

DIAGNOSIS PENYAKIT DEMAM BERDARAH DENGUE MENERAPKAN ALGORITMA *K-NEAREST NEIGHBOR*

I. G. Haga Olas Tyamarta^{1*}, I. N. Sukajaya², Putu Kartika Dewi³

^{1,2}Matematika dan Ilmu Pengetahuan Alam; Jalan Udayana (Kampus Tengah), Singaraja-Bali, 81116, Indonesia; (0362) 22570 / 25735

Keywords:

Diagnosis, Demam Berdarah Dengue, *K-Nearest Neighbor*, Penambangan Data

Correspondent Email:

nyoman.sukajaya@undiksha.ac.id

Abstrak. Demam berdarah dengue merupakan salah satu penyakit menular yang menjadi permasalahan kesehatan serius di seluruh wilayah tropis Indonesia, dan Kabupaten Klungkung merupakan salah satu daerah yang turut terdampak. Diagnosis dini terhadap penyakit ini sangat penting guna mencegah terjadinya komplikasi yang berakibat fatal. Penelitian ini bertujuan untuk mendiagnosis penyakit demam berdarah dengue dengan menerapkan algoritma *K-Nearest Neighbor* serta mengevaluasi performa model yang telah dibangun. Data yang digunakan merupakan data historis pasien dari Rumah Sakit Umum Daerah Kabupaten Klungkung tahun 2021 hingga 2023 sebanyak 312 data. Variabel input yang digunakan dalam penelitian meliputi usia, jenis kelamin, durasi demam, suhu tubuh, sakit kepala, mual dan muntah, nyeri otot dan persendian, hasil uji tourniquet, jumlah trombosit, jumlah leukosit, dan jumlah hematokrit. Metode preprocessing data dilakukan melalui tahap pembersihan, seleksi, dan transformasi data. Pembagian data dilakukan menggunakan metode *Aggregated Hold-Out* (Agghoo) sebanyak tiga iterasi dengan rasio 80% sebagai data latih dan 20% sebagai data uji. Model terbaik diperoleh pada nilai $K = 15$ dengan hasil evaluasi akurasi sebesar 0,9894, sensitivitas sebesar 0,9915, dan spesifisitas sebesar 0,9861. Model tersebut kemudian diimplementasikan ke dalam sebuah aplikasi berbasis website guna mempermudah proses diagnosis serta pencatatan data pasien.



Copyright © [JITET](#) (Jurnal Informatika dan Teknik Elektro Terapan). This article is an open access article distributed under terms and conditions of the Creative Commons Attribution (CC BY NC)

Abstract. Dengue hemorrhagic fever is an infectious disease that remains a serious public health problem across tropical regions of Indonesia, including Klungkung Regency. Early diagnosis of this disease is essential to prevent severe complications that may lead to fatal outcomes. This study aims to diagnose dengue hemorrhagic fever by applying the *K-Nearest Neighbor* algorithm and to evaluate the performance of the developed model. The dataset used in this study consists of 312 historical patient records obtained from the Regional General Hospital of Klungkung Regency during the period 2021–2023. The input variables include age, gender, duration of fever, body temperature, headache, nausea and vomiting, muscle and joint pain, tourniquet test results, platelet count, leukocyte count, and hematocrit level. Data preprocessing was conducted through data cleaning, feature selection, and data transformation stages. The data were divided using the *Aggregated Hold-Out* (Agghoo) method over three iterations, with 80% allocated as training data and 20% as testing data. The best model was obtained at $K = 15$, achieving an accuracy of 0.9894, sensitivity of 0.9915, and specificity of 0.9861. The resulting model was subsequently implemented into a web-based application to facilitate the diagnostic process and patient data management.

1. PENDAHULUAN

Penyakit demam berdarah dengue merupakan salah satu masalah kesehatan masyarakat yang masih menjadi ancaman serius di berbagai negara tropis, termasuk Indonesia. Penyakit ini disebabkan oleh infeksi virus

dengue yang ditularkan melalui gigitan nyamuk *Aedes Aegypti* dan *Aedes Albopictus*, dengan *Aedes Aegypti* sebagai vektor utama penularan. Virus dengue sendiri merupakan virus RNA berantai tunggal dari famili *Flaviviridae* yang terdiri dari atas empat serotipe, yaitu DENV-1,

DENV-2, DENV-3, dan DENV-4. Infeksi primer umumnya menyebabkan demam dengue, sedangkan infeksi sekunder oleh serotipe berbeda berpotensi menimbulkan demam berdarah dengue (*secondary heterologous infection*) [1], [2].

Indonesia merupakan salah satu negara dengan tingkat penyebaran dengue tertinggi di Asia Tenggara. Data *World Health Organization* (WHO) menunjukkan bahwa jumlah kasus dengue di Indonesia cenderung tinggi setiap tahunnya. Organisasi Pencegahan dan Pengendalian Penyakit Menular Kementerian Kesehatan Republik Indonesia (P2PM Kemenkes RI) melaporkan bahwa dalam lima tahun terakhir ditemukan rata-rata 121.191 kasus demam berdarah dengue per tahun. Pada tahun 2023, jumlah kasus meningkat menjadi 141.435 dengan 894 angka kematian dan 50,6 per 100.000 penduduk angka *incidence rate*. Berdasarkan data Badan Pusat Statistik Provinsi Bali, *incidence rate* kasus demam berdarah dengue di Provinsi Bali pada tahun 2023 mencapai 152 per 100.000 penduduk atau 6.707 kasus. Kabupaten Klungkung merupakan kabupaten dengan *incidence rate* tertinggi di Provinsi Bali, yaitu 327,57 per 100.000 penduduk atau 684 kasus pada tahun yang sama [3], [4], [5].

Tingginya jumlah kasus demam berdarah dengue di Kabupaten Klungkung menjadi tantangan bagi fasilitas pelayanan kesehatan, khususnya dalam proses diagnosis. Penetapan diagnosis memerlukan ketelitian tinggi karena variasi gejala klinis antar pasien, sehingga dokter membutuhkan waktu yang lama setelah hasil laboratorium diterima, bahkan pada kondisi tercepat sekitar satu jam, untuk melakukan pemeriksaan fisik, evaluasi gejala klinis, dan analisis hasil laboratorium secara menyeluruh. Kondisi ini berpotensi menimbulkan keterlambatan penanganan, terutama ketika jumlah pasien meningkat, yang dapat meningkatkan risiko komplikasi.

Kondisi tersebut menunjukkan perlunya pendekatan berbasis teknologi untuk mendukung proses diagnosis secara lebih cepat dan akurat. Perkembangan teknologi memungkinkan pemanfaatan data klinis melalui teknik data mining atau knowledge discovery in database (KDD) guna menggali pola dari data berskala besar. Dalam konteks diagnosis demam berdarah dengue, teknik ini

memanfaatkan atribut klinis dan parameter laboratorium pasien. Penerapan pendekatan ini diharapkan dapat membantu tenaga medis dalam pengambilan keputusan diagnosis secara efisien dan tepat.[6].

Salah satu metode yang banyak digunakan dalam *data mining* untuk proses diagnosis adalah Algoritma *K-Nearest Neighbor*. Algoritma ini menentukan diagnosis data baru dengan mengidentifikasi nilai K pada data pelatihan yang memiliki jarak paling dekat. Algoritma *K-Nearest Neighbor* dikenal sederhana, memiliki sedikit parameter, namun mampu memberikan performa diagnosis yang baik pada berbagai dataset medis [7].

Amilia dan Prasetyo menunjukkan bahwa algoritma *K-Nearest Neighbor* mampu menghasilkan tingkat akurasi yang tinggi dalam diagnosis penyakit demam berdarah dengue [8]. Penelitian tersebut menggunakan 225 data dengan lima variabel, yaitu usia, jenis kelamin, hemoglobin, trombosit, dan hematokrit, serta memperoleh akurasi sebesar 94,22%. Namun, penelitian tersebut masih menggunakan variabel input yang terbatas sehingga belum melibatkan gejala klinis dan hasil pemeriksaan laboratorium secara bersamaan dalam proses diagnosis demam berdarah dengue. Oleh karena itu, penelitian ini mengusulkan model diagnosis dengan memperluas variabel input yang digunakan dengan mengombinasikan gejala klinis pasien dan hasil pemeriksaan laboratorium. Variabel yang digunakan meliputi usia pasien, jenis kelamin, durasi demam, suhu tubuh, sakit kepala, mual dan muntah, nyeri otot dan persendian, hasil uji tourniquet, jumlah trombosit, jumlah leukosit, dan jumlah hematokrit [9]. Pendekatan ini diharapkan mampu mendukung proses diagnosis yang lebih cepat dan akurat.

Berdasarkan kondisi tersebut, penelitian ini dilakukan untuk mengimplementasikan Algoritma *K-Nearest Neighbor* dalam mendiagnosis penyakit demam berdarah dengue di Kabupaten Klungkung dengan menggunakan 312 data dari data historis pasien RSUD Kabupaten Klungkung tahun 2021 hingga 2023. Penelitian ini juga menganalisis performa serta optimalisasi model melalui metode pembagian data *Aggregated Hold-Out* dengan tiga pengulangan guna memperoleh evaluasi yang lebih optimal dan representatif. Melalui penelitian ini, diharapkan diperoleh

model diagnosis yang dapat mendukung proses diagnosis penyakit demam berdarah dengue secara lebih cepat dan akurat.

2. TINJAUAN PUSTAKA

2.1 Penyakit Demam Berdarah Dengue

Penyakit demam berdarah dengue merupakan penyakit tular vektor yang masih menjadi masalah kesehatan serius di Indonesia, disebabkan oleh virus dengue yang ditularkan melalui gigitan nyamuk *Aedes Aegypti* dan *Aedes Albopictus*. Virus dengue dari famili *Flaviviridae* memiliki empat serotipe, dan infeksi sekunder oleh serotipe berbeda dapat memicu peningkatan permeabilitas kapiler yang menyebabkan kebocoran plasma serta pendarahan. Infeksi dengue berlangsung dalam tiga fase, yaitu fase febris, fase kritis, dan fase pemulihan. Fase kritis merupakan fase paling berbahaya karena suhu tubuh menurun tetapi risiko syok meningkat akibat trombositopenia, leukopenia, dan peningkatan hematokrit. Jika tidak ditangani dengan cepat, kondisi ini dapat menyebabkan kerusakan organ vital dan kematian meskipun pada fase pemulihan keadaan pasien mulai membaik [9], [10].

2.2 Diagnosis

Diagnosis merupakan proses penting dalam medis yang bertujuan menentukan penyakit atau kondisi pasien melalui gejala klinis, pemeriksaan fisik, dan hasil laboratorium. Proses ini berfungsi untuk mengelompokkan kondisi pasien sehingga tenaga medis dapat menetapkan tindakan atau terapi yang tepat. Diagnosis juga menjadi alat untuk mengenali serta memprediksi keadaan tertentu melalui pengamatan dan pengukuran yang sistematis. Dalam pelaksanaannya, diagnosis menuntut ketelitian, pemahaman instrumen pemeriksaan, serta penerapan etika profesional seperti menjaga kerahasiaan data pasien, dan memastikan keputusan berbasis data objektif. Dengan demikian, diagnosis memiliki peran vital sebagai dasar pengambilan keputusan klinis dan penentuan arah penatalaksanaan penyakit [11].

2.3 Algoritma *K-Nearest Neighbor*

Algoritma *K-Nearest Neighbor* merupakan algoritma yang menentukan diagnosis data baru

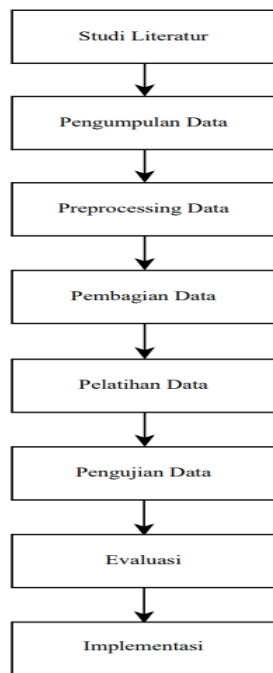
dengan mengidentifikasi nilai *K* pada data pelatihan yang memiliki jarak paling dekat. Algoritma ini bekerja dengan menghitung jarak antara data uji dengan data latih menggunakan beberapa rumus jarak seperti *Euclidean*, *Manhattan*, *Minkowski*, dan *Chebyshev*. Nilai *K* biasanya di tetapkan dalam jumlah ganjil untuk mencegah terjadinya hasil prediksi yang bias. Setelah jarak dihitung, algoritma memilih *K* tetangga terdekat dan menetapkan kelas berdasarkan mayoritas dari tetangga tersebut. Dengan demikian, Algoritma *K-Nearest Neighbor* dapat memberikan diagnosis secara efektif melalui analisis kemiripan pola antara data baru dan data latih yang sudah memiliki label atau kelas [12], [13].

3. METODE PENELITIAN

Penelitian ini merupakan penelitian eksperimen berbasis data mining yang menggunakan data pasien dari Rumah Sakit Umum Daerah Kabupaten Klungkung. Proses diagnosis dilakukan dengan menerapkan Algoritma *K-Nearest Neighbor* menggunakan bahasa pemrograman *Python* untuk mengukur performa algoritma tersebut dalam mendiagnosis pasien apakah menderita penyakit demam berdarah dengue atau tidak. Adapun prosedur penelitian terdiri dari beberapa tahapan, yaitu studi literatur, pengumpulan data, pra pengolahan dan pembersihan data, seleksi data, transformasi data, pembagian data, pelatihan model, evaluasi, dan implementasi. Metode analisis data yang digunakan adalah Algoritma *K-Nearest Neighbor* dengan evaluasi akhir menggunakan *confusion matrix* melalui perhitungan akurasi, sensitivitas, dan spesifisitas. Tahap akhir berupa implementasi model dilakukan dengan menerapkan model diagnosis terbaik untuk memprediksi kelas diagnosis pada data uji sehingga dapat memberikan gambaran kinerja model dalam kondisi nyata. Tahapan metode penelitian pada studi ini dijelaskan lebih lanjut pada sub bagian berikut.

3.1 Prosedur Penelitian

Alur penelitian diagnosis penyakit demam berdarah dengue di Kabupaten Klungkung dengan menerapkan Algoritma *K-Nearest Neighbor* dapat dilihat pada Gambar 1.



Gambar 1. Prosedur Penelitian

3.1.1 Studi Literatur

Pada tahap ini dilakukan studi literatur melalui artikel, jurnal, dan buku yang membahas penyakit demam berdarah dengue, *data mining*, serta Algoritma *K-Nearest Neighbor*. Dari kajian tersebut diperoleh bahwa Algoritma *K-Nearest Neighbor* mampu mendiagnosis penyakit demam berdarah dengue dengan akurasi yang baik, sebagaimana ditunjukkan pada penelitian Amilia dan Prasetyo yang mencapai akurasi 94,22%.

3.1.2 Pengumpulan Data

Data penelitian ini merupakan data historis pasien demam berdarah dengue tahun 2021-2023 yang diperoleh dari divisi rekam medik dan medikolegal RSUD Kabupaten Klungkung, mencakup variabel gejala klinis dan hasil laboratorium seperti usia, jenis kelamin, durasi demam, suhu tubuh, mual dan muntah, uji tourniquet, trombosit, leukosit, dan hematokrit. Data tersebut digunakan sebagai dasar dalam pembangunan dan pengujian model diagnosis menggunakan Algoritma *K-Nearest Neighbor*.

3.1.3 Pra-Pemrosesan Data

Pengolahan data pada sistem *data mining* tidak dapat dilakukan secara langsung karena diperlukan beberapa tahap persiapan sebelum

data siap untuk proses pelatihan dan pengujian model. Tahap ini meliputi pembersihan data untuk menghilangkan data yang tidak lengkap atau relevan, seleksi data untuk memilih data berkualitas, serta transformasi data agar seluruh variabel berada dalam format dan skala yang sesuai. Proses ini bertujuan meningkatkan kualitas data sehingga model komputasi dapat menghasilkan hasil diagnosis yang lebih akurat [14].

3.1.4 Pembagian Data

Pembagian data pada penelitian ini menggunakan metode *Aggregated Hold-Out*, yaitu membagi dataset menjadi dua bagian (80% menjadi data latih dan 20% menjadi data uji) secara acak dan dilakukan pengulangan sebanyak tiga iterasi. Pada setiap iterasi, model dilatih dan diuji kemudian hasil evaluasinya digabung dan dirata-ratakan untuk memperoleh performa yang lebih stabil dan representatif. Metode ini diterapkan untuk mengurangi bias yang muncul pada pembagian data tunggal konvensional (*hold-out*) [15].

3.1.5 Pelatihan Data

Pada tahap ini dilakukan proses pelatihan untuk mempelajari pola antara fitur input dan kelas output pada dataset yang telah disegmentasi menjadi dua label, yaitu positif dan negatif demam berdarah dengue. Sebanyak 80% data digunakan sebagai data latih yang kemudian diproses menggunakan Algoritma *K-Nearest Neighbor* untuk membangun model diagnosis.

3.1.6 Pengujian Data

Pada tahap ini akan dilaksanakan pengujian terhadap hasil dari pelatihan data yang dilakukan pada tahap sebelumnya. Data yang digunakan pada tahap ini adalah data uji atau *testing data* dengan rasio 20% pada dataset yang sama. Model akan diuji dengan data uji dengan tujuan untuk mengetahui apakah hasil dari pelatihan data yang telah dilaksanakan sebelumnya memiliki performa dan akurasi yang optimal.

3.1.7 Evaluasi

Confusion matrix digunakan untuk mengevaluasi performa model melalui empat komponen utama, yaitu *true positive*, *true negatif*, *false positif*, dan *false negative*, yang masing-masing menggambarkan kesesuaian antara hasil diagnosis dan kondisi aktual. Berdasarkan matriks tersebut dihitung metrik sensitivitas, spesifisitas, dan akurasi yang menunjukkan kemampuan model dalam mengidentifikasi kasus positif, negatif, dan keseluruhan diagnosis secara tepat [16].

3.1.8 Implementasi

Tahap implementasi merupakan tahap terakhir dalam prosedur penelitian, di mana model yang telah melalui proses evaluasi diimplementasi ke dalam aplikasi sistem diagnosis penyakit demam berdarah dengue berbasis website. Aplikasi website tersebut dibangun menggunakan bahasa pemrograman *Python* dengan *framework django*. Pengujian sistem dilakukan menggunakan metode *black box testing* untuk memastikan setiap fungsi pada sistem berjalan sesuai dengan kebutuhan dan spesifikasi yang dirancang [17], [18].

3.2 Metode Analisa Data

3.2.1 Algoritma *K-Nearest Neighbor*

Algoritma *K-Nearest Neighbor* merupakan algoritma yang menentukan diagnosis data baru dengan mengidentifikasi nilai *K* pada data pelatihan yang memiliki jarak paling dekat [19]. Perhitungan jarak data uji dengan data latih pada masing-masing iterasi pada penelitian ini menggunakan rumus jarak Euclidean. Jarak Euclidean merupakan jarak antara dua titik dalam ruang fitur, jarak ini dihitung sebagai akar kuadrat dari selisih kuadrat setiap atribut dari dua titik data. Dalam penelitian ini ruang fitur dapat diartikan sebagai ruang yang terbentuk dari kumpulan atribut pada data.

$$d(x^i, y^i) = \sqrt{\sum_{k=1}^n (x_k^i - y_k^i)^2}$$

Dengan keterangan.

$d(x^i, y^i)$: Hasil jarak antara titik data uji dan data latih

x^i : Data uji ke-*i*

y^i : Data latih ke-*i*

n : Jumlah dimensi (atribut data)

Berikut beberapa tahapan yang dilakukan dalam proses perhitungan algoritma *K-Nearest Neighbor* yaitu.

1. Menentukan parameter *K* (jumlah tetangga terdekat).
2. Menghitung jarak antara data uji dengan data pelatihan menggunakan rumus *Euclidean distance*.
3. Mengurutkan hasil perhitungan jarak *Euclidean* dari jarak terkecil hingga terbesar.
4. Menentukan kelas diagnosis akhir, yaitu kelas yang paling banyak muncul (majority voting) dari *K* tetangga terdekat. Kelas terbanyak tersebut ditetapkan sebagai hasil prediksi terhadap data uji.

3.2.2 Metode Pengujian Model

Confusion matrix merupakan sebuah tabel yang diterapkan untuk melakukan penilaian terhadap suatu pemodelan diagnosis dengan membandingkan nilai prediksi dengan nilai sebenarnya dari data uji. Nilai sensitivitas, spesifisitas dan akurasi dapat diperoleh melalui perhitungan menggunakan *confusion matrix* pada Tabel 1.

Tabel 1. *Confusion Matrix*

		Actual Values	
		Positive	Negative
Predictive Values	Positive	True positive	False Positive
	Negative	False Negative	True Negative

Confusion Matrix terdiri atas empat komponen utama, yaitu *true positive*, *true negatif*, *false positif*, dan *false negative*, yang digunakan untuk mengevaluasi kinerja model diagnosis. Berdasarkan komponen tersebut, adapun beberapa metrik evaluasi model yang dapat dihitung untuk mengevaluasi model [20].

1. Akurasi

Akurasi mengukur proporsi prediksi yang benar, baik untuk kelas positif maupun negatif terhadap seluruh data yang diuji. Nilai akurasi dihitung menggunakan rumus berikut.

$$\text{Akurasi} = \left(\frac{TP + TN}{\text{total sampel}} \right) \times 100\%$$

2. Sensitivitas

Sensitivitas mengukur kemampuan model dalam memprediksi kasus positif dengan benar. Ini dihitung sebagai rasio antara jumlah prediksi positif yang benar (*True Positive*) terhadap total jumlah data kasus positif demam berdarah dengue (TP + FN). Dengan rumus sensitivitas yaitu:

$$\text{Sensitivitas} = \left(\frac{TP}{TP + FN} \right) \times 100\%$$

3. Spesifisitas

Spesifisitas mengukur kemampuan model dalam memprediksi kasus negatif dengan benar. Ini dihitung sebagai rasio antara jumlah prediksi negatif yang benar (*True Negative*) terhadap total jumlah data kasus negatif demam berdarah dengue (TN + FP). Dengan rumus spesifisitas yaitu:

$$\text{Spesifisitas} = \left(\frac{TN}{TN + FP} \right) \times 100\%$$

3.3 Implementasi Hasil Penelitian

Implementasi hasil penelitian ini dilakukan melalui pengembangan aplikasi website berbasis *Django* yang memanfaatkan algoritma *K-Nearest Neighbor* untuk mendiagnosis penyakit demam berdarah dengue. Proses implementasi meliputi instalasi *Django*, pra-pemrosesan data, pemodelan, evaluasi model, integrasi model ke aplikasi website, pembuatan antarmuka, pengujian aplikasi menggunakan *black box testing*, serta penyediaan fitur interaksi pengguna. Setiap tahap dilaksanakan secara berurutan untuk memastikan data diproses dengan benar, model bekerja secara optimal, dan aplikasi dapat berfungsi sesuai dengan kebutuhan. Hasil akhirnya adalah aplikasi website yang memungkinkan pengguna memasukkan data dan memperoleh hasil diagnosis secara langsung melalui antarmuka yang telah disediakan.

4. HASIL DAN PEMBAHASAN

4.1 Hasil Penelitian

4.1.1 Hasil Analisis Algoritma *K-Nearest Neighbor*

Algoritma *K-Nearest Neighbor* mendiagnosis data baru dengan membandingkan jarak antara data baru dengan data latih yang telah memiliki kelas. Proses diagnosis dilakukan dengan memilih nilai *K* terdekat berdasarkan jarak terkecil, kemudian

menentukan kelas mayoritas dari tetangga tersebut sebagai hasil diagnosis. Tahapan dalam proses perhitungan algoritma *K-Nearest Neighbor* dapat dijelaskan sebagai berikut.

a. Tahap *Encoding Dataset*

Tahap *encoding* dilakukan untuk mengubah atribut kategori menjadi nilai numerik agar dapat diolah oleh algoritma *K-Nearest Neighbor*. Pada penelitian ini, dataset terdiri dari 11 atribut data, meliputi usia pasien, jenis kelamin pasien, durasi demam, suhu tubuh, sakit kepala, mual dan muntah, nyeri otot dan persendian, hasil uji tourniquet, jumlah trombosit, jumlah leukosit, dan jumlah hematokrit, serta satu label target yaitu diagnosis. Atribut yang bersifat kategorik dilakukan proses *encoding* dengan pemberian nilai numerik tertentu. Sebagai contoh, atribut jenis kelamin diencoding dengan nilai 1 untuk laki-laki dan 0 untuk perempuan, sedangkan atribut sakit kepala di *encoding* dengan nilai 1 untuk kondisi ya dan 0 untuk kondisi tidak. Proses *encoding* ini bertujuan untuk memastikan seluruh atribut data berada dalam format numerik sehingga siap diproses oleh algoritma *K-Nearest Neighbor*.

b. Tahap Normalisasi Data

Normalisasi data dilakukan untuk menyamakan rentang nilai antar atribut numerik sehingga tidak terjadi dominasi atribut tertentu dalam proses pembelajaran model. Pada penelitian ini, normalisasi diterapkan menggunakan metode *Min-Max Normalization*, yang mentransformasikan nilai atribut ke dalam rentang 0 hingga 1 dengan persamaan sebagai berikut.

$$X_{scaled} = \frac{X - X_{min}}{X_{max} - X_{min}}$$

Proses normalisasi diterapkan pada data hasil *encoding* dengan menentukan nilai minimum dan maksimum pada setiap atribut numerik, kemudian seluruh data ditransformasi menggunakan persamaan tersebut. Contoh hasil normalisasi disajikan pada Gambar 2.

No	Usia	Jenis Kelamin	Durasi Demam	Suhu Tubuh	Sakit Kepala	Mual Muntah
1	0.19	0.0	0.87	0.24	1.0	1.0
2	0.19	0.0	0.11	0.12	0.0	0.0
3	0.09	1.0	0.49	0.64	0.0	1.0
4	0.05	1.0	0.62	0.48	1.0	1.0
5	0.05	1.0	0.07	0.00	1.0	0.0
6	0.23	1.0	0.49	0.50	1.0	0.0
7	0.23	1.0	0.11	0.02	0.0	0.0
8	0.12	1.0	0.62	0.71	1.0	1.0

Gambar 2. Hasil Normalisasi

c. Menentukan Nilai K Terbaik

Penentuan nilai K terbaik dilakukan dengan mengimplementasikan algoritma *K-Nearest Neighbor* menggunakan bahasa pemrograman *Python*. Pengujian dilakukan pada nilai K = 3 hingga 17 dengan skema pembagian data menggunakan metode *Aggregated Hold-Out* pada iterasi 2 hingga iterasi 10. Kinerja model dievaluasi berdasarkan metrik akurasi, sensitivitas, dan spesifisitas. Hasil pengujian menunjukkan bahwa nilai K = 15 secara konsisten memberikan kinerja terbaik dibandingkan nilai K lainnya. Selain itu, iterasi 3 menghasilkan nilai evaluasi tertinggi sehingga digunakan sebagai dasar penentuan hasil akhir penelitian.

d. Pembagian Data dengan Metode *Aggregated Hold-Out*

Pembagian data pada penelitian ini dilakukan menggunakan metode *Aggregated Hold-Out*, yaitu metode yang membagi dataset secara acak ke menjadi data latih dan data uji secara berulang, kemudian hasil evaluasi dari setiap iterasi digabungkan dan dihitung rata-rata dari seluruh iterasi guna memperoleh performa yang lebih optimal dan representatif. Proses pembagian data diimplementasikan menggunakan bahasa pemrograman *Python* dengan rasio 80% data latih dan 20% data uji. Pembagian data dilakukan sebanyak tiga iterasi, di mana pada setiap iterasi dataset diacak kembali sebelum dipilih menjadi data latih dan data uji. Contoh data hasil pembagian disajikan pada Gambar 3.

DATA LATIH (ITERASI 1)							
Suhu Tubuh	Sakit Kepala	Mual Muntah	Nyeri Otot Persendian	Hasil Uji Tourniquet	Jumlah Trombosit	Jumlah Leukosit	
0.02	0.0	0.0	0.0	0.0	0.21	0.07	
0.02	0.0	1.0	0.0	0.0	0.29	1.00	
0.45	0.0	0.0	0.0	1.0	0.30	0.10	
DATA UJI (ITERASI 1)							
Suhu Tubuh	Sakit Kepala	Mual Muntah	Nyeri Otot Persendian	Hasil Uji Tourniquet	Jumlah Trombosit	Jumlah Leukosit	
0.02	1.0	0.0	0.0	0.0	0.26	0.13	
0.14	1.0	1.0	0.0	0.0	0.76	0.18	
0.69	0.0	1.0	1.0	1.0	0.10	0.12	
DATA LATIH (ITERASI 2)							
Suhu Tubuh	Sakit Kepala	Mual Muntah	Nyeri Otot Persendian	Hasil Uji Tourniquet	Jumlah Trombosit	Jumlah Leukosit	
0.02	0.0	0.0	0.0	0.0	0.21	0.07	
0.48	1.0	1.0	1.0	1.0	0.12	0.07	
0.83	1.0	1.0	1.0	1.0	0.14	0.05	
DATA UJI (ITERASI 2)							
Suhu Tubuh	Sakit Kepala	Mual Muntah	Nyeri Otot Persendian	Hasil Uji Tourniquet	Jumlah Trombosit	Jumlah Leukosit	

Gambar 3. Hasil Pembagian *Aggregated Hold-Out*

e. Perhitungan Jarak dengan Algoritma *K-Nearest Neighbor*

Perhitungan jarak data uji dengan data latih pada masing-masing iterasi pada penelitian ini menggunakan rumus jarak Euclidean. Jarak Euclidean merupakan jarak antara dua titik dalam ruang fitur, jarak ini dihitung sebagai akar kuadrat dari selisih kuadrat setiap atribut dari dua titik data. Dalam penelitian ini ruang fitur dapat diartikan sebagai ruang yang terbentuk dari kumpulan atribut pada data. Perhitungan ini diterapkan antara setiap data uji dan seluruh data latih pada masing-masing iterasi, kemudian hasil dari perhitungan jarak yang diperoleh dievaluasi untuk mengukur performa model.

f. Evaluasi Performa Model dengan *Confusion Matrix*

Evaluasi performa model dilakukan menggunakan *confusion matrix* dengan tiga metrik utama, yaitu akurasi, sensitivitas, dan spesifisitas. Evaluasi diterapkan pada setiap iterasi pengujian untuk menilai kemampuan model dalam melakukan diagnosis. Selanjutnya, nilai akurasi, sensitivitas, dan spesifisitas dari seluruh iterasi digabungkan untuk memperoleh nilai rata-rata sebagai representasi performa model secara keseluruhan. Hasil evaluasi menunjukkan bahwa model *K-Nearest Neighbor* dengan nilai K=15 menghasilkan nilai akurasi, sensitivitas, dan spesifisitas yang paling optimal pada

seluruh iterasi pengujian. Hasil evaluasi performa model dengan confusion matrix disajikan pada Tabel 3.

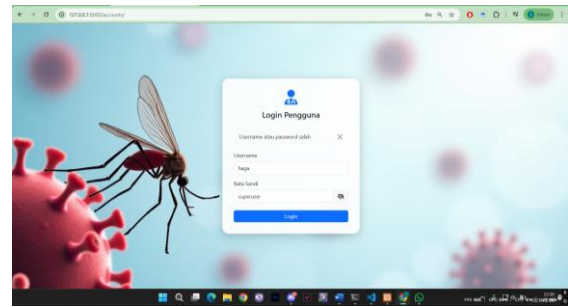
Tabel 2. Hasil Evaluasi Performa Model

	Akurasi	Sensitivitas	Spesifisitas
Iterasi 1	1	1	1
Iterasi 2	0,9683	0,9744	0,9583
Iterasi 3	1	1	1
Rata-Rata	0,9894	0,9915	0,9861

Berdasarkan nilai rata-rata tersebut, dapat disimpulkan bahwa performa model diagnosis pada dataset penelitian ini berada pada kondisi optimal, hal tersebut ditunjukkan oleh nilai akurasi, sensitivitas, dan spesifisitas yang tinggi pada seluruh evaluasi.

4.1.2 Implementasi Model Algoritma *K-Nearest Neighbor*

Implementasi sistem diagnosis demam berdarah dengue dilakukan menggunakan bahasa pemrograman *python* untuk pengolahan data dan pembangunan model *K-Nearest Neighbor*, yang diintegrasikan ke dalam aplikasi berbasis website menggunakan *framework django*. Sistem dirancang dengan pembagian peran pengguna (*role-based access*), yaitu *superuser*, *admin*, dan *user*, untuk mendukung pengelolaan dan penggunaan aplikasi secara terstruktur. Antarmuka *superuser* menyediakan akses penuh terhadap sistem, meliputi pemantauan riwayat diagnosis, pengelolaan data pengguna, serta pengawasan aktivitas sistem melalui dashboard dan log aktivitas. Antarmuka *admin* berfungsi untuk membantu pengelolaan data user, dan pemantauan hasil diagnosis dengan kewenangan yang lebih terbatas dibandingkan *superuser*. Sementara itu, antarmuka user difokuskan pada fitur diagnosis, penyajian hasil diagnosis, serta penyimpanan dan peninjauan riwayat diagnosis *user* itu sendiri.



Gambar 4 Antarmuka Aplikasi Diagnosis Demam Berdarah Dengue

4.1.3 Pengujian *Black Box*

Pengujian *black box* dilakukan untuk memvalidasi fungsi-fungsi sistem berdasarkan kesesuaian antara fitur yang telah dirancang dan kinerja aplikasi sesuai spesifikasi fungsional, tanpa meninjau struktur internal atau kode program. Pengujian ini diterapkan untuk memastikan seluruh fitur sistem beroperasi secara konsisten dengan rancangan yang telah ditetapkan. Pengujian dilakukan pada seluruh antarmuka pengguna, meliputi *superuser*, *admin*, dan *user*. Pada antarmuka *superuser* dan *admin*, pengujian mencakup fungsi autentikasi, pengelolaan data pengguna, pemantauan riwayat diagnosis, serta pencatatan aktivitas sistem sesuai dengan hak akses masing-masing. Sementara itu, pengujian pada antarmuka *user* difokuskan pada kemudahan penggunaan, responsivitas sistem, proses input data diagnosis, serta penyajian hasil dan riwayat diagnosis.

4.1.4 Hasil Pengujian *Black Box*

Pengujian sistem dilakukan menggunakan metode *black box testing* untuk memastikan bahwa seluruh fungsi pada sistem diagnosis penyakit demam berdarah dengue berjalan sesuai dengan kebutuhan fungsional yang telah ditetapkan. Pengujian ini berfokus pada kesesuaian keluaran sistem terhadap masukan yang diberikan oleh pengguna tanpa memperhatikan struktur internal sistem. Selanjutnya, hasil pengujian fungsional sistem disajikan berdasarkan peran pengguna sebagai berikut:

a. Hasil Pengujian Antarmuka Superuser

N o	Skenario Pengujian	Hasil Pengujian	Kesimpula n
1	<i>Login Superuser</i>	Sistem menerima data autentikasi yang valid dan menolak data tidak valid.	Valid
2.	Fitur manajemen Superuser	Sistem menampilkan dan mengelola dashboard, riwayat diagnosis, data pengguna, serta aktivitas pengguna secara akurat; seluruh fungsi tambah, edit, hapus, pencarian, dan filter berjalan sesuai kebutuhan.	Valid
3.	<i>Log out Superuser</i>	Sistem mengakhiri sesi pengguna dan menolak akses ke dashboard tanpa autentikasi ulang.	Valid

b. Hasil Pengujian Antarmuka Admin

N o	Skenario Pengujian	Hasil Pengujian	Kesimpula n
1	<i>Login Admin</i>	Sistem menerima data autentikasi yang valid dan menolak data tidak valid.	Valid
2.	Fitur manajemen Admin	Sistem menampilkan dan mengelola dashboard, riwayat diagnosis, data pengguna, serta aktivitas pengguna secara akurat; seluruh fungsi detail, tambah, edit, hapus, pencarian, filter, serta visualisasi grafik berjalan sesuai kebutuhan fungsional sistem.	
3.	<i>Log out Admin</i>	Sistem mengakhiri sesi admin dan menolak akses ke dashboard tanpa autentikasi ulang.	Valid

c. Hasil Pengujian Antarmuka User

N o	Skenario Pengujian	Hasil Pengujian	Kesimpulan
1	<i>Login User</i>	Sistem menerima data autentikasi yang valid dan menolak data tidak valid.	Valid
2.	Fitur manajemen User	Sistem menampilkan dashboard, melakukan diagnosis, menampilkan hasil dan riwayat diagnosis, serta halaman informasi aplikasi secara akurat; seluruh fungsi tambah, detail, hapus, pencarian, filter, dan visualisasi grafik berjalan sesuai dengan kebutuhan fungsional sistem.	Valid
3.	<i>Log out User</i>	Sistem mengakhiri sesi user dan menolak akses ke sistem tanpa autentikasi ulang.	Valid

Berdasarkan hasil pengujian menggunakan metode *black box testing*, seluruh fungsi pada antarmuka *superuser*, *admin*, dan *user* telah

berjalan sesuai dengan kebutuhan fungsional sistem. Proses autentikasi, pengelolaan data, serta pengakhiran sesi pengguna dapat dilakukan dengan baik tanpa ditemukan kesalahan fungsional. Secara keseluruhan, sistem menunjukkan kinerja yang stabil, responsif, dan siap digunakan.

4.1.5 Karakteristik Aplikasi Website

Aplikasi sistem diagnosis demam berdarah dengue dikembangkan dalam bentuk aplikasi website untuk membantu tenaga medis dalam melakukan diagnosis secara cepat dan akurat. Aplikasi ini menggunakan algoritma *K-Nearest Neighbor* sebagai metode diagnosis berdasarkan gejala klinis dan hasil pemeriksaan laboratorium pasien. Sistem dibangun menggunakan bahasa pemrograman *python* dengan *framework django* dengan antarmuka berbasis *HTML dan CSS*. Aplikasi telah melalui pengujian *black box* untuk memastikan seluruh fitur berfungsi sesuai kebutuhan pengguna. Dengan penerapan algoritma *K-Nearest Neighbor*, aplikasi ini mampu menghasilkan diagnosis yang akurat dan konsisten, sehingga sangat membantu dalam proses pengambilan keputusan diagnosis secara cepat dan tepat. Selain itu, aplikasi sistem diagnosis penyakit demam berdarah dengue juga dapat memberikan kontribusi dalam meningkatkan efisiensi kerja tenaga medis serta mendukung pengelolaan data pasien secara sistematis dan terstruktur.

5. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa algoritma *K-Nearest Neighbor* memiliki performa yang sangat baik dalam mendiagnosis penyakit demam berdarah dengue menggunakan data historis pasien RSUD Kabupaten Klungkung periode 2021–2023. Penelitian ini menggunakan 312 data pasien dengan penerapan metode *Aggregated Hold-Out* sebanyak tiga iterasi dan rasio pembagian data latih sebesar 80% serta data uji sebesar 20% guna memperoleh hasil evaluasi yang representatif.

Evaluasi performa model menggunakan confusion matrix menunjukkan nilai rata-rata akurasi sebesar 0,9894, sensitivitas sebesar

0,9915, dan spesifisitas sebesar 0,9861, yang mengindikasikan bahwa model mampu mengklasifikasikan pasien positif dan negatif demam berdarah dengue dengan tingkat kesalahan yang sangat rendah. Hasil tersebut menegaskan efektivitas algoritma *K-Nearest Neighbor* dalam mendukung proses diagnosis penyakit demam berdarah dengue.

Model yang dikembangkan selanjutnya diimplementasikan ke dalam sistem diagnosis berbasis website menggunakan bahasa pemrograman *Python* dengan *framework Django*. Pengujian sistem menggunakan metode *black box testing* menunjukkan bahwa seluruh fungsionalitas aplikasi berjalan sesuai dengan rancangan. Dengan demikian, sistem yang dihasilkan berpotensi menjadi alat bantu bagi tenaga medis dalam melakukan diagnosis penyakit demam berdarah dengue secara cepat dan akurat.

UCAPAN TERIMA KASIH

Terima kasih kepada pihak RSUD Kabupaten Klungkung yang telah bersedia meluangkan waktu dan menyediakan data untuk keperluan penelitian ini.

DAFTAR PUSTAKA

- [1] F. Abdul Masykur, "Hubungan Antara Lama Demam dengan Hasil Pemeriksaan Profil Darah pada Pasien Demam Berdarah Dengue," *Jurnal Ilmu Medis Indonesia*, vol. 1, no. 2, pp. 53–58, Mar. 2022, doi: 10.35912/jimi.v1i2.912.
- [2] Ariyanti Melly and Anggraini Debie, "Aspek Klinis dan Pemeriksaan Laboratorium untuk Diagnosis Demam Berdarah Dengue," vol. 1, 2022.
- [3] World Health Organization (WHO), "Dengue - Global situation," World Health Organization (WHO). Accessed: Apr. 27, 2024. [Online]. Available: <https://www.who.int/emergencies/disease-outbreak-news/item/2023-DON498>
- [4] P2PM KEMENKES RI, "Infografis Demam Berdarah Dengue di Indonesia." Accessed: Apr. 30, 2024. [Online]. Available: <https://p2pm.kemkes.go.id/publikasi/infografis/info-dbd-terkini-hingga-minggu-ke-12>
- [5] Badan Pusat Statistik Provinsi Bali, "Jumlah Kasus Penyakit Menurut Jenis Penyakit Menurut Kabupaten/Kota di Provinsi Bali, 2019-2023." Accessed: May 03, 2024. [Online]. Available: <https://bali.bps.go.id/indicator/12/28/1/proyek-si-penduduk-provinsi-bali-menurut-jenis-kelamin-dan-kabupaten-kota.html>
- [6] N. Safitri and C. Bella, "Penggunaan Algoritma Apriori dalam Penerapan Data Mining untuk Analisis Pola Pembelian Pelanggan (Studi Kasus: Toko Diengva Bandar Jaya)," 2022.
- [7] K. Huda, S. D. Pohan, and Y. Herlina, "Penerapan Pembobotan Term Frequency-Inverse Document Frequency dan Algoritma K-Nearest Neighbor untuk Analisis Ulasan Hotel di Situs TripAdvisor," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 12, no. 3, Aug. 2024, doi: 10.23960/jitet.v12i3.4800.
- [8] R. Amilia and E. Prasetyo, "Klasifikasi Diagnosa Penyakit Demam Berdarah Dengue pada Anak Menggunakan Metode K-Nearest Neighbor Studi Kasus Rumah Sakit PKU Muhammadiyah Ujung Pangkah Gresik," *INDEXIA: Informatic and Computational Intelegent Journal*, vol. 2, no. 2, pp. 1–10, 2020.
- [9] E. Nugraheni, D. Rizqoh, and M. Sundari, "Manifestasi Klinis Demam Berdarah Dengue (DBD)," *Jurnal Kedokteran dan Kesehatan : Publikasi Ilmiah Fakultas Kedokteran Universitas Sriwijaya*, vol. 10, no. 3, pp. 267–274, Aug. 2023, doi: 10.32539/jkk.v10i3.21425.
- [10] I. Made Kusuma Wijaya, I. Adi Wibowo, and K. Indra Purnomo, *Pelatihan Siswa Pemantau Jentik di Kecamatan Sukasada*. 2021.
- [11] S. Fatimah, "Evaluasi Ketepatan Penulisan Diagnosis Akhir berdasarkan Kebijakan Standar Kode Diagnosa, Prosedur, Definisi, Simbul, dan Singkatan di RS Premier Surabaya (Studi Lapangan)," *Stikes Yayasan RS DR. Soetomo, Surabaya*, 2021.
- [12] G. A. Pradnyana, I. K. A. Suryantara, and I. G. M. Darmawiguna, "Impression Classification of Endek (Balinese Fabric) Image Using K-Nearest Neighbors Method," *Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control*, pp. 213–220, Apr. 2018, doi: 10.22219/kinetik.v3i3.611.
- [13] W. Wahyono, I. N. P. Trisna, S. L. Sariwening, M. Fajar, and D. Wijayanto, "Comparison of distance measurement on k-nearest neighbour in textual data classification," *Jurnal Teknologi dan Sistem Komputer*, vol. 8, no. 1, pp. 54–58, Jan. 2020, doi: 10.14710/jtsiskom.8.1.2020.54-58.
- [14] I. P. Suwiryana, I. M. Candiasa, and G. R. Dantes, "Evaluation of ATM Location Placement Using the K-Means Clustering in

- BNI Denpasar Regional Office,” *Journal of Computer Networks, Architecture and High Performance Computing*, vol. 4, no. 2, pp. 158–168, Jul. 2022, doi: 10.47709/cnahpc.v4i2.1580.
- [15] G. Maillard, S. Arlot, and M. Lerasle, “Aggregated Hold-Out,” 2021. [Online]. Available: <http://jmlr.org/papers/v22/19-624.html>.
- [16] M. Artama, I. N. Sukajaya, and G. Indrawan, “Classification of official letters using TF-IDF method,” in *Journal of Physics: Conference Series*, Institute of Physics Publishing, Jun. 2020. doi: 10.1088/1742-6596/1516/1/012001.
- [17] I. M. Candiasa, N. M. Sri Mertasar, G. Ratnaya, N. Santiyadnya, and K. Widiartin, “Pelatihan Koding dan Kecerdasan Artifisial untuk Guru-Guru Penggerak pada Balai Guru Penggerak Provinsi Bali,” 2025.
- [18] N. Tri *et al.*, “Rekayasa Sistem Informasi Dievaluasi berbasis Model Waterfall: Eksperimen Keandalan Black-Box dan Optimalisasi UX melalui UEQ,” *Kumpulan Artikel Mahasiswa Pendidikan Teknik Informatika (KARMAPATI)*, vol. 14, no. 2, 2025.
- [19] I. Bagus, A. Indra Iswara, G. Anandita, and M. Dahul, “Comparative Analysis of Naïve Bayes and K-Nearest Neighbor (KNN) Algorithms in Stroke Classification,” *Architecture and High Performance Computing*, vol. 6, no. 3, 2024, doi: 10.47709/cnape.v6i3.4395.
- [20] D. Valero-Carreras, J. Alcaraz, and M. Landete, “Comparing two SVM models through different metrics based on the confusion matrix,” *Comput Oper Res*, vol. 152, Apr. 2023, doi: 10.1016/j.cor.2022.106131.