

BAB I

PENDAHULUAN

1.1 Latar Belakang

Lumpy Skin Disease (LSD) adalah penyakit pada ternak sapi yang disebabkan oleh virus dari famili *Poxviridae* genus *Capripox* (PKG) [1], dengan penyebaran virusnya berhubungan dengan faktor lingkungan meliputi curah hujan dan interaksinya dengan suhu [2]. Penyakit ini pertama kali diidentifikasi di Afrika dan kemudian menyebar ke Timur Tengah, Eropa, dan Asia, termasuk Indonesia [1]. Indonesia sebagai negara tropis memiliki kondisi lingkungan yang mendukung perkembangan vektor penular LSD meliputi nyamuk dan lalat. LSD pertama kali terdeteksi di Indonesia pada tahun 2022, tepatnya di Provinsi Riau dan wabahnya dilaporkan menyebar ke selatan serta ke berbagai bagian Pulau Jawa [2]. Identifikasi dini LSD masih dilakukan secara manual oleh para ahli dan memerlukan waktu yang lama yaitu 12 hingga 21 hari untuk mendapatkan hasil yang akurat [3]. Hal ini dapat menyebabkan dampak signifikan pada sektor peternakan serta perdagangan daging sapi meliputi penurunan harga hewan yang terinfeksi, kematian, keguguran, dan penurunan produksi susu [2]. Oleh karena itu, diperlukan model klasifikasi yang lebih cepat dan akurat untuk mengidentifikasi LSD sejak dini. Pendekatan yang dapat digunakan adalah dengan menerapkan metode kecerdasan buatan dalam klasifikasi penyakit sehingga memungkinkan identifikasi lebih efisien dan tepat waktu.

Salah satu tantangan utama dalam pengembangan sistem klasifikasi otomatis LSD adalah ketidakseimbangan data antara sampel hewan terinfeksi dan tidak terinfeksi. Data ini terdiri dari 24.803 sampel, 21.764 sampel tidak terinfeksi dan 3.039 sampel terinfeksi. Ketidakseimbangan ini dapat menyebabkan bias dalam hasil klasifikasi, di mana model cenderung memprediksi kelas mayoritas. Dalam menangani ketidakseimbangan kelas terdapat beberapa pendekatan umum yaitu menerapkan teknik *Resampling* menggunakan *Random Undersampling* (RUS) dan *Synthetic Minority Oversampling Technique* (SMOTE) [1]. Terdapat beberapa penelitian yang menangani ketidakseimbangan kelas dengan RUS atau SMOTE. Penelitian pertama menggunakan *Random Undersampling* (RUS) dalam klasifikasi

serangan pada perangkat *IoT* menggunakan model *Random Forest*. Penelitian tersebut mendapatkan hasil akurasi yang besarnya mencapai 99.73% [4], penelitian tersebut menggunakan 46.686.545 record dengan 34 atribut. Selanjutnya, terdapat penelitian yang membandingkan RUS dan SMOTE dalam klasifikasi pasien terinfeksi *COVID-19* menggunakan model *Random Forest*. AUC yang didapatkan sebesar 94.52% pada RUS dan nilai AUC pada SMOTE sebesar 99.32% [5]. Penelitian tersebut menggunakan *dataset* berupa gambar sebanyak 402 record pasien Covid-19. Penelitian ketiga melakukan perbandingan antara RUS dan SMOTE untuk menangani ketidakseimbangan kelas pada *dataset* LSD. Hasilnya, SMOTE memiliki nilai *Recall* lebih tinggi yaitu 97.7% dan AUC sebesar 0.996 daripada RUS yang memiliki nilai *Recall* 96%. Hasil penelitian ini menunjukkan bahwa SMOTE memberikan performa yang lebih baik dibandingkan dengan RUS. Selain itu, AUC digunakan untuk menilai seberapa baik model yang sudah diimbangi dengan teknik SMOTE [1]. RUS cocok jika ukuran *dataset* besar dan kelas mayoritas sangat dominan [6] dan SMOTE menyeimbangkan data tanpa menghapus sampel dari kelas mayoritas, sehingga informasi penting dari data tetap terjaga [7]. Berdasarkan hal tersebut, dalam penelitian ini akan menggunakan RUS dan SMOTE dalam menangani ketidakseimbangan kelas. SMOTE bekerja dengan menghasilkan salinan data dan melihat kemiripan antara sample kelas minoritas yang berdekatan dimana setiap data akan diperbanyak berdasarkan pola dari tetangga terdekatnya, sedangkan RUS bekerja dengan cara mengurangi kelas mayoritas secara acak supaya jumlah dari kelas mayoritas sama dengan jumlah kelas minoritas [1].

Dalam proses klasifikasi, berbagai algoritma *Machine Learning* telah digunakan untuk mengidentifikasi penyakit LSD. Beberapa di antaranya termasuk *Logistic Regression*, *Decision Tree*, *Gradient Boosting*, dan *Support Vector Machine* (SVM). Berdasarkan penelitian terdahulu, algoritma *Support Vector Machine* hanya mencapai akurasi sebesar 73.36% dalam konteks klasifikasi kemampuan adaptasi siswa [7], sementara pada kasus klasifikasi LSD, *Logistic Regression* hanya mencapai akurasi 89.68% [6]. Hasil tersebut menunjukkan bahwa algoritma-algoritma tersebut belum memberikan kinerja optimal dalam permasalahan klasifikasi yang kompleks dan ketidakseimbangan data. Disisi lain,

Random Forest menunjukkan performa yang lebih baik dan konsisten dalam berbagai penelitian. Sebuah studi menyatakan bahwa *Random Forest* memiliki akurasi sebesar 97.9% dalam klasifikasi biaya pendidikan mahasiswa, lebih tinggi dibanding *Multilayer Perceptron* (95.54%) dan *Naïve Bayes* (68.36%) [6]. Pada kasus klasifikasi LSD, *Random Forest* menunjukkan performa akurasi sebesar 98% dan *F1-score* sebesar 98% ketika dikombinasikan dengan metode SMOTE [17]. Keunggulan *Random Forest* terletak pada kemampuannya dalam mengelola data dalam jumlah besar serta menghasilkan prediksi yang stabil dengan menggabungkan beberapa pohon keputusan [5].

Metode *Random Forest* bekerja dengan mengkombinasikan beberapa pohon keputusan untuk menghasilkan keputusan dalam menangani permasalahan yang kompleks [8]. Sebuah penelitian yang membandingkan empat teknik klasifikasi, yaitu *Random Forest*, *Multilayer Perceptron*, *Logistic Regretion*, dan *Naïve Bayes* dalam menentukan klasifikasi biaya pendidikan mahasiswa menunjukkan bahwa *Random Forest* memiliki tingkat keakuratan paling tinggi sebesar 97.9%, diikuti oleh *Multilayer Perceptron* (95.54%), *Logistic Regretion* (89.68%), dan *Naïve Bayes* (68.36%) [9]. Penelitian ini menggunakan data pendaftar Seleksi Bersama Masuk Politeknik Negeri yang diterima mulai tahun 2019 sampai dengan tahun 2021 pada Politeknik Elektronika Negeri Surabaya sebanyak 873 sampel dengan 31 atribut. Penelitian lain melakukan membandingkan metode *Random Forest* dengan *Support Vector Machine* dalam klasifikasi pada kemampuan level beradaptasi siswa. Hasilnya menunjukan tingkat akurasi *Random Forest* mencapai akurasi 91.5%, sedangkan akurasi dari *Support Vector Machine* hanya mencapai 73.36% yang menunjukan bahwa *Random Forest* lebih unggul daripada *Support Vector Machine* dalam penelitian tersebut [10]. Algoritma *Random Forest* banyak digunakan untuk klasifikasi karena efisiensinya [11]. Namun, *Random Forest* memiliki kendala terkait dalam menentukan *Hyperparameter* optimal yang dapat meningkatkan kinerja model, sehingga memerlukan waktu yang lama [11]. Oleh karena itu, diperlukan *Hyperparameter tuning* untuk meningkatkan kinerja *Random Forest* [12].

Random Search adalah salah satu metode *Hyperparameter tuning* yang diterapkan untuk mengoptimalkan kinerja model. Dalam penelitian ini, proses

Hyperparameter tuning yang digunakan adalah *Random Search*. Berbeda dengan *Grid Search*, yang menjelajahi seluruh kombinasi *Hyperparameter* sehingga memerlukan banyak waktu, *Random Search* memilih kombinasi *Hyperparameter* secara acak tanpa perlu menjelajahi keseluruhan ruang *parameter* sehingga penggunaannya lebih cepat [13]. Penelitian sebelumnya menunjukkan bahwa penerapan *Random Search* pada metode *Random Forest* dapat meningkatkan akurasi model, seperti yang terlihat dalam penelitian yang memprediksi kegagalan jantung dengan 299 sampel dengan 13 atribut. Nilai AUC meningkat dari 0.866 menjadi 0.906 setelah proses optimalisasi *Hyperparameter* dilakukan [12]. Penelitian serupa juga mendeteksi keberlangsungan hidup menggunakan data medis 299 pasien gagal jantung. Penelitian tersebut menunjukan bahwa penggunaan *Random Search* untuk mengoptimalkan *Hyperparameter* pada *Random Forest* dapat meningkatkan akurasi prediksi dari 80% menjadi 83,33% [11]. Hal ini menunjukan bahwa metode ini efektif dalam meningkatkan kinerja model dibandingkan tanpa *Hyperparameter* [11].

Penelitian ini memiliki kemiripan dengan penelitian yang dilakukan oleh Suparyati (2022). Penelitian tersebut juga menggunakan *dataset* LSD dari sumber yang sama dan menerapkan metode *Random Forest* untuk klasifikasi serta membandingkan metode *Resampling* terhadap kinerja klasifikasi [1]. Perbedaan utama antara penelitian ini dengan penelitian tersebut terletak pada penerapan langkah penyetelan *Hyperparameter* menggunakan *Random Search* untuk mengoptimalkan *parameter* model *Random Forest*, yang diharapkan dapat menghasilkan kinerja klasifikasi yang lebih baik. Penelitian sebelumnya lebih fokus pada perbandingan metode *Resampling* terhadap kinerja klasifikasi, sementara penelitian ini menggabungkan optimasi *Hyperparameter* sebagai tambahan untuk meningkatkan performa model. Berdasarkan perbedaan pendekatan ini, penelitian ini memiliki hipotesis bahwa kinerja model *Random Forest* akan mengalami peningkatan yang signifikan setelah dilakukan *Hyperparameter tuning*, dibandingkan dengan model *Random Forest* tanpa optimasi.

Berdasarkan pemaparan di atas, penelitian ini bertujuan untuk mengembangkan model klasifikasi LSD menggunakan algoritma *Random Forest* yang dioptimalkan dengan *Random Search* dan ditingkatkan dengan teknik

penyeimbangan data menggunakan RUS dan SMOTE. Penelitian ini berkontribusi dalam menggabungkan pendekatan *balancing* dan *Hyperparameter tuning* secara sistematis untuk menghasilkan model klasifikasi yang tidak hanya akurat namun juga efisien dalam mengatasi masalah ketidakseimbangan kelas. Dengan demikian, diharapkan model yang dihasilkan mampu mengidentifikasi penyakit LSD secara dini dan tepat, sehingga dapat digunakan sebagai sistem pendukung keputusan dalam pengendalian wabah di sektor peternakan, terutama di Indonesia.

1.2 Rumusan Masalah

Rumusan masalah pada penelitian ini sebagai berikut:

1.2.1 Identifikasi Masalah

Berdasarkan penelitian sebelumnya yang menggunakan *dataset* dari sumber sama untuk mengklasifikasikan *Lumpy Skin Disease* (LSD) pada hewan ternak sapi berdasarkan faktor lingkungan, berfokus pada perbandingan kedua metode *Resampling* dan belum mengoptimalkan model *Random Forest*. Pada penelitian terbaru ini membandingkan kedua metode *Resampling*, yaitu *Synthetic Minority Oversampling Technique* (SMOTE) dan *Random Undersampling* (RUS), serta menerapkan *Hyperparameter tuning* pada model *Random Forest* guna meningkatkan akurasi model dalam klasifikasi LSD.

1.2.2 Metode Usulan

Untuk meningkatkan performa model *Random Forest*, penelitian ini menerapkan *Random Forest* dengan *Hyperparameter tuning* menggunakan *Random Search*. Berdasarkan ulasan beberapa penelitian, *Random Search* terbukti lebih efisien dibandingkan dengan *Grid Search* sehingga dapat meningkatkan performa klasifikasi *Lumpy Skin Disease* (LSD). Selain itu, SMOTE dan RUS diterapkan untuk mengatasi ketidakseimbangan kelas dalam *dataset*.

1.2.3 Pertanyaan Penelitian

Berdasarkan permasalahan dan usulan yang telah dibahas, pada penelitian ini didapatkan pertanyaan yaitu berapa tingkat akurasi algoritma *Random Forest* dalam mengklasifikasikan *Lumpy Skin Disease* (LSD) berdasarkan faktor lingkungan dengan dan tanpa optimasi menggunakan

Random Search, serta dengan penerapan teknik *Resampling* SMOTE dan *RUS*?

1.3 Tujuan dan Manfaat

Adapun tujuan dan manfaat dalam penelitian ini yaitu:

1.3.1 Tujuan Penelitian

Tujuan dari penelitian ini adalah untuk mengetahui akurasi algoritma *Random Forest* dalam mengklasifikasikan *LSD* berdasarkan faktor lingkungan, baik dengan maupun tanpa optimasi menggunakan *Random Search* serta dengan penerapan teknik *Resampling* SMOTE dan *RUS*.

1.3.2 Manfaat Penelitian

Manfaat yang diharapkan adalah untuk memperoleh wawasan mengenai tingkat keakuratan model *Random Forest* yang telah dioptimalkan melalui penerapan *Hyperparameter tuning* dengan *Random Search*, sehingga dapat dijadikan referensi dalam mengembangkan sistem klasifikasi dalam mengklasifikasikan *Lumpy Skin Disease* (*LSD*) secara dini.

1.4 Batasan Masalah

Peneliti telah menerapkan batasan-batasan berdasarkan permasalahan yang tertulis untuk memfokuskan penelitian ini, batasan-batasan tersebut yakni:

1. Sumber data yang peneliti gunakan didapat dari Mendeley *Dataset* pada tautan <https://data.mendeley.com/datasets/7pyhbzb2n9/1> yang dikumpulkan dari tahun 2011
2. Jumlah keseluruhan data yang akan peneliti gunakan yaitu sebanyak 24.803, dengan fitur sebanyak 19 fitur dan 1 target “*Lumpy*” yang digunakan sebagai kelas yang terdiri dari kelas *Lumpy* (1) dan *No lumpy* (0).
3. Metode yang digunakan untuk klasifikasi adalah *Random Forest*.
4. *Hyperparameter* yang digunakan adalah *Random Search*.
5. *Dataset* termasuk tidak seimbang dengan kelas *No lumpy* (0) sebanyak 21.764 data dan *Lumpy* (1) sebanyak 3.039 data, peneliti menggunakan *Synthetic Minority Oversampling Technique* (SMOTE) dan *Random Undersampling* (RUS) untuk menyeimbangkan *dataset*.
6. Klasifikasi pada penelitian ini difokuskan pada ternak sapi.

1.5 Sistematika Penulisan

Struktur penyusunan penelitian ini terbagi ke dalam beberapa bab, di antaranya sebagai berikut:

BAB I PENDAHULUAN

Bab ini mencakup pendahuluan penelitian yang membahas latar belakang pentingnya penelitian yang dilakukan untuk klasifikasi *Lumpy Skin Disease* pada hewan ternak berdasarkan faktor lingkungan menggunakan *Random Forest*, permasalahan yang di angkat, metode usulan yaitu *Random Forest* menggunakan *Hyperparameter Random Search*, tujuan dan manfaat dari penelitian, dan batasan-batasan dalam penelitian, serta sistematika penulisan penelitian.

BAB II TEORI PENUNJANG

Bab ini menguraikan teori-teori yang menjadi dasar yang sesuai sebagai landasan ilmiah dalam penelitian, meliputi algoritma *Random Forest*, *Synthetic Minority Oversampling Technique* (SMOTE) dan *Random Undersampling* (RUS) yang digunakan untuk teknik *Resampling*, dan *Hyperparameter tuning* yang digunakan yaitu *Random Search*.

BAB III METODOLOGI PENELITIAN

Bab ini berisi alur penelitian dan metodologi yang diterapkan dengan menjelaskan langkah yang akan dilakukan untuk mencapai tujuan penelitian, meliputi pengolahan data dari sumber data, menerapkan teknik *Resampling*, pemodelan algoritma *Random Forest* dengan *Hyperparameter* menggunakan *Random Search*.

BAB VI HASIL DAN PEMBAHASAN

Bab ini menguraikan hasil klasifikasi hipertensi yang didapatkan menggunakan metode *Random Forest* baik dengan dan tanpa *Hyperparameter tuning* menggunakan *Random Search* dan hasil dari pengujian sesuai dengan skenario yang telah ditentukan

BAB V PENUTUP

Bab ini berisi kesimpulan yang didapatkan dari penelitian Klasifikasi Penyakit *Lumpy Skin Disease* (LSD) Pada Ternak Sapi Berdasarkan Faktor Lingkungan

menggunakan Metode *Random Forest* baik dengan dan tanpa *Hyperparameter tuning* menggunakan *Random Search* serta saran untuk pengembangan penelitian selanjutnya.