahwa sebanyak 40 data negatif berhasil diprediksi dengan benar sebagai negatif. Namun demikian, masih terdapat 6 data negatif yang salah diklasifikasikan sebagai positif. Pada kelas positif, seluruh data sebanyak 22 berhasil diprediksi dengan tepat sebagai positif, dan tidak terdapat data positif yang keliru diprediksi sebagai negatif. Secara keseluruhan, hasil ini menunjukkan bahwa penerapan metode *Edited Nearest Neighbor* (ENN) sebelum algoritma *XGBoost* mampu meningkatkan performa klasifikasi secara signifikan, khususnya dengan menghilangkan kesalahan prediksi pada kelas positif dan menekan kesalahan pada kelas negatif, sehingga model menjadi lebih optimal dan stabil.

Berdasarkan Tabel 2 hasil pengujian metode, dapat diketahui bahwa penerapan metode *Edited Nearest Neighbor* (ENN) memberikan peningkatan kinerja yang signifikan pada algoritma *Random Forest* maupun *XGBoost*. Algoritma *Random Forest* (RF) tanpa *preprocessing* ENN menghasilkan nilai akurasi sebesar 0.59, *precision* sebesar 0.65, *recall* sebesar 0.74, dan *F1-score* sebesar 0.69. Hasil ini menunjukkan bahwa meskipun RF cukup baik dalam mendeteksi kelas positif, tingkat akurasi dan keseimbangan performa antarmetrik masih tergolong sedang. Setelah diterapkan *preprocessing* ENN pada *Random Forest* (ENN-RF), terjadi peningkatan performa yang sangat jelas, di mana nilai akurasi meningkat menjadi 0.91, *precision* menjadi 0.79, *recall* mencapai 1.00, dan *F1-score* meningkat menjadi 0.88. Nilai *recall* yang sempurna menunjukkan bahwa seluruh data kelas positif berhasil diklasifikasikan dengan benar, sementara peningkatan *F1-score* menandakan keseimbangan yang lebih baik antara *precision* dan *recall*.

Pada algoritma *XGBoost* tanpa ENN, diperoleh nilai akurasi sebesar 0.60, *precision* sebesar 0.65, *recall* sebesar 0.76, dan *F1-score* sebesar 0.70. Hasil ini menunjukkan pola yang hampir serupa dengan *Random Forest* tanpa ENN, yaitu model cukup baik dalam mendeteksi kelas positif, tetapi masih menghasilkan kesalahan klasifikasi yang cukup tinggi sehingga berdampak pada nilai akurasi dan *F1-score*. Selanjutnya, penerapan ENN pada *XGBoost* (ENN-XGBoost) juga memberikan peningkatan kinerja yang signifikan. Nilai akurasi meningkat menjadi 0.91, *precision* menjadi 0.79, *recall* mencapai 1.00, dan *F1-score* sebesar 0.88. Hasil ini mengindikasikan bahwa proses pembersihan data menggunakan ENN mampu mengurangi *noise* dan data yang tidak konsisten, sehingga membantu model *XGBoost* dalam menghasilkan klasifikasi yang lebih akurat dan stabil.

Secara keseluruhan, hasil pengujian menunjukkan bahwa penggunaan *preprocessing Edited Nearest Neighbor* (ENN) sangat berpengaruh dalam meningkatkan performa klasifikasi, baik pada algoritma *Random Forest* maupun *XGBoost*. Penelitian ini sejalan dengan penelitian yang dilakukan oleh [1] dan [2], yang menunjukkan bahwa penerapan metode *Edited Nearest Neighbours* (ENN) terbukti mampu meningkatkan akurasi, memperbaiki keseimbangan antara *precision* dan *recall*, serta menghasilkan nilai *F1-score* yang lebih tinggi, sehingga model

menjadi lebih andal dalam melakukan klasifikasi, seperti yang ditunjukkan pada Tabel 3.

Tabel 3. Hasil Pengujian Metode

Metode	Accuracy	Precision	Recall	F1-Score
RF	0.59	0.65	0.74	0.69
ENN-RF	0.91	0.79	1.00	0.88
XGBoost	0.60	0.65	0.76	0.70
ENN-XGBoost	0.91	0.79	1.00	0.88

4. KESIMPULAN

Penelitian ini berhasil menunjukkan bahwa penerapan metode reduksi *noise Edited Nearest Neighbours* (ENN) efektif dalam meningkatkan performa model klasifikasi penyakit scabies pada sapi berdasarkan gejala klinis. Hasil pengujian membuktikan bahwa pembersihan data dari sampel yang tidak konsisten mampu mengoptimalkan kinerja algoritma *Random Forest* dan *XGBoost* secara signifikan. Hal ini terlihat dari peningkatan akurasi kedua algoritma dari kisaran 0.59–0.60 menjadi 0.91 setelah melalui tahap *preprocessing* ENN. Selain itu, penggunaan ENN berhasil menghasilkan nilai *recall* yang sempurna (1.00), yang menunjukkan bahwa sistem mampu mendeteksi seluruh sampel positif scabies dengan tepat, sehingga memberikan tingkat keandalan yang lebih tinggi bagi praktisi kesehatan hewan dalam melakukan diagnosis dini.

Saran pengembangan penelitian ini ke depannya adalah menggunakan dataset yang lebih besar dan bervariasi dari berbagai wilayah guna meningkatkan kemampuan generalisasi model. Selain itu, penelitian selanjutnya dapat mengeksplorasi penggabungan teknik penanganan ketidakseimbangan kelas lainnya, seperti *Synthetic Minority Over-sampling Technique* (SMOTE) dengan ENN, serta mengintegrasikan model ke dalam aplikasi berbasis *mobile* atau *web* agar dapat dimanfaatkan secara praktis oleh peternak di lapangan untuk menekan kerugian ekonomi akibat penyakit scabies.

DAFTAR PUSTAKA

- [1] I. Indah dan N. Nurmayanti, “Identifikasi Kasus *Sarcoptes scabiei* pada Hewan Peliharaan di UPTD Pusat Kesehatan Hewan Makassar,” *Filogeni: Jurnal Mahasiswa Biologi*, vol. 3, no. 2, pp. 106–111, 2023.
- [2] S. Marwah, R. Astuti, dan F. M. Basysyar, “Implementasi Data Mining Menggunakan Algoritma Naïve Bayes untuk Diagnosis Penyakit Kulit Scabies pada Hewan,” *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 7, no. 6, pp. 3892–3897, Feb. 7, 2024. DOI: [10.36040/jati.v7i6.8276](https://doi.org/10.36040/jati.v7i6.8276).
- [3] A. P. Ramadhani, “Analisis Performa Algoritma Support Vector Machine dan Algoritma K-Nearest Neighbors untuk Kasus Penyakit Mulut dan Kuku pada Sapi di Jawa Timur,” *Journal Zetroem*, vol. 6, no. 1, pp. 73–78, Mar. 8, 2024. DOI: [10.36526/ztr.v6i1.3489](https://doi.org/10.36526/ztr.v6i1.3489).
- [4] C. Sukmawati, “Algoritma Naïve Bayes, Random Forest dan Support Vector Machine untuk Klasifikasi Penyakit Sapi,” *Techno Xplore : Jurnal Ilmu Komputer dan Teknologi Informasi*, vol. 9, no. 1, pp. 44–50, May 20, 2024. DOI: [10.36805/technoxplore.v9i1.6480](https://doi.org/10.36805/technoxplore.v9i1.6480).
- [5] M. I. Ichsan, “Implementasi Machine Learning untuk Deteksi Penyakit pada Kucing Menggunakan Random Forest,” *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 13, pp. 2000–2010, 3S1 Oct. 19, 2025. DOI: [10.23960/jitet.v13i3S1.8164](https://doi.org/10.23960/jitet.v13i3S1.8164).
- [6] S. Das, R. K. Roy, dan T. Bezboruah, “Machine Learning in Animal Healthcare: A Comprehensive Review,” *International Journal of Recent Engineering Science*, vol. 11, no. 3, pp. 89–93, Jun. 30, 2024. DOI: [10.14445/23497157/IJRES-V11I3P109](https://doi.org/10.14445/23497157/IJRES-V11I3P109).
- [7] I. D. Sulistyowati, S. Sunarno, dan D. Djuniadi, “Penerapan machine learning dengan algoritma support vector machine untuk prediksi kelembapan udara rata-rata,” id, *Jurnal Sistem Informasi, Teknologi Informatika dan Komputer*, vol. 15, no. 1, pp. 284–290, Sep. 30, 2024. DOI: [10.24853/justit.15.1.284-290](https://doi.org/10.24853/justit.15.1.284-290).

- [8] R. K. Halder et al., “Enhancing K-Nearest Neighbor Algorithm: A Comprehensive Review and Performance Analysis of Modifications,” en, *Journal of Big Data*, vol. 11, no. 1, p. 113, Aug. 11, 2024. DOI: [10.1186/s40537-024-00973-y](https://doi.org/10.1186/s40537-024-00973-y).
- [9] A. S. Salsabila, F. Sthevanie, dan K. N. Ramadhan, “Scabies Classification in Animal Using Uniform Local Binary Patterns,” in *2020 12th International Conference on Information Technology and Electrical Engineering (ICITEE)*, Yogyakarta, Indonesia: IEEE, Oct. 6, 2020, pp. 356–361. DOI: [10.1109/ICITEE49829.2020.9271720](https://doi.org/10.1109/ICITEE49829.2020.9271720).
- [10] M. F. Syam, M. N. Y. Utomo, dan E. Tungadi, “Cattle Disease Diagnosis System Using Random Forest Classification Method,” en, *Journal of Informatics and Computer Engineering Research*, vol. 2, no. 1, pp. 1–8, Aug. 7, 2025. DOI: [10.31963/jicer.v2i1.5563](https://doi.org/10.31963/jicer.v2i1.5563).
- [11] X. E. K. Permana, N. A. R. Rais, dan M. Muqorobin, “Classification of Cattle Diseases in Semin District Using Convolutional Neural Network (CNN),” en, *International Journal of Computer and Information System (IJCIS)*, vol. 5, no. 2, pp. 125–131, Jun. 30, 2024. DOI: [10.29040/ijcis.v5i2.172](https://doi.org/10.29040/ijcis.v5i2.172).
- [12] C. Eirene et al., “Classification of Learning Styles of Junior High School Students Using Random Forest & XGBoost Algorithm,” *Jurnal Bumigora Information Technology (BITe)*, vol. 7, no. 1, pp. 15–24, Jun. 19, 2025. DOI: [10.30812/bite.v7i1.4913](https://doi.org/10.30812/bite.v7i1.4913).
- [13] D. Priyanto et al., “Optimization of Random Forest for Health Data Classification Using PCA and K-Means SMOTE-ENN,” en, *Engineering, Technology & Applied Science Research*, vol. 15, no. 5, pp. 27646–27652, Oct. 6, 2025. DOI: [10.48084/etasr.12976](https://doi.org/10.48084/etasr.12976).
- [14] M. K. Ihsan, D. S. C. Saputri, dan N. Sulistianingsih, “Perbandingan algoritma boosting untuk klasifikasi gaya belajar siswa sekolah menengah kejuruan,” id, *Jutisi : Jurnal Ilmiah Teknik Informatika dan Sistem Informasi*, vol. 14, no. 2, pp. 963–974, Aug. 6, 2025. DOI: [10.35889/jutisi.v14i2.2701](https://doi.org/10.35889/jutisi.v14i2.2701).
- [15] F. Brishti et al., “Imbalanced Classification with Label Noise: A Systematic Review and Comparative Analysis,” en, *ICT Express*, vol. 11, no. 6, pp. 1127–1145, Dec. 2025. DOI: [10.1016/j.icte.2025.09.011](https://doi.org/10.1016/j.icte.2025.09.011).