

BAB I

PENDAHULUAN

1.1 Latar Belakang

Obesitas merupakan masalah kesehatan masyarakat yang semakin meningkat dan dipengaruhi oleh beragam faktor yang saling berkaitan, seperti pola konsumsi, tingkat aktivitas fisik, faktor genetik, kondisi lingkungan, tingkat pendidikan, hingga status sosial ekonomi. Kompleksitas faktor-faktor tersebut menunjukkan bahwa obesitas bukanlah kondisi yang dapat dijelaskan secara sederhana. Temuan penelitian dalam beberapa tahun terakhir juga menegaskan bahwa obesitas tidak dapat dijelaskan hanya oleh satu atau dua faktor saja, melainkan merupakan hasil kombinasi dari berbagai pengaruh yang bekerja secara bersamaan [1].

Kompleksitas tersebut menghadirkan tantangan tersendiri dalam proses analisis data. Banyaknya variabel yang terlibat serta adanya hubungan non-linear dan interaksi kompleks antarvariabel membuat metode analisis tradisional, seperti regresi, sering kali tidak mampu menangkap pola hubungan secara optimal. Oleh karena itu, diperlukan pendekatan analitik yang lebih fleksibel dan mampu mengakomodasi struktur data yang rumit.

Seiring dengan perkembangan teknologi data dan peningkatan kapasitas komputasi, pendekatan berbasis machine learning kemudian menawarkan alternatif yang menjanjikan untuk mengkaji fenomena obesitas secara lebih komprehensif. Salah satu metode yang semakin banyak digunakan adalah Random Forest, yaitu algoritma yang membangun sejumlah pohon keputusan dan menggabungkan hasilnya untuk menghasilkan prediksi yang lebih stabil dan akurat [2]. Keunggulan utama Random Forest terletak pada kemampuannya mengolah data dengan banyak variabel serta tidak memerlukan asumsi statistik yang ketat sebagaimana metode regresi.

Dalam konteks penelitian obesitas, Random Forest menjadi semakin relevan karena tidak hanya berfungsi sebagai alat prediksi, tetapi juga mampu mengidentifikasi variabel yang memiliki kontribusi paling besar terhadap risiko obesitas melalui analisis *variable importance*. Informasi ini sangat bermanfaat bagi

peneliti maupun pembuat kebijakan dalam menentukan prioritas intervensi yang lebih efektif. Sejumlah studi menunjukkan bahwa Random Forest memiliki performa prediktif yang lebih baik dibandingkan metode lain, seperti Support Vector Machine dan regresi logistik, khususnya ketika data memiliki pola hubungan non-linear [3].

Selain keunggulan prediktif, Random Forest juga relatif praktis diterapkan dalam penelitian kesehatan. Metode ini mampu menangani data yang tidak lengkap, mengolah variabel kategorikal dan numerik secara bersamaan, serta relatif tahan terhadap overfitting apabila diterapkan dengan tepat. Penelitian terbaru menunjukkan bahwa Random Forest dapat mengidentifikasi faktor dominan penyebab obesitas, seperti pola makan, aktivitas fisik, dan kondisi sosial ekonomi, dengan lebih jelas dibandingkan pendekatan analisis tradisional [4].

Dalam konteks Indonesia, permasalahan obesitas menjadi semakin relevan mengingat angka kejadiannya yang terus meningkat pada berbagai kelompok usia. Kondisi ini dipengaruhi oleh perubahan pola makan, meningkatnya konsumsi makanan cepat saji, rendahnya aktivitas fisik, urbanisasi, serta minimnya edukasi gizi. Penelitian di tingkat nasional juga menunjukkan bahwa usia, jenis kelamin, pola makan, dan tingkat pendidikan memiliki hubungan yang signifikan dengan risiko obesitas [5]. Mengingat banyaknya faktor yang terlibat, penerapan Random Forest menjadi sangat potensial untuk mengidentifikasi faktor-faktor yang paling dominan memengaruhi obesitas di Indonesia.

Berdasarkan uraian tersebut, penggunaan Random Forest dalam penelitian ini dinilai sebagai pendekatan yang tepat. Dengan kemampuannya dalam menganalisis data yang kompleks serta menentukan variabel yang berpengaruh secara akurat, metode ini diharapkan dapat memberikan pemahaman yang lebih menyeluruh mengenai faktor dominan penyebab obesitas dan mendukung perumusan strategi intervensi yang lebih efektif.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, penelitian ini akan menjawab beberapa pertanyaan utama yaitu:

1. Faktor-faktor apa saja yang berpengaruh terhadap kejadian obesitas pada populasi penelitian?
2. Variabel mana yang menjadi faktor dominan atau memiliki kontribusi terbesar dalam menentukan status obesitas berdasarkan hasil analisis Random Forest?
3. Bagaimana performa metode Random Forest dalam menganalisis dan memprediksi status obesitas berdasarkan variabel-variabel yang tersedia?

1.3 Tujuan

Tujuan penelitian ini adalah sebagai berikut:

1. Menganalisis faktor-faktor yang berpengaruh terhadap obesitas berdasarkan data empiris.
2. Mengidentifikasi faktor paling dominan penyebab obesitas menggunakan metode *Random Forest*.
3. Menguji tingkat akurasi model Random Forest dalam memprediksi risiko obesitas.

1.4 Manfaat

Adapun manfaat yang dapat dihasilkan dari penelitian ini adalah:

4. Memberikan gambaran faktor utama penyebab obesitas berdasarkan hasil analisis menggunakan algoritma *Random Forest* sebagai dasar upaya pencegahan dan pengendalian obesitas.
5. Menjadi acuan bagi penelitian selanjutnya yang ingin menggunakan pendekatan *machine learning* untuk analisis masalah kesehatan.

1.5 Batasan Masalah

Ruang lingkup penelitian ini dibatasi pada beberapa hal berikut:

1. Penelitian hanya menganalisis faktor-faktor yang berpengaruh terhadap obesitas berdasarkan variabel yang tersedia dalam dataset.
2. Metode yang digunakan dibatasi pada algoritma Random Forest sebagai teknik utama dalam proses analisis dan pemodelan.
3. Data yang digunakan berasal dari satu sumber dataset, sehingga hasil penelitian hanya mencerminkan kondisi sesuai data tersebut.
4. Penelitian tidak membahas aspek klinis atau intervensi medis, melainkan hanya berfokus pada analisis faktor dan pembuatan model prediksi.
5. Hasil penelitian dibatasi pada identifikasi faktor dominan dan evaluasi akurasi model, tanpa melakukan perbandingan dengan algoritma lain.

1.6 Keterbaruan

Terdapat beberapa penelitian sebelumnya yang berkaitan dengan penerapan algoritma Random Forest dalam memprediksi obesitas serta menganalisis faktor-faktor yang memengaruhi risiko obesitas. Berikut adalah ringkasan penelitian terdahulu:

1. [6] meneliti faktor-faktor yang memengaruhi status gizi anak (underweight, overweight, obesity) menggunakan berbagai algoritma *ensemble*. Hasilnya menunjukkan bahwa Random Forest memiliki performa terbaik untuk memprediksi status obesitas anak pada usia 3–14 tahun.
2. [7] Romo-Muñoz et al. (2024) mengombinasikan Random Forest dengan pendekatan ekonometrik untuk menganalisis determinan overweight dan obesity pada populasi dewasa. RF terbukti mampu menangkap pola non-linier serta mengidentifikasi variabel paling berpengaruh melalui *variabel importance*.
3. [8] Alanazi et al. (2024) membandingkan Random Forest dengan model regresi dan decision tree dalam memprediksi usia munculnya obesitas anak. Random Forest menunjukkan akurasi prediksi tertinggi dan menyoroti faktor-faktor signifikan yang memengaruhi onset obesitas.

4. [9] Shin et al. (2023) membangun model prediksi sindrom metabolik (kondisi yang sangat berkorelasi dengan obesitas) menggunakan metode ML, termasuk Random Forest. Studi ini menunjukkan bahwa RF efektif memprediksi risiko kesehatan tanpa data invasif, sehingga relevan untuk penelitian obesitas berbasis kuesioner.
5. [10] Mohan et al. (2024) menggunakan beberapa metode ML, termasuk Random Forest, untuk mengevaluasi faktor-faktor penyebab overweight dan obesitas pada remaja Korea. Temuan menunjukkan bahwa Random Forest berhasil mengidentifikasi fitur-fitur paling penting terkait gaya hidup dan faktor lingkungan.
6. [11] Peng et al. (2025) memanfaatkan Random Forest dan XGBoost dengan teknik interpretabilitas SHAP untuk menganalisis faktor perilaku yang dapat dimodifikasi pada remaja obesitas. Pendekatan ini membantu meningkatkan transparansi model dan menunjukkan bahwa RF dapat menjelaskan faktor dominan secara jelas.
7. [12] Muliawan et al. (2024) menerapkan Random Forest pada dataset lokal di Indonesia untuk memprediksi risiko obesitas berdasarkan pola makan. Hasil penelitian menunjukkan bahwa Random Forest memiliki akurasi paling baik dibanding KNN dan Decision Tree.
8. [13] Utiahman & Pratama (2024) mengidentifikasi bahwa faktor gaya hidup dan pola makan, seperti aktivitas fisik, konsumsi makanan cepat saji, frekuensi konsumsi sayuran, dan asupan kalori, memiliki pengaruh signifikan terhadap tingkat obesitas.