

BAB I

PENDAHULUAN

1.1. Latar Belakang

Diabetes Melitus Gestasional (DMG) merupakan salah satu komplikasi kehamilan umum yang dapat mengancam kesehatan ibu dan janin yang dikandungnya (Chan et al., 2023). Organisasi Kesehatan Dunia atau *World Health Organization* (WHO) dan Federasi Internasional Ginekologi dan Obstetri (FIGO) menyatakan bahwa diabetes melitus gestasional dapat diklasifikasikan sebagai hiperglikemia dalam kehamilan yang dapat terjadi selama periode kehamilan dan tidak diharapkan berlangsung setelah melahirkan (IDF Diabetes Atlas Committee, 2021). Sebanyak 21,3 juta atau 16,2% kelahiran global berhubungan dengan hiperglikemia dalam kehamilan dimana diabetes melitus gestasional menyumbang 86,4% dari seluruh kasus yang ada (Adli, 2021).

Perkumpulan Endokrinologi Indonesia (PERKENI) menghimpun data yang menunjukkan adanya indikasi peningkatan prevalensi diabetes melitus gestasional di berbagai wilayah dunia menjadi lebih dari 10% dari populasi ibu hamil (PB PERKENI, 2021). Hal ini diperkuat melalui data Federasi Diabetes Internasional (IDF) yang menunjukkan bahwa pada tahun 2021, terdapat peningkatan prevalensi diabetes melitus gestasional di wilayah Timur Tengah dan Afrika Utara (MENA), Asia Tenggara (SEA), Pasifik Barat (WP), dan Afrika (AFR) dengan angka sebesar 27,6%, 20,8%, 14,7%, dan 14,2% (H. Wang et al., 2022). Malaysia menjadi salah satu negara pada regional Asia Tenggara yang memiliki prevalensi diabetes melitus gestasional yang tinggi dengan angka sebesar 27,9%. Sedangkan, Indonesia mencatatkan angka prevalensi sebesar 1,9 – 3,6% pada studi yang dilakukan

sebelum tahun 2000. Kecilnya hasil yang didapat disebabkan oleh sedikitnya studi yang berkaitan dengan angka prevalensi di Indonesia hingga saat ini (PB PERKENI, 2021).

Minimnya studi yang berkaitan dengan diabetes melitus gestasional tidak lantas membuat penyakit ini dapat diremehkan. Seorang ibu yang terindikasi menderita diabetes melitus gestasional berpotensi mengidap beberapa komplikasi dikemudian hari, seperti hipertensi, dislipidemia, inflamasi kronis, penyakit ginjal kronis, disfungsi vascular, hingga diabetes tipe 2 (Saravanan et al., 2020). Sementara itu, bayi yang lahir dari ibu pengidap diabetes melitus gestasional memiliki risiko yang lebih tinggi untuk menderita komplikasi jangka pendek, seperti makrosomia, persalinan prematur, hipoglikemia neonatal, distosia bahu, hingga kesulitan pernafasan (Johns et al., 2018).

Deteksi dini merupakan suatu proses yang dilakukan sebagai langkah identifikasi awal untuk mencegah risiko yang lebih besar terjadi dikemudian hari. Informasi awal yang berkaitan dengan penderita diabetes nantinya digunakan untuk membantu proses pencegahan penyakit yang relevan (Mucholladin et al., 2021). Melalui identifikasi awal yang baik diharapkan ibu dan janin yang dikandungnya dapat mendapatkan penanganan terbaik untuk menghindari komplikasi serius dalam jangka pendek maupun jangka panjang (Bashir et al., 2019). Selain itu, pemanfaatan rekam medis elektronik dalam melakukan analisis prediktif mampu memberikan hasil yang cukup menjanjikan terhadap pelayanan klinis (Lee et al., 2020).

Oleh karena itu, pemanfaatan teknologi dilakukan melalui pengolahan data rekam medis dengan menggunakan pembelajaran mesin untuk menentukan apakah

pasien dapat termasuk dalam kategori normal atau dicurigai menderita diabetes melitus gestasional. Hal ini dijalankan melalui analisis prediktif dengan cara klasifikasi berdasarkan dengan atribut yang terdapat dalam *dataset*. *Dataset* yang digunakan berupa data pemeriksaan laboratorium dan rekam medis yang diperoleh dari Rumah Sakit Islam (RSI) Surabaya Jemursari.

Data yang diperoleh nantinya akan dilakukan pengolahan menggunakan kerangka kerja *Cross Industry Process Model for Data Mining* (CRISP-DM) dengan melalui 6 tahapan, yaitu *business understanding*, *data understanding*, *data preparation*, *modeling*, *evaluation*, dan *deployment*. CRISP-DM digunakan atas kemampuan pendekatannya yang berbasis tujuan sehingga masih digunakan secara luas untuk proyek penambangan data hingga saat ini (Solano et al., 2022). Algoritma *Random Forest* (RF) nantinya akan diaplikasikan untuk pembuatan model atas dasar kelebihanannya yang mampu memberikan hasil akhir klasifikasi yang baik dengan error yang lebih rendah, serta performa yang baik pada jumlah data yang banyak (Junus et al., 2023). *LightGBM* akan digunakan sebagai pembanding karena kemampuannya yang baik dalam mengolah data dengan kelas yang tidak seimbang (Septiana Rizky et al., 2022).

Hyperparameter Tuning menggunakan *grid search* juga akan diimplementasikan pada model. Hal ini dilakukan karena optimasi menggunakan *hyperparameter tuning* ternyata mampu memberikan performa terbaik dengan akurasi tertinggi (Olisah et al., 2022). Selain itu, skenario penanganan *class imbalance* juga diimplementasikan pada beberapa model yang dibuat untuk melihat dampak atau pengaruh terhadap performanya. Beberapa skenario akan dibuat untuk mencari model dengan performa terbaik. Evaluasi dilakukan terhadap seluruh

model yang telah dibangun menggunakan *confusion matrix* untuk melihat seberapa bagus performa model yang telah dibuat dalam melakukan klasifikasi. Pengujian penggunaan cpu dan waktu pemrosesan juga dibandingkan untuk melihat keunggulan dan kekurangan dari masing – masing model. Model dengan performa terbaik akan diimplementasikan dalam suatu sistem website dengan menggunakan bantuan Flask *micro-service*. Sehingga, sistem tersebut diharapkan mampu dalam membantu fasilitas kesehatan dalam melakukan peningkatan pelayanan terhadap pasien khususnya ibu hamil.

1.2. Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan sebelumnya maka dapat dirumuskan beberapa permasalahan pada penelitian seperti berikut:

1. Bagaimana pengaruh penanganan *class imbalance* pada performa model yang digunakan dalam melakukan klasifikasi?
2. Bagaimana hasil pengujian beban cpu dan waktu pemrosesan pada setiap model yang telah dibuat?
3. Bagaimana hasil performa model dengan algoritma *random forest* dan *lightgbm* dalam melakukan klasifikasi pada dataset penyakit diabetes melitus gestasional berdasarkan data rekam medis?
4. Bagaimana hasil performa model yang dipilih untuk melakukan klasifikasi pada sistem website yang dibuat menggunakan *Flask*?

1.3. Batasan Masalah

Adapun batasan masalah yang digunakan untuk menghindari penyimpangan pembahasan pada penelitian adalah sebagai berikut:

1. Dataset diperoleh dari rekam medis Rumah Sakit Islam Jemursari (RSI) Surabaya dengan jumlah sebanyak 270 baris dan 20 atribut terlampir.
2. Data yang digunakan berasal dari poli kandungan dengan rentang waktu 2 Januari 2023 hingga 31 Maret 2023.
3. Algoritma yang digunakan untuk pembuatan model adalah *random forest* dan *lightgbm*.
4. Proses pembuatan setiap model akan melibatkan beberapa skenario pembagian data, penanganan *class imbalance* dan penyetelan *hyperparameter*.
5. Evaluasi setiap model dilakukan menggunakan *confusion matrix*, *cpu usage*, *processing time*, dan pembuatan *learning curve*.
6. Proses pengolahan data hingga evaluasi model dilakukan menggunakan bahasa pemrograman *python*.
7. Sistem website akan dibangun dengan bantuan Flask *micro-service*.

1.4. Tujuan

Berdasarkan rumusan masalah yang telah dibuat sebelumnya maka tujuan penelitian dapat dijabarkan sebagai berikut:

1. Mengetahui seberapa besar pengaruh penanganan *class imbalance* pada performa model yang digunakan dalam melakukan klasifikasi.
2. Mengetahui keunggulan dari dua algoritma yang digunakan sebagai basis model dari sisi pengujian beban cpu dan waktu pemrosesan.
3. Mendapatkan informasi mengenai model terbaik yang dihasilkan dalam melakukan klasifikasi pada dataset diabetes melitus gestasional berdasarkan data rekam medis.

4. Mendapatkan informasi mengenai performa yang dihasilkan oleh sistem website dalam melakukan klasifikasi dengan menggunakan model terbaik.

1.5. Sistematika Penulisan

Sistematika penulisan ini berisikan mengenai bagian bagian yang akan diuraikan selama penulisan skripsi. Adapun hal ini dilakukan guna membantu penyusunan laporan supaya tidak menyimpang dari tujuan penulisan yang diharapkan. Tahapan dari penyusunan skripsi ini adalah sebagai berikut:

BAB I PENDAHULUAN

Bab Pendahuluan berisikan mengenai gambaran umum penelitian yang akan dibuat. Bagian ini memiliki komposisi latar belakang, rumusan masalah, batasan masalah, tujuan, dan sistematika penulisan.

BAB II TINJAUAN PUSTAKA

Bab Tinjauan Pustaka berisikan mengenai dasar teori dan penelitian terdahulu yang menjadi serupa dengan penelitian yang dilakukan. Dasar teori yang akan dijelaskan pada bagian ini berkaitan dengan diabetes melitus gestasional, *data mining*, *klasifikasi*, *CRISP-DM*, *random forest*, *LightGBM*, *hyperparameter tuning*, *confusion matrix* dan flask

BAB III METODOLOGI PENELITIAN

Bab Metodologi Penelitian berisikan mengenai tahapan ataupun metode penelitian yang akan digunakan. Tahapan ini dimulai dari pemahaman bisnis, pemahaman terhadap data, persiapan data, pemodelan, evaluasi, dan penyebaran.

BAB IV HASIL DAN PEMBAHASAN

Bab Hasil dan Pembahasan berisikan mengenai penjelasan ataupun pembahasan dari penelitian yang telah dilakukan. Bagian ini akan menjelaskan

mengenai hasil pengumpulan data, hasil pemahaman terhadap data, persiapan data, pemodelan, evaluasi, hingga dilakukan penyebaran.

BAB V PENUTUP

Bab Penutup berisikan mengenai kesimpulan dari hasil penelitian yang telah dilaksanakan dan saran terhadap penelitian selanjutnya.

DAFTAR PUSTAKA

Bagian ini berisikan mengenai literatur yang digunakan sebagai sumber dalam penyusunan skripsi.

LAMPIRAN

Bagian ini berisikan tentang data penunjang maupun dokumen lain yang digunakan dalam penyusunan skripsi.