

BAB I

PENDAHULUAN

1.1 Latar Belakang

Demam berdarah dengue (DBD) merupakan penyakit yang disebabkan oleh virus Dengue yang ditularkan dari orang ke orang melalui gigitan nyamuk *Aedes Aegypti* [1]. Perjalanan penyakit ini diawali dengan nyamuk yang belum terinfeksi virus dengue menghisap darah orang yang telah terinfeksi virus tersebut. Kemudian nyamuk yang terinfeksi dapat mentransmisikan virus dengue tersebut ke manusia sehat yang digigitnya. Pada awal tahun 2020 *World Health Organization* (WHO) kemudian memasukan dengue sebagai salah satu ancaman kesehatan global di antara 10 penyakit lainnya [2]. Dalam 3 dekade terakhir penyakit ini meningkat insidennya di berbagai belahan dunia terutama daerah tropis dan sub-tropis banyak ditemukan di wilayah urban dan semi-urban [3]. Faktor iklim dengan curah hujan tinggi menyebabkan populasi nyamuk *aedes aegypti* berkembang pesat. Selain itu, mobilitas penduduk yang tinggi serta kepadatan penduduk yang meningkat turut berkontribusi terhadap penyebaran penyakit ini. Menerapkan Pola Hidup Bersih dan Sehat dalam setiap aktivitas sangat penting untuk menjaga kesehatan dan keselamatan. Jika PHBS tidak dijalankan, dapat berisiko terhadap kesehatan bahkan nyawa seseorang [4]. Kesadaran akan pentingnya PHBS bisa dimulai dari langkah-langkah sederhana, seperti mencegah adanya genangan air yang dapat menjadi sarang nyamuk penyebab penyakit.

Penerapan teknologi dalam dunia kesehatan semakin menjadi perhatian utama guna meningkatkan efisiensi, akurasi, dan kualitas pelayanan. Salah satu terobosan teknologi yang mendominasi era kesehatan modern adalah *Machine Learning* (ML) atau pembelajaran mesin. ML memiliki potensi besar untuk mengoptimalkan berbagai aspek operasional khususnya pada sektor kesehatan. Penelitian yang berjudul Pemanfaatan Machine Learning Di Bidang Kesehatan [4]. Yang bertujuan untuk mengulas penerapan *machine learning* dalam diagnosis dan prediksi penyakit melalui metode *literature review*. Metode analisis dilakukan terhadap berbagai

algoritma *machine learning* seperti *Logistic Regression*, *Random Forest*, *Neural Network*, *K-Nearest Neighbor* (KNN), dan *Support Vector Machine* (SVM). Hasil

penelitian mampu memberikan solusi efektif dalam diagnosis dini, prediksi penyakit, dan klasifikasi data Kesehatan, seperti diabetes, hipertensi, penyakit jantung, kanker paru-paru, gagal ginjal dan *Alzheimer*.

Penyakit berbahaya seperti DBD memiliki risiko tinggi dan harus segera ditangani dengan tepat. Jika penanganan terlambat atau terjadi kesalahan, kondisi pasien bisa memburuk dan bahkan berakibat fatal. Untuk membantu menentukan apakah seseorang terkena DBD atau tidak, teknologi seperti *machine learning*, khususnya dalam bidang klasifikasi, dapat digunakan sebagai alat pendukung dalam proses diagnosa..

Klasifikasi *machine learning* merupakan suatu Teknik pembelajaran yang dapat digunakan untuk mengelompokkan nilai sekumpulan atribut dalam mendeskripsikan kelas yang tujuannya untuk memprediksi kelas objek yang belum diketahui kelasnya. Penelitian yang dilakukan pada klasifikasi kualitas air Sungai Daerah Istimewa Yogyakarta menggunakan algoritma *random forest*, menunjukkan bahwa model yang dibuat oleh peneliti mampu mengklasifikasi kualitas air Sungai di Daerah Istimewa Yogyakarta dengan hasil akurasi yang paling tinggi adalah 97.93% [5]. Hal ini menunjukkan bahwa klasifikasi mampu mengelompokkan objek berdasarkan kelas yang dimiliki.

Dalam konteks medis, klasifikasi digunakan untuk membantu proses diagnosis dan prediksi penyakit berdasarkan data pasien. Klasifikasi penyakit merupakan salah satu pendekatan yang penting untuk mendukung diagnosis dini dan pengambilan keputusan klinis. Dalam upaya meningkatkan efektivitas diagnosis dan penanganan, diperlukan metode klasifikasi yang mampu mengidentifikasi pasien berdasarkan gejala dan faktor resiko yang ada. Penerapan klasifikasi dalam berbagai penelitian telah banyak digunakan dan terbukti memiliki hasil yang sesuai. Seperti penelitian yang telah dilakukan pada Penerapan *Algoritma Naive Bayes* untuk Klasifikasi Demam Berdarah di RSUD dr. Rumah Sakit Achmad Darwis, menghasil analisis menyatakan model *Naive Bayes* berhasil mengklasifikasikan 12 pasien menjadi positif DBD dan 22 pasien negatif DBD berdasarkan data testing.

Didapatkan hasil akurasi dari model sebesar 91,18% dimana nilai ini menunjukkan bahwa model yang dibangun sangat baik untuk mengklasifikasi pasien DBD [6]. Penelitian tentang klasifikasi penyakit DBD juga dilakukan pada Klasifikasi Diagnosa Penyakit Demam Berdarah Dengue Pada Anak Menggunakan Metode *K-Nearest Neighbor* Studi Kasus Rumah Sakit PKU Muhammadiyah Ujung Pangkah Gresik, berdasarkan hasil perhitungan *confusion matrix* untuk pengujian 150 data dan latih 75 data uji sistem ini memiliki nilai K terbaik yaitu 94,22 % pada K=3, K=5, K=7, K=4, K=6 dan K=8, nilai K terbaik dapat digunakan untuk acuan klasifikasi selanjutnya [7].

Selain klasifikasi tentang penyakit DBD, pada penelitian Klasifikasi Obesitas dengan Algoritma C5.0 Berdasarkan Pola Makan dan Kondisi fisik, dataset yang digunakan dalam penelitian ini terdiri dari 2.111 sampel dengan 17 variabel, model C5.0 yang dibangun menunjukkan hasil yang sangat baik dengan akurasi mencapai 94,78% pada data uji. Hasil ini menunjukkan bahwa algoritma C5.0 dapat menjadi alat yang efektif untuk mendukung sistem pendukung keputusan dalam mendeteksi resiko obesitas, yang pada akhirnya dapat membantu dalam pengembangan strategi pencegahan dan *intervensi* yang lebih efektif untuk meningkatkan kesehatan masyarakat [8]. Pada penelitian ini menggunakan algoritma C5.0.

Decision Tree merupakan metode yang umum digunakan dalam klasifikasi untuk menyelesaikan berbagai masalah [9]. Penelitian menunjukkan bahwa algoritma *Decision Tree* C5.0 merupakan salah satu algoritma yang sangat efektif dalam melakukan klasifikasi data. Algoritma ini dianggap sangat bermanfaat karena kemampuannya untuk menghasilkan struktur pohon keputusan yang mudah dipahami. Melalui struktur ini, hubungan antar fitur dalam data dapat dijelaskan secara jelas dan transparan. Setiap cabang dalam pohon keputusan merepresentasikan aturan *if-then* yang membantu dalam mengidentifikasi karakteristik data yang diklasifikasikan. Kejelasan dan *interpretabilitas* dari struktur pohon keputusan ini memudahkan pengguna untuk melakukan penggalian informasi dan analisis yang mendalam terhadap data [10]. Pada penelitian yang telah dilakukan pada Penerapan Algoritma C5.0 Untuk Menentukan Baik atau Buruknya Hasil Audit 5s dari hasil penelitian dapat disimpulkan bahwa dalam

mengatasi hasil audit 5S, algoritma C5.0 dapat digunakan dan akurasi yang dihasilkan adalah 98,21% [11].

Pada penelitian terkait juga membandingkan penerapan algoritma C5.0 dengan metode *machine learning* lainnya, seperti pada penelitian Perbandingan Algoritma C4.5 dan C5.0 dalam Klasifikasi Status Gizi Balita Stunting didapatkan bahwa algoritma C5.0 menghasilkan akurasi klasifikasi yang lebih tinggi dibandingkan algoritma C4.5. Algoritma C5.0 memberikan akurasi rata-rata sebesar 83%. Sedangkan algoritma C4.5 memberikan akurasi rata-rata sebesar 79%. Dapat disimpulkan bahwa algoritma C5.0 lebih efektif dalam mengklasifikasikan status gizi balita *stuntin* [12]. Pada penelitian lain pada Perbandingan Algoritma C5.0 dan Regresi Linear untuk Prediksi Kelulusan Mahasiswa, hasil perbandingan R2 score dari algoritma C5.0 mencapai 96.85% (training) dan 93.72% (testing) sedangkan hasil dari Regresi Linear mencapai 33.31% (training) dan 40.30% (testing) [13]. Pada kasus tentang klasifikasi penyakit DBD metode C5.0 juga pernah dilakukan Penerapan Algoritma C5.0 dalam Klasifikasi Faktor-faktor Penyebab Terjadinya DBD, berdasarkan analisis algoritma C5.0 terdapat lima variabel diantaranya variabel umur, jenis kelamin, riwayat pernah DBD, riwayat penyakit penyerta, serta status pekerjaan. Dari kelima variabel tersebut variabel umur merupakan atribut yang memberikan kontribusi terbesar [14]. Pada penelitian yang berjudul Klasifikasi penyakit demam berdarah dengue (DBD) menggunakan algoritma C5.0 berbasis Binary Particle Swarn Optimization (BPSO), hasil dari penelitian tersebut disimpulkan model yang lebih baik pada penelitian ini adalah klasifikasi algoritma C5.0 berbasis BPSO yang mampu mereduksi fitur dari 12 fitur menjadi 9 fitur dengan akurasi sebesar 86% dibandingkan klasifikasi dengan algoritma C5.0 saja yang menghasilkan akurasi sebesar 71% [15].

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk menerapkan algoritma *Decision Tree* C5.0 dalam klasifikasi penyakit Demam Berdarah Dengue (DBD). Algoritma C5.0 dipilih karena kemampuannya dalam menghasilkan model klasifikasi yang tidak hanya sederhana dan mudah dipahami, tetapi juga memiliki tingkat akurasi yang tinggi. Dengan menggunakan algoritma ini, diharapkan dapat dibuat sebuah model prediktif yang mampu memetakan risiko pasien secara lebih efisien dan andal berdasarkan data yang tersedia. Model ini diharapkan dapat

membantu tenaga medis dan pihak terkait untuk mendeteksi pasien yang memiliki risiko tinggi terkena DBD, sehingga langkah pencegahan atau penanganan dapat dilakukan lebih awal. Selain itu, penerapan algoritma C5.0 juga memungkinkan proses analisis data kesehatan menjadi lebih terstruktur dan sistematis, yang pada akhirnya dapat meningkatkan efektivitas pengambilan keputusan. Dengan demikian, penelitian ini berupaya memberikan kontribusi nyata dalam bidang kesehatan, khususnya dalam upaya mendukung penanganan dan pengelolaan kasus DBD secara lebih baik, baik di tingkat individu maupun populasi secara umum.

1.2 Rumusan Masalah

Berdasarkan latar belakang di atas, rumusan masalah didapatkan seperti berikut:

1.2.1 Permasalahan

Penyakit Demam Berdarah Dengue (DBD) menjadi salah satu masalah Kesehatan yang mengkhawatirkan, penyakit DBD ini perlu segera ditangani agar tidak menyebar lebih luas. Gejala yang timbul biasanya ditandai dengan demam, sakit/nyeri pada ulu hati terus-menerus, pendarahan pada hidung, mulut, gusi atau memar pada kulit. Penyakit yang berbahaya dan beresiko kematian seperti penyakit DBD harus di tangani dengan cepat dan tepat. Kesalahan dan keterlambatan penanganan dapat menyebabkan penyakit semakin parah dan bisa berujung pada kematian. Oleh sebab itu perlu adanya model klasifikasi yang dapat membantu proses diagnosis penyakit DBD ini.

Kemudian dari berbagai penelitian terkait, salah satu metode yang tepat digunakan untuk klasifikasi penyakit DBD adalah Algoritma C5.0. Algoritma C5.0 dipilih karena kemampuannya dalam menghasilkan model klasifikasi yang tidak hanya sederhana dan mudah dipahami, tetapi juga memiliki tingkat akurasi yang tinggi.

1.2.2 Metode Usulan

Peneliti menggunakan algoritma Decision Tree C5.0 sebagai proses klasifikasi penyakit DBD yang berada di Puskesmas Singgahan.

1.2.3 Pertanyaan penelitian

1. Berapa nilai akurasi yang dihasilkan oleh algoritma C5.0 dalam klasifikasi penyakit Demam Berdarah Dengue?

2. Sejauh mana penggunaan metode validasi K-Fold Cross Validation dapat meningkatkan akurasi model klasifikasi?

1.3 Tujuan dan Manfaat Penelitian

1.3.1 Tujuan Penelitian

1. Mengetahui nilai akurasi yang dihasilkan oleh algoritma C5.0 dalam klasifikasi penyakit Demam Berdarah Dengue.
2. Mengetahui peningkatan akurasi model dalam penggunaan metode validasi K-Fold cross validation.

1.3.2 Manfaat Penelitian

Adapun manfaat pada penelitian ini yaitu:

1. Pada penelitian ini membantu tenaga medis dalam mendiagnosis penyakit Demam Berdarah Dengue dengan lebih cepat dan akurat melalui model klasifikasi yang di hasilkan.
2. Hasil penelitian ini dapat digunakan untuk mendeteksi dini penyakit Demam Berdarah Dengue, sehingga pengobatan bisa dilakukan lebih awal dan dapat mengurangi resiko komplikasi atau kematian.
3. Penelitian ini memberikan kontribusi dalam pengembangan ilmu data mining, khususnya dalam penerapan algoritma C5.0 untuk klasifikasi di bidang Kesehatan, sehingga dapat menjadi refrensi untuk penelitian selanjutnya.

1.4 Batasan Masalah

Batas permasalahan pada penelitian yang akan dilakukan meliputi:

1. Penelitian ini menggunakan data yang di ambil Puskesmas Singgahan.
2. Penelitian ini menggunakan data pasien yang berjumlah 200 data dengan rentang tahun 2020 – 2024.
3. Penelitian ini menggunakan algoritma C5.0.
4. Pembagian kelas atribut pada penelitian ini dibagi menjadi 2 kategori yaitu DBD dan SEHAT.
5. Penelitian ini memiliki data 99 yang Tekena DBD dan 101 yang sehat.
6. Penelitian ini menggunakan data yang tersedia dalam dataset, yaitu jumlah trombosit, hematokrit, status serologi (IgG dan IgM), jenis kelamin, usia, serta gejala klinis seperti ruam, nyeri kepala, nyeri otot, dan demam.

7. Penelitian ini hanya fokus pada variabel-variabel yang terdapat dalam data set. Faktor lain seperti Riwayat medis atau kondisi lingkungan, tidak dimasukkan dalam analisis.

1.5 Sitematika Penulisan

Sistematika penulisan dalam penelitian ini disusun sebagai berikut.

BAB I PENDAHULUAN

Bab ini membahas latar belakang masalah yang meliputi ruang lingkup masalah, solusi yang diusulkan untuk mengatasi permasalahan, tujuan dan manfaat dari penerapan solusi tersebut, batasan – batasan, serta penyusunan proposal yang sistematis.

BAB II KAJIAN PUSTAKA

Teori – teori dasar yang digunakan dalam penelitian tugas akhir ini disajikan dalam bab ini yang meliputi DBD, klasifikasi, *preprocessing*, *K-Fold Cross Validation*, *C5.0*, evaluasi model, serta penelitian terkait.

BAB III METODOLOGI PENELITIAN

Pada penelitian ini secara mendalam mengulas tentang dataset yang digunakan sekaligus arsitektur sistem dan tahapan yang akan diterapkan dalam penelitian.

BAB IV HASIL DAB PEMBAHASAN

Pada bab ini berisi pembahasan atau penjelasan yang lebih rinci mengenai implementasi model klasifikasi DBD berdasarkan atribut – atribut yang terdapat dalam data dengan menggunakan metode *C5.0* yang telah diusulkan pada bab sebelumnya dan hasil pengujian dari model yang dihasilkan. Pengujian dilakukan untuk mengukur seberapa besar akurasi dalam mengklasifikasikan penyakit DBD.

BAB V PENUTUP

Bab ini menyajikan rangkuman atau kesimpulan dari implementasi yang telah dilakukan, serta saran rekomendasi agar penelitian ini bisa dikembangkan pada penelitian mendatang