Jurnal Penelitian dan Pengabdian Masyarakat
4 (4) , November 2026 Page : 3176–3183 / e-ISSN : 2986-5212
Creative Commons - Attribution 4.0 International - CC BY 4.0
DOI : https://doi.org/10.61231/2bpzph26
Received: 02-10-2026 Accepted: 06-11-2026 Publication: 17-11-2026

Analisis Variabel K Pada Penerapan Metode *K-Nearest Neighbors* (K-NN) Untuk Klasifikasi Penyakit Demam *Tifoid* (Tifus)

Nanda Nur'Azmi Utami[1)], Rendra Gustriansyah[2)], Nazori Suhandi[3)]

[1,2,3] Universitas Indo Global Mandiri, Indonesia

Email: 2022110008@students.uigm.ac.id [1], rendra@uigm.ac.id [2], nazori@uigm.ac.id[3]

Correspondence : 2022110008@students.uigm.ac.id*

Abstract: *Clinical diagnosis of typhoid fever frequently encounters challenges due to overlapping symptomatic characteristics with other infectious diseases, thereby increasing the risk of misdiagnosis. This study implements the K-Nearest Neighbor (K-NN) algorithm for typhoid fever classification, focusing on optimizing the selection of the K parameter value through a probabilistic Bayesian Likelihood approach. The dataset consists of 203 primary medical records from Dempo Health Center, Palembang, for the 2024–2026 period, with input features comprising fever duration, gastrointestinal symptoms, neurological symptoms, skin manifestations, and Widal test results. Model testing was conducted using a 70:30 data-splitting scheme (142 training data and 61 testing data). Evaluation results from both Bayesian Likelihood and the Elbow Method consistently determined K = 3 as the most optimal nearest-neighbor parameter. The K-NN model with K=3 achieved excellent classification performance, with an Accuracy of 96.72%, a Precision of 100.00%, and a Recall of 89.47%. These results indicate that the model is highly feasible as a clinical decision support system to assist medical personnel in detecting typhoid fever quickly, objectively, and accurately.*

Abstrak: *Diagnosis demam tifoid secara klinis sering menghadapi kendala akibat kemiripan karakteristik gejala dengan penyakit infeksi lainnya, sehingga meningkatkan risiko kesalahan diagnosis awal. Penelitian ini menerapkan algoritma K-Nearest Neighbor (K-NN) untuk klasifikasi status penyakit demam tifoid, dengan fokus pada optimalisasi penentuan nilai parameter K menggunakan pendekatan probabilistik Bayesian Likelihood. Dataset yang digunakan merupakan data primer rekam medis dari Puskesmas Dempo Palembang periode 2024–2026 sebanyak 203 data, dengan fitur masukan berupa durasi demam, gejala gastrointestinal, gejala neurologis, manifestasi kulit, dan hasil uji Widal. Pengujian model menggunakan skema pembagian data (data splitting) rasio 70:30 (142 data latih dan 61 data uji). Hasil evaluasi Bayesian Likelihood dan Elbow Method secara konsisten menetapkan K = 3 sebagai parameter tetangga terdekat paling optimal. Model K-NN dengan K=3 menghasilkan performa klasifikasi yang sangat baik dengan Accuracy 96,72%, Precision 100,00%, dan Recall 89,47%. Hasil ini menunjukkan bahwa model tersebut layak digunakan sebagai sistem pendukung keputusan klinis untuk membantu tenaga medis mendeteksi demam tifoid secara cepat, objektif, dan akurat.*

Keywords : *Demam Tifoid, K-Nearest Neighbor, Variabel K, Pendekatan Bayesian, Klasifikasi*

PENDAHULUAN

Perkembangan teknologi informasi dan komunikasi telah membawa perubahan besar pada berbagai aspek kehidupan, termasuk sektor kesehatan. Pemanfaatan teknologi informasi kini menjadi bagian penting dari praktik medis dan sistem pelayanan kesehatan secara keseluruhan (Zahra et al.,

2021). Demam *tifoid* atau tifus merupakan penyakit infeksi yang disebabkan oleh bakteri *Salmonella typhi* yang umumnya menyerang sistem pencernaan manusia dan menular melalui mekanisme fekal-oral akibat sanitasi lingkungan yang kurang baik (Erald Giovanny Hasiholan Simatupang, 2023). Diperkirakan terdapat sekitar 500.000 hingga 1.000.000 kasus tifus di Indonesia setiap tahun, dengan tren kasus yang kembali meningkat pada tahun 2024 seiring memburuknya kualitas air pada musim hujan (Tobing, 2024).

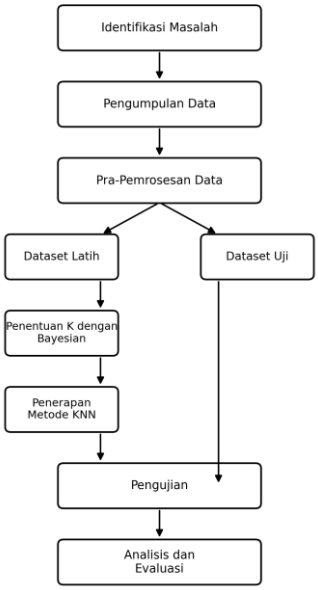
Pengenalan gejala klinis secara tepat menjadi kunci utama dalam menegakkan diagnosis demam *tifoid*. Namun, proses diagnosis kerap terkendala oleh kemiripan gejala *tifoid* dengan berbagai penyakit infeksi lain, sehingga berisiko menimbulkan kesalahan diagnosis (Fahmi Ginting et al., 2023). Kondisi ini mendorong perlunya pendekatan berbasis teknologi, salah satunya metode klasifikasi K-*Nearest Neighbor* (K-NN), yang banyak dipilih karena kesederhanaan konsep dan efektivitasnya dalam mengelompokkan data berdasarkan kemiripan karakteristik. Meski demikian, performa K-NN sangat bergantung pada pemilihan nilai parameter K, sehingga penelitian ini difokuskan pada analisis variasi nilai K guna memperoleh nilai K yang optimal.

Penelitian terdahulu telah banyak memanfaatkan pendekatan pembelajaran mesin untuk diagnosis *tifoid* maupun penyakit infeksi sejenis. Fariha Ahmed Nishat dan Istiak Mahmud mengembangkan model *ensemble* yang menggabungkan SVM, *Gaussian Naive Bayes*, dan *Decision Tree* untuk diagnosis *tifoid* dengan *precision* hingga 99% dan *recall* 100% (Fariha Ahmed Nishat & Istiak Mahmud, 2026). Sementara itu, penerapan K-NN pada kasus demam berdarah *dengue* (DBD) oleh Latuconsina memperoleh akurasi terbaik 75,12% pada K=3 ,sedangkan Matdoan memperoleh akurasi 72,31% pada kasus serupa (Latuconsina, 2022) . Fauzi menunjukkan bahwa variasi nilai K pada klasifikasi penyakit liver menghasilkan perbedaan akurasi yang signifikan, dengan K=5 menghasilkan 66,67% dan K=15 menghasilkan 76,06% sebagai nilai optimal (Fauzi, 2025) .

Dari uraian penelitian terdahulu, dapat disimpulkan bahwa meskipun K-NN telah banyak diaplikasikan dalam klasifikasi data kesehatan, kajian mengenai pengaruh variasi nilai K terhadap akurasi klasifikasi—khususnya pada kasus demam *tifoid*—masih terbatas, dan sebagian besar penelitian tersebut masih mengandalkan metode evaluasi K-*Fold Cross Validation*. Oleh karena itu, penelitian ini bertujuan menganalisis pengaruh variasi nilai K menggunakan pendekatan *Bayesian* pada metode K-NN dalam klasifikasi penyakit demam *tifoid* dengan pendekatan data splitting 70:30, guna memperoleh nilai K optimal yang menghasilkan tingkat akurasi terbaik serta memberikan referensi pemilihan nilai K yang tepat pada penerapan K-NN di bidang kesehatan.

METODE

Penelitian ini dilaksanakan melalui serangkaian tahapan yang meliputi identifikasi masalah, pengumpulan data, pra-pemrosesan data, pembagian dataset, penentuan nilai K menggunakan pendekatan *Bayesian*, penerapan metode K-NN, pengujian, hingga analisis dan evaluasi hasil. Kerangka kerja penelitian disajikan pada Gambar 1.



Gambar 1. Kerangka kerja penelitian

Dataset yang digunakan merupakan data primer rekam medis pasien Puskesmas Dempo Palembang periode 2024–2026 sebanyak 203 data. Dataset terdiri atas lima atribut prediktor, yaitu *Fever Duration (Days)*, *Gastrointestinal Symptoms*, *Neurological Symptoms*, *Skin Manifestations*, dan *Widal Test*, serta satu atribut target *Typhoid Status*. Dari total data, 63 data (31,03%) berlabel *Typhoid* dan 140 data (68,97%) berlabel *No Typhoid*. Cuplikan data disajikan pada Tabel 1.

Tabel 1. Cuplikan data gejala pasien Puskesmas Dempo

No	Fever (Days)	GI Symptoms	Neuro Symptoms	Skin Manif.	Widal Test	Status
1	0	Diarrhea	Headache	Yes	High O&H	No Typhoid
2	16	Constipation	Headache	Yes	High O&H	Typhoid
3	17	None	Headache	Yes	Low O&H	Typhoid
4	13	Diarrhea	Confusion	No	Low O&H	Typhoid
5	0	Abdominal Pain	Delirium	Yes	Low O&H	No Typhoid
…	…	…	…	…	…	…
203	0	Diarrhea	None	Yes	High O&H	No Typhoid

HASIL DAN PEMBAHASAN

Pemeriksaan kualitas data menunjukkan nilai kosong pada atribut *Gastrointestinal Symptoms* sebanyak 57 data (28,08%) dan *Neurological Symptoms* sebanyak 51 data (25,12%), yang ditangani dengan mengisi kategori "*None*" tanpa menghapus baris data. Seluruh atribut kategorik selanjutnya dikonversi menjadi representasi numerik melalui *Label Encoding* agar dapat diproses oleh algoritma K-NN yang berbasis perhitungan jarak, sebagaimana ditunjukkan pada Tabel 2.

Tabel 2. Hasil label encoding atribut kategorik

Atribut	Kategori	Kode
Gastrointestinal	None / Diarrhea / Constipation / Abdominal Pain	0–3
Neurological	None / Headache / Confusion / Delirium	0–3
Skin Manifestations	No / Yes	0–1
Widal Test	High O&H / Low O&H Antibody	0–1
Typhoid Status	No Typhoid / Typhoid	0–1

Dataset yang telah melalui proses *encoding* dibagi menjadi data latih dan data uji dengan rasio 70:30 secara acak (*random_state* = 42), menghasilkan 142 data latih (70%) dan 61 data uji (30%) sebagaimana Tabel 3.

Tabel 3. Rekapitulasi pembagian data latih dan data uji

Rasio	Jenis Himpunan Data	Jumlah	%
70:30	Data Latih (*Training*)	142	70%
70:30	Data Uji (*Testing*)	61	30%
Total	-	203	100%

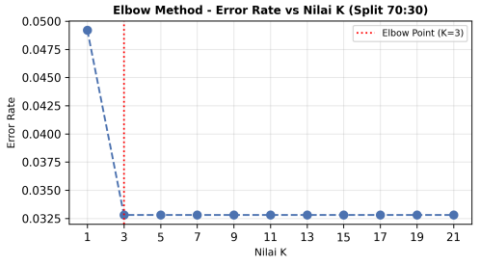
Nilai K optimal ditentukan dengan mengevaluasi likelihood (akurasi) model K-NN pada setiap kandidat nilai K ganjil (K = 1 hingga 21) menggunakan data uji. Ringkasan hasil evaluasi disajikan pada Tabel 4.

Tabel 4. Evaluasi *likelihood Bayesian* - skenario 70:30

Nilai K	Benar (TP+TN)	Total Uji	Likelihood
K=1	58	61	95,08%
K=3	59	61	96,72%
K=5 – K=17	59	61	96,72%
K=19, K=21	59	61	96,72%

Setelah nilai K optimal diperoleh, algoritma K-NN diimplementasikan menggunakan pustaka *Scikit-Learn* dengan parameter n_*neighbors* = 3. Proses klasifikasi mengikuti tahapan: (1) menghitung jarak *Euclidean* antara data uji dan seluruh data latih, (2) mengurutkan jarak dari terkecil ke terbesar, (3) memilih K=3 tetangga terdekat, dan (4) menentukan kelas melalui *majority voting*.

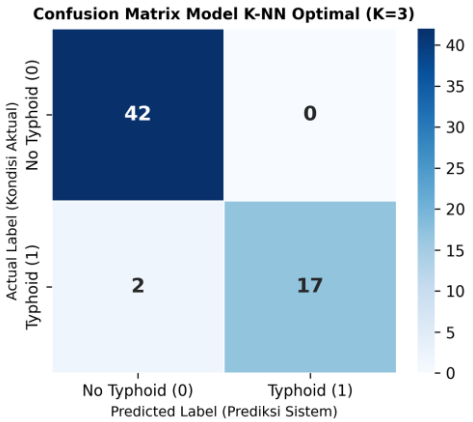
Berdasarkan Tabel 4, performa model meningkat dari K=1 (95,08%) menjadi 96,72% pada K=3, kemudian mendatar (*plateau*) hingga K=21 dengan nilai *likelihood* yang identik. Untuk memvalidasi hasil tersebut secara visual, dilakukan analisis *Elbow Method* terhadap *error rate* (1-*accuracy*) pada setiap nilai K, sebagaimana ditunjukkan pada Gambar 2.



Gambar 2. Kurva *Elbow Method - Error Rate* vs Nilai K

Grafik pada Gambar 2 menunjukkan titik siku *(elbow point)* yang jelas pada K=3, konsisten dengan hasil evaluasi *Bayesian Likelihood*. K=3 dipilih sebagai nilai K optimal tunggal karena merupakan nilai ganjil terkecil pada rentang performa optimal, sehingga menghindarkan model dari risiko *overfitting* pada K=1 sekaligus meminimalkan beban komputasi dibandingkan nilai K yang lebih besar.

Model K-NN dengan K=3 diuji pada 61 data uji skenario 70:30, menghasilkan *confusion matrix* pada Gambar 3.



Gambar 3. *Confusion matrix* model K-NN optimal (K=3)

Hasil pengujian menunjukkan *True Negative* (TN) = 42, *False Positive* (FP) = 0, *False Negative* (FN) = 2, dan *True Positive* (TP) = 17. Berdasarkan nilai tersebut, diperoleh *Accuracy* sebesar (17+42)/61 = 96,72%, *Precision* sebesar 17/17 = 100,00%, dan *Recall* sebesar 17/19 = 89,47%, sebagaimana dirangkum pada Tabel 5.

Tabel 5. Hasil evaluasi model K-NN (K=3)

Metrik	Nilai (%)	Keterangan
Accuracy	96,72	59 dari 61 data diklasifikasikan benar
Precision	100,00	17 dari 17 prediksi Typhoid benar
Recall	89,47	17 dari 19 pasien Typhoid terdeteksi

Konsistensi hasil antara pendekatan *Bayesian Likelihood* dan *Elbow Method* yang sama-sama menunjuk K=3 memperkuat validitas bahwa nilai tersebut merupakan parameter tetangga terdekat yang optimal dan *robust* untuk karakteristik dataset Puskesmas Dempo. Nilai *precision* sempurna (100,00%) menunjukkan bahwa model tidak pernah menghasilkan kesalahan *False Positive*, yang sangat *krusial* dalam domain medis karena mencegah alarm palsu berupa pasien sehat yang salah didiagnosis tifus. Di sisi lain, *recall* sebesar 89,47% menunjukkan bahwa masih terdapat 2 dari 19 pasien *Typhoid* yang tidak terdeteksi (*False Negative*), yang mengindikasikan peluang perbaikan lebih lanjut, misalnya melalui pembobotan fitur atau penyeimbangan kelas.

Dibandingkan dengan penelitian terdahulu yang menerapkan K-NN pada kasus DBD dengan akurasi 72,31%–75,12% [6][7] maupun klasifikasi penyakit liver dengan akurasi 66,67%–76,06% [8], model K-NN pada penelitian ini menghasilkan akurasi yang jauh lebih tinggi (96,72%). Hal ini menunjukkan bahwa kombinasi antara data primer rekam medis yang representatif dan pendekatan *Bayesian Likelihood* dalam menentukan nilai K mampu meningkatkan performa klasifikasi demam *tifoid* secara signifikan, sehingga model ini berpotensi diimplementasikan sebagai sistem pendukung keputusan klinis (*clinical decision support system*) bagi tenaga medis di Puskesmas Dempo. pendampingan

KESIMPULAN

Penelitian ini berhasil menerapkan metode K-*Nearest Neighbor* (K-NN) untuk klasifikasi penyakit demam *tifoid* berdasarkan data gejala pasien Puskesmas Dempo. Pendekatan probabilistik *Bayesian Likelihood*, yang divalidasi secara konsisten oleh *Elbow Method*, menetapkan K=3 sebagai parameter tetangga terdekat paling optimal. Pada skenario pembagian data 70:30, model K-NN (K=3) menghasilkan *Accuracy* 96,72%, *Precision* sempurna 100,00%, dan *Recall* 89,47%, yang membuktikan bahwa model sangat layak digunakan untuk membantu tenaga medis melakukan deteksi dini demam

tifoid secara cepat, objektif, dan akurat. Penelitian selanjutnya disarankan untuk memperbanyak dan menyeimbangkan jumlah data pasien positif tifus, menguji skema pembobotan jarak pada algoritma K-NN untuk mengurangi kesalahan *False Negative*, serta membandingkan metode penentuan nilai K otomatis lain berbasis kecerdasan buatan dengan pendekatan *Bayesian* yang digunakan pada penelitian ini masyarakat

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada dosen pembimbing atas arahan, masukan, dan bimbingan selama proses penelitian serta penyusunan artikel ini, serta kepada pihak Puskesmas Dempo Palembang atas izin dan bantuan akses data rekam medis pasien yang digunakan dalam penelitian ini. baik

DAFTAR PUSTAKA

A. C. Prihandoko, E. Susanto, and W. R., "Evaluasi Model Klasifikasi *Machine Learning* untuk Diagnosis Penyakit Berbasis Data Rekam Medis," J. Inform. J. Pengemb. IT, vol. 9, no. 1, pp. 12–19, 2024.

D. Zahra, A. Nurdin, U. Fitria, K. A. Dinen, and R. Kurnia, "Pemanfaatan Teknologi dalam Bidang Kesehatan Masyarakat," *Public Health* J., 2021.

E. Fahmi Ginting, M. Hutasuhut, M. Roza Fitri, and S. Triguna Dharma, "Sistem Cerdas Mendiagnosa Penyakit Demam *Tifoid* dengan *Metode Case Based Reasoning*," J. Sci. Soc. Res., 2023.

Erald Giovanny Hasiholan Simatupang et al., "*Epidemiologi dan Resistensi Antibiotik Salmonella typhi dan Paratyphi* A pada Kasus Demam *Tifoid* di Jakarta: *A Systematic Literature Review*," J. Ilmu Psikol. dan Kesehat., 2023.

Fanny Ramadhani, Andy Satria, and Indah Purnama Sari, "Implementasi Metode *Fuzzy K-Nearest Neighbor* dalam Klasifikasi Penyakit Demam Berdarah," *Hello World J.* Ilmu Komput., vol. 2, no. 2, pp. 58–62, 2023

F. A. Nishat and I. Mahmud, "*Enhancing Typhoid Fever Diagnosis Based on Clinical Data Using a Lightweight Machine Learning Metamodel*," vol. 15, no. 5, 2026.

Gustriansyah, R., Alie, J., Sanmorino, A., Heriansyah, R. and Noor, M.M., 2022. *Noor (Machine Learning for Regencies-Cities Clustering Based on Inflation and Poverty Rates in Indonesia) Machine Learning for Regencies-Cities Clustering Based on Inflation and Poverty Rates in Indonesia. Indonesian Journal of Information Systems (IJIS)*, .

Gustriansyah, R., Suhandi, N., Puspasari, S. and Sanmorino, A., 2024. Machine Learning Method to Predict the Toddlers' Nutritional Status. *JURNAL INFOTEL*, 16(1). https://doi.org/10.20895/infotel.v15i4.988.

H. Hafid, "Penerapan K-*Fold Cross Validation* untuk Menganalisis Kinerja Algoritma K-*Nearest Neighbor* pada Data Kasus Covid-19 di Indonesia," J. Math., 2023.

H. Wang, X. Z., and W. Pedrycz, "*An Overview on the Roles of Fuzzy Set Techniques in Big Data Processing: Trends, Challenges and Opportunities," Knowledge-Based Syst.*, pp. 569–589, 2023.

H. Sarker, "*Machine Learning: Algorithms, Real-World Applications and Research Directions,"* SN Comput. Sci., vol. 2, no. 3, p. 160, 2021.

Ibrahim Akbar, M., Heriansyah, R., Haviz Irfani, M., Indo Global Mandiri, U., Jend Sudirman Km, J., Jurusan Teknik Informatika, P. and Ilmu Komputer Dan Sains, F., 2026. Pengaruh Splitting Data terhadap Akurasi Klasifikasi Demam Berdarah Dengue Menggunakan K-Nearest Neighbors. *Jurnal Algoritme*, 6(2), pp.244–256. https://doi.org/10.35957/algoritme.v6i2.13274.

I. Fauzi and H. C., "Analisis Kinerja Algoritma K-*Nearest Neighbor* dalam Klasifikasi Penyakit Liver," J. Tek. Inform. UNIKA Santo Thomas, vol. 10, no. 2, pp. 357–364, 2025.

J. F. J. Tobing, "Demam *Tifoid*," 2024, doi: 10.37817/ikraith-humaniora.v8i2

M. Y. Matdoan, "Penerapan Metode K-*Nearest Neighbor* untuk Mengklasifikasi Penyebaran Kasus Demam Berdarah *Dengue* (DBD) di Kabupaten Maluku Tenggara," Square J. Math. Math. Educ., vol. 4, no. 2, pp. 75–82, 2022, doi: 10.21580/square.2022.4.2.13056.

Nazori Suhandi, Rendra Gustriansyah and Destria, A., 2025. Klasifikasi Penyakit TBC Menggunakan Metode UMAP dan K-NN. *bit-Tech*, 7(3), pp.843–852. https://doi.org/10.32877/bt.v7i3.2227.

N. M. Tangkelayuk and E. Mailoa, "Evaluasi Performa Model Klasifikasi Menggunakan *Confusion Matrix* pada *Dataset* Penyakit," J. Tek. Inform. dan Sist. Inf., vol. 8, no. 2, pp. 401–410, 2022.

S. Latuconsina and M. Y. Matdoan, "Prediksi Penyebaran Kasus Demam Berdarah *Dengue* (DBD) Dengan Metode K-*Nearest Neighbor* di Kabupaten Maluku Tengah," J. Ilm. Mat. dan Terap., vol. 19, no. 2, pp. 254–261, 2022.

Suhandi, N., Gustriansyah, R., Destria, A., Amalia, M. and Kris, V., 2024. Prediksi Kualitas Susu Menggunakan Metode K-Nearest Neighbors Milk Quality Prediction Using The K-Nearest Neighbors Method. [online] 14(2). https://doi.org/10.30700/sisfotenika.v14i2.430.