



Klasifikasi Penyakit Demam Tifoid Menggunakan Algoritme K-Nearest Neighbor

*Hidayatul Nur Hanifah¹, Henny Dwi Bhakti^{2,3}, Nadya Husenti³

^{1,2,3}Informatika, Teknik, Universitas Muhammadiyah Gresik, Gresik, Jawa Timur, Indonesia, 61121

Email: hidayatulnh04@gmail.com^{1*}, hennydwi@umg.ac.id², nadyahusenti@umg.ac.id³

ABSTRAK

Demam tifoid merupakan penyakit infeksi yang disebabkan oleh bakteri *Salmonella typhi*. Penularan demam tifoid dapat terjadi melalui fecal dan oral dari konsumsi makanan dan minuman yang terkontaminasi, pola makan yang tidak teratur dan kurang bersih pada tempat pengolahan makanan. Gejala pada pasien dilakukan sebuah analisa oleh dokter dengan dilakukan pemeriksaan untuk mendapatkan diagnosis. Diagnosis demam tifoid pada pasien agar dapat dilakukan penanganan secara cepat salah satu tindakan yang bisa dilakukan dengan mengembangkan sistem gejala penyakit demam tifoid. Sistem tersebut dibentuk dengan model klasifikasi dari teknik data mining untuk mendiagnosis penyakit demam tifoid pada pasien. Proses klasifikasi dalam menyelesaikan permasalahan diagnosis pasien dengan gejala demam tifoid menggunakan metode K-Nearest Neighbor (KNN). Berdasarkan dari implementasi dan pengujian sistem dengan algoritme K-Nearest Neighbor mendapatkan kesimpulan dalam analisis setiap gejala pasien yang di uji dengan sistem, dapat memberikan hasil diagnosis demam tifoid pada pasien berdasarkan pengujian dengan data training. Analisis akurasi tertinggi evaluasi algoritme K-Nearest Neighbor dari pengujian 20 data testing dan 10 data training terdapat pada K=9 sebesar 75%, recall sebesar 66,67%, dan precision 75%.

Kata Kunci : Demam tifoid, K-Nearest Neighbor, Klasifikasi

ABSTRACT

Typhoid fever is an infectious disease caused by Salmonella typhi bacteria. Transmission of typhoid fever can occur through fecal and oral consumption of contaminated food and beverages, irregular eating patterns and lack of cleanliness in food processing sites. Symptoms in patients are analyzed by a doctor with an examination to get a diagnosis. Diagnosis of typhoid fever in patients in order to be handled quickly one of the actions that can be done by developing a system of symptoms of typhoid fever. The system was formed with a classification model from data mining techniques to diagnose typhoid fever in patients. Classification process in solving the problem of diagnosis of patients with symptoms of typhoid fever using the K-Nearest Neighbor (KNN) method. Based on the implementation and testing of the system with K-Nearest Neighbor algorithm to obtain conclusions in the analysis of each patient's symptoms in the test with the system, can provide results of diagnosis of typhoid fever in patients based on testing with training data. Analysis of the highest accuracy of K-Nearest Neighbor algorithm evaluation of 20 testing data and 10 training data contained in K=9 at 75%, recall at 66.67%, and precision 75%.

Keywords: Typhoid fever, K-Nearest Neighbor, Classification

Submitted : *Revision :* *Accepted :*

PENDAHULUAN

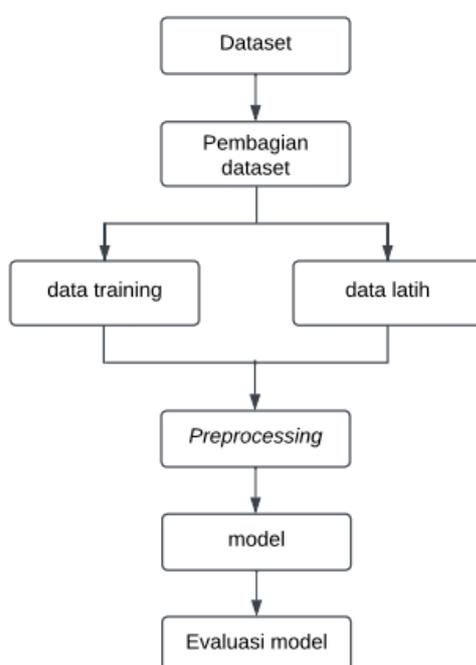
Demam tifoid merupakan penyakit yang disebabkan oleh bakteri *Salmonella typhi* melalui fecal dan oral dari konsumsi makanan dan minuman yang terkontaminasi, pola makan yang tidak teratur dan kurang bersih pada tempat pengolahan makanan (Nurul Eliza et al., 2023). Penderita demam tifoid memiliki gejala

yang berkaitan dengan masalah pencernaan antara lain mual, muntah, dan nyeri perut. Gejala demam tifoid lainnya adalah demam dengan suhu naik turun selama tiga sampai satu minggu, demam yang disertai menggigil, nyeri persendian atau otot dan nyeri kepala. Beberapa gejala pada pasien dilakukan sebuah analisa oleh dokter dengan dilakukan pemeriksaan

dengan diagnosis penyakit. Algoritme K-Nearest Neighbor (KNN) dalam beberapa penelitian mendapatkan hasil yang optimal dalam pemecahan permasalahan, diantaranya penelitian dalam klasifikasi penyakit diabetes melitus dapat memberikan diagnosis diabetes melitus dari gejala pasien (Dwi Fasnuari et al., 2022). Diagnosis demam tifoid pada pasien agar dapat dilakukan penanganan secara cepat salah satu tindakan yang bisa dilakukan dengan mengembangkan sistem gejala penyakit demam tifoid. Sistem tersebut dibentuk dengan model klasifikasi dari teknik data mining untuk mendiagnosis penyakit demam tifoid pada pasien (A'yuniyah & Reza, 2023). Proses klasifikasi dalam menyelesaikan permasalahan diagnosis pasien dengan gejala demam tifoid menggunakan metode K-Nearest Neighbor (KNN). Proses klasifikasi demam tifoid dilakukan dengan uji coba perhitungan jarak data training dan data testing, jarak yang terdekat atau terkecil akan di urutkan berdasarkan nilai k yang telah ditentukan (Harun et al., 2020). Penelitian ini bertujuan untuk memberikan hasil diagnosis demam tifoid pada pasien berdasarkan pengujian gejala pasien.

METODOLOGI

Langkah-langkah dalam penelitian gejala pasien menggunakan metode *K-Nearest Neighbor* pada Gambar dibawah ini :



Gambar 1. Alur Penelitian *K-Nearest Neighbor*

Pada Gambar 1 merupakan gambaran tahapan dari penelitian data gejala pasien dengan metode *K-Nearest Neighbor*. Pada penelitian ini menggunakan data gejala pasien dari sebuah Pelayanan Kesehatan Praktek Bersama dr. Ahmad Izzudin Afif-drg. Desyllia Tanjung. Dataset pada penelitian ini terdiri dari 127 data pasien, melalui preprocessing data terisa 93 data gejala dan di uji menggunakan metode K-Nearest Neighbor (KNN). Langkah pertama adalah pembagian dataset 10 data *training* dan 20 data *testing*. Langkah kedua yaitu memproses data sebelum di uji dengan normalisasi data MinMax Scaler. Langkah ketiga pengujian data menggunakan metode *K-Nearest Neighbor*. Setelah diperoleh hasil pengujian dari 20 data *testing* selanjutnya di evaluasi model dengan akurasi, *recall*, dan *precision*.

Preprocessing

Preprocessing merupakan teknik untuk mengubah data yang telah dikumpulkan dari beberapa sumber yang akan dilakukan pengolahan lebih lanjut. Terdapat beberapa metode dalam preprocessing data salah satu dari metode tersebut adalah MinMax Scaler. MinMax Scaler adalah transformasi fitur dengan menskala secara individual dengan rentang tertentu.(Prasetyo et al., 2019). Preprocessing MinMax Scaler melakukan proses pengurangan pada data yang akan dilakukan normalisasi dengan data terkecil pada fitur dan dilakukan pembagian dari hasil pengurangan nilai terbesar pada fitur dengan nilai terkecil pada fitur (Azzahra Nasution et al., 2019).

Klasifikasi

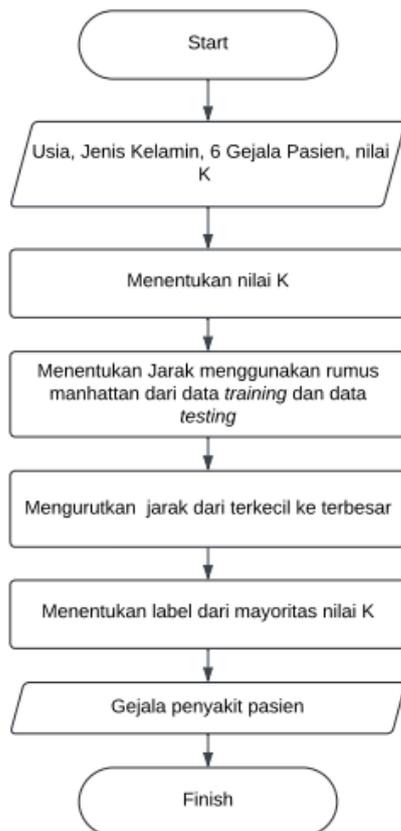
Klasifikasi merupakan pelabelan sebuah data dengan cara memisahkan sekumpulan data sesuai dengan jenis label atau kelas yang sudah di tentukan (Widaningsih, 2019). Klasifikasi dalam data mining digunakan untuk pengelompokan data dengan membangun model atau algoritme yang dapat memprediksi kelas dari data baru untuk mengambil keputusan berdasarkan pola yang teridentifikasi dalam data (Witten et al., 2016). Keunggulan penerapan klasifikasi menggunakan algoritme dapat dinilai dari kebenaran klasifikasi model terhadap data sebenarnya, dan tepatnya penerapan model dalam prediksi kelas klasifikasi (Dinata et al., 2020). Pada penelitian

ini klasifikasi digunakan untuk melabel gejala pasien untuk menghasilkan diagnosis pasien.

K-Nearest Neighbor

K-Nearest Neighbor (K-NN) merupakan metode yang digunakan untuk klasifikasi suatu objek yang memiliki jarak terdekat, dengan mencoba data baru yang belum diketahui kelas atau label berdasarkan hasil K (Br Sinuhaji et al., 2024).

Kelebihan dalam penggunaan metode K-Nearest Neighbor (K-NN) adalah mudah dalam proses implementasi, dapat dilihat dari langkah-langkah metode K-Nearest Neighbor (K-NN) dalam prosesnya menggunakan dua parameter yaitu nilai K dan fungsi jarak. Kelebihan metode K-Nearest Neighbor (K-NN) lainnya adalah tidak memerlukan training sebelum prediksi sehingga saat ada data baru dapat dilakukan tanpa mengurangi nilai keakuratan. Adapun kekurangan pada metode K-Nearest Neighbor (K-NN) apabila terdapat data yang besar performa algoritme akan menurun dan diperlukan standarisasi terlebih dahulu dengan cara menormalisasikan data sebelum penerapan algoritme K-Nearest Neighbor (K-NN) (Kumar et al., 2018). Berikut merupakan alur pada K-Nearest Neighbor :



Gambar 1. Alur Metode K-Nearest Neighbor

Confusion Matrix

Confusion Matrix merupakan model evaluasi yang digunakan untuk mengukur kinerja dari proses klasifikasi.

Tabel 1. Confusion Matrix

True label	Negative	TN	FP
	Positive	FN	TP
		Negative	Positive
		Predict Label	

Pada Tabel 1 merupakan tabel dari Confusion Matrix. TN (True Negative) dalam gambar tersebut berarti data negatif yang diprediksi benar, TP (True Positive) berarti data positif di prediksi benar saat di uji , FN (False Negative) berarti data positif yang diprediksi negatif, dan FP (False Positive) berarti data negatif di prediksi positif. Pengukuran kinerja tersebut menggunakan akurasi, recall, dan precission.

$$Akurasi = \frac{(TP+TN)}{(TP+TN+FP+FN)} \times 100\%$$

$$Recall = \frac{TP}{(TP+FP)} \times 100\%$$

$$Precision = \frac{TP}{(TP+FN)} \times 100\%$$

HASIL DAN PEMBAHASAN

Langkah awal dalam penelitian ini adalah memproses sebuah dataset sejumlah 127 data gejala pasien. Data pasien yang telah dikumpulkan akan melalui preprocessing. Pada beberapa data tersebut, dilakukan pembersihan data untuk memperbaiki data agar sesuai dengan gejala penyakit demam tifoid. Dari total 127 data pasien setelah melalui proses pembersihan data tersisa 93 data yang sesuai dengan gejala demam tifoid. Berikut merupakan Tabel 2 memaparkan 93 data pasien:

Tabel 2. Dataset

No	usia	jk	G1	...	G5	G6	Label
1	20	0	0	...	0	0	+
2	15	1	0	...	0	0	+
3	19	1	0	...	1	0	+
4	13	1	0	...	0	0	+
5	30	0	0	...	0	1	+
6	56	0	0	...	0	1	+
7	69	0	0	...	0	0	+
8	12	1	0		0	0	+
....
93	11	1	1	...	1	0	+

Keterangan Tabel :

Usia : Usia Pasien

- jk : Jenis Kelamin
- G01 : Gejala Demam (Nilai 1 = >7 Hari dan Nilai 0 = 3 - & Hari)
- G02 : Gejala demam disertai menggigil (Nilai 1 = Ya dan Nilai 0 = Tidak)
- G03 : Gejala Sakit Kepala (Nilai 1 = Ya dan Nilai 0 = Tidak)
- G04 : Gejala Mual dan Muntah (Nilai 1 = Ya dan Nilai 0 = Tidak)
- G05 : Gejala Nyeri perut (Nilai 1 = Ya dan Nilai 0 = Tidak)
- G06 : Gejala Nyeri Persendian (Nilai 1 = Ya dan Nilai 0 = Tidak)
- Label : Berlabel (+) Pasien terdiagnosis demam tifoid dan berlabel (-) Pasien tidak terdiagnosis demam Tifoid

Dari data tersebut di random secara acak menghasilkan 10 data *training* yang tersaji dalam Tabel 3 :

Tabel 3. Data *Training*

No	usia	jk	G1	...	G5	G6	Label
1	19	1	0	...	1	0	+
2	14	0	1	...	0	0	-
3	20	0	0	...	1	0	-
4	49	0	0	...	1	0	-
5	25	1	0	...	1	0	+
6	28	0	0	...	0	0	+
7	27	1	1	...	0	1	-
8	21	1	0	...	0	1	-
9	9	1	0	...	1	0	-
10	39	1	0	...	1	1	+

Pengujian model *K-Nearest Neighbor* membutuhkan data *testing* untuk uji coba terhadap data *training* yang telah terpilih berikut merupakan 20 data *testing* untuk pengujian klasifikasi gejala pasien :

Tabel 4. Data *Testing*

No	usia	jk	G1	...	G5	G6	Label
1	19	1	0	...	1	0	+
2	14	0	1	...	0	0	-
3	20	0	0	...	1	0	-
4	49	0	0	...	1	0	-
5	25	1	0	...	1	0	+
6	28	0	0	...	0	0	+
7	27	1	1	...	0	1	-
8	21	1	0	...	0	1	-
...
20	11	1	1	...	1	0	+

Dalam pengujian menggunakan metode data usia dinormalisasikan menggunakan *MinMax Scaler*. Contoh perhitungan nilai usia menggunakan rumus *MinMax Scaler* dengan

data pasien nomor 1 berusia 20 tahun , nilai terkecil (X_{max}) pada usia pasien adalah 1 tahun dan nilai terbesar (X_{max}) usia pasien adalah 72 pada Tabel 1. Berikut adalah perhitungan usia pada data nomor 1 :

$$X_{new} = \frac{(20 - 1)}{(72 - 1)}$$

$$X_{new} = \frac{19}{71}$$

$$X_{new} = 0,268$$

Perhitungan selanjutnya yaitu memasukkan nilai data *training* dan data *testing* di uji menggunakan rumus *Manhattan*. Berikut merupakan hasil dari pengujian 20 data *testing* terhadap data *training* $K=1$

Tabel 5. Hasil Klasifikasi KNN $K=1$

No Data Ke-	Kelas Asli	Hasil Prediksi
1	+	+
2	-	+
3	-	+
4	-	+
5	+	+
6	+	+
7	-	+
8	-	-
9	-	-
10	+	+
11	+	-
12	+	-
13	-	+
14	-	-
15	-	-
16	-	-
17	-	-
18	+	+
19	+	+
20	+	+

Dalam pengujian klasifikasi menggunakan metode *K-Nearest Neighbor* peneliti menggunakan 5 nilai K diantaranya $K=1$, $K=3$, $K=5$, $K=7$, dan $K=9$. Evaluasi kinerja metode menggunakan confusion matrix berikut merupakan contoh evaluasi kinerja metode *K-Nearest Neighbor* nilai $K=1$:

Tabel 6. Confusion Matrix $K=1$

True Label	Negative	6	5
	Positive	2	7
		Negative	Positive
		Predict Label	

Hasil dari kinerja algoritme *K-Nearest Neighbor* $K=1$ menghasilkan 7 (True Positif) yang berarti model tersebut diprediksi positif

dan benar, 6 (True Negative) model yang diuji diprediksi negatif dan benar, 5 (False Positive) yang berarti label asli adalah negatif dan diprediksi positif, dan 2 (False Negative) yang berarti kelas asli positif tetapi dari 2 sample tersebut diprediksi negatif. Setelah mengisi dari tabel confusion matrix pengukuran kinerja model dapat menggunakan akurasi, *recall*, dan *precision* :

$$\text{Akurasi} = \frac{7+6}{7+6+5+2} \times 100\% = 65\%$$

$$\text{Recall} = \frac{7}{(7+2)} \times 100\% = 77,78\%$$

$$\text{Precision} = \frac{7}{(7+5)} \times 100\% = 58,3\%$$

Hasil Klasifikasi K-Nearest Neighbor

Pengujian 20 data testing dari data gejala pasien bulan Januari - April pada menggunakan algoritme K-Nearest Neighbor dengan nilai K 1, 3, 5, 7, dan 9. Data testing tersebut diproses dengan data training untuk perhitungan jarak rumus Manhattan dan menghasilkan akurasi dari tiap pengujian nilai K sebagai berikut :

Tabel 7. Evaluasi Kinerja K-Nearest Neighbor

Nilai K	Akurasi	Recall	Precision
1	65%	77,78%	58,33%
3	65%	62,5%	55,56%
5	65%	55,56%	62,5%
7	70%	55,56%	71,43%
9	75%	66,67%	75%

Pada tabel 7 memaparkan hasil evaluasi KNN yang meliputi akurasi, recall, dan precision dari pengujian 10 data training dan 20 data testing. Pengujian tersebut meliputi tiap data dari gejala pasien yang di input oleh dokter untuk mengetahui gejala demam tifoid. Hasil evaluasi tersebut menghasilkan peningkatan akurasi dari setiap nilai K yang di uji. Pengujian akurasi pada algoritme digunakan untuk memprediksi data yang benar dari seluruh data yang di uji. Pada hasil akurasi tersebut akurasi tertinggi pada nilai K=9 sebesar 75%, hasil tersebut menunjukkan bahwa model dapat memprediksi sebagian data dengan benar. Pada recall cenderung lebih tinggi pada nilai K=1, saat nilai K meningkat ke K=3, K=5, dan K=7 nilai recall menurun, hal tersebut dapat mempengaruhi deteksi data positif yang salah di klasifikasikan sebagai negatif. Pada pengujian nilai K=9 algoritme dapat memperbaiki kinerja prediksi dalam kelas positif. Pada pengujian

precision algoritme KNN hasil evaluasi cenderung meningkat seiring bertambahnya nilai K, hal ini menunjukkan bahwa algoritme dapat mengurangi kesalahan prediksi kelas positif pada data yang di uji. Precision tertinggi pada pengujian data dengan algoritme KNN diperoleh pada K=9 sebesar 75%.

KESIMPULAN

Bedasarkan dari implementasi dan pengujian sistem dengan algoritme K-Nearest Neighbor mendapatkan kesimpulan dalam analisis setiap gejala pasien yang di uji dengan sistem, dapat memberikan hasil diagnosis demam tifoid pada pasien berdasarkan pengujian dengan data training pada Praktek Bersama dr. Ahmad Izzudin Afif – drg. Desyllia Tanjung.

UCAPAN TERIMA KASIH

Terima kasih kepada dosen pembimbing 1 dan 2 atas dukungan dan bimbingan yang dapat menyelesaikan penulisan penelitian ini, beserta bapak ibu dokter yang telah memberikan izin terhadap pengambilan data pasien untuk kebutuhan penelitian.

DAFTAR PUSTAKA

A'yuniyah, Q., & Reza, M. (2023). Penerapan Algoritma K-Nearest Neighbor Untuk Klasifikasi Jurusan Siswa Di Sma Negeri 15 Pekanbaru. *IJIRSE (Indonesian Journal of Informatic Research and Software Engineering)*, 3(1), 39–45.

Azzahra Nasution, D., Khotimah, H. H., & Chamidah, N. (2019). Perbandingan Normalisasi Data Untuk Klasifikasi Wine Menggunakan Algoritma K-NN. *Jurnal Universitas Negeri Medan*, 4(1), 2502–7131.

Br Sinuhaji, O., Nofriansyah, D., Mariami, I., Informasi, S., & Triguna Dharma, S. (2024). Sistem Pakar Mendiagnosa Penyakit Demam Tifoid (Typhoid Fever) Menggunakan Metode K-Nearest Neighbor. *Jurnal Sistem Informasi Tgd*, 3(2), 283–292. <https://ojs.trigunadharna.ac.id/index.php/jsi>

Dinata, R. K., Akbar, H., & Hasdyna, N. (2020). Algoritma K-Nearest Neighbor dengan Euclidean Distance dan Manhattan

- Distance untuk Klasifikasi Transportasi Bus. *ILKOM Jurnal Ilmiah*, 12(2), 104–111.
<https://doi.org/10.33096/ilkom.v12i2.539>.
104-111
- Dwi Fasnuari, H. A., Yuana, H., & Chulkamdi, M. T. (2022). Penerapan Algoritma K-Nearest Neighbor Untuk Klasifikasi Penyakit Diabetes Melitus. *Antivirus : Jurnal Ilmiah Teknik Informatika*, 16(2), 133–142.
<https://doi.org/10.35457/antivirus.v16i2.2445>
- Harun, R., Chandra Pelangi, K., & Lasena, Y. (2020). Penerapan Data Mining Untuk Menentukan Potensi Hujan Harian Dengan Menggunakan Algoritma K Nearest Neighbor (KNN). *Jurnal Manajemen Informatika & Sistem Informasi*, 3(1), 2614–1701. <http://ejournal.stmiklombok.ac.id/index.php/misi>
- Kumar, M., Chhabra, P., & Garg, N. K. (2018). An efficient content based image retrieval system using BayesNet and K-NN. *Multimedia Tools and Applications*, 77(16), 21557–21570.
<https://doi.org/10.1007/s11042-017-5587-8>
- Nurul Eliza, Y., Hasta Handayani Idrus, K., Irsandy Syahrudin, F., Mangarengi, Y., & Bamahry, A. R. (2023). *Pengaruh Edukasi Pola Makan dan Pola Hidup Bersih dan Sehat terhadap Penyembuhan Pasien Demam Tifoid di RS Ibnu Sina YW-UMI Makassar*.
<https://fmj.fk.umi.ac.id/index.php/fmj>
- Prasetyo, V. R., Hartanto, B., & Mulyono, A. A. (2019). Penentuan Pembimbing Tugas Akhir Mahasiswa Jurusan Teknik Informatika Universitas Surabaya Dengan Metode Dice Coefficient. *Teknika*, 8(1), 44–51.
<https://doi.org/10.34148/teknika.v8i1.147>
- Widaningsih, S. (2019). Perbandingan Metode Data Mining Untuk Prediksi Nilai Dan Waktu Kelulusan Mahasiswa Prodi Teknik Informatika Dengan Algoritma C4,5, Naive Bayes, KNN dan SVM. *Jurnal Tekno Insektif*, 13(1), 16–25.
<https://doi.org/10.36787/jti.v13i1.78>
- Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). *Data mining: Practical machine learning tools and techniques* (4th ed.). Morgan Kaufmann.