

BAB I

PENDAHULUAN

1.1 Latar Belakang

Hingga saat ini, demam berdarah dengue (DBD) masih merupakan masalah kesehatan yang signifikan. *Virus dengue* adalah penyebab penyakit ini dan menyebar melalui gigitan nyamuk *Aedes aegypti*. Masalah DBD sangat sulit, terutama dalam hal pengendalian sehingga belum mencapai hasil yang optimal. Kasus DBD masih ada dan tersebar luas (Tri & Sukesu, 2018).

Dilansir dari laman *website* kementerian kesehatan pada tahun 2024 pada minggu ke-17 menunjukkan adanya peningkatan kasus demam berdarah di Indonesia, diperkirakan 88.593 orang menderita penyakit demam berdarah dan angka kematian di Indonesia hingga minggu ke-17 mencapai 621 angka kematian dan angka tersebut terus meningkat dari pekan-pekan sebelumnya (Rokom, 2024). Pada tahun 2022, jumlah kasus penyakit demam berdarah dengue (DBD) yang tercatat di Puskesmas I Kecamatan Mengwi berjumlah 300 kasus dan pada tahun 2023 kasus dengan penyakit ini mengalami peningkatan menjadi 600-an kasus.

Dengan tingginya jumlah penderita penyakit ini, salah satu langkah yang dapat diambil untuk membantu Dokter atau tenaga medis untuk mendiagnosa penyakit ini adalah dengan menggunakan *data mining*. Menurut (Yoo et al., 2012) menerangkan penerapan *data mining* yang berhasil memberikan pengetahuan medis dan perawatan kesehatan baru yang dapat digunakan secara efektif untuk mendukung pengambilan keputusan. Berbagai metode dapat diterapkan untuk

mengklasifikasikan penyakit demam berdarah, antara lain Decision Tree, K-Nearest Neighbor (KNN), Support Vector Machine (SVM), Random Forest, Naïve Bayes dan lain-lain.

Terdapat penelitian yang membandingkan antara algoritma K-Nearest Neighbor dengan algoritma Naïve Bayes. Berdasarkan hasil yang diperoleh, algoritma KNN menunjukkan tingkat akurasi sebesar 72,31% dengan nilai $K = 2$ (Matdoan, 2022) dan akurasi sebesar 57,97% dengan nilai $K = 3$ (Matdoan, 2022). Sementara itu, penelitian yang menerapkan algoritma Naïve Bayes menghasilkan akurasi sebesar 67% (Andiaini et al., n.d.) dan pada penelitian lainnya menghasilkan akurasi sebesar 64.02% (Setiadi et al., 2021). Meskipun metode KNN dan Naïve Bayes berhasil melakukan klasifikasi, namun hasil akurasi dari masing-masing algoritma masih kurang memuaskan, dan persentase yang diperoleh belum optimal. Oleh karena itu, dilakukan kolaborasi dengan menggunakan *ensemble learning* melalui pendekatan *stacking*.

Penelitian lain sebelumnya yang meneliti mengenai penggunaan *stacking ensemble learning* untuk meningkatkan *accuracy* yang diperoleh dari penggunaan metode *single classifier* pada *Implementation of Stacking Ensemble Classifier for Multi-class Classification of COVID-19 Vaccines Topics on Twitter* ini menggunakan *base learner* Logistic Regression *algorithm*, Support Vector Machine (SVM) dan Random Forest sedangkan Logistic Regression sebagai *meta learner*. Dalam penelitian tersebut didapatkan hasil yang lebih baik ketika menggunakan *stacking* daripada penggunaan *single algorithm*. Alasan penggunaan Logistic Regression sebagai *meta learner* adalah karena Logistic Regression ini

mampu menangani *biner classification* atau *multi-class classification* dan cocok dengan kriteria dataset yang digunakan dalam penelitian klasifikasi demam berdarah ini (Jayapermana et al., 2022).

Ensemble learning adalah sebuah metode yang mengintegrasikan atau menggabungkan beberapa metode dasar untuk meningkatkan akurasi. Metode ini lebih tahan terhadap *noise* daripada menggunakan satu pengklasifikasi. Contoh teknik *ensemble learning* antara lain *Bagging*, *Boosting*, dan *Stacking* (Ayu Nandia Lestari et al., 2023).

Stacking merupakan suatu metode pada algoritma *ensemble learning* yang di dalam proses pembelajarannya menggunakan *base-model* atau model dasar dalam melakukan klasifikasi secara mandiri. Dalam hasil akhir pembelajaran mandiri dari setiap *base-learner* ini menghasilkan prediksi, dimana semua prediksi dari setiap *base-model* akan digabungkan untuk menghasilkan dataset baru. Dataset baru ini akan digunakan untuk melakukan pembelajaran kembali pada model selanjutnya atau sering disebut sebagai *meta-learner*. Hasil dari *meta-learner* ini menjadi prediksi atau keputusan terakhir untuk memberikan label pada suatu data baru (Putri & Suparwito, 2023).

Oleh karena itu dilaksanakan penelitian dengan judul “Klasifikasi Penyakit Demam Berdarah Menggunakan Algoritma Stacking Ensemble Learning”. Penelitian ini bertujuan untuk melakukan klasifikasi penyakit demam berdarah menggunakan metode *stacking ensemble learning*.

1.2 Identifikasi Masalah

Dari latar belakang masalah yang sudah dipaparkan, penting untuk mengidentifikasi permasalahan yang akan diteliti. Komparasi dilakukan diantara penggunaan *single classifier* dengan penggunaan *stacking ensemble learning* untuk mengetahui perbandingan akurasi yang dihasilkan diantara penggunaan kedua metode tersebut dikarenakan pada beberapa referensi metode *stacking* ini tidak selalu mengalami peningkatan akurasi bisa saja terjadi sebaliknya.

1.3 Batasan Masalah

Batasan yang ditetapkan untuk masalah penelitian ini antara lain:

- A. Kumpulan data yang digunakan terdiri dari data mulai dari gejala hingga diagnosis penyakit demam berdarah, yang disesuaikan dengan data yang dikumpulkan dari Puskesmas I Kecamatan Mengwi bagian Rekam Medis.
- B. Parameter yang dipakai untuk penelitian ini meliputi Umur, Jenis Kelamin, Sistole, Diastole, Trombosit, Leukosit, Hematokrit, Bintik Merah pada Tubuh, Nyeri, Suhu Tubuh, serta Pendarahan pada Hidung dan Pendarahan pada Gusi.
- C. Dalam penelitian ini, digunakan KNN, Naïve Bayes, dan Decision Tree sebagai model dasar (*base learner*) sedangkan untuk *meta-learner* digunakan Logistic regression.

1.4 Rumusan Masalah

Berdasarkan latar belakang yang telah dipaparkan, rumusan masalah yang bisa diidentifikasi yaitu bagaimana membandingkan hasil akurasi antara *single*

classifier dan penggunaan metode *stacking ensemble learning* dalam klasifikasi penyakit demam berdarah apakah mengalami peningkatan akurasi atau sebaliknya.

1.5 Tujuan Penelitian

Dari rumusan masalah yang sudah disebutkan, tujuan penelitian ini adalah untuk melakukan klasifikasi serta membandingkan akurasi terhadap dataset demam berdarah, baik menggunakan *single classifier* maupun metode *stacking*.

1.6 Manfaat Penelitian

Beberapa manfaat yang didapatkan dari penelitian ini diantara lainnya adalah teoritis serta praktis:

A. Manfaat Teoretis

Hasil penelitian ini diharapkan dapat meningkatkan pemahaman peneliti tentang metode serta memberikan kontribusi ide kepada pembaca tentang studi metode *stacking ensemble learning* untuk identifikasi penyakit demam berdarah.

B. Manfaat Praktis

Diharapkan studi ini dapat menjadi sumber informasi untuk memprediksi demam berdarah melalui penerapan teknik *data mining*.