

BAB I

PENDAHULUAN

Bab ini membahas mengenai latar belakang, rumusan masalah, manfaat dan tujuan, ruang lingkup, dan sistematika penulisan skripsi terkait Implementasi Algoritma *LightGBM* Untuk Klasifikasi Obesitas.

1.1 Latar Belakang

Pada saat ini, obesitas merupakan salah satu masalah serius yang sedang dihadapi oleh banyak negara di seluruh dunia. Data menunjukkan angka kejadian obesitas terus meningkat dan ditemukan pada beberapa kelompok umur baik dari usia anak-anak hingga dewasa, bahkan pada usia lanjut Hermawan et al. (2015) Menurut *World Health Organization* (WHO), obesitas merupakan suatu penyakit kompleks yang ditandai dengan timbunan lemak yang berlebihan yang dapat mengganggu kesehatan. Obesitas dapat mempengaruhi penyakit lainnya seperti peningkatan risiko diabetes tipe 2, penyakit jantung, kesehatan tulang dan reproduksi, serta dapat meningkatkan risiko kanker. Diagnosis kelebihan berat badan dan obesitas dapat dilihat dengan menghitung *Body Mass Index* (BMI) yaitu dengan membagi berat badan (Kg) dengan tinggi badan (m), BMI juga merupakan salah satu penanda yang dapat membantu melakukan diagnosis terhadap obesitas. Pada tahun 2022, 2,5 miliar orang dewasa berusia diatas 18 tahun keatas mengalami kelebihan berat badan, termasuk 890 juta orang dewasa yang menderita obesitas. Angka ini setara dengan 43% orang dewasa berusia 18 tahun keatas yang kelebihan berat badan. Selain itu, pada tahun 2022 diperkirakan terdapat 37 juta anak dibawah usia 5 tahun mengalami kelebihan berat badan, selain itu 390 juta anak-anak dan remaja diantara umur 5-19 tahun mengalami kelebihan berat badan pada tahun 2022.

Menurut Hermawan et al. (2015) obesitas merupakan suatu penyakit yang tidak boleh dianggap remeh, karena obesitas adalah awal dari beberapa penyakit *degenerative* yang sangat mematikan seperti penyakit kardiovaskular, jantung, *stroke*, hipertensi, *diabetes mellitus*, *cancer*, *alzheimer*, *sleep apnea*, dan lain-lain, sehingga tidak jarang ditemukan obesitas menjadi salah satu faktor utama penyebab kematian dini atau kematian pada usia muda. Selain pada gangguan kesehatan, obesitas juga dapat menimbulkan gangguan psikis terhadap penderitanya. Pada banyak penelitian yang telah dilakukan menyatakan bahwa penderita

obesitas memiliki tingkat kepercayaan diri yang rendah, sebagai contoh adalah penelitian yang dilakukan oleh Sumiyati & Irianti (2021) menjelaskan bahwa sebagian besar siswa-siswi pada SMP Negeri 2 Pagetan yang mengalami obesitas tipe I memiliki *p-value* 0,03 yang dapat diartikan bahwa siswa-siswi dengan tingkat obesitas tipe 1 memiliki tingkat harga diri yang rendah sehingga dapat menyebabkan gangguan psikologis seperti depresi dan kurang percaya diri.

Telah banyak riset dalam bidang kedokteran yang telah dilakukan untuk mengklasifikasi penyakit, dan salah satu metode yang paling populer adalah *machine learning*. *Machine learning* memungkinkan komputer untuk mempelajari data dan melakukan klasifikasi atau prediksi berdasarkan informasi yang diperoleh Russel & Norvig (2016).

Beberapa penelitian telah dilakukan dengan menggunakan metode klasifikasi untuk mengidentifikasi individu yang memiliki risiko terkena penyakit obesitas. Sebagai contoh, penelitian yang dilakukan oleh Wie & Siddik (2022) dalam mengklasifikasi obesitas pada pria memperoleh tingkat akurasi 84,15% dengan menggunakan algoritma *Naïve Bayes*. Selain itu, keberhasilan metode klasifikasi dalam melakukan identifikasi pada obesitas juga telah dijelaskan pada penelitian yang dilakukan oleh Dewi & Dwidasmara (2020) dalam melakukan klasifikasi obesitas pada dataset *obesity levels* memperoleh tingkat akurasi tertinggi sebesar 78,98% dengan menggunakan algoritma *K-Nearest Neighbor* (KNN). *LightGBM* termasuk kedalam metode *gradient boosting* berbasis *decision tree* yang memiliki kemampuan untuk memproses data secara cepat, terdistribusi dan memiliki *high performance*, selain itu *LightGBM* merupakan salah satu metode *ensemble learning* yang melakukan prediksi dari beberapa *decision tree*. Sejumlah penelitian telah mengungkapkan bahwa algoritma *LightGBM* sangat cocok digunakan untuk melakukan klasifikasi pada dataset dengan akurasi yang tinggi. Sebagai contoh pada penelitian yang dilakukan oleh Rufo et al. (2021) dalam melakukan klasifikasi pada penyakit diabetes mellitus penggunaan algoritma *LightGBM* menghasilkan tingkat akurasi sebesar 98,1%. Penelitian lain yang dilakukan oleh Wardhana et al. (2022) dalam melakukan klasifikasi pada kacang kering dengan menggunakan algoritma *Gradient Boosting Machine*, *Random Forest*, dan *Light GBM* tingkat akurasi yang didapat oleh algoritma *Light GBM* mencapai 99,9%. Namun, terdapat permasalahan terkait penurunan pada parameter *learning rate*, namun pada parameter *estimator* mampu mencapai tingkat akurasi tersebut.

LightGBM termasuk kedalam metode *gradient boosting* berbasis *decision tree* yang memiliki kemampuan untuk memproses data secara cepat, terdistribusi dan memiliki *high performance*, selain itu *LightGBM* merupakan salah satu metode *ensemble learning* yang

melakukan prediksi dari beberapa *decision tree*. Sejumlah penelitian telah mengungkapkan bahwa algoritma *LightGBM* sangat cocok digunakan untuk melakukan klasifikasi pada dataset dengan akurasi yang tinggi. Sebagai contoh pada penelitian yang dilakukan oleh Rufo et al. (2021) dalam melakukan klasifikasi pada penyakit diabetes mellitus penggunaan algoritma *LightGBM* menghasilkan tingkat akurasi sebesar 98,1%. Penelitian lain yang dilakukan oleh Wardhana et al. (2022) dalam melakukan klasifikasi pada kacang kering dengan menggunakan algoritma *Gradient Boosting Machine*, *Random Forest*, dan *Light GBM* tingkat akurasi yang didapat oleh algoritma *Light GBM* mencapai 99,9%. Namun, terdapat permasalahan terkait penurunan pada parameter *learning rate*, namun pada parameter *estimator* mampu mencapai tingkat akurasi tersebut.

Pada penelitian yang dilakukan kali ini, terdapat beberapa perpustakaan yang digunakan seperti, *NumPy*, *Pandas*, *Matplotlib*, *seaborn*, *SciPy*, *warnings* dan *Scikit-learn*. Setiap *library* memiliki peran unik tersendiri dalam mengelola data, memanipulasi data, serta mengimplementasi dan evaluasi model klasifikasi. Selain itu, dalam penelitian yang dilakukan kali ini, *tools* utama yang digunakan yaitu *Jupyter Notebook*, sebagai suatu wadah interaktif dalam melakukan penulisan serta menjalankan kode untuk membuat model, dan *Kaggle*, sebagai sumber didatakannya data dan akses dataset yang relevan dengan penelitian yang dilakukan.

Berdasarkan permasalahan yang telah diuraikan, maka penelitian ini melakukan klasifikasi tingkat obesitas menggunakan *LightGBM* dengan fitur yang ada pada dataset *obesity levels*. Dengan mengimplementasikan algoritma *LightGBM* diharapkan dapat melakukan klasifikasi serta prediksi tingkat obesitas pada dataset *obesity levels*. Selama proses pelatihan model, berbagai macam pengaturan *hyperparameter* dalam algoritma *LightGBM* dieksplorasi. Hal ini penting karena pengaturan *hyperparameter*. Kemudian hasil dari klasifikasi tersebut menghasilkan suatu nilai akurasi perhitungan dari model *LightGBM* dalam mengklasifikasi data *obesity levels*.

1.2 Rumusan Masalah

Berdasarkan latar belakang tersebut, rumusan masalah yang dijadikan landasan pada penelitian ini adalah bagaimana mengoptimalkan implementasi dan kinerja yang dihasilkan dari algoritma *LightGBM* dalam mengklasifikasi tingkat obesitas ?

1.3 Tujuan dan Manfaat Penelitian

Tujuan dari penelitian ini adalah untuk mengoptimalkan implementasi dan melakukan evaluasi kinerja dari algoritma *LightGBM Classifier* dalam mengklasifikasikan tingkat obesitas.

Manfaat yang diharapkan dari dilakukannya penelitian ini adalah diperoleh suatu model yang dapat digunakan untuk melakukan klasifikasi pada tingkat obesitas.

1.4 Ruang Lingkup

Ruang Lingkup dari implementasi algoritma *Light Gradient Boosting Machine Classifier* untuk melakukan klasifikasi pada tingkat obesitas, yaitu :

1. Dataset yang digunakan pada penelitian ini adalah dataset *Obesity Level* yang diperoleh dari *Kaggle* yang berisi data individu dari negara Mexico, Peru, dan Colombia.
2. Model *Light Gradient Boosting Machine Classifier* dibangun menggunakan Bahasa pemrograman *python* dengan memanfaatkan library *Scikit-learn*.
3. Pelatihan serta pengujian model dilakukan menggunakan *Google Collab*.
4. Evaluasi model dilakukan menggunakan *Confusion Matrix*.
5. Validasi model dilakukan menggunakan *K-Fold Cross-Validation*.
6. *Hyperparameter* yang akan dicari kombinasi terbaiknya yaitu *n_estimators*, *num_leaves*, *min_child_samples*, *learning_rate*, *colsample_bytree*, *reg_alpha*, *reg_lambda*, *n_jobs*, *max_bin* dan *verbose*.

1.5 Sistematika Penulisan

Sistematika penulisan dalam penelitian yang dilakukan secara urut dengan cara dibagi kedalam beberapa bab. Sistematika penulisan laporan penelitian tugas akhir yang dijelaskan dari masing-masing bab yaitu sebagai berikut:

BAB I PENDAHULUAN

Bab ini membahas mengenai latar belakang, rumusan masalah, manfaat dan tujuan, ruang lingkup, dan sistematika penulisan skripsi terkait Implementasi Algoritma *LightGBM* Untuk Klasifikasi Obesitas.

BAB II TINJAUAN PUSTAKA

Pada bab ini membahas hasil studi literatur terkait teori yang berhubungan dengan topik ataupun masalah yang dibahas pada penelitian dengan judul

implementasi algoritma *LightGBM* untuk klasifikasi tingkat obesitas. Pada bab ini terdiri dari beberapa subbab antar lain *literatur review*, pengertian obesitas, *obesity levels*, pengertian klasifikasi, *ensemble learning*, *LightGBM*, *balancing data*, normalisasi *min-max*, evaluasi model klasifikasi, metode *cross-validation*, *tools* dan *library*.

BAB III METODOLOGI PENELITIAN

Pada bab ini menyajikan informasi terkait metodologi penelitian yang digunakan dalam melakukan Implementasi Algoritma *LightGBM* untuk Klasifikasi Tingkat Obesitas. Tahapan yang dilakukan untuk menyelesaikan masalah pada penelitian kali ini diantaranya adalah alat dan bahan yang digunakan untuk penelitian, garis besar penyelesaian masalah, pengumpulan data.

BAB IV HASIL DAN PEMBAHASAN

Pada pembahasan ini menyajikan informasi terkait hasil penelitian yang didapatkan, skenario pelatihan, dan menyajikan pembahasan dari analisis hasil penelitian dalam implementasi algoritma *LightGBM* pada tingkat obesitas. Bagian ini menguraikan tentang skenario pelatihan dan pengujian model *LightGBM*, hasil dan analisa skenario pelatihan dan pengujian model *LightGBM*, dan analisis hasil pengujian skenario.

BAB V PENUTUP

Bab ini berisi tentang kesimpulan dari bab-bab yang telah dijelaskan dan berisi saran sebagai bahan masukan untuk pengembangan atau penelitian yang mungkin dilakukan di masa yang akan datang.