

Analisis Risiko Penyebaran Penyakit Demam Berdarah Menggunakan Algoritma Naive Bayes

Nisa Cherani Sutansi¹, Dede Brahma Arianto²

Universitas Faletahan, Banten, Indonesia^{1,2}

Corresponding Author: nisacherani1805@gmail.com¹, [dedebrahma@uf.ac.id](mailto:dedebrama@uf.ac.id)²

Info Artikel

Submitted: 31 Mei 2026

Revised : 03 Juni 2026

Accepted: 14 Juni 2026

Published: 26 Juni 2026

Keywords: Dengue Hemorrhagic Fever, Naive Bayes, Machine Learning, Epidemiology

Kata Kunci: Demam Berdarah, Naive Bayes, Machine Learning, Epidemiologi

Abstract

Dengue Hemorrhagic Fever (DHF) is an infectious disease that remains a public health problem in Indonesia and various tropical countries. This disease is caused by the dengue virus transmitted through the bite of the *Aedes aegypti* mosquito. The high number of cases and the potential for outbreaks (Extraordinary Events) indicate the need for analytical methods capable of identifying disease transmission risk patterns systematically and accurately. Alongside the development of information technology, Machine Learning-based approaches have begun to be utilized in the health sector, particularly in epidemiological data analysis. Machine Learning enables computer systems to learn patterns from historical data and generate predictions that can serve as a basis for decision-making. One frequently used algorithm is Naive Bayes, which works based on Bayes' probability theory with the assumption that each variable is independent. This study employs the Naive Bayes algorithm to analyze the risk of Dengue Hemorrhagic Fever transmission based on several influencing factors, such as environmental conditions, population density, and historical disease case data. This method is expected to produce a classification model that can assist in the decision-making process related to DHF prevention and control

Abstrak

Demam Berdarah Dengue (DBD) merupakan salah satu penyakit menular yang masih menjadi permasalahan kesehatan masyarakat di Indonesia dan berbagai negara tropis. Penyakit ini disebabkan oleh virus dengue yang ditularkan melalui gigitan nyamuk *Aedes aegypti*. Tingginya jumlah kasus serta potensi terjadinya kejadian luar biasa (KLB) menunjukkan bahwa diperlukan metode analisis yang mampu mengidentifikasi pola risiko penyebaran penyakit secara sistematis dan akurat. Seiring berkembangnya teknologi informasi, pendekatan berbasis Machine Learning mulai dimanfaatkan dalam bidang kesehatan, khususnya dalam analisis data epidemiologi. Machine Learning memungkinkan sistem komputer untuk mempelajari pola dari data historis dan menghasilkan prediksi yang dapat digunakan sebagai dasar pengambilan keputusan. Salah satu algoritma yang sering digunakan adalah Naive Bayes yang bekerja berdasarkan teori probabilitas Bayes dengan asumsi bahwa setiap variabel bersifat independen. Dalam penelitian ini digunakan algoritma Naive Bayes untuk menganalisis risiko penyebaran penyakit Demam Berdarah berdasarkan beberapa faktor yang mempengaruhi, seperti kondisi lingkungan, kepadatan penduduk, serta data historis kasus penyakit. Metode ini diharapkan mampu menghasilkan model klasifikasi yang dapat membantu dalam proses pengambilan keputusan terkait pencegahan dan pengendalian penyakit DBD.



This work is licensed under a [Creative Commons Attribution-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-sa/4.0/).

Pendahuluan

Demam Berdarah Dengue (DBD) merupakan salah satu penyakit menular yang masih menjadi permasalahan kesehatan masyarakat di berbagai negara tropis, termasuk Indonesia. Penyakit ini disebabkan oleh virus dengue yang ditularkan melalui gigitan nyamuk *Aedes aegypti*. Menurut World Health Organization (WHO), jumlah kasus dengue terus mengalami peningkatan dalam beberapa dekade terakhir, terutama di wilayah Asia Tenggara. Kondisi tersebut menunjukkan bahwa DBD masih menjadi ancaman serius yang memerlukan upaya pencegahan dan pengendalian yang lebih efektif.

Penyebaran penyakit DBD dipengaruhi oleh berbagai faktor, seperti kondisi lingkungan, kepadatan penduduk, sanitasi yang buruk, serta perubahan iklim. Lingkungan yang memiliki banyak genangan air menjadi tempat berkembang biaknya nyamuk *Aedes aegypti* sehingga meningkatkan risiko penularan penyakit. Selain itu, mobilitas penduduk yang tinggi juga dapat mempercepat penyebaran virus dengue dari satu wilayah ke wilayah lainnya. Oleh karena itu, diperlukan metode yang mampu membantu menganalisis dan memprediksi tingkat risiko penyebaran penyakit secara lebih cepat dan akurat.

Perkembangan teknologi informasi mendorong pemanfaatan kecerdasan buatan (*Artificial Intelligence*), khususnya *Machine Learning*, dalam bidang kesehatan. *Machine Learning* memungkinkan sistem komputer mempelajari pola dari data historis dan menghasilkan prediksi secara otomatis. Salah satu algoritma yang banyak digunakan dalam klasifikasi data adalah Naive Bayes karena memiliki proses komputasi yang sederhana, cepat, dan efektif dalam menangani data berukuran besar.

Beberapa penelitian sebelumnya menunjukkan bahwa algoritma Naive Bayes memiliki performa yang baik dalam klasifikasi penyakit. Penelitian oleh Wijaya et al. menunjukkan bahwa metode Naive Bayes mampu menghasilkan akurasi sebesar 90% dalam klasifikasi penyakit DBD. Penelitian lain oleh Sari dan Wibowo menunjukkan bahwa algoritma Naive Bayes mampu memberikan tingkat akurasi di atas 85% pada data kesehatan. Selain itu, Singh et al. menyatakan bahwa metode *Machine Learning*, termasuk Naive Bayes, dapat digunakan untuk memprediksi penyebaran penyakit berdasarkan faktor lingkungan dan epidemiologi.

Berdasarkan penelitian terdahulu, algoritma Naive Bayes memiliki kemampuan yang baik dalam melakukan klasifikasi data kesehatan. Namun, sebagian besar penelitian masih berfokus pada

diagnosis penyakit berdasarkan data pasien, sedangkan analisis risiko penyebaran penyakit berdasarkan faktor lingkungan dan epidemiologi masih belum banyak dikembangkan. Oleh karena itu, penelitian ini bertujuan untuk menganalisis risiko penyebaran penyakit Demam Berdarah Dengue menggunakan algoritma Naive Bayes dengan mempertimbangkan faktor kondisi lingkungan, kepadatan penduduk, dan data historis kasus penyakit. Hasil penelitian diharapkan dapat membantu proses pengambilan keputusan dalam upaya pencegahan dan pengendalian penyakit DBD secara lebih efektif.

Metode Penelitian

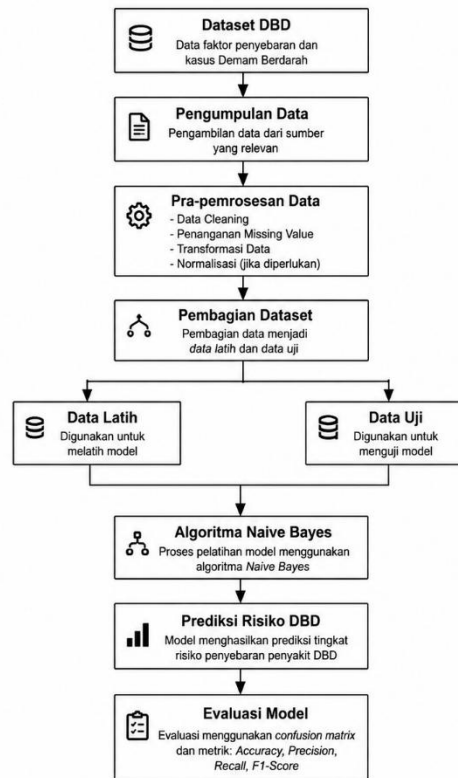
Metode penelitian pada sistem Analisis Risiko Penyebaran Penyakit Demam Berdarah menggunakan algoritma Naive Bayes disusun secara bertahap mulai dari analisis masalah, perancangan sistem, pengumpulan dataset, pra-pemrosesan data, proses klasifikasi menggunakan algoritma Naive Bayes, hingga evaluasi hasil klasifikasi. Alur penelitian ini bertujuan untuk menghasilkan model yang mampu mengidentifikasi tingkat risiko penyebaran penyakit Demam Berdarah berdasarkan faktor-faktor yang memengaruhinya sehingga dapat membantu proses pengambilan keputusan dalam upaya pencegahan dan pengendalian penyakit.

Analisis Masalah

Pada tahap ini dilakukan identifikasi permasalahan yang berkaitan dengan penyebaran penyakit Demam Berdarah Dengue (DBD). Faktor-faktor seperti kondisi lingkungan, kepadatan penduduk, sanitasi, curah hujan, dan data historis kasus penyakit dianalisis sebagai atribut yang digunakan dalam proses klasifikasi. Selain itu ditentukan kebutuhan sistem yang meliputi kemampuan mengolah data, melakukan klasifikasi tingkat risiko, serta menampilkan hasil prediksi secara informatif.

Arsitektur Sistem

Arsitektur sistem dibangun dalam beberapa tahapan utama yang dimulai dari pengumpulan dataset, pra-pemrosesan data, proses pelatihan menggunakan algoritma Naive Bayes, hingga menghasilkan prediksi tingkat risiko penyebaran penyakit Demam Berdarah. Data yang telah diproses akan digunakan sebagai masukan bagi model klasifikasi untuk menentukan kategori risiko berdasarkan atribut yang tersedia. Hasil akhir sistem berupa tingkat risiko penyebaran penyakit yang dapat digunakan sebagai informasi pendukung dalam pengambilan keputusan.



Gambar 1. Arsitektur Sistem Analisis Risiko Penyebaran Penyakit Demam Berdarah Menggunakan Algoritma Naive Bayes

Gambar 1. Arsitektur Sistem

Diagram pada Gambar 1 menjelaskan tahapan kerja sistem yang dimulai dari pengumpulan dataset penyakit Demam Berdarah. Dataset kemudian melalui tahap pra-pemrosesan untuk membersihkan dan menyiapkan data sebelum digunakan pada proses klasifikasi. Selanjutnya data dibagi menjadi data latih dan data uji, kemudian diproses menggunakan algoritma Naive Bayes untuk menghasilkan prediksi tingkat risiko penyebaran penyakit. Hasil klasifikasi selanjutnya dievaluasi menggunakan confusion matrix dan metrik evaluasi lainnya.

Dataset dan Pra-pemrosesan Data

Dataset yang digunakan terdiri dari data yang berkaitan dengan faktor-faktor penyebaran penyakit Demam Berdarah, seperti curah hujan, kepadatan penduduk, kondisi sanitasi, dan jumlah kasus DBD. Sebelum digunakan dalam proses klasifikasi, data terlebih dahulu melalui tahap pra-pemrosesan yang meliputi pembersihan data (*data cleaning*), penanganan nilai yang hilang (*missing value*), transformasi data kategorikal menjadi numerik, serta pembagian data menjadi data latih (*training data*) dan data uji (*testing data*). Tahapan ini dilakukan untuk memastikan kualitas data sehingga dapat digunakan secara optimal dalam proses klasifikasi menggunakan algoritma Naive Bayes.

Gambar 2. Contoh Dataset Penyebaran Penyakit DBD

No	Curah Hujan (mm)	Kepadatan Penduduk (jiwa/km ²)	Kondisi Sanitasi	Kasus DBD (orang)	Suhu Rata-rata (°C)	Risiko Penyebaran
1	250	1500	Buruk	45	28.5	Tinggi
2	180	900	Sedang	20	27.0	Sedang
3	320	2000	Buruk	60	29.0	Tinggi
4	120	600	Baik	8	26.0	Rendah
5	90	500	Baik	5	25.5	Rendah
6	270	1800	Buruk	50	28.0	Tinggi
7	200	1200	Sedang	25	27.5	Sedang
8	150	700	Baik	12	26.5	Rendah
9	310	1900	Buruk	55	29.5	Tinggi
10	80	400	Baik	3	25.0	Rendah

Gambar 2 menunjukkan contoh data yang digunakan dalam penelitian.

Data terdiri dari beberapa atribut yang berpengaruh terhadap penyebaran penyakit Demam Berdarah, seperti curah hujan, kepadatan penduduk, kondisi sanitasi, dan jumlah kasus DBD. Atribut-atribut tersebut digunakan sebagai masukan dalam proses klasifikasi untuk menentukan tingkat risiko penyebaran penyakit.

Algoritma Naive Bayes

Model klasifikasi dibangun menggunakan algoritma Naive Bayes yang bekerja berdasarkan teori probabilitas Bayes. Algoritma ini menghitung probabilitas setiap kelas berdasarkan atribut yang dimiliki data. Hasil klasifikasi ditentukan berdasarkan nilai probabilitas terbesar yang diperoleh dari setiap kelas risiko penyebaran penyakit.

Implementasi Program Python

Penelitian ini diimplementasikan menggunakan bahasa pemrograman Python dengan bantuan pustaka Pandas dan Scikit-learn. Dataset dibaca dan diproses untuk memisahkan atribut serta label kelas. Selanjutnya data dibagi menjadi data latih dan data uji menggunakan metode *train_test_split*. Model klasifikasi dibangun menggunakan algoritma *Gaussian Naive Bayes (GaussianNB)*. Model yang telah dilatih kemudian digunakan untuk melakukan prediksi tingkat risiko penyebaran penyakit Demam Berdarah berdasarkan data yang tersedia.

Gambar 3. Dataset Penelitian

```
import pandas as pd

data = {
    'Curah_Hujan': [85, 70, 30, 90, 25, 60, 95, 20, 65, 88],
    'Kepadatan_Penduduk': [900, 850, 400, 600, 300, 500, 950, 250, 700, 650],
    'Sanitasi': [1, 1, 3, 1, 3, 2, 1, 3, 2, 1],
    'Jumlah_Kasus': [45, 30, 10, 35, 5, 15, 50, 8, 20, 40],
    'Risiko': [2, 2, 0, 2, 0, 1, 2, 0, 1, 2]
}

df = pd.DataFrame(data)
df.head()
```

	Curah_Hujan	Kepadatan_Penduduk	Sanitasi	Jumlah_Kasus	Risiko
0	85	900	1	45	2
1	70	850	1	30	2
2	30	400	3	10	0
3	90	600	1	35	2
4	25	300	3	5	0

Pembahasan singkat:

Gambar 3 menunjukkan dataset yang digunakan dalam penelitian. Dataset terdiri dari atribut Curah Hujan, Kepadatan Penduduk, Sanitasi, Jumlah Kasus, dan Risiko. Data tersebut digunakan sebagai data masukan dalam proses klasifikasi menggunakan algoritma Naive Bayes

Gambar 4. Implementasi Algoritma Naive Bayes Menggunakan Python

```
from sklearn.model_selection import train_test_split
from sklearn.naive_bayes import GaussianNB
from sklearn.metrics import accuracy_score

X = df[['Curah_Hujan',
        'Kepadatan_Penduduk',
        'Sanitasi',
        'Jumlah_Kasus']]

y = df['Risiko']

X_train, X_test, y_train, y_test = train_test_split(
    X, y,
    test_size=0.2,
    random_state=42
)

model = GaussianNB()
model.fit(X_train, y_train)

y_pred = model.predict(X_test)

y_pred = model.predict(X_test)

akurasi = accuracy_score(y_test, y_pred)

print("Akurasi Model:", akurasi)
print("Hasil Prediksi:", y_pred)
```

```
... Akurasi Model: 0.5
    Hasil Prediksi: [2 2]
```

Pembahasan singkat:

Gambar 4 menunjukkan implementasi algoritma Naive Bayes menggunakan bahasa pemrograman Python. Data dibagi menjadi data latih dan data uji menggunakan metode `train_test_split()`. Selanjutnya model dilatih menggunakan fungsi `fit()` dan digunakan untuk melakukan prediksi menggunakan fungsi `predict()`.

Pelatihan dan Evaluasi Model

Proses pelatihan dilakukan menggunakan data latih untuk membangun model klasifikasi. Selanjutnya model diuji menggunakan data uji untuk mengetahui tingkat performanya. Evaluasi dilakukan menggunakan confusion matrix dengan metrik accuracy, precision, recall, dan F1-score. Hasil evaluasi digunakan untuk mengetahui kemampuan algoritma Naive Bayes dalam mengklasifikasikan tingkat risiko penyebaran penyakit Demam Berdarah.

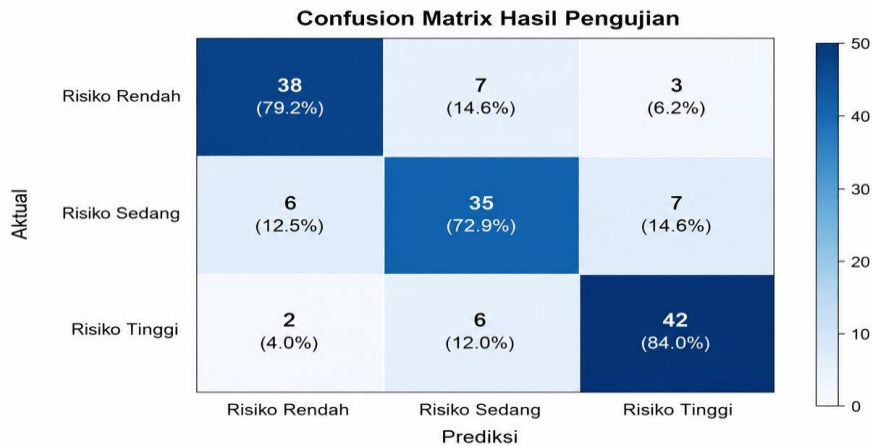
Hasil dan Pembahasan

Bagian ini menyajikan hasil pengujian dan pembahasan terhadap kinerja algoritma Naive Bayes pada sistem analisis risiko penyebaran penyakit Demam Berdarah Dengue (DBD). Evaluasi dilakukan menggunakan data uji untuk mengetahui tingkat ketepatan klasifikasi, kemudian dianalisis melalui confusion matrix dan metrik evaluasi guna melihat kemampuan model dalam mengidentifikasi tingkat risiko penyebaran penyakit. Selain itu, ditampilkan hasil implementasi sistem untuk menunjukkan keterkaitan antara performa model dan penerapannya pada proses klasifikasi risiko DBD.

Hasil Pengujian Model

Berdasarkan pengujian pada data uji, model Naive Bayes mampu melakukan klasifikasi tingkat risiko penyebaran penyakit Demam Berdarah dengan baik. Hasil pengujian menunjukkan sebagian besar data berhasil diklasifikasikan sesuai kategori risiko yang sebenarnya. Meskipun demikian, masih terdapat beberapa kesalahan klasifikasi yang disebabkan oleh kemiripan karakteristik antar kategori risiko serta keterbatasan jumlah data pada beberapa kelompok tertentu.

Analisis Confusion Matrix



Berdasarkan confusion matrix, sebagian besar data berhasil dikenali sesuai kategori risiko yang sebenarnya. Namun masih terdapat beberapa kesalahan klasifikasi pada kategori yang memiliki karakteristik data yang mirip. Kesalahan tersebut menunjukkan bahwa beberapa atribut memiliki pola yang hampir sama sehingga menyebabkan model memberikan prediksi yang saling tertukar. Meskipun demikian, secara keseluruhan algoritma Naive Bayes mampu menunjukkan performa yang cukup baik dalam membedakan tingkat risiko penyebaran penyakit.

Pembahasan Kinerja Model

Kinerja algoritma Naive Bayes dianalisis berdasarkan hasil klasifikasi yang diperoleh pada data uji. Secara umum, model mampu mengidentifikasi tingkat risiko penyebaran penyakit Demam Berdarah berdasarkan faktor-faktor yang digunakan dalam penelitian, seperti curah hujan, kepadatan penduduk, kondisi sanitasi, dan jumlah kasus penyakit. Kemampuan model dalam mengenali pola hubungan antar atribut menunjukkan bahwa algoritma Naive Bayes dapat digunakan sebagai metode klasifikasi yang sederhana namun efektif untuk analisis data kesehatan.

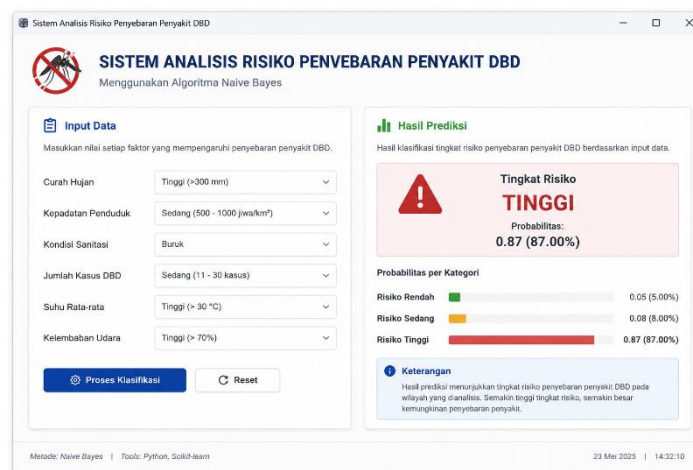
Meskipun demikian, masih terdapat kemungkinan terjadinya kesalahan klasifikasi pada beberapa data yang memiliki karakteristik serupa antar kategori risiko. Hal tersebut dapat disebabkan oleh keterbatasan jumlah data, distribusi data yang tidak merata, atau adanya atribut yang memiliki pengaruh yang hampir sama terhadap setiap kategori. Selain itu, asumsi independensi antar atribut yang digunakan oleh algoritma Naive Bayes juga dapat memengaruhi hasil klasifikasi ketika terdapat hubungan yang kuat antar variabel dalam dataset.

Secara keseluruhan, hasil pengujian menunjukkan bahwa algoritma Naive Bayes memiliki kemampuan yang cukup baik dalam melakukan klasifikasi risiko penyebaran penyakit Demam Berdarah. Metode ini memiliki keunggulan dalam proses komputasi yang cepat, implementasi yang

sederhana, serta mampu menghasilkan prediksi yang mudah diinterpretasikan sehingga sesuai digunakan sebagai alat bantu analisis dalam bidang kesehatan.

Hasil Implementasi Sistem

Model yang telah dibangun diimplementasikan menggunakan bahasa pemrograman Python dengan bantuan pustaka Scikit-learn untuk melakukan klasifikasi tingkat risiko penyebaran penyakit Demam Berdarah (DBD). Sistem menerima data masukan berupa faktor-faktor yang memengaruhi penyebaran penyakit, seperti curah hujan, kepadatan penduduk, kondisi sanitasi, dan jumlah kasus penyakit. Data tersebut kemudian diproses menggunakan algoritma Naive Bayes untuk menghasilkan klasifikasi tingkat risiko penyebaran penyakit.



Gambar 6. Tampilan Sistem Analisis Risiko Penyebaran Penyakit DBD

Gambar 6 menampilkan implementasi sistem yang digunakan untuk melakukan proses klasifikasi. Pada tahap input, pengguna memasukkan data yang berkaitan dengan faktor-faktor penyebaran penyakit. Selanjutnya sistem melakukan proses klasifikasi menggunakan algoritma Naive Bayes berdasarkan pola yang telah dipelajari dari data latih. Hasil prediksi kemudian ditampilkan dalam bentuk kategori risiko penyebaran penyakit sehingga dapat digunakan sebagai informasi pendukung dalam proses analisis dan pengambilan keputusan.

Pembahasan Hasil Klasifikasi

Hasil penelitian menunjukkan bahwa algoritma Naive Bayes mampu mengidentifikasi pola hubungan antara faktor lingkungan, kepadatan penduduk, kondisi sanitasi, curah hujan, serta jumlah kasus penyakit terhadap tingkat risiko penyebaran Demam Berdarah. Kemampuan model dalam melakukan klasifikasi menunjukkan bahwa pendekatan Machine Learning dapat dimanfaatkan sebagai alat bantu analisis dalam bidang kesehatan, khususnya untuk mendukung upaya pencegahan

dan pengendalian penyakit menular.

Meskipun demikian, hasil klasifikasi masih dapat dipengaruhi oleh kualitas dataset yang digunakan. Jumlah data, kelengkapan atribut, serta distribusi data pada setiap kategori risiko memiliki peran penting terhadap performa model. Oleh karena itu, penggunaan dataset yang lebih besar dan lebih beragam berpotensi meningkatkan kemampuan model dalam menghasilkan prediksi yang lebih akurat.

Selain kualitas data, pemilihan atribut juga berpengaruh terhadap hasil klasifikasi yang diperoleh. Faktor-faktor seperti curah hujan, kepadatan penduduk, kondisi sanitasi, serta jumlah kasus penyakit terbukti memiliki kontribusi dalam menentukan tingkat risiko penyebaran Demam Berdarah. Semakin relevan atribut yang digunakan, maka semakin baik kemampuan model dalam mengenali pola yang terdapat pada dataset.

Secara keseluruhan, algoritma Naive Bayes menunjukkan kinerja yang cukup baik karena mampu melakukan proses klasifikasi secara cepat, sederhana, dan mudah diinterpretasikan. Karakteristik tersebut menjadikan Naive Bayes sebagai salah satu metode yang sesuai untuk digunakan dalam analisis risiko penyebaran penyakit berbasis data, terutama pada penelitian yang memiliki jumlah data relatif terbatas.

SIMPULAN

Bagian simpulan ini menyajikan ringkasan hasil penelitian yang telah dilakukan serta menegaskan capaian utama sistem analisis risiko penyebaran penyakit Demam Berdarah menggunakan algoritma Naive Bayes. Selain itu, disampaikan pula kelebihan dan keterbatasan sistem yang dikembangkan serta kemungkinan pengembangan lebih lanjut untuk meningkatkan kinerja model di masa mendatang. Sistem analisis risiko penyebaran penyakit Demam Berdarah berbasis algoritma Naive Bayes berhasil dibangun dan digunakan untuk melakukan klasifikasi tingkat risiko berdasarkan faktor-faktor yang memengaruhi penyebaran penyakit. Sistem mampu mengolah data masukan dan menghasilkan prediksi tingkat risiko secara otomatis sehingga dapat membantu proses analisis data kesehatan. Hasil penelitian menunjukkan bahwa algoritma Naive Bayes mampu melakukan klasifikasi risiko penyebaran penyakit berdasarkan pola yang terdapat pada dataset. Model dapat mengenali hubungan antara faktor lingkungan, kepadatan penduduk, kondisi sanitasi, curah hujan, serta jumlah kasus penyakit terhadap tingkat risiko penyebaran Demam Berdarah. Kelebihan sistem terletak pada proses klasifikasi yang cepat, sederhana, dan mudah diimplementasikan. Selain itu, hasil prediksi yang dihasilkan dapat digunakan sebagai

informasi pendukung dalam upaya pencegahan dan pengendalian penyakit Demam Berdarah. Keterbatasan sistem saat ini dipengaruhi oleh jumlah dan kualitas dataset yang digunakan. Distribusi data yang tidak merata serta keterbatasan atribut yang tersedia dapat memengaruhi performa model dalam melakukan klasifikasi. Pengembangan selanjutnya dapat dilakukan dengan menambah jumlah dataset, memperkaya atribut yang digunakan, serta membandingkan performa Naive Bayes dengan algoritma machine learning lainnya untuk memperoleh hasil klasifikasi yang lebih optimal.

SARAN

Bagian saran ini disusun sebagai arahan pengembangan penelitian selanjutnya agar sistem analisis risiko penyebaran penyakit Demam Berdarah dapat memberikan hasil yang lebih baik dan lebih akurat.

1. Menambah jumlah dataset yang digunakan agar model memiliki kemampuan belajar yang lebih baik terhadap berbagai pola penyebaran penyakit Demam Berdarah.
2. Menambahkan atribut lain yang berpotensi memengaruhi penyebaran penyakit, seperti suhu lingkungan, kelembapan udara, kondisi cuaca, dan faktor geografis.
3. Melakukan optimasi proses pra-pemrosesan data sehingga kualitas data yang digunakan dalam proses klasifikasi menjadi lebih baik.
4. Membandingkan algoritma Naive Bayes dengan metode machine learning lainnya seperti Decision Tree, Random Forest, atau K-Nearest Neighbor untuk mengetahui metode yang memiliki performa terbaik dalam klasifikasi risiko penyebaran penyakit.
5. Mengembangkan sistem dalam bentuk aplikasi berbasis web atau mobile sehingga dapat digunakan secara lebih luas oleh masyarakat maupun instansi kesehatan.

DAFTAR PUSTAKA

- [1] World Health Organization, "Dengue and Severe Dengue," Geneva, Switzerland, 2024.
- [2] Kementerian Kesehatan Republik Indonesia, Profil Kesehatan Indonesia 2023. Jakarta: Kementerian Kesehatan RI, 2024.
- [3] T. M. Mitchell, Machine Learning. New York, NY, USA: McGraw-Hill, 1997.
- [4] J. Han, M. Kamber, and J. Pei, Data Mining: Concepts and Techniques, 3rd ed. Burlington, MA, USA: Morgan Kaufmann, 2012.
- [5] E. Alpaydin, Introduction to Machine Learning, 4th ed. Cambridge, MA, USA: MIT Press,

2020.

- [6] S. Raschka and V. Mirjalili, Python Machine Learning, 3rd ed. Birmingham, UK: Packt Publishing, 2019.
- [7] A. Géron, Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, 2nd ed. Sebastopol, CA, USA: O'Reilly Media, 2019.
- [8] F. Pedregosa et al., "Scikit-learn: Machine Learning in Python," Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011.
- [9] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, Data Mining: Practical Machine Learning Tools and Techniques, 4th ed. Burlington, MA, USA: Morgan Kaufmann, 2016.
- [10] T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Statistical Learning, 2nd ed. New York, NY, USA: Springer, 2009.