

BAB I

PENDAHULUAN

Bab ini menjelaskan mengenai latar belakang, rumusan masalah, tujuan dan manfaat, serta ruang lingkup tugas akhir mengenai Implementasi Algoritma *Light Gradient Boosting Machine Classifier* untuk Prediksi Risiko Obesitas.

1.1 Latar Belakang

Menurut WHO, obesitas adalah suatu penyakit yang disebabkan oleh penimbunan lemak di dalam tubuh yang dapat meningkatkan risiko penyakit lainnya, seperti diabetes, masalah jantung, serta tulang. WHO juga menjelaskan bahwa orang dengan *body mass index* (BMI) yang bernilai lebih dari 30 dapat dikategorikan sebagai orang yang terkena obesitas. Pada tahun 2019, terdapat sebanyak 5 juta kematian karena penyakit tidak menular disebabkan oleh BMI di atas rata-rata (WHO, 2023). Berdasarkan studi yang dilakukan oleh NCD Risk Factor Collaboration (NCD-RisC), ditemukan bahwa sejak tahun 1990 hingga tahun 2022, masalah berat badan meningkat sebanyak dua kali pada dewasa dan empat kali pada anak-anak. Tercatat bahwa pada tahun 2022, sebanyak 43% orang dewasa di seluruh dunia kelebihan berat badan (Phelps dkk., 2023).

Penelitian yang dilakukan oleh Al-Jawaldeh dan Abbas (2022) membuktikan bahwa gaya hidup dan diet yang sangat berpengaruh terhadap kesehatan. Dengan gaya hidup dan diet yang buruk, risiko obesitas pada seseorang akan semakin tinggi. Obesitas tersebut kemudian dapat menyebabkan munculnya penyakit-penyakit lain, seperti darah tinggi, dan akhirnya berujung pada kematian.

Jindal dkk. (2018) telah berhasil melakukan prediksi terhadap risiko obesitas pasien berdasarkan gaya hidup dan diet dengan algoritma *Random Forest* dan mendapatkan nilai akurasi sebesar 89,68%. Selain itu, Mondal dkk. (2023) juga mendalami permasalahan prediksi risiko obesitas ini menggunakan algoritma *Random Forest* dan *ADA Boosting* dan berhasil mendapatkan nilai akurasi sebesar 70,3% dengan *Random Forest* dan 70,3% dengan *ADA Boosting*.

Penelitian yang dilakukan oleh Jindal dkk. (2018) masih memiliki keterbatasan, yaitu tidak adanya *feature selection* dan penggunaan algoritma *ensemble learning* dengan waktu komputasi yang cukup lama. Keterbatasan ini dapat menyebabkan nilai akurasi yang tidak optimal dan tingginya kompleksitas model sehingga algoritma tidak cocok jika diterapkan pada *dataset* yang besar. Tentunya hal ini dapat berpengaruh pada kinerja model secara keseluruhan. Untuk mengatasi keterbatasan pada penelitian sebelumnya, dapat ditangani dengan adanya *feature selection* untuk menilai atribut mana yang memiliki korelasi atau pengaruh terhadap target serta penggunaan algoritma *ensemble learning* dengan kompleksitas komputasi yang rendah sehingga dapat memangkas waktu yang dibutuhkan untuk melakukan prediksi. Penelitian yang dilakukan oleh Mondal dkk. (2023) memiliki keterbatasan dimana *dataset* yang kecil dan tidak seimbang menyebabkan akurasi yang kurang optimal. Keterbatasan ini dapat diatasi dengan penerapan *data balancing* serta penggunaan algoritma yang *robust* dan dapat meng-handle *dataset* yang kurang baik.

Gradient Boosting Machine (GBM) adalah kumpulan teknik *machine learning* yang unggul dan kuat karena dapat diaplikasikan berdasarkan kebutuhan aplikasi. Hal ini disebabkan karena adanya penggabungan dari beberapa model (*ensemble*) atau *loss function* yang berbeda. GBM mampu memberikan *insight* tambahan pada model, sehingga menghasilkan analisis dan investigasi yang lebih spesifik (Natekin & Knoll, 2013).

Namun, penelitian yang dilakukan oleh Natekin dan Knoll (2013) menunjukkan bahwa algoritma GBM membutuhkan memori dan waktu yang cukup signifikan dalam penerapannya karena algoritma tersebut akan melakukan beberapa iterasi untuk membuat dan mengevaluasi masing-masing *learner* pada *ensemble*. Maka, walaupun penggunaan GBM terbukti lebih akurat, terdapat suatu *tradeoff* antara kompleksitas model dengan keakuratan tersebut.

Untuk mengatasi permasalahan tersebut, Ke dkk. (2019) membuat suatu algoritma baru yang dapat mengurangi waktu komputasi yang diperlukan oleh algoritma GBM biasa. *Light Gradient Boosting Machine* (LightGBM) adalah suatu *framework open-source* yang dibuat oleh Microsoft berdasarkan penelitian oleh Ke dkk. (2019). LightGBM menggunakan algoritma berbasis pohon dan umumnya

digunakan untuk masalah klasifikasi, ranking, dan lainnya (Microsoft, 2016). Berdasarkan penelitian yang dilakukan Abou Omar (2018), pohon yang dibentuk oleh algoritma LightGBM terbentuk secara *leaf wise* sehingga dapat meminimalkan *loss function* dan menghasilkan akurasi yang lebih tinggi tanpa adanya *tradeoff* dengan kompleksitas komputasi.

Dalam beberapa penelitian, *Light Gradient Boosting Machine* mampu memprediksi berbagai jenis penyakit dengan tepat dan memiliki nilai akurasi yang tinggi. Pada penelitian yang dilakukan Rufo dkk. (2021), *LightGBM Classifier* mampu memprediksi penyakit diabetes melitus dengan akurasi sebesar 98,1%, jauh lebih baik dibandingkan dengan algoritma lainnya yang diuji. Penelitian yang dilakukan oleh Zhang dan Gong (2020) juga membuktikan bahwa penggunaan *LightGBM Classifier* dapat menghasilkan nilai akurasi yang lebih tinggi dibandingkan algoritma *gradient boosting* lainnya, di mana *LightGBM Classifier* mendapat nilai akurasi sebesar 76%, XGBoost 75%, dan algoritma lainnya hanya sekitar 67%.

Berdasarkan uraian yang telah dijelaskan, peneliti ingin membuat dan menganalisis hasil suatu program yang menerapkan algoritma *Light Gradient Boosting Machine Classifier* untuk memprediksi risiko obesitas dengan label *output* tidak obesitas dan obesitas.

1.2 Rumusan Masalah

Berdasarkan pada latar belakang yang telah dijelaskan, dapat dirumuskan beberapa permasalahan yang dapat diambil sebagai berikut.

1. Bagaimana mengimplementasikan algoritma *Light Gradient Boosting Machine Classifier* untuk memprediksi risiko obesitas.
2. Bagaimana kinerja algoritma *Light Gradient Boosting Machine Classifier* untuk memprediksi risiko obesitas.

1.3 Tujuan dan Manfaat

Tujuan dari penelitian skripsi ini adalah mengimplementasikan algoritma *Light Gradient Boosting Machine Classifier* untuk memprediksi risiko obesitas serta menilai

kinerja algoritma *Light Gradient Boosting Machine Classifier* dalam penerapannya untuk memprediksi risiko obesitas.

Manfaat yang diharapkan adalah dapat diperoleh suatu model yang dapat memprediksi risiko obesitas dengan akurat sehingga dapat digunakan oleh masyarakat umum dalam lingkup kesehatan untuk memprediksi risiko obesitas dan memonitor kesehatan.

1.4 Ruang Lingkup

Ruang lingkup pengimplementasian algoritma *Light Gradient Boosting Machine Classifier* untuk memprediksi risiko obesitas adalah *dataset* yang digunakan adalah data estimasi tingkat obesitas untuk rentang usia 14 sampai 61 tahun yang diperoleh dari Kaggle.

1.5 Sistematika Penulisan

Sistematika penulisan yang digunakan dalam skripsi ini terbagi dalam beberapa pokok bahasan, yaitu:

1. BAB I PENDAHULUAN

Bab ini membahas latar belakang, rumusan masalah, tujuan dan manfaat, ruang lingkup, dan sistematika penulisan skripsi mengenai implementasi algoritma *Light Gradient Boosting Machine Classification* untuk memprediksi obesitas.

2. BAB II LANDASAN TEORI

Bab ini memberikan kajian Pustaka yang berhubungan dengan tema skripsi sebagai landasan untuk perumusan dan analisis permasalahan pada skripsi. Kajian pustaka yang digunakan meliputi obesitas, kategori obesitas, klasifikasi, tahapan *preprocessing* data, metode *gradient boosting*, *Light Gradient Boosting Machine*, Python, metrik evaluasi sistem, serta *tools* dan *library* yang digunakan dalam penelitian.

3. BAB III METODOLOGI PENELITIAN

Bab ini menjelaskan mengenai tahapan dalam penyelesaian masalah skripsi. Tahapan tersebut meliputi *data input*, *preprocessing*, algoritma *Light Gradient Boosting Machine Classification*, *confusion matrix*, serta analisis dan penilaian hasil prediksi.

4. BAB IV HASIL DAN PEMBAHASAN

Bab ini menyajikan hasil dan pembahasan dalam penerapan algoritma *Light Gradient Boosting Machine Classification* untuk memprediksi obesitas. Hasil dan pembahasan terdiri dari hasil implementasi model serta analisis dari hasil implementasi.

5. BAB V PENUTUP

Bab ini berisi tentang kesimpulan dari penulisan tugas akhir dan saran untuk pengembangan selanjutnya.