



Analisis Perbandingan Naive Bayes, Regresi Logistik Biner, dan Support Vector Machine untuk Prediksi Kasus Demam Berdarah di Purwokerto

Rosa Ratri Kusuma Hariningsih¹ , Diwahana Mutiara Candrasari² , Endang Setyawati³ , Syamsu Wahidin⁴ , Jevon Nataniel Putra⁵ 

^{1,2,3,4,5} Sekolah Tinggi Ilmu Komputer Yos Sudarso, Indonesia

Abstract: Dengue Hemorrhagic Fever (DHF) remains a significant public health issue in Purwokerto, with the increasing number of cases influenced by environmental factors such as temperature, humidity, rainfall, and population density. Accurate and adaptive predictive methods are essential to anticipate the spread of DHF, one of which involves the application of machine learning algorithms. This study aims to compare the performance of three algorithms, namely Naïve Bayes, Binary Logistic Regression, and Support Vector Machine (SVM), in predicting DHF risk in Purwokerto. Secondary data were obtained from the Health Office, Meteorology Agency (BMKG), and Statistics Bureau (BPS), covering DHF case records and environmental factors. The analysis was conducted using a quantitative predictive approach, employing 5-Fold Cross Validation and evaluation metrics including accuracy, precision, recall, and F1-score. The results indicate that the SVM model demonstrated the highest performance with an accuracy of 82% and a high recall rate for the positive class, making it effective for DHF risk mapping. The Naïve Bayes model showed adequate sensitivity but lower precision, while the Binary Logistic Regression model produced the lowest overall performance. This study recommends SVM as the most effective algorithm to support early warning systems and risk mitigation for DHF based on environmental data in Purwokerto.

ARTIKEL HISTORI

Received 14/09/2025

Publish 19/12/2025

KEYWORDS:

Dengue Fever, Classification, Support Vector Machine, Disease Prediction

Pendahuluan

Demam Berdarah Dengue (DBD) masih menjadi salah satu tantangan kesehatan masyarakat yang signifikan di wilayah Purwokerto. Peningkatan kasus dari tahun ke tahun menunjukkan bahwa penyebaran DBD dipengaruhi oleh dinamika faktor lingkungan, terutama suhu udara, kelembapan, curah hujan, dan kepadatan penduduk (Kularatne & Dalugama, 2022; Sutriyawan et al., 2024). Hubungan antar variabel tersebut bersifat kompleks dan tidak linear, sehingga metode analisis konvensional sering kali kurang mampu menggambarkan pola penyebaran DBD secara akurat. Kondisi ini mendorong perlunya pendekatan analitis yang lebih adaptif dan sensitif terhadap perubahan variabel lingkungan.

Dalam beberapa tahun terakhir, machine learning telah banyak dimanfaatkan untuk memprediksi penyakit berbasis data historis karena kemampuannya dalam mengolah pola data yang variatif dan kompleks. Sejumlah penelitian terdahulu melaporkan bahwa algoritma seperti Naïve Bayes, Regresi Logistik, dan Support Vector Machine (SVM) mampu memberikan hasil prediktif yang menjanjikan pada berbagai kasus kesehatan. Rahman et al. (2021), misalnya, menemukan bahwa SVM memiliki kinerja yang konsisten dalam memetakan potensi penyebaran vektor *Aedes aegypti*. Sementara itu, Rocha & Giesbrecht (2022) melaporkan bahwa Naïve Bayes

Corresponding author:

Rosa Ratri Kusuma Hariningsih, Sekolah Tinggi Ilmu Komputer Yos Sudarso, Indonesia, Email: rosaratri23@gmail.com

Received : 14-09-2025, Accepted : 19-12-2025. DOI: <https://doi.org/10.31316/j.derivat.v12i3.8408>

memiliki sensitivitas tinggi dalam deteksi risiko dengue, meskipun precision-nya cenderung rendah. Di sisi lain, penelitian Gupta et al. (2023) menegaskan bahwa performa model sangat dipengaruhi karakteristik data, termasuk ketidakseimbangan kelas serta variasi faktor lingkungan.

Meskipun berbagai penelitian telah dilakukan, sebagian besar studi tersebut berfokus pada wilayah skala besar, seperti provinsi atau kota metropolitan, dan belum banyak yang mengkaji performa algoritma machine learning berdasarkan karakteristik lingkungan mikro di Purwokerto. Padahal, kondisi lokal seperti variasi curah hujan antar kecamatan atau perbedaan kepadatan penduduk dapat memengaruhi pola kejadian DBD secara signifikan. Selain itu, penelitian terdahulu cenderung menitikberatkan pada akurasi sebagai metrik utama, sementara pada konteks penyakit seperti DBD, recall terhadap kasus positif justru lebih penting untuk mendukung sistem peringatan dini dan upaya mitigasi.

Berdasarkan gap tersebut, penelitian ini dilakukan untuk mengevaluasi secara komprehensif performa tiga algoritma machine learning—Naïve Bayes, Regresi Logistik Biner, dan Support Vector Machine—dalam memprediksi risiko DBD berdasarkan variabel lingkungan di Purwokerto. Penelitian ini tidak hanya menilai akurasi model, tetapi juga precision, recall, dan F1-score untuk memberikan gambaran yang lebih menyeluruh mengenai keandalan model, terutama dalam mendeteksi kelas positif.

Dengan demikian, penelitian ini bertujuan untuk menentukan algoritma mana yang memberikan performa paling optimal untuk memprediksi risiko DBD di Purwokerto berdasarkan faktor lingkungan. Temuan dari penelitian ini diharapkan dapat menjadi dasar ilmiah bagi pengembangan sistem pendukung keputusan dan sistem peringatan dini yang lebih efektif di tingkat lokal.

Dalam klasifikasi data kesehatan, Naïve Bayes dikenal sebagai algoritma yang sederhana namun efektif, dengan asumsi independensi antar fitur (Anagora et al., n.d.). Metode ini cepat dan efisien, tetapi performanya menurun jika fitur saling berkorelasi. Regresi Logistik Biner adalah metode statistik klasik yang banyak digunakan untuk memodelkan variabel biner (Chernozhukov et al., n.d.), menawarkan interpretasi yang mudah namun terbatas pada relasi linier (Khikmah et al., 2022).

Support Vector Machine (SVM) di sisi lain merupakan algoritma klasifikasi yang mampu menangani data non-linier dengan kompleksitas tinggi melalui teknik kernel (Fadli et al., 2021). SVM dikenal memiliki akurasi tinggi dan sangat baik dalam generalisasi model (Shahzad et al., 2022), meskipun proses pelatihannya relatif lebih lama. Berbagai studi terdahulu menunjukkan bahwa dalam konteks prediksi penyakit menular, SVM sering kali menghasilkan performa lebih tinggi dibandingkan metode lainnya, terutama jika data memiliki dimensi dan variabilitas tinggi (Arhin & Gatiba, 2020).

Dengan mempertimbangkan teori dan temuan empiris tersebut, penelitian ini membandingkan ketiga algoritma untuk menentukan metode mana yang paling tepat diterapkan pada kasus DBD di Purwokerto. Fokus utama terletak pada bagaimana masing-masing model mengelola data dengan karakteristik kompleks, serta sejauh mana keandalan prediksi yang dihasilkan berdasarkan indikator evaluasi kinerja seperti akurasi, presisi, recall, dan F1-score.

Metode Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksploratif dan prediktif untuk menganalisis pola faktor lingkungan serta membangun model prediksi risiko Demam Berdarah Dengue (DBD). Metode eksploratif digunakan untuk memahami karakteristik awal data, sedangkan metode prediktif diterapkan untuk membangun dan mengevaluasi model machine learning.

Lokasi dan Sumber Data

Penelitian dilakukan pada empat kecamatan di wilayah Purwokerto, yaitu Purwokerto Selatan, Barat, Utara, dan Timur. Data sekunder dikumpulkan dari:

- a. Dinas Kesehatan: data historis kasus DBD bulanan,
- b. BMKG: data suhu udara, kelembapan, dan curah hujan.

Seluruh data mencakup periode pengamatan dari Januari 2020 hingga Desember 2023.

Variabel Penelitian

- Variabel dependen: Klasifikasi risiko DBD (0 = rendah, 1 = tinggi), yang ditetapkan berdasarkan median jumlah kasus bulanan.
- Variabel independen: suhu udara ($^{\circ}\text{C}$), kelembapan (%), curah hujan (mm), jumlah kasus bulanan, dan bulan pengamatan.

Preprocessing data

Tahapan preprocessing dilakukan untuk memastikan data siap digunakan untuk pemodelan. Langkah-langkahnya meliputi:

- Pemeriksaan kelengkapan data: *Missing values* pada curah hujan dan kelembapan ditangani menggunakan metode imputasi mean.
- Deteksi dan penanganan *outlier*: *Outlier* pada curah hujan (nilai ekstrem) dievaluasi menggunakan metode IQR, dan nilai ekstrem yang tidak wajar disesuaikan dengan batas *upper-lower fence*.
- Normalisasi data numerik: Fitur suhu, kelembapan, dan curah hujan dinormalisasi menggunakan *Min-Max Scaling* agar selaras dengan algoritma SVM dan *Logistic Regression*.
- Encoding variabel kategori: Variabel “bulan” diubah menjadi representasi numerik 1–12.
- Pembentukan label kelas: Kelas risiko dibagi dua berdasarkan median jumlah kasus (*high-risk vs low-risk*) untuk mengurangi ketidakseimbangan distribusi

Pembagian Data dan Validasi Model

Dataset dibagi menggunakan teknik 5-Fold Cross Validation, di mana seluruh data dibagi menjadi lima subset, dan proses pelatihan–pengujian dilakukan secara bergiliran. Teknik ini digunakan untuk:

- meningkatkan reliabilitas model pada dataset yang kecil,
- mencegah *overfitting* pada salah satu subset data,
- memvalidasi performa model secara lebih stabil.

Algoritma yang Digunakan

Tiga algoritma dibandingkan dalam penelitian ini, yaitu:

- Naïve Bayes* (*Multinomial*) digunakan karena kesederhanaan dan kemampuan menangani fitur dengan distribusi probabilitistik.
- Regresi Logistik Biner digunakan sebagai pembanding model statistik klasik untuk klasifikasi biner.
- Support Vector Machine* (SVM) digunakan karena efektif menangani data non-linear dan memiliki performa baik pada dataset kecil.

Proses Pelatihan dan Tuning Model

- Naïve Bayes*: tanpa parameter tuning khusus karena sifat model yang sederhana.
- Regresi Logistik: menggunakan regularisasi L2 untuk mencegah *overfitting*.
- SVM: dilakukan tuning parameter (C dan gamma) menggunakan *Grid Search* untuk memperoleh *hyperplane* optimal

Metrik Evaluasi

Model dievaluasi berdasarkan:

- Accuracy
- Precision
- Recall
- F1-Score
- Confusion Matrix

Metrik *precision-recall* diprioritaskan pada kelas positif karena penting dalam konteks deteksi dini kasus DBD. F1-score digunakan sebagai indikator keseimbangan *precision* dan *recall*.

Hasil dan Pembahasan

Analisis Statistik Deskriptif

Data lingkungan dan jumlah kasus DBD dianalisis terlebih dahulu untuk memahami pola dasar sebelum dilakukan pemodelan. Suhu udara di Purwokerto relatif stabil dengan kisaran 25,6–28,5°C, sedangkan kelembapan berada pada tingkat tinggi (81–85%). Curah hujan menunjukkan variasi yang besar, dari 1 mm hingga 695 mm, yang menggambarkan perbedaan musim yang cukup tajam. Jumlah kasus DBD per bulan berkisar antara 1 hingga 30 kasus.

Tabel berikut menyajikan ringkasan statistik deskriptif dari variabel-variabel utama yang digunakan dalam analisis:

Tabel 1. Ringkasan Statistik Deskriptif Dataset Penelitian

Variabel	Mean	Min	Max	Standar Deviasi
Bulan	6.50	1	12	3.48
Suhu(°C)	27.69	25.6	28.5	0.82
Kelembapan (%)	82.83	81	85	1.22
Curah Hujan (mm)	230.17	1	695	218.15
Jumlah Kasus DBD	9.21	1	30	6.28

Evaluasi Model Machine Learning

Hasil evaluasi tiga algoritma ditampilkan pada Tabel 4.2.

Tabel 2. Evaluasi Model

Model	Accurac y	Precision (o)	Recall (o)	F1-Score (o)	Precision (i)	Recall (i)	F1- Score (i)	Macro Avg Precision	Macro Avg Recall	Macro Avg F1- Score
SVM Tuning	0.82	0.97	0.81	0.88	0.46	0.86	0.60	0.71	0.83	0.74
Logistic Regression	0.48	0.85	0.46	0.60	0.17	0.57	0.26	0.51	0.52	0.43
Multinomial Naive Bayes	0.55	0.90	0.51	0.66	0.22	0.71	0.33	0.56	0.61	0.49

Secara umum, SVM menghasilkan performa terbaik dengan akurasi 82%, disusul Multinomial Naïve Bayes (55%), dan Regresi Logistik (48%).

Support Vector Machine (SVM)

SVM memperoleh nilai recall tinggi untuk kelas positif (0.86), yang sangat penting dalam konteks prediksi penyakit. Hal ini menunjukkan bahwa model mampu menangkap pola variabel lingkungan yang berasosiasi dengan peningkatan kasus DBD. Temuan ini konsisten dengan penelitian Rahman et al. (2021) yang menyebutkan bahwa SVM unggul pada dataset non-linear dan memiliki kemampuan generalisasi yang baik.

Precision untuk kelas positif pada SVM relatif rendah (0.46), yang menunjukkan masih adanya false positive. Namun, trade-off ini dapat diterima dalam konteks kesehatan masyarakat, karena lebih baik memprediksi “berpotensi terjadi DBD” meskipun beberapa prediksi salah, dibandingkan gagal mendeteksi kasus yang sebenarnya (false negative).

Regresi Logistik

Regresi Logistik menunjukkan performa terendah. Rendahnya precision dan F1-score pada kelas positif mengindikasikan bahwa model ini tidak mampu memisahkan kelas secara efektif. Hal ini dapat disebabkan oleh sifat hubungan antar variabel yang tidak linear, sedangkan Regresi Logistik bekerja optimal pada hubungan linier antara variabel prediktor dan logit probabilitas. Temuan ini

sejalan dengan penelitian Arista (2022), yang menunjukkan bahwa Logistic Regression sering gagal menangani data lingkungan yang kompleks dan non-linear.

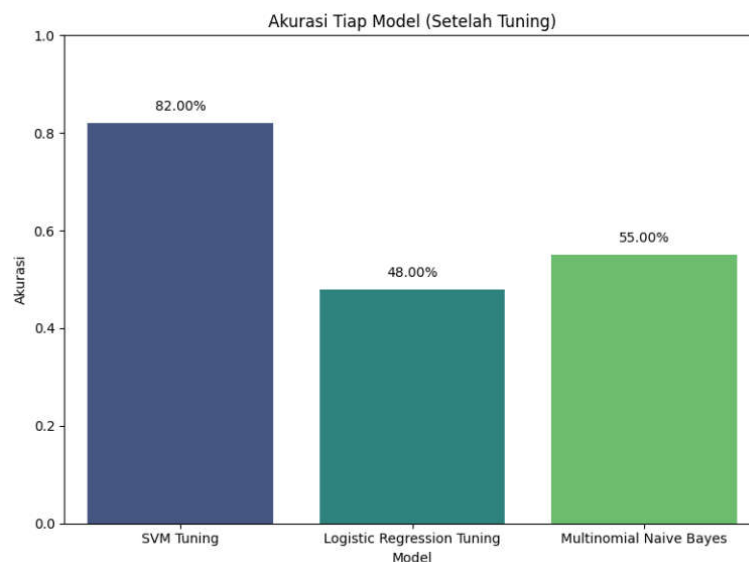
Multinomial Naïve Bayes

Hasil Naïve Bayes berada di antara dua model lainnya. Model ini berhasil mencapai recall 0.71 pada kelas positif, menunjukkan sensitivitas yang cukup baik. Namun, precision rendah (0.22) menandakan bahwa model menghasilkan banyak false positive. Hal ini sesuai dengan karakteristik Naïve Bayes yang mengasumsikan independensi antar fitur, padahal dalam kasus DBD, variabel seperti suhu, kelembapan, dan curah hujan saling berhubungan (Rocha & Giesbrecht, 2022)..

Pembahasan Hasil dan Perbandingan Model

Perbedaan performa ketiga model mencerminkan bagaimana masing-masing algoritma menangani karakteristik data. SVM memiliki keunggulan dalam memetakan data yang memiliki batas pemisah non-linear, sehingga mampu mempelajari hubungan kompleks antara variabel cuaca dan kejadian DBD. Naïve Bayes menunjukkan performa moderat karena kesederhanaannya, sementara Regresi Logistik kurang mampu menangani hubungan non-linear antar variabel.

Gambar 1. Akurasi Tiap Model



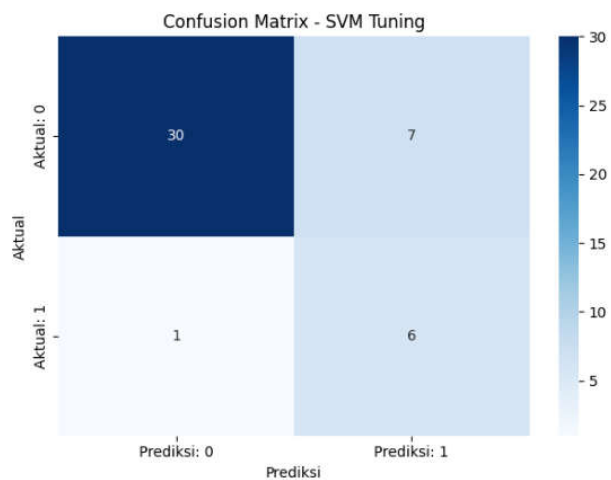
Jika dilihat dari keseluruhan hasil, SVM adalah model paling efektif untuk konteks prediksi penyakit. Hasil ini konsisten dengan penelitian terdahulu seperti Bari Antor et al. (2021) dan Rahman et al. (2021), yang menunjukkan bahwa SVM unggul pada data lingkungan dan epidemiologi.

Temuan ini juga memperkuat bahwa evaluasi model tidak bisa hanya mengandalkan akurasi. Pada konteks kesehatan, recall tinggi terhadap kelas positif lebih penting untuk mendeteksi potensi wabah lebih awal. Dengan demikian, meskipun precision SVM masih bisa ditingkatkan, model ini tetap memiliki performa terbaik untuk sistem peringatan dini di Purwokerto

Confusion Matrix

Confusion Matrix SVM Tuning

Gambar 2. Confusion Matrix Support Vector Machine



Keterangan:

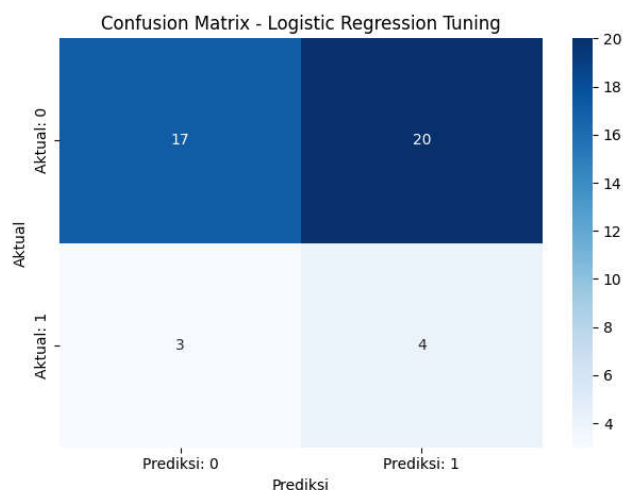
- a) *True Negative (TN)* = 30: Data aktual 0 diprediksi 0 → benar
- b) *False Positive (FP)* = 7: Data aktual 0 diprediksi 1 → salah
- c) *False Negative (FN)* = 1: Data aktual 1 diprediksi 0 → salah
- d) *True Positive (TP)* = 6: Data aktual 1 diprediksi 1 → benar

Analisis:

- a) Akurasi cukup tinggi.
- b) Model cukup baik dalam mendeteksi kelas 1
(*Recall* untuk kelas 1 = $6 / (6 + 1) = 0.857$).
- c) Namun masih cukup banyak *false positive* (7 kasus kelas 0 salah dikira 1).

Confusion Matrix Logistic Regression Tuning

Gambar 3. Confusion Matrix Regresi Logistik



Keterangan:

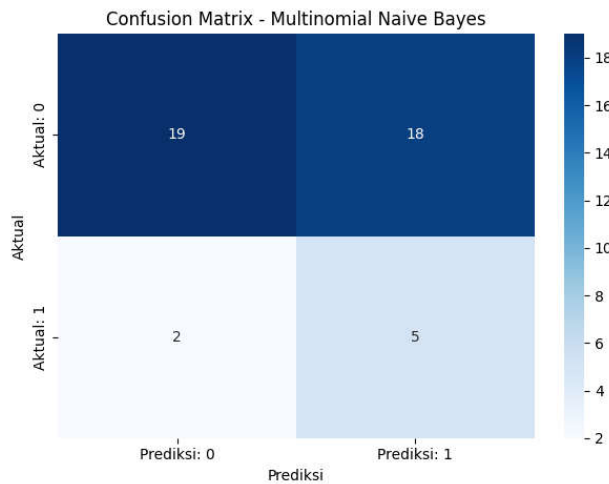
- a) *True Negative (TN)* = 17, Data aktual 0 diprediksi 0 → benar
- b) *False Positive (FP)* = 20, Data aktual 0 diprediksi 1 → salah
- c) *False Negative (FN)* = 3, Data aktual 1 diprediksi 0 → salah
- d) *True Positive (TP)* = 4, Data aktual 1 diprediksi 1 → benar

Analisis:

- Akurasi menurun, karena banyak kesalahan dalam memprediksi kelas 0 (20 kasus *false positive*).
- Recall* kelas 1 masih cukup ($4 / (4 + 3) = 0.571$).
- Model ini cenderung lebih sering mengklasifikasikan data sebagai kelas 1, menyebabkan banyak *false positive*.

Confusion Matrix Multinomial Naïve Bayes

Gambar 4. Confusion Matrix Multinomial Naive Bayes



Keterangan:

- True Negative (TN)* = 19, Data aktual 0 diprediksi 0 → benar
- False Positive (FP)* = 18, Data aktual 0 diprediksi 1 → salah
- False Negative (FN)* = 2, Data aktual 1 diprediksi 0 → salah
- True Positive (TP)* = 5, Data aktual 1 diprediksi 1 → benar

Analisis:

- Masih banyak false positive (18 kasus).
- Recall* kelas 1 = $5 / (5 + 2) = 0.714$.
- Akurasi sedikit lebih baik dari Regresi Logistik, tapi masih lebih lemah dibanding SVM.

Secara keseluruhan, hasil evaluasi menunjukkan adanya perbedaan performa yang cukup jelas antara ketiga algoritma yang diuji. Berdasarkan metrik akurasi, SVM merupakan model dengan performa tertinggi dengan nilai 82%, sedangkan Naïve Bayes berada pada kategori menengah dengan akurasi 55%, dan Regresi Logistik menunjukkan performa terendah dengan 48%. Pola yang sama juga terlihat pada metrik F1-score dan recall untuk kelas positif, di mana SVM consistently menghasilkan nilai tertinggi, sehingga menunjukkan kemampuan model dalam mendeteksi bulan dengan risiko DBD yang lebih tinggi.

Meskipun Naïve Bayes memiliki akurasi yang lebih rendah dibandingkan SVM, model ini tetap menunjukkan sensitivitas yang cukup baik melalui nilai recall 0.71 untuk kelas positif. Sebaliknya, Regresi Logistik cenderung gagal membedakan kelas secara optimal, ditunjukkan oleh nilai precision, recall, dan F1-score yang rendah pada kedua kelas.

Secara umum, overview hasil ini menunjukkan bahwa:

- SVM memberikan performa paling stabil dan unggul pada seluruh metrik, terutama recall kelas positif.
- Naïve Bayes bersifat moderat, sensitif tetapi memiliki banyak false positive.
- Regresi Logistik kurang sesuai untuk karakteristik data ini, khususnya karena hubungan antar variabel bersifat non-linear.

Temuan ini memberikan gambaran awal bahwa karakteristik data lingkungan di Purwokerto lebih sesuai dimodelkan dengan algoritma yang mampu menangani struktur data non-linear, yang kemudian diperkuat oleh pembahasan pada bagian selanjutnya..

Table 1. Descriptive statistical analysis of pretest and posttest data

Variable	Stage	Mean	Median	Variance	Std. Deviation	Minimum	Maximum
mathematical communication	Pretes	30,14	31,30	43,08	6,56	10,40	41,70
	Postes	66,62	66,70	51,54	7,18	50,00	77,10

Keterangan untuk judul tabel diletakkan di atasnya. Ukuran font di dalam tabel 10 pt, namun dapat kurang dari 10 pt jika ukuran tidak mencukupi. Sedangkan untuk keterangan gambar/grafik diletakkan di bawah gambar/grafik tersebut. Setiap tabel dan gambar harus ada kalimat yang merujuk pada tabel atau gambar tersebut. Misalnya, hasil penelitian disajikan pada Tabel 1. Model pengembangan yang digunakan disajikan pada Gambar 1.

Figure 1. Dick & Carey System Approach Model

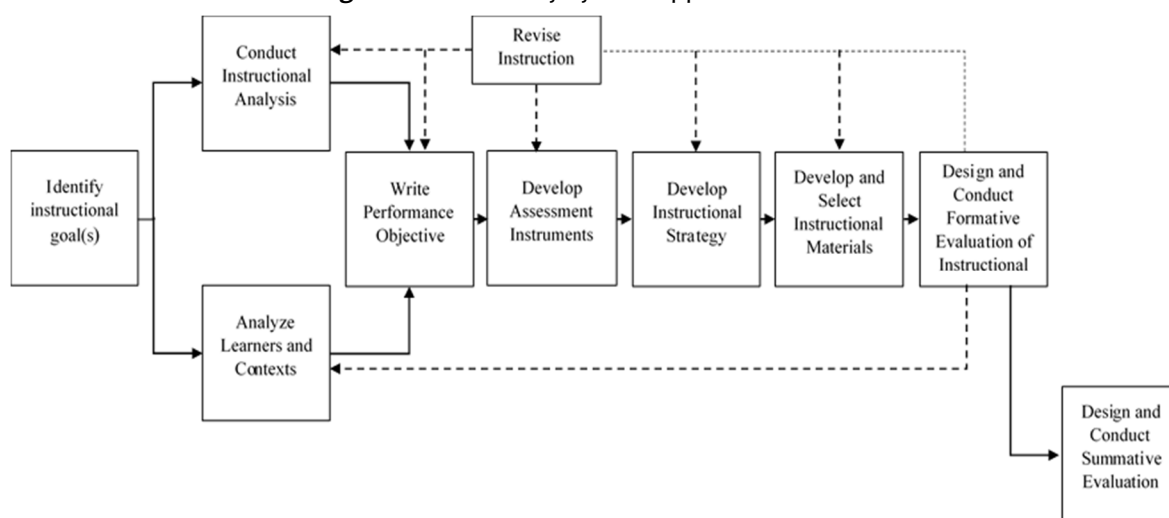


Table 2. N gain score analysis based on pretest and posttest data

Variable	Pretes	Postes	N- Gain Scor	Criteria
mathematical communication	30,14	66,62	0,52	medium

The results of the N-gain score test (table 2) show that problem posing learning models that accommodate local cultural contexts can improve mathematical communication with moderate criteria.

Kesimpulan

Penelitian ini membandingkan performa tiga algoritma machine learning, yaitu Naïve Bayes (NB), Regresi Logistik Biner (LR), dan Support Vector Machine (SVM) untuk memprediksi risiko Demam Berdarah Dengue (DBD) berdasarkan variabel lingkungan di Purwokerto. Hasil analisis menunjukkan bahwa SVM memiliki performa terbaik dengan akurasi 82% dan recall tertinggi pada kelas positif. Hal ini menunjukkan bahwa SVM lebih mampu memetakan hubungan non-linear antara variabel lingkungan dan kejadian DBD. Naïve Bayes menunjukkan performa moderat dengan recall tinggi namun precision rendah, sedangkan Regresi Logistik memiliki performa paling rendah karena tidak mampu menangani kompleksitas data dengan baik.

Temuan ini menegaskan bahwa model prediksi penyakit pada wilayah dengan karakteristik cuaca yang dinamis lebih sesuai menggunakan algoritma yang mampu menangkap pola non-linear. Dengan demikian, SVM direkomendasikan sebagai model yang paling efektif digunakan untuk sistem peringatan dini DBD di Purwokerto. Penelitian selanjutnya disarankan untuk menggunakan

dataset yang lebih besar, menambahkan variabel sosial-ekonomi, serta mengimplementasikan metode penyeimbangan kelas untuk meningkatkan kualitas prediksi.

References

- Anagora, R., Taufiq, R., Dedi Jubaedi, A., Wirawan, R., & Syah Putra, A. (n.d.). The Classification of Phishing Websites using Naive Bayes Classifier Algorithm. *International Journal Of Science*.
- Arhin, S. A., & Gatiba, A. (2020). Predicting crash injury severity at unsignalized intersections using support vector machines and naïve Bayes classifiers. *Transportation Safety and Environment*, 2(2), 120–132. <https://doi.org/10.1093/tse/tdaa012>
- Arista, A. (2022). Comparison Decision Tree and Logistic Regression Machine Learning Classification Algorithms to determine Covid-19. *Sinkron*, 7(1), 59–65. <https://doi.org/10.33395/sinkron.v7i1.11243>
- Bari Antor, M., Jamil, A. H. M. S., Mamtaz, M., Monirujjaman Khan, M., Aljahdali, S., Kaur, M., Singh, P., & Masud, M. (2021). A Comparative Analysis of Machine Learning Algorithms to Predict Alzheimer's Disease. *Journal of Healthcare Engineering*, 2021. <https://doi.org/10.1155/2021/9917919>
- Bhatia, S., & Malhotra, J. (2021). Naïve bayes classifier for predicting the novel coronavirus. *Proceedings of the 3rd International Conference on Intelligent Communication Technologies and Virtual Mobile Networks, ICICV 2021*, 880–883. <https://doi.org/10.1109/ICICV50876.2021.9388410>
- Chernozhukov, V., Newey, W., Quintas-Martínez, V., & Syrgkanis, V. (n.d.). *RieszNet and ForestRiesz: Automatic Debaised Machine Learning with Neural Nets and Random Forests*.
- Fadli, S., Ashari, M., Studi Sistem Informasi, P., & Lombok, S. (2021). *JISA (Jurnal Informatika dan Sains) Optimization of Support Vector Machine Method Using Feature Selection to Improve Classification Results*.
- Fallo, S. I. (n.d.). *ABSTRACT SUPPORT VECTOR MACHINE, NA"IVENA" NA"IVE BAYES CLASSIFIER AND ORDINAL LOGISTIC REGRESSION IN WEATHER PREDICTION*.
- Gupta, G., Khan, S., Guleria, V., Almjally, A., Alabdullah, B. I., Siddiqui, T., Albahlal, B. M., Alajlan, S. A., & AL-subaie, M. (2023). DDPM: A Dengue Disease Prediction and Diagnosis Model Using Sentiment Analysis and Machine Learning Algorithms. *Diagnostics*, 13(6). <https://doi.org/10.3390/diagnostics13061093>
- Khikmah, K. N., Indahwati, I., Fitrianto, A., Erfiani, E., & Amelia, R. (2022). Backwards Stepwise Binary Logistic Regression for Determination Population Growth Rate Factor in Java Island. *Jambura Journal of Mathematics*, 4(2), 177–187. <https://doi.org/10.34312/jjom.v4i2.13529>
- Kularatne, S. A., & Dalugama, C. (2022). Dengue infection: Global importance, immunopathology and management. *Clinical Medicine, Journal of the Royal College of Physicians of London*, 22(1), 9–13. <https://doi.org/10.7861/clinmed.2021-0791>
- Rahman, M. S., Pientong, C., Zafar, S., Ekalaksananan, T., Paul, R. E., Haque, U., Rocklöv, J., & Overgaard, H. J. (2021). Mapping the spatial distribution of the dengue vector *Aedes aegypti* and predicting its abundance in northeastern Thailand using machine-learning approach. *One Health*, 13. <https://doi.org/10.1016/j.onehlt.2021.100358>

- Rocha, F. P., & Giesbrecht, M. (2022). Machine learning algorithms for dengue risk assessment: a case study for São Luís do Maranhão. *Computational and Applied Mathematics*, 41(8). <https://doi.org/10.1007/s40314-022-02101-z>
- Shahzad, N., Ding, X., & Abbas, S. (2022). A Comparative Assessment of Machine Learning Models for Landslide Susceptibility Mapping in the Rugged Terrain of Northern Pakistan. *Applied Sciences (Switzerland)*, 12(5). <https://doi.org/10.3390/app12052280>
- Sutriyawan, A., Kurniati, N., Novianti, Farida, U., Yusanti, L., Destriani, S. N., & Saputra, M. K. F. (2024). ANALYSIS OF TEMPERATURE, HUMIDITY, RAINFALL, AND WIND VELOCITY ON DENGUE HEMORRHAGIC FEVER IN BANDUNG MUNICIPALITY. *Russian Journal of Infection and Immunity*, 14(1), 155–162. <https://doi.org/10.15789/2220-7619-AOT-2110>