

BABI PENDAHULUAN

I.1 Latar Belakang

Secara umum, Diabetes dikenal sebagai penyakit epidemi yang berdampak di hampir setiap negara, kelompok usia, dan faktor ekonomi di seluruh dunia. Berdasarkan data dari *International Diabetes Federation* (IDF) tahun 2015, diperkirakan sebanyak 415 juta jiwa mengidap penyakit diabetes. Angka ini diperkirakan juga dapat meningkat hingga 640 juta pada tahun 2040. (Papatheodorou dkk., 2018). Selain itu, persentase kematian penderita diabetes sebelum usia 70 tahun lebih tinggi terjadi di negara-negara yang memiliki penghasilan rendah dan menengah dibandingkan dengan negara-negara berpenghasilan tinggi. (Kementerian Kesehatan, 2018)

Di Indonesia sendiri, berdasarkan data dari IDF tahun 2020, dari total populasi orang dewasa sebanyak 172 juta jiwa, 10.5 juta jiwa mengidap diabetes. prevalensi penderita Diabetes tertinggi berdasarkan status pendidikan merupakan tamatan pendidikan setingkat D1/D2/D3/PT, dan untuk status pekerjaan, yang paling banyak mengidap Diabetes adalah masyarakat yang berstatus sebagai PNS/TNI/Polri/BUMN/BUMD. Rentang penderita Diabetes tersebut paling banyak berada pada usia 55-64 tahun dan 65-74 tahun. Data tersebut didapat dari hasil klasifikasi pemeriksaan gula darah pada Riskesdas 2018. (Kementerian Kesehatan, 2018)

Kementerian Kesehatan telah melakukan beberapa usaha untuk mengendalikan diabetes seperti membentuk 13.500 Pos Pembinaan Terpadu (Posbindu) agar memudahkan warga untuk melakukan *Medical Diagnosis* atau Diagnosis Medis. (Kementerian Kesehatan, 2018). Ketika merancang *software* atau perangkat lunak diagnosis medis, prediksi penyakit menjadi salah satu tugas yang memakan banyak waktu dikarenakan terdapat banyak aspek yang masing-masing harus diperiksa berdasarkan data yang ada. Teknik pembelajaran mesin atau *Machine Learning* pun digunakan sebagai alat bantu melakukan diagnosis medis. (Choudhury & Gupta, 2019)

Machine Learning adalah salah satu domain penting dalam penelitian yang memiliki tujuan untuk memprediksi dan melakukan tinjauan sistematis. (Choudhury & Gupta, 2019). Terdapat tiga kelas pembelajaran pada *Machine Learning* yaitu *Supervised Learning*, *Unsupervised Learning*, dan *Reinforcement Learning*. (Carleo dkk., 2019)

Supervised Learning digunakan untuk melakukan klasifikasi dan prediksi keluaran nilai data, *Unsupervised Learning* digunakan untuk mendapatkan pola pada sekumpulan data atau yang biasa disebut *clustering*, dan yang terakhir *Reinforcement Learning* digunakan untuk mempelajari suatu kebijakan atau membuat strategi di suatu lingkungan yang ditentukan untuk mendapatkan hasil yang paling baik. (Carleo dkk., 2019)

Machine learning berbasis *supervised learning* dapat digunakan untuk mengklasifikasikan dan mendeteksi penyakit diabetes. Beberapa penelitian sebelumnya telah dilakukan, salah satunya adalah penelitian yang dilakukan oleh Deepti Sisodia dan Dilip Singh Sisodia dengan judul *Prediction of Diabetes using Classification Algorithms*. Penelitian tersebut memprediksi diabetes dengan menggunakan algoritma klasifikasi *Naive Bayes*, *Support Vector Machine (SVM)* dan *Decision Tree* dengan hasil akurasi masing-masing algoritma 76.3% untuk *Naive Bayes*, 65.1% untuk SVM, dan 73.82% untuk *Decision Tree* (Sisodia & Sisodia, 2018)

Dari penelitian tersebut, Algoritma *AdaBoost* dan Algoritma *LightGBM* belum digunakan sebagai metode pengklasifikasian diabetes. Selain itu, kedua algoritma tersebut dikenal memiliki tingkat akurasi yang tinggi, seperti pada penelitian yang dilakukan oleh Agustin Yoga Handoko, Kusri, dan Luthfi Emha Taufiq dengan judul “Klasifikasi Penerimaan Mahasiswa Baru Menggunakan Algoritma C4.5 Dan *Adaboost*” tingkat akurasi Algoritma *AdaBoost* memiliki hasil yang sama dengan algoritma C4.5 sebesar 90.28% (Agustin dkk., 2017) dan penelitian yang dilakukan oleh Wang Bijun, Wang Yulong, Qin Kun, Xia Qizhi dengan judul “*Detecting Transportation Modes Based on LightGBM Classifier from GPS Trajectory Data*” yang membandingkan akurasi dari algoritma *Decision Tree*,

LightGBM, dan *XGBoost*. Algoritma *LightGBM* unggul dengan tingkat akurasi sebesar 90.61% (B. Wang dkk., 2018).

Berdasarkan permasalahan dan penelitian yang telah ada sebelumnya, penulis akan melakukan penelitian yang membandingkan tingkat akurasi Algoritma *AdaBoost* dengan Algoritma *LightGBM* secara spesifik dalam mengklasifikasikan penyakit diabetes.

I.2 Perumusan Masalah

Berdasarkan uraian-uraian pada latar belakang, maka dapat disimpulkan rumusan masalah yang ada yaitu:

1. Bagaimana tingkat akurasi Algoritma *AdaBoost* dalam mengklasifikasikan penyakit diabetes?
2. Bagaimana tingkat akurasi Algoritma *LightGBM* dalam mengklasifikasikan penyakit diabetes?
3. Bagaimana hasil perbandingan tingkat akurasi dari Algoritma *AdaBoost* dengan *LightGBM* dalam mengklasifikasikan penyakit diabetes?

I.3 Tujuan Penelitian

Penelitian ini bertujuan untuk:

1. Mengetahui tingkat akurasi Algoritma *AdaBoost* dalam mengklasifikasikan penyakit diabetes.
2. Mengetahui tingkat akurasi Algoritma *LightGBM* dalam mengklasifikasikan penyakit diabetes.
3. Mengetahui hasil perbandingan tingkat akurasi dari Algoritma *AdaBoost* dengan *LightGBM* dalam mengklasifikasikan penyakit diabetes.

I.4 Batasan Penelitian

Adapun batasan dalam melakukan penelitian ini adalah sebagai berikut:

1. Data yang digunakan adalah *Pima Indians Diabetes Database (PIDD)* yang didapat dari *UCI Benchmark Dataset*

2. Algoritma yang digunakan untuk perbandingan klasifikasi adalah *AdaBoost* dan *LightGBM*
3. Hasil akhir dari penelitian ini berupa perbandingan nilai akurasi dari Algoritma *AdaBoost* dengan *LightGBM* dan evaluasi model dari Algoritma *AdaBoost* dan *LightGBM*.

I.5 Manfaat Penelitian

Adapun manfaat yang didapatkan dari penelitian ini adalah sebagai berikut:

Manfaat Teoritis:

1. Mengetahui tingkat akurasi masing-masing algoritma dalam mengklasifikasikan penyakit diabetes.
2. Memahami penerapan masing-masing algoritma dalam mengklasifikasikan penyakit diabetes.

Manfaat Akademis:

1. Sebagai bahan studi dan rujukan dalam hal klasifikasi penyakit diabetes atau sebagai dokumen bagi mahasiswa Fakultas Rekayasa Industri terutama Jurusan Sistem Informasi.
2. Sebagai rujukan mengenai algoritma mana yang lebih baik digunakan pada penelitian yang akan datang untuk klasifikasi atau prediksi penyakit diabetes.

Manfaat Praktis

Dapat menghasilkan model klasifikasi untuk membuat model prediksi. Model prediksi tersebut dapat digunakan untuk membantu instansi kesehatan dalam mendeteksi dini penyakit diabetes.

I.6 Sistematika Penulisan

Penelitian ini diuraikan dengan sistematika penulisan sebagai berikut:

Bab I Pendahuluan

Pada bab ini berisi uraian mengenai konteks permasalahan, latar

belakang penelitian, perumusan masalah, tujuan penelitian, batasan penelitian, manfaat penelitian, dan sistematika penulisan.

Bab II Tinjauan Pustaka

Bab ini berisi literatur yang relevan dengan permasalahan yang diteliti dan dibahas pula hasil-hasil penelitian terdahulu. Minimal terdapat lebih dari satu metodologi/metode/kerangka kerja yang disertakan pada bab ini untuk menyelesaikan permasalahan atau meminimalisir gap antara kondisi eksisting dengan target. Pada akhir bab, analisis pemilihan metodologi/metode/kerangka kerja harus dijelaskan untuk menentukan metodologi/metode/kerangka kerja yang akan digunakan di penelitian ini.

Bab III Metodologi Penelitian

Metodologi penelitian merupakan strategi dan langkah-langkah (*plan of attack*) yang akan dilakukan di penelitian dalam rangka menjawab rumusan masalah yang disusun sebelumnya. Penyusunan metodologi penelitian harus dilakukan secara kritis apakah metode atau teknik yang dipilih memang tepat sesuai tujuan penelitian. Pada bab ini dijelaskan langkah-langkah penelitian secara rinci meliputi: tahap merumuskan masalah penelitian, merumuskan hipotesis, mengembangkan model penelitian, mengidentifikasi dan melakukan operasionalisasi variabel penelitian, menyusun kuesioner penelitian, merancang pengumpulan dan pengolahan data, melakukan uji instrumen, merancang analisis pengolahan data.

Bab IV Analisis dan Perancangan

Pada bab ini, disajikan analisis dan perancangan pada data yang akan diimplementasikan pada algoritma yang akan penulis lakukan dalam penelitian.

Bab V Implementasi dan Pengujian

Pada bab ini, disajikan hasil rancangan, temuan, analisis dan pengolahan data. Selain itu bab ini juga berisi tentang validasi atau verifikasi hasil dari penelitian, sehingga hasil tersebut apakah telah benar-benar menyelesaikan masalah atau menurunkan gap antara kondisi eksisting dan target yang akan dicapai. Analisis sensitivitas juga dapat digunakan di bab ini untuk lebih mengetahui hasil penelitian dapat diterapkan baik secara khusus di konteks penelitian maupun secara umum di konteks serupa (misal perusahaan di sektor serupa). Selain itu metode-metode evaluasi yang lain dapat diterapkan untuk memvalidasi hasil TA sesuai dengan kebutuhan.

Secara keseluruhan bab ini membahas secara mendetail mengenai hasil dari penelitian dan refleksinya terhadap tujuan penelitian. Untuk penelitian yang berfokus pada merancang sistem informasi/ aplikasi maka penamaan bab ini mengikuti tahapan penerapan SDLC yang digunakan dalam penelitian.

Bab VI Kesimpulan dan Saran

Pada bab ini dijelaskan kesimpulan dari penelitian yang dilakukan serta jawaban dari pertanyaan penelitian yang disajikan di pendahuluan. Saran penelitian dikemukakan pada bab ini untuk penelitian selanjutnya.