

BAB I

PENDAHULUAN

1.1. Latar Belakang

Obesitas adalah salah satu permasalahan kesehatan paling umum yang ditandai dengan penumpukan lemak tubuh secara abnormal sehingga dapat mengganggu fungsi dan kesehatan secara keseluruhan [1]. Kondisi ini telah terbukti memberikan dampak negatif terhadap kesejahteraan fisik, metabolik, dan psikologis individu. Tidak hanya menjadi isu kesehatan di negara maju, obesitas kini telah berkembang menjadi masalah kesehatan global yang memengaruhi lebih dari 1,9 miliar penduduk dunia dari berbagai latar belakang, usia, dan kelompok sosial [2]-[4].

Secara klinis, tingkat obesitas umumnya diukur menggunakan *Body Mass Index* (BMI), yang dihitung dengan membagi berat badan (kg) terhadap tinggi badan dalam meter yang dikuadratkan. BMI menjadi standar internasional dalam menentukan kategori berat badan, termasuk berat badan kurang, normal, berat badan berlebih, dan obesitas. Asupan kalori berlebih yang tidak seimbang dengan pengeluaran energi menyebabkan akumulasi lemak yang berlebihan dan berkontribusi terhadap berbagai risiko kesehatan [5]-[8].

Dalam lima dekade terakhir, prevalensi obesitas mengalami peningkatan signifikan. Di Amerika Serikat, misalnya, angka obesitas tercatat meningkat hingga tiga kali lipat, serta diperkirakan bahwa pola ini akan berlanjut di masa depan meskipun berbagai program penanggulangan telah dilakukan [9]-[12]. Penelitian sebelumnya juga menunjukkan bahwa obesitas tidak hanya dialami oleh kelompok tertentu, tetapi mencakup seluruh spektrum masyarakat meliputi berbagai tingkat sosial ekonomi, etnis, wilayah geografis, dan kelompok usia [13]-[15].

Hingga September 2021, diperkirakan obesitas berkontribusi sekitar 2,8 juta kematian global setiap tahun [16]. Terlepas dari kemungkinan perubahan angka ini, obesitas tetap dikaitkan dengan peningkatan risiko penyakit infeksi tingkat sedang hingga kronis, serta berperan besar dalam perkembangan berbagai penyakit metabolik seperti diabetes tipe 2, hipertensi, resistensi insulin, *dislipidemia*, penyakit kardiovaskular, hingga kanker [17]-[20]. Dampak ini dapat memicu timbulnya penyakit berat lainnya, termasuk penyakit hati berlemak, gangguan neurodegeneratif,

serta kondisi kardiometabolik lainnya [21]-[23]. Temuan tersebut menunjukkan bahwa obesitas memiliki mekanisme *patofisiologis* yang kompleks dan saling berkaitan dengan berbagai penyakit kronis.

Dengan meningkatnya prevalensi obesitas dan keterbatasan metode konvensional dalam memprediksi risikonya, pendekatan komputasional seperti *machine learning* (ML) semakin relevan untuk dikembangkan. ML telah digunakan untuk memprediksi, menilai, dan mengurangi risiko obesitas melalui analisis pola aktivitas fisik, kebiasaan nutrisi, serta faktor-faktor penting lainnya. Sebagai contoh, penelitian oleh Cheng *et al.* [27] menggunakan algoritma ML pada dataset NHANES untuk mengidentifikasi karakteristik individu yang berpotensi menjadi kelebihan berat badan atau obesitas. Dari 11 algoritma yang diuji, pendekatan *ensemble random subset* menghasilkan akurasi terbaik sebesar 70,01% dan AUC sebesar 63,3%. Hal ini menunjukkan peran ML yang semakin penting dalam pencegahan obesitas, terutama di negara menengah ke bawah yang memiliki keterbatasan sumber daya kesehatan [24]-[26].

Penelitian ini bertujuan untuk memperkuat basis pengetahuan dalam bidang prediksi obesitas dengan mengintegrasikan teknik seleksi fitur terbaru, yaitu SpFSR (*Simultaneous Perturbation Feature Selection and Ranking*), ke dalam ML. Melalui penerapan SpFSR, jumlah fitur dikurangi dari 17 menjadi 10 variabel utama. Pengurangan fitur dilakukan dengan mempertimbangkan keseimbangan antara kompleksitas model, interpretabilitas, serta akurasi prediksi. Pendekatan ini merujuk pada penelitian-penelitian sebelumnya yang menekankan bahwa reduksi fitur dapat meningkatkan efisiensi komputasi dan performa model [31]-[32].

Selain menurunkan risiko *overfitting*, SpFSR memastikan bahwa fitur yang dipilih relevan, signifikan, dan beragam, sehingga model lebih mampu menangkap pola-pola penting terkait obesitas. Oleh karena itu, penelitian ini menghadirkan kontribusi baru dalam peningkatan tingkat akurasi klasifikasi obesitas serta mendukung pengembangan sistem pendukung keputusan yang lebih efektif di bidang kesehatan.

1.2. Identifikasi Masalah

Mengacu pada latar belakang penelitian yang telah dipaparkan sebelumnya, masih terdapat sejumlah permasalahan yang perlu diidentifikasi. Masalah yang akan diidentifikasi dalam proposal tesis ini adalah:

1. Belum diketahui algoritma *machine learning* mana yang memberikan performa terbaik dalam mengklasifikasikan tingkat obesitas berdasarkan *dataset* yang digunakan.
2. Variasi performa antar model (*Support Vector Machine, Naïve Bayes, Decision Tree, Logistic Regression, K-Nearest Neighbors, dan Random Forest*) menyebabkan perlunya evaluasi komprehensif untuk menentukan model yang paling akurat dan stabil.
3. Adanya potensi ketidakseimbangan kelas dalam *dataset* obesitas mengharuskan pemilihan model yang tidak hanya tinggi akurasi, tetapi juga memiliki *precision, recall, dan F1-score* yang seimbang.
4. Belum diketahui peran relatif masing-masing fitur dalam memprediksi tingkat obesitas, sehingga diperlukan model yang dapat memberikan informasi tingkat kepentingan fitur.
5. Kompleksitas model dan jumlah fitur yang banyak menimbulkan kebutuhan untuk membandingkan model fitur lengkap dengan model yang menggunakan fitur hasil seleksi (SpFSR) untuk memastikan efisiensi tanpa mengorbankan performa.
6. Belum tersedia *baseline* model yang kuat untuk menilai efektivitas proses tingkat kepentingan fitur atau pengembangan model selanjutnya.

1.3. Tujuan Penelitian

Dengan mempertimbangkan permasalahan yang telah diidentifikasi sebelumnya, penelitian ini disusun dengan tujuan untuk memperoleh solusi yang sistematis. Penelitian ini bertujuan untuk mencapai beberapa hal berikut:

1. Menentukan model *machine learning* terbaik dari beberapa algoritma (*Support Vector Machine, Naïve Bayes, Decision Tree, Logistic Regression, K-Nearest Neighbors, dan Random Forest*) untuk klasifikasi tingkat obesitas.

2. Mengevaluasi performa tiap model menggunakan metrik akurasi, presisi, *recall*, dan *F1-score* untuk mendapatkan pemahaman yang menyeluruh mengenai kemampuan klasifikasi mereka.
3. Mengidentifikasi model yang paling stabil dan konsisten dalam menghadapi distribusi kelas yang tidak seimbang.
4. Menganalisis fitur yang paling berpengaruh terhadap prediksi obesitas melalui mekanisme tingkat kepentingan fitur, khususnya pada model *Random Forest*.
5. Membangun *Full Model* akhir menggunakan algoritma terbaik dan menjadikannya *baseline model* untuk dibandingkan dengan model hasil tingkat kepentingan fitur SpFSR.
6. Menilai efektivitas metode SpFSR dengan membandingkan performa model fitur lengkap dan model hasil seleksi fitur untuk memperoleh model yang optimal dan efisien.

Dengan tercapainya tujuan-tujuan tersebut, penelitian ini diharapkan mampu memberikan kontribusi akademik terhadap pengembangan metodologi *machine learning* untuk klasifikasi obesitas, serta manfaat praktis berupa *pipeline* yang dapat diterapkan dalam bidang kesehatan untuk deteksi dini dan penilaian risiko obesitas dengan tingkat interpretabilitas yang tinggi.

1.4. Ruang Lingkup Penelitian

Untuk menjaga fokus penelitian dan mencegah perluasan pembahasan di luar tujuan yang dirumuskan, penelitian ini dibatasi oleh ruang lingkup tertentu. Beberapa diantaranya yaitu:

1. Penelitian dibatasi pada penggunaan *dataset* obesitas yang telah melalui proses pembersihan data, normalisasi, dan pengolahan awal yang diperlukan sebelum tahap pemodelan.
2. Algoritma *machine learning* yang dibandingkan hanya mencakup enam model, yaitu *Support Vector Machine (SVM)*, *Naïve Bayes*, *Decision Tree*, *Logistic Regression*, *K-Nearest Neighbors (K-NN)*, dan *Random Forest*.
3. Seluruh evaluasi performa model dilakukan menggunakan metrik akurasi, *precision*, *recall*, dan *F1-score*, tanpa mempertimbangkan metrik tambahan seperti AUC-ROC atau MCC.

4. Pemilihan model terbaik dilakukan pada model dengan fitur lengkap sebelum dilakukan perbandingan dengan model berbasis fitur hasil seleksi SpFSR.
5. Proses tingkat kepentingan fitur difokuskan hanya pada metode SpFSR (*Sparse-Feature-Selection Regularization*) dan tidak membahas metode seleksi fitur lainnya.
6. Lingkup analisis interpretabilitas terbatas pada identifikasi fitur penting yang diberikan oleh algoritma *Random Forest*, tanpa menggunakan teknik interpretasi lanjutan seperti SHAP atau LIME.
7. Penelitian ini hanya berfokus pada aspek analitis dan tidak mencakup pengembangan sistem aplikasi atau implementasi model ke dalam *platform* produksi.
8. Model yang dipilih sebagai *Full Model* digunakan sebagai *baseline* evaluasi, sehingga penelitian tidak membahas *hyperparameter tuning* secara mendalam pada seluruh algoritma, kecuali jika diperlukan untuk meningkatkan stabilitas model.

1.5. Sistematika Penulisan

Proposal tesis ini disusun secara sistematis dalam tiga bab utama, yang masing-masing berkaitan erat dengan penelitian yang akan dilakukan. Format penulisan adalah sebagai berikut:

BAB I PENDAHULUAN

Bab ini memaparkan latar belakang penelitian, yang menjelaskan fenomena, masalah, dan alasan pentingnya penelitian dilakukan. Selanjutnya disajikan identifikasi masalah yang muncul dari latar belakang serta batasan atau ruang lingkup penelitian yang ditetapkan agar pembahasan tidak melebar. Bab ini juga memuat rumusan masalah yang disusun dalam bentuk pertanyaan penelitian, diikuti dengan tujuan penelitian yang menjadi jawaban dari rumusan masalah tersebut. Terakhir, disertakan sistematika penulisan yang menggambarkan susunan bab dalam laporan penelitian secara keseluruhan.

BAB II LANDASAN / KERANGKA PEMIKIRAN

Bab ini menyajikan landasan atau kerangka pemikiran berisi alur berpikir yang menjelaskan bagaimana penelitian disusun secara sistematis, mulai dari masalah yang ditemukan, teori-teori yang mendukung, hingga tahapan logis yang menghubungkan variabel dan metode yang digunakan. Bagian ini menggambarkan hubungan antar konsep, menjelaskan dasar pemilihan model atau metode penelitian, serta menunjukkan arah analisis yang akan dilakukan. Kerangka pemikiran menjadi dasar teoritis sekaligus panduan logika penelitian, sehingga pembaca dapat memahami bagaimana peneliti berpindah dari perumusan masalah menuju tujuan penelitian melalui landasan teori, metode, dan pendekatan yang dipilih.

BAB III METODOLOGI PENELITIAN

Bab ini menjelaskan langkah-langkah sistematis yang digunakan untuk mencapai tujuan penelitian. Bagian ini mencakup jenis penelitian, sumber dan karakteristik data, teknik pengumpulan serta pengolahan data, dan metode analisis. Pada penelitian ini, metodologi meliputi proses pengumpulan *dataset* obesitas, tahap pra-pemrosesan data seperti pembersihan, normalisasi, dan pemilihan fitur, serta penggunaan beberapa algoritma seperti *Support Vector Machine (SVM)*, *Naïve Bayes*, *Decision Tree*, *Logistic Regression*, *K-Nearest Neighbors (K-NN)*, dan *Random Forest*. Selanjutnya, dilakukan evaluasi performa model menggunakan metrik akurasi, presisi, *recall*, dan *F1-score* untuk menentukan model terbaik. Metodologi ini juga mencakup proses pemilihan model berbasis SpFSR serta perbandingan antara *Full Model* dan model hasil seleksi fitur.

BAB IV HASIL PENELITIAN DAN PEMBAHASAN

Bab ini berisi penyajian hasil analisis yang diperoleh dari proses pengolahan data serta interpretasi mendalam terhadap temuan tersebut. Pada bagian ini, ditampilkan hasil evaluasi performa setiap model *machine learning* yang digunakan, meliputi akurasi, *precision*, *recall*, dan *F1-score*, untuk menggambarkan kemampuan masing-masing algoritma dalam mengklasifikasikan tingkat obesitas. Setelah itu, hasil perbandingan antar model diinterpretasikan guna menentukan model terbaik, sekaligus menjelaskan alasan pemilihan *Random Forest* sebagai *Full Model* berdasarkan stabilitas dan keunggulan nilai metriknya. Bab ini juga membahas hasil analisis tingkat kepentingan fitur guna mengidentifikasi faktor-faktor yang memiliki pengaruh paling

signifikan terhadap prediksi obesitas. Selain itu, dilakukan pembahasan mengenai perbandingan performa antara *Full Model* dan model hasil seleksi fitur menggunakan SpFSR, sehingga dapat dinilai efektivitas pengurangan fitur terhadap kualitas prediksi.

BAB V PENUTUP

Bab ini memuat rangkuman temuan penelitian dan kesimpulan yang disusun dari hasil analisis serta pembahasan pada bab terdahulu. Pada bagian ini, peneliti menyampaikan poin-poin utama yang menjawab rumusan masalah dan menunjukkan pencapaian tujuan penelitian secara ringkas, jelas, dan sistematis. Selain itu, bab penutup juga memuat saran yang dapat menjadi acuan bagi peneliti berikutnya, praktisi, atau pihak yang berkepentingan terkait topik penelitian. Saran yang diberikan mencakup rekomendasi pengembangan metode, perluasan *dataset*, peningkatan akurasi model, maupun penelitian lanjutan yang dapat memperkuat temuan penelitian ini.