

BAB 1

PENDAHULUAN

1.1 Latar Belakang Masalah

Obesitas merupakan salah satu masalah kesehatan dunia serius yang disebabkan oleh berlebihnya timbunan lemak pada tubuh [3]. *World Health Organization* (WHO) menyatakan bahwa tingkat obesitas dunia telah mengalami peningkatan sebanyak 3 kali lipat dalam 40 tahun terakhir [4]. Dari 1.9 miliar orang dewasa yang mengalami kelebihan berat badan, 650 juta di antaranya mengalami obesitas [4].

Obesitas sebagian besar disebabkan oleh pola hidup yang tidak sehat, seperti makan terlalu banyak dan kurangnya aktivitas fisik yang pada akhirnya menimbulkan penimbunan lemak yang berlebih [3]. Obesitas juga dapat meningkatkan risiko berbagai penyakit lain yang lebih berbahaya, seperti diabetes tipe-2, kanker, penyakit jantung, hipertensi, dan stroke [4, 5]. Dengan demikian, obesitas seharusnya tidak hanya dianggap sebagai akibat dari gaya hidup yang tidak sehat, tetapi sebagai penyakit yang dapat menimbulkan faktor risiko penyakit lain yang lebih berbahaya. Dengan demikian, deteksi dini dari tingkat obesitas penting untuk dilakukan [3].

Saat ini, telah terdapat indikator pengukuran tingkat obesitas dengan menggunakan *Body Mass Index* (BMI). Tetapi, telah diketahui bahwa BMI telah melebih-lebihkan obesitas bagi yang memiliki massa otot yang tinggi, seperti para atlet dan meremehkan obesitas bagi yang memiliki massa otot yang rendah, seperti para lansia [6]. Hal ini dikarenakan, BMI hanya memperhitungkan dan bergantung pada tinggi serta berat badan saja, bukan pada komposisi tubuh dan keseharian seseorang [6].

Pendekatan alternatif lainnya dapat diaplikasikan untuk menutupi kekurangan dari BMI dalam melakukan pengukuran tingkat obesitas. Penerapan *machine learning* dalam berbagai bidang, termasuk studi medis sudah sangat populer karena dapat digunakan untuk proses klasifikasi, *clustering*, dan deteksi anomali data [7]. Pemanfaatannya telah berkembang dan berkontribusi terhadap ilmu pengetahuan, termasuk bidang kesehatan dan kedokteran [7]. Klasifikasi dengan *machine learning* dapat digunakan untuk mengklasifikasikan tingkat obesitas berdasarkan gaya hidup dan pola makan sehari-hari [3].

Dalam penelitian terdahulu yang mengimplementasikan algoritma *K-Nearest Neighbor* (KNN) untuk mengklasifikasikan tingkat obesitas menggunakan "PIMA Indian Diabetes" *dataset*, didapatkan tingkat akurasi sebesar 78.98% dengan sarannya adalah menggunakan algoritma lain untuk meningkatkan akurasi model dalam mengklasifikasikan tingkat obesitas agar sistem klasifikasi yang dibuat dapat menjadi lebih handal dan akurat lagi [3]. Dalam penelitian terdahulu lainnya yang membandingkan algoritma *Random Forest*, *Decision Tree*, *Support Vector Machine*, dan *K-Nearest Neighbor* dalam mengklasifikasikan penyakit diabetes, didapatkan bahwa *Random Forest* menghasilkan performa yang terbaik dengan 97.5% akurasi, 97.4% *precision*, 96.6% *recall*, dan 97% *f1-score* [8]. Dari hasil evaluasi tersebut, dapat disimpulkan bahwa algoritma *Random Forest* memiliki performa yang sangat baik dalam mengklasifikasikan penyakit diabetes.

Dalam upaya meningkatkan akurasi model berdasarkan saran pada penelitian terdahulu pertama, maka akan digunakan algoritma *Random Forest Classifier* untuk mengklasifikasikan tingkat obesitas menggunakan "PIMA Indian Diabetes" *dataset* sebagai pengembangannya. Hal ini dikarenakan, algoritma tersebut memiliki performa yang sangat baik dalam mengklasifikasikan penyakit diabetes dari penelitian terdahulu kedua, yang berhubungan erat dengan obesitas karena merupakan salah satu penyakit yang dapat ditimbulkan oleh obesitas. Algoritma *Random Forest* sendiri merupakan salah satu algoritma yang populer dan banyak digunakan untuk proses klasifikasi medis [7]. Beberapa keunggulannya adalah menghasilkan hasil klasifikasi yang baik, memiliki kemampuan untuk menangani *dataset* campuran, memiliki tingkat efisiensi tinggi dan akurasi yang sangat baik, serta menghasilkan tingkat *error* yang rendah [9, 10]. Harapannya adalah agar nantinya sistem klasifikasi tingkat obesitas yang dibuat dapat menghasilkan performa yang sangat baik juga seperti pada penelitian terdahulu kedua.

Dalam proses pembangunannya, akan dilakukan *model selection* dengan mengimplementasikan teknik pemilihan fitur dan *hyperparameter* terbaik. Model-model tersebut akan dibandingkan satu sama lain untuk mendapatkan model terbaik yang menghasilkan akurasi yang tertinggi. Setelah itu, model terpilih akan diimplementasikan ke dalam sebuah sistem untuk membantu dan mempermudah pendeteksian dini tingkat obesitas dengan lebih akurat. Berdasarkan seluruh penjelasan dan penjabaran yang telah diuraikan, maka penelitian ini akan difokuskan pada perancangan dan pembangunan sistem klasifikasi tingkat obesitas menggunakan algoritma *Random Forest Classifier* dengan tujuan untuk

memberikan hasil klasifikasi dengan tingkat akurasi yang tinggi.

1.2 Rumusan Masalah

1. Bagaimana cara merancang dan membangun sistem untuk mengklasifikasikan tingkat obesitas dengan menggunakan algoritma *Random Forest Classifier*?
2. Bagaimana cara melakukan perbandingan model untuk mendapatkan model terbaik berdasarkan hasil *model selection* dengan menerapkan pemilihan fitur dan *hyperparameter* terbaik menggunakan algoritma *Random Forest Classifier* untuk mengklasifikasikan tingkat obesitas, serta mengevaluasi performa model terpilih berdasarkan nilai akurasi, *precision*, *recall*, dan *f1-score*?

1.3 Batasan Permasalahan

1. Penelitian hanya dilakukan berdasarkan satu *dataset* yang berjumlah 2111 data dengan 16 fitur dengan rentang usia dari 14-61 tahun [11].
2. Penelitian hanya berfokus pada penggunaan algoritma *Random Forest Classifier* untuk mengklasifikasikan tingkat obesitas melalui pembuatan sistem.

1.4 Tujuan Penelitian

1. Mengetahui cara untuk merancang dan membangun sistem untuk mengklasifikasikan tingkat obesitas dengan menggunakan algoritma *Random Forest Classifier*.
2. Mengetahui cara untuk melakukan perbandingan model untuk mendapatkan model terbaik berdasarkan hasil *model selection* dengan menerapkan pemilihan fitur dan *hyperparameter* terbaik menggunakan algoritma *Random Forest Classifier* untuk mengklasifikasikan tingkat obesitas, serta mengevaluasi performa model terpilih berdasarkan nilai akurasi, *precision*, *recall*, dan *f1-score*.

1.5 Manfaat Penelitian

1. Menghasilkan sistem klasifikasi tingkat obesitas yang lebih akurat.

2. Menambah wawasan terkait penggunaan algoritma *Random Forest Classifier* untuk mengklasifikasikan tingkat obesitas.
3. Mengetahui manfaat dari pengimplementasian teknik pemilihan fitur dan *hyperparameter* terbaik dalam meningkatkan akurasi model *Random Forest Classifier*, serta mengevaluasi performa algoritma ini dalam mengklasifikasikan tingkat obesitas melalui nilai akurasi, *precision*, *recall*, dan *f1-score*.

1.6 Sistematika Penulisan

Sistematika penulisan laporan adalah sebagai berikut:

- Bab 1 PENDAHULUAN
Bab 1 terdiri dari penjelasan terkait latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, serta sistematika penulisan.
- Bab 2 LANDASAN TEORI
Bab 2 terdiri dari penjelasan terkait teori-teori yang digunakan pada laporan skripsi ini. Teori-teori tersebut antara lain adalah obesitas, pembelajaran mesin, *decision tree*, *random forest*, *randomized search*, serta *confusion matrix*.
- Bab 3 METODOLOGI PENELITIAN
Bab 3 terdiri dari penjelasan terkait alur dari penelitian yang dilakukan, yaitu pengumpulan data, pembangunan model, serta pada tahap perancangan dan pembangunan sistem.
- Bab 4 HASIL DAN DISKUSI
Bab 4 terdiri dari penjelasan terkait hasil implementasi dari hasil perancangan. Bab ini mencakup penjelasan hasil implementasi pembangunan model *machine learning* dan pembangunan sistem.
- Bab 5 SIMPULAN DAN SARAN
Bab 5 terdiri dari penjelasan terkait kesimpulan dari penelitian yang telah dilakukan serta saran untuk pengembangan dan/ atau penelitian selanjutnya.