

BAB I

PENDAHULUAN

A. Latar Belakang

Demam Berdarah *Dengue* (DBD) merupakan salah satu penyakit menular yang masih menjadi permasalahan serius di negara tropis seperti Indonesia. Diagnosis DBD umumnya dilakukan melalui observasi gejala klinis yang khas seperti demam tinggi, nyeri sendi, dan ruam kulit. Namun, gejala tersebut sering kali serupa dengan penyakit infeksi lainnya, sehingga diperlukan pemeriksaan laboratorium sebagai pendukung diagnosis. Pemeriksaan laboratorium yang paling relevan adalah pemeriksaan darah, karena komponen darah seperti trombosit, leukosit, hemoglobin, dan hematokrit mengalami perubahan signifikan akibat infeksi virus *dengue* [1].

Beberapa penelitian menunjukkan bahwa sekitar 63,3% pasien DBD mengalami leukopenia dan 90% mengalami trombositopenia. Peningkatan hematokrit menjadi indikator kebocoran plasma, sedangkan penurunan hemoglobin dapat menandakan adanya perdarahan internal. Oleh karena itu, pemantauan parameter hematologi sangat penting tidak hanya untuk memastikan diagnosis DBD, tetapi juga untuk menentukan tingkat keparahan penyakit. Perbedaan antara *Dengue Fever* (DF) dan *Dengue Hemorrhagic Fever* (DHF) pun dapat diidentifikasi melalui fluktuasi parameter darah tersebut[2].

Sayangnya, di Indonesia proses diagnosis masih sangat bergantung pada pemeriksaan manual oleh dokter yang memiliki keterbatasan waktu dan risiko *human error*. Kesalahan dalam interpretasi hasil hematologi dapat berakibat fatal bagi pasien jika penanganan tidak segera dan tepat. Salah satu contoh kasus terjadi di Jakarta, di mana dua pasien Demam Berdarah *Dengue* (DBD) meninggal dunia akibat salah diagnosis oleh pihak rumah sakit. Dalam kasus ini, kedua pasien sempat didiagnosis hanya mengalami tifus, padahal keduanya menunjukkan gejala khas DBD yang tidak segera tertangani secara optimal sehingga menyebabkan keterlambatan penanganan medis hingga akhirnya berujung pada kematian[3].

Seiring dengan perkembangan teknologi, pemanfaatan *machine learning* dalam bidang kesehatan menjadi solusi potensial untuk mengatasi keterbatasan

diagnosis manual. *Machine learning* memiliki kemampuan untuk menganalisis data medis dalam jumlah besar dengan cepat dan akurat. Teknologi ini memungkinkan pengenalan pola-pola kompleks dalam data hematologi sehingga dapat membantu klasifikasi Non-DBD, DF, dan DHF secara otomatis. Implementasi *machine learning* juga dapat membantu dokter dalam mengambil keputusan medis yang lebih tepat waktu dan berbasis data [4].

Berbagai algoritma *machine learning* seperti *Support Vector Machine* (SVM), *Random Forest*, dan *Gradient Boosting* telah banyak diterapkan untuk prediksi dan klasifikasi penyakit. Ketiga algoritma ini dikenal memiliki performa tinggi dalam analisis data medis dengan tingkat akurasi yang baik. Model yang dikembangkan dievaluasi melalui metrik seperti akurasi, sensitivitas, spesifisitas, dan *F1-score* untuk memastikan keandalannya dalam aplikasi klinis. Dengan pendekatan ini, diharapkan dapat meminimalisasi kesalahan diagnosis yang selama ini masih sering terjadi dalam praktik medis konvensional [4].

Berdasarkan kondisi tersebut, penelitian ini bertujuan membangun model klasifikasi DBD menggunakan parameter hematologi dengan algoritma *machine learning*. Algoritma yang digunakan dalam penelitian ini adalah *Random Forest*, *Gradient Boosting*, dan *Support Vector Machine*. Pemanfaatan teknologi ini diharapkan mampu meningkatkan akurasi diagnosis serta menjadi sistem pendukung keputusan bagi dokter dalam menangani pasien DBD. Selain itu, pengembangan ini juga menjadi bagian dari transformasi digital di bidang kesehatan yang sejalan dengan inisiatif pemerintah untuk meningkatkan kualitas pelayanan kesehatan di Indonesia.

B. Rumusan Masalah

Rumusan masalah yang menjadi fokus utama dalam penelitian ini adalah Algoritma mana yang paling optimal untuk diterapkan dalam sebuah web sederhana dalam mengklasifikasikan data hematologi untuk membedakan antara *Dengue Fever* (DF) dan *Dengue Hemorrhagic Fever* (DHF)?

C. Batasan Masalah

Penelitian ini lebih terfokus dan terarah, maka ditetapkan beberapa batasan sebagai berikut:

1. Data yang digunakan dalam penelitian ini adalah data sekunder, yaitu data hematologi pasien yang didiagnosis dengan Demam Berdarah *Dengue*.
2. Variabel yang digunakan adalah indeks darah yang komprehensif seperti haemoglobin, jumlah sel darah putih total, jumlah sel darah putih diferensial, sel darah merah (RBC), dan jumlah trombosit.
3. Pada penelitian ini tidak membahas masalah vektor pembawa penyakit maupun faktor lain yang mempengaruhi tingkat resiko terjangkit demam berdarah selain indeks darah.
4. Penelitian ini mengklasifikasikan *Dengue Fever* (DF) dan *Dengue Hemorrhagic Fever* (DHF).
5. Evaluasi model menggunakan metrik yaitu: *accuracy*, *precision*, *recall*, dan *F1-score* untuk menentukan model terbaik.
6. Algoritma yang diuji yaitu *Random Forest*, *Gradient Boosting*, dan *Support Vector Machine*.

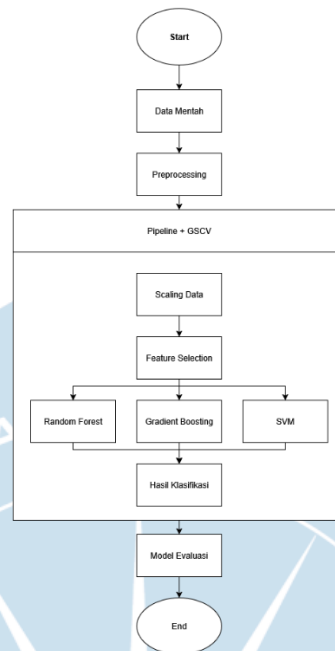
D. Tujuan Penelitian

Tujuan dari penelitian ini adalah untuk Mengetahui algoritma yang optimal dengan membandingkan metrik evaluasi dalam mengklasifikasikan data hematologi untuk membedakan *Dengue Fever* (DF) dan *Dengue Hemorrhagic Fever* (DHF).

E. Metode Penelitian

Penelitian ini akan dilakukan menggunakan jenis penelitian kuantitatif, dengan pendekatan eksperimental dan deskriptif. Dalam penelitian ini, akan menguji dan membandingkan kinerja model *machine learning* dalam memprediksi penyakit demam berdarah berdasarkan dataset hematologi. Penelitian ini dapat dilakukan dengan menggunakan data pelatihan dan pengujian untuk mengevaluasi akurasi dan performa model. Selain itu, dalam penelitian ini juga akan menggambarkan karakteristik dataset hematologi yang akan diatur sesuai dengan data hematologi di Indonesia, dataset ini digunakan dan menganalisis hasil klasifikasi berdasarkan metode *machine learning*. Penelitian ini akan memberikan wawasan tentang seberapa baik *machine learning* dalam memprediksi kemungkinan terjadinya penyakit demam berdarah berdasarkan variabel yang tersedia dalam dataset Hematologi.

Penelitian ini akan dilaksanakan melalui beberapa tahapan yang disusun dalam kerangka kerja yang terstruktur. Kerangka kerja ini dirancang untuk memastikan bahwa proses penelitian berjalan secara sistematis dan terarah. Setiap tahapan saling terkait satu sama lain, seperti yang digambarkan pada Gambar 1.1.



Gambar 1.1 Alur Penelitian

1. **Data Mentah**

Pengumpulan data hematologi pasien yang terdiagnosis Demam Berdarah *Dengue* (DBD). Data yang dikumpulkan adalah data sekunder berupa parameter hematologi seperti hemoglobin (Hb), jumlah sel darah putih, sel darah merah (RBC), trombosit, dan diferensial leukosit yang didapat dari Kaggle.

2. ***Preprocessing***

Data yang terkumpul akan melalui tahap *preprocessing*, yang mencakup:

- a. Pembersihan data (menghilangkan nilai yang hilang atau data *outliers*).
- b. Normalisasi atau standarisasi data.
- c. Melakukan penyeimbangan dataset terhadap kelas minoritas pada data train / data latih menggunakan ADASYN (*Adaptive Synthetic Sampling Approach*) agar menghindari data *leakage*.

3. ***Pipeline + GridSearchCV***

Pada tahapan ini untuk menggabungkan alur proses dari *preprocessing* sampai ke pelatihan model dalam satu alur, dan

menggunakan *GridSearchCV* untuk mencari kombinasi *hyperparameter* yang terbaik berdasarkan *cross-validation*.

4. **Data Scaling**

Pada Tahapan ini akan menormalisasikan skala dari setiap parameter agar mempermudah dalam tahapan *training* dan mendapatkan akurasi yang baik. Data *Scaling* yang digunakan adalah *Standardscaler* untuk menormalisasikan skala dari setiap parameter.

5. **Feature Selection**

Tahapan ini adalah tahapan dalam memilih fitur secara otomatis berdasarkan fitur dari model tertentu dan fitur yang tidak penting akan dieliminasi. Pemilihan fitur yang digunakan adalah menggunakan *SelectFromModel* dengan menggunakan *estimator*nya adalah *Radom Forest*. Terdapat perbedaan di setiap model yaitu 2 skenario, pada skenario 1 parameter yang mengatur seleksi fitur sudah ditetapkan, sedangkan pada skenario 2 menggunakan *hyperparameter* yang nantinya dicari parameter paling optimal untuk pemilihan fitur tersebut.

6. **Pelatihan Model**

Pelatihan model akan menggunakan 3 model yaitu *Random Forest*, *Gradient Boosting* dan SVM.

7. **Evaluasi Model**

Pada Tahapan ini akan mengevaluasi dari model yang sudah di *training* dengan metrik score menggunakan akurasi untuk seberapa banyak prediksi yang benar dari semua prediksi yang dibuat, *precision* untuk seberapa tepat prediksi positif, *recall* untuk mengukur seberapa banyak kasus nyata yang berhasil ditemukan, f1-score adalah keseimbangan dari *precision* dan *recall*.

F. Sistematika Penulisan

Sistematika penulisan tugas akhir dapat diuraikan menjadi beberapa bagian yang tertulis secara sistemik, meliputi:

1. BAB I PENDAHULUAN

Bab ini terdiri atas subbab yang ditulis secara urut berupa latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, metode penelitian dan sistematika penulisan. Latar belakang penelitian menjabarkan mengenai konteks dan kondisi yang mendorong dilakukannya penelitian. Rumusan masalah dan batasan masalah menjelaskan tentang masalah yang hendak diangkat dalam penelitian beserta batasan-batasannya dalam ruang lingkup penelitian. Tujuan penelitian menjabarkan tentang tujuan yang hendak dicapai dalam penelitian dan metode penelitian menjabarkan tentang metode yang digunakan pada penelitian.

2. BAB II TINJAUAN PUSTAKA

Penulisan bab ini berisi penelitian-penelitian terdahulu yang pernah dilakukan dan menyerupai topik penelitian yang dilakukan. Selain itu, terdapat 5 tabel perbandingan antara penelitian yang dilakukan sekarang dengan penelitian terdahulu.

3. BAB III LANDASAN TEORI

Bab ini menjelaskan teori-teori dasar yang menjadi pijakan dalam penelitian, serta konsep-konsep yang digunakan untuk mendukung pengembangan dan analisis pada penelitian ini. Setiap teori yang dibahas memiliki relevansi langsung terhadap metode atau pendekatan yang diterapkan.

4. BAB IV DATASET DAN PENGEMBANGAN MODEL

Bab ini menjelaskan tahapan pengolahan dataset, termasuk proses pengondisian dataset agar sesuai untuk permasalahan klasifikasi multikelas. Selain itu, dipaparkan proses pengembangan dan tuning terhadap tiga model pembelajaran mesin yang digunakan, masing-masing diuji dengan dua skenario untuk mengevaluasi performa model secara menyeluruh.

5. BAB V IMPLEMENTASI MODEL DAN HASIL

Bab ini menyajikan implementasi dari ketiga model yang telah dikembangkan serta hasil evaluasi kinerjanya berdasarkan dua skenario

pengujian pada masing-masing model. Dengan begitu didapatkan model yang digunakan dalam penerapan ke dalam *Streamlit*.

6. BAB VI PENUTUP

Bab ini berisi kesimpulan dari hasil penelitian serta saran yang dapat dijadikan masukan untuk pengembangan penelitian di masa mendatang. Kesimpulan disusun berdasarkan hasil analisis dan temuan yang diperoleh, sedangkan saran ditujukan untuk pengembangan lebih lanjut baik dari sisi metode, data, maupun aplikasi penelitian.

