

# BAB I

## PENDAHULUAN

### 1.1 Latar belakang masalah

Obesitas telah menjadi salah satu masalah kesehatan global yang paling kompleks dan berdampak luas pada abad ke-21[1]. Berbagai studi kesehatan global menunjukkan bahwa prevalensi obesitas dunia terus meningkat secara signifikan dalam dua dekade terakhir dan berkaitan erat dengan adaptasi gaya hidup modern, khususnya pola asupan tinggi energi dan rendah intensitas kegiatan fisik [2]. Pendekatan berbasis data gaya hidup dengan metode *machine learning* telah banyak digunakan untuk memahami pola dan risiko obesitas secara lebih komprehensif [3]. Dan pengembangan model prediksi berbasis *machine learning* telah terbukti memberikan hasil optimasi yang signifikan dalam berbagai bidang klasifikasi dan peramalan data yang kompleks[4]. Selain itu, obesitas terbukti berdampak pada penurunan produktivitas dan peningkatan risiko penyakit kronis pada populasi usia kerja, sebagaimana ditunjukkan pada studi populasi dewasa di Asia Tenggara [5]. Perkembangan sistem prediksi dan visualisasi risiko obesitas berbasis *machine learning* semakin menegaskan pentingnya pendekatan berbasis data dalam mendukung kebijakan kesehatan preventif [6].

Indonesia mengalami tren peningkatan obesitas yang konsisten dalam beberapa tahun terakhir, khususnya pada penduduk usia dewasa dan usia produktif[7]. Penelitian di Indonesia menunjukkan bahwa pemanfaatan *machine learning* untuk prediksi obesitas memberikan hasil yang bervariasi tergantung pada algoritma yang digunakan, sehingga pemilihan model menjadi faktor krusial [8]. Selain itu, perbandingan antara algoritma *machine learning* dan *deep learning* menunjukkan perbedaan performa yang signifikan dalam mengklasifikasikan tingkat obesitas [9]. Studi lain menegaskan pentingnya evaluasi komparatif antar-model untuk memperoleh model prediksi obesitas yang optimal dan aplikatif [10]. Kondisi ini menunjukkan bahwa pendekatan prediktif berbasis analisis risiko individual masih sangat diperlukan untuk mendukung upaya pencegahan obesitas di Indonesia [11].

Faktor gaya hidup dan antropometri merupakan determinan utama dalam terjadinya obesitas dan menjadi dasar penting dalam pemodelan prediksi. Pola makan tidak seimbang, kebiasaan sedentari, dan minim kegiatan fisik berperan secara relevan

terhadap kenaikan risiko obesitas [12]. Dalam konteks pemodelan, Logistic Regression dan XGBoost merupakan dua algoritma yang sering diterapkan dan dibandingkan pada penelitian prediksi obesitas [13]. Logistic Regression mempunyai keistimewaan dalam interpretabilitas dan kemudahan analisis faktor risiko, namun mempunyai limitasi dalam mengatasi relasi non-linear antarvariabel [14]. Berbagai penelitian juga menegaskan bahwa aktivitas fisik dan kebiasaan nutrisi merupakan variabel penting dalam pemodelan prediksi tingkat obesitas berbasis machine learning [15].

Selain itu, kombinasi fitur gaya hidup dan karakteristik fisik terbukti mampu meningkatkan kemampuan model dalam mengklasifikasikan tingkat obesitas secara lebih akurat [16]. Perkembangan kerangka kerja machine learning yang mampu menangani data kesehatan yang kompleks dan berdimensi tinggi semakin memperkuat relevansi pendekatan ini dalam penelitian obesitas [17]. Studi lain menunjukkan bahwa aktivitas fisik memiliki pengaruh signifikan dalam klasifikasi tingkat obesitas menggunakan berbagai algoritma machine learning [18]. Untuk meningkatkan kinerja model, optimasi fitur juga menjadi aspek penting karena terbukti mampu mengurangi kompleksitas data dan meningkatkan performa prediksi [19].

Penelitian terdahulu menunjukkan bahwa algoritma ensemble seperti XGBoost memiliki performa yang unggul dalam memodelkan risiko obesitas berbasis pola hidup [20]. Namun, sebagian besar penelitian tersebut masih berfokus pada satu algoritma dan belum banyak menekankan evaluasi komparatif dengan mempertimbangkan aspek interpretabilitas model. Maka penelitian ini dirancang guna mengintegrasikan aspek gaya hidup dan antropometri serta mengkomparasi Logistic Regression dan XGBoost dalam prediksi tingkat obesitas. Penelitian ini diharapkan mampu menghasilkan model prediksi yang tepat dan interpretatif, serta menjadi referensi metodologis bagi pengembangan studi prediksi obesitas di masa mendatang .

## **1.2 Rumusan Masalah**

Peningkatan prevalensi obesitas menuntut pendekatan prediktif berbasis data karena metode statistik konvensional memiliki keterbatasan dalam menangkap hubungan kompleks antarvariabel. Oleh karena itu, efektivitas algoritma Logistic Regression dan XGBoost dalam hal memprediksi tingkat obesitas berbasis faktor gaya hidup dan antropometri perlu dievaluasi dengan komparatif.

Berdasarkan uraian tersebut, rumusan masalah penelitian ini adalah

1. Bagaimana karakteristik faktor gaya hidup dan antropometri dalam memprediksi tingkat obesitas?
2. Bagaimana kinerja model prediksi tingkat obesitas berbasis Logistic Regression dan XGBoost berdasarkan faktor gaya hidup dan antropometri?
3. Bagaimana hasil evaluasi kinerja model Logistic Regression dan XGBoost dalam mengklasifikasikan tingkat obesitas berdasarkan metrik evaluasi yang digunakan?
4. Model prediksi manakah yang memiliki kinerja paling optimal dalam mengklasifikasikan tingkat obesitas berdasarkan hasil evaluasi Logistic Regression dan XGBoost?

## **1.3 Batasan Masalah**

Penelitian ini memiliki beberapa batasan masalah, di antara lain:

1. Batasan Topik dan Metode Penelitian

Berfokus pada pemodelan prediksi tingkat obesitas menggunakan algoritma Logistic Regression dan XGBoost, tanpa membahas aspek klinis, intervensi medis, maupun analisis kausalitas.

2. Batasan Data Penelitian

Menggunakan data sekunder yang memuat faktor gaya hidup dan antropometri, tanpa melibatkan pengumpulan data primer.

3. Aspek yang Diteliti

Penelitian ini mencakup pra-pemrosesan data, pembangunan model prediksi, serta evaluasi kinerja menggunakan metrik klasifikasi.

#### 4. Aspek yang Tidak Diteliti

Penelitian ini tidak membahas aspek klinis obesitas, hubungan sebab-akibat, pengembangan aplikasi, maupun algoritma machine learning selain Logistic Regression dan XGBoost.

### **1.4 Tujuan Penelitian**

Penelitian ini dimaksud untuk mengkaji karakteristik faktor gaya hidup dan antropometri dalam pemodelan prediksi tingkat obesitas. Selain itu, penelitian ini bertujuan untuk menganalisis serta membandingkan kinerja model Logistic Regression dan XGBoost berdasarkan faktor gaya hidup dan antropometri guna menentukan model prediksi tingkat obesitas yang paling optimal.

### **1.5 Manfaat Penelitian**

Penelitian ini memiliki beberapa manfaat, di antara lain:

#### 1. Kontribusi Teoretis

Memperkaya literatur di bidang data science kesehatan terkait penerapan dan perbandingan metode statistik klasik dan machine learning dalam prediksi obesitas.

#### 2. Manfaat Metodologis

Menjadi referensi ilmiah dalam pemilihan model prediksi obesitas berbasis faktor gaya hidup dan antropometri.

#### 3. Manfaat Praktis

Memberikan dasar awal bagi pengembangan sistem skrining obesitas berbasis data di bidang kesehatan masyarakat.

#### 4. Manfaat Kebijakan

Mendukung pengambilan keputusan berbasis data dalam perencanaan strategi pencegahan obesitas.