

BAB I

PENDAHULUAN

1.1 Latar Belakang

Sindrom ovarium polikistik (Polycystic Ovary Syndrome/PCOS) merupakan salah satu gangguan endokrin yang umum terjadi pada wanita usia reproduksi [1]. Kondisi ini ditandai dengan sekresi androgen yang lebih tinggi di ovarium wanita dengan PCOS dibandingkan dengan wanita normal [2]. Ketidakseimbangan ini menyebabkan berbagai gejala, seperti menstruasi yang tidak teratur dan terbentuknya polikistik ovarium (PCOS), yaitu kumpulan ovum yang tidak matang di dalam ovarium [3]. Selain itu, PCOS juga berkaitan erat dengan berbagai komplikasi kesehatan lain, termasuk obesitas, diabetes tipe 2, serta peningkatan risiko penyakit kardiovaskular [4]. Beberapa faktor risiko signifikan yang dapat meningkatkan kemungkinan terjadinya PCOS antara lain adalah kelebihan status gizi, kurangnya aktivitas fisik, serta kondisi kesehatan mental yang buruk, seperti stres dan depresi [5]. Faktor-faktor ini berkontribusi terhadap ketidakseimbangan hormonal yang menjadi ciri khas utama dari PCOS, termasuk produksi androgen yang berlebihan dan kegagalan ovarium dalam melepaskan telur secara teratur [6].

Hingga saat ini, PCOS dikenal sebagai gangguan heterogen dengan etiologi yang belum sepenuhnya dipahami. Namun, berbagai penelitian menunjukkan adanya interaksi kompleks antara faktor genetik, lingkungan, dan perilaku dalam perkembangan sindrom ini [7]. Dalam praktiknya, gejala PCOS sering kali menjadi alasan utama pasien untuk mencari bantuan medis, terutama karena gangguan siklus menstruasi, kesulitan untuk hamil (infertilitas), masalah berat badan berlebih (obesitas), serta kelainan lain seperti hirsutisme (pertumbuhan rambut berlebihan) dan jerawat [5].

Penelitian ini bertujuan untuk melakukan klasifikasi terhadap penyakit PCOS menggunakan algoritma Random Forest. Algoritma tersebut dipilih karena mampu mengklasifikasikan data dalam jumlah besar, selain itu memiliki akurasi

tinggi dan kesalahan rendah serta mampu menangani dataset training yang besar [8]. Ketidakseimbangan pada dataset biasanya akan mempengaruhi hasil, maka digunakan Synthetic Minority Oversampling Technique (SMOTE) untuk menyeimbangkan data tersebut. Melalui SMOTE, jumlah data pada kelas minoritas ditingkatkan dengan menghasilkan data sintesis baru yang dibentuk berdasarkan pada kemiripan antar sampel data dari kelas minoritas [9]. Penelitian ini akan dilakukan melalui tahapan, eda, preprocessing, pemilihan fitur, SMOTE, modeling, dan evaluasi. Selanjutnya, model terbaik akan di-deploy dalam bentuk web sederhana.

1.2 Rumusan Masalah

Berdasarkan latar belakang diatas, maka hal-hal yang perlu dirumuskan adalah sebagai berikut:

1. Bagaimana tahapan exploratory data analysis (EDA), preprocessing data, seleksi fitur, pemodelan, dan evaluasi dalam membangun model klasifikasi penyakit Polycystic Ovary Syndrome (PCOS) berdasarkan data klinis dan biologis pasien menggunakan algoritma Random Forest?
2. Bagaimana tingkat akurasi dan performa model machine learning berbasis Random Forest dalam mengklasifikasikan kemungkinan seseorang menderita PCOS berdasarkan data klinis dan biologis pasien?
3. Bagaimana pengaruh penerapan metode Synthetic Minority Oversampling Technique (SMOTE) terhadap performa klasifikasi penyakit PCOS pada dataset klinis dan biologis yang tidak seimbang?
4. Bagaimana implementasi model klasifikasi penyakit PCOS yang telah dibangun menggunakan Random Forest dan SMOTE ke dalam bentuk antarmuka sederhana?

1.3 Batasan Masalah

Untuk menghindari pelebaran pokok masalah agar penelitian dapat terarah, maka dalam penelitian ini terdapat beberapa Batasan masalah yaitu:

1. Penelitian ini hanya menggunakan data klinis pasien PCOS yang tersedia

dalam dataset terpilih yang diambil dari platform Kaggle.

2. Teknik penyeimbangan data yang digunakan dalam penelitian ini menggunakan metode SMOTE.
3. Algoritma yang digunakan untuk membangun model klasifikasi penyakit PCOS hanya menggunakan algoritma Random Forest, tanpa membandingkannya dengan algoritma machine learning lain.
4. Proses pemodelan dilakukan menggunakan Google Colab.
5. Evaluasi kinerja model hanya difokuskan pada metrik kinerja model seperti akurasi, presisi, recall, dan F1-score ROC dan AUC tanpa melibatkan uji klinis atau validasi oleh tenaga medis profesional.
6. Implementasi sistem klasifikasi hanya dibangun dalam bentuk antarmuka web sederhana sebagai prototipe yang bertujuan untuk mendemonstrasikan hasil pemodelan dan pengujian sistem.

1.4 Tujuan Penelitian

Tujuan yang akan dicapai oleh peneliti dalam penelitiannya adalah:

1. Menjelaskan tahapan exploratory data analysis (EDA), preprocessing data, seleksi fitur, pemodelan, dan evaluasi dalam membangun model klasifikasi penyakit Polycystic Ovary Syndrome (PCOS) berdasarkan data klinis pasien menggunakan algoritma Random Forest.
2. Mengetahui tingkat akurasi dan performa model machine learning yang dibangun dalam mengklasifikasikan kemungkinan seseorang menderita PCOS berdasarkan data klinis pasien.
3. Menganalisis pengaruh penerapan metode Synthetic Minority Oversampling Technique (SMOTE) terhadap performa klasifikasi pada dataset yang tidak seimbang.
4. Mengimplementasikan model klasifikasi PCOS yang telah dibangun ke dalam bentuk antarmuka sederhana sebagai media demonstrasi hasil penelitian.

1.5 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat teoritis dan manfaat praktis. Secara teoritis, sistem yang dikembangkan menjadi referensi tambahan bagi penelitian selanjutnya yang ingin mengembangkan atau menguji model klasifikasi penyakit PCOS dengan pendekatan atau algoritma berbeda. Secara praktis, penelitian ini memberikan alternatif pendekatan berbasis data dalam melakukan klasifikasi kasus PCOS menggunakan data klinis dan biologis pasien, serta menyajikan hasil analisis performa algoritma klasifikasi pada kondisi data yang tidak seimbang, yang dapat dimanfaatkan oleh peneliti atau pengembang sistem sebagai bahan evaluasi dan pengembangan model klasifikasi selanjutnya.

1.6 Sistematika Penulisan

Sistematika penulisan skripsi ini disusun dalam lima bab utama yang berisi uraian secara garis besar mengenai isi dari masing-masing bab, sebagai berikut :

BAB I PENDAHULUAN

Berisi gambaran umum mengenai penelitian yang dilakukan. Di dalamnya mencakup latar belakang masalah yang menjelaskan alasan dilakukannya penelitian, rumusan masalah yang memaparkan pertanyaan penelitian, batasan masalah untuk memperjelas ruang lingkup penelitian, serta tujuan penelitian yang ingin dicapai. Selain itu, bab ini juga memuat manfaat penelitian baik secara teoritis maupun praktis, dan sistematika penulisan yang menjelaskan susunan tiap bab dalam skripsi.

BAB II TINJAUAN PUSTAKA

Berisi dasar teori dan studi literatur penelitian terdahulu yang relevan dengan topik. Teori-teori yang dibahas mencakup konsep dasar seperti PCOS (Polycystic Ovary Syndrome), machine learning, algoritma Random Forest, Gird Search, serta teknik oversampling SMOTE. Selain itu juga dibahas teori mengenai evaluasi model klasifikasi seperti confusion matrix, ROC, dan AUC.

BAB III METODE PENELITIAN

Berisi rancangan dan tahapan penelitian yang dilakukan dalam penelitian ini. Pembahasan mencakup alur penelitian, sumber dataset yang digunakan, serta alat dan bahan penelitian. Selain itu, pada bab ini juga dijelaskan tahapan penelitian secara sistematis, mulai dari pengumpulan dan pemahaman data, preprocessing data, penanganan outlier, penyeimbangan data menggunakan metode SMOTE, pemilihan fitur, pemodelan menggunakan algoritma Random Forest, evaluasi model, hingga proses deployment model ke dalam aplikasi berbasis web sebagai bentuk implementasi hasil penelitian.

BAB IV HASIL DAN PEMBAHASAN

Bab ini merupakan tahapan yang dilakukan penulis dalam mengembangkan dan menguji model klasifikasi PCOS. Pembahasan meliputi hasil eksplorasi data, preprocessing data, penanganan outlier, penerapan metode SMOTE, serta hasil pemodelan menggunakan algoritma Random Forest. Selain itu, pada bab ini juga dibahas hasil evaluasi kinerja model menggunakan metrik evaluasi seperti confusion matrix, akurasi, presisi, recall, F1-score, ROC, dan AUC, serta hasil implementasi model ke dalam aplikasi web sederhana sebagai tahap deployment dan pengujian sistem.

BAB V PENUTUP

Bab ini berisi kesimpulan yang dapat peneliti rangkum berdasarkan hasil penelitian dan pembahasan pada bab sebelumnya, serta saran yang diberikan sebagai bahan pertimbangan dan pengembangan untuk penelitian selanjutnya.