ISSN: 2339-2541
JURNAL GAUSSIAN, Volume 15, Nomor 1, Tahun 2026, Halaman 255 - 263
Online di: https://ejournal3.undip.ac.id/index.php/gaussian/

IMPLEMENTASI ALGORITMA *FUZZY K-NEAREST NEIGHBOR* UNTUK KLASIFIKASI PENYAKIT DIARE

Nur Dihyah[1*], Budi Warsito[2], Iut Tri Utami[3]
[1,2,3] Departemen Statistika, Fakultas Sains dan Matematika, Universitas Diponegoro
***e-mail**: nurdihyah25@gmail.com

DOI: 10.14710/j.gauss.15.1.255-263

Article Info:
Received: 2024-09-29
Accepted: 2025-12-30
Available Online: 2026-08-20

Keywords:
Diarrhea; Data Mining; Classification; Fuzzy K-Nearest Neighbor

Abstract: Diarrhea is digestive disruption retrieved by defecation that become more fluid and occur over three times a day. The prevalence of diarrhea in Indonesia is public health problem with high cases. Diarrhea management is carried out with rehydration efforts by administering oral rehydration salts at puskesmas. Diarrhea is classified into 2 types, namely acute diarrhea (mild) and chronic diarrhea (severe). Puskesmas only handles mild diarrhea so that a method is needed to classify diagnosis of diarrhea that occurs at puskesmas appropriately so that diagnosis of acute diarrhea is not misclassified into chronic diarrhea. This research implements Fuzzy K-Nearest Neighbor procedure for Diarrhea Classification. Fuzzy K-Nearest Neighbor incorporates fuzzy logic and K-Nearest Neighbor in the classification practice. The advantage of Fuzzy K-Nearest Neighbor is data will have membership value in each data class so that it further strengthens reason for data to enter predicted class. Data is processed by applying Shiny Package in Rstudio to create Graphical User Interface (GUI-R) so that it makes it easier for researchers to process data. The results obtained highest classification accuracy at K = 3 with accuracy of 80.19% and specificity of 85.45% so that Fuzzy K-Nearest Neighbor was able to classify diarrhea well.

1. PENDAHULUAN

Diare merupakan penyakit gangguan pencernaan yang diindikasikan dengan defekasi yang memicu feses berubah lebih cair serta berlangsung lebih dari tiga kali sehari. Prevalensi diare termasuk dalam kasus kesehatan di masyarakat Indonesia dengan frekuensi yang tinggi dan masih menyumbang tingkat kematian yang cukup tinggi terutama bagi balita. Jumlah pelayanan kasus diare pada semua kelompok umur sebesar 2.473.081 kasus dari sasaran yang ditetapkan pada tahun 2021. Penanganan diare dilakukan dengan upaya rehidrasi dengan menjalankan program Lima Langkah Tuntaskan Diare (Lintas Diare) yang dijalankan melalui pengarahan dan pemberian oralit di puskesmas. Diare pada umumnya dikelompokkan menjadi 2 jenis, yaitu diare akut yang masih termasuk kategori ringan dan diare kronik. Diare akut diklasifikasikan lagi menjadi koleriform yaitu jenis diare yang gejalanya ringan dan disentriform yaitu diare yang kondisi fesesnya terdapat darah dan lendir (Irwan, 2017). Puskesmas pada umumnya hanya menangani kasus diare ringan sehingga dibutuhkan sebuah metode yang mampu mengelompokkan diagnosis diare yang terjadi di puskesmas dengan tepat agar diagnosis diare akut tidak salah diklasifikasikan ke dalam diare kronik.

Salah satu metode yang dapat digunakan untuk menemukan model berdasarkan jenis kelas data adalah klasifikasi. Penelitian mengenai klasifikasi penyakit diare dilakukan menggunakan salah satu algoritma klasifikasi yaitu *Fuzzy K-Nearest Neighbor*. Algoritma *Fuzzy K-Nearest Neighbor* merepresentasikan perbaikan kinerja *K-Nearest Neighbor* dengan menggabungkan konsep tetangga terdekat dan logika *fuzzy* untuk memprediksi kelas data uji. Diare pada jenis koleriform dan disentriform ada yang memiliki gejala yang sama

sehingga bisa menimbulkan kesamaran dalam pemilihan kelas data. Keunggulan *Fuzzy K-Nearest Neighbor* adalah data memiliki nilai keanggotaan pada setiap kelas data sehingga lebih memperkuat alasan data untuk masuk ke dalam kelas yang diprediksi. Jenis data diare pada penelitian ini adalah data numerik dan kategorik yang bisa diformulasikan menggunakan konsep kedekatan jarak masing-masing variabel berdasarkan algoritma *Fuzzy K-Nearest Neighbor* yang memiliki formula sesuai tipe data (Prasetyo, 2014). Penelitian sebelumnya yang dicantumkan untuk menjadi rujukan pada penelitian ini yakni penelitian oleh Nugraha *et al.* (2017) berupa perbandingan algoritma *Fuzzy K-Nearest Neighbor* dan *K-Nearest Neighbor* yang diaplikasikan untuk mengategorikan status gizi balita. Hasil penelitian yang diperoleh yaitu *Fuzzy K-Nearest Neighbor* memiliki akurasi yang lebih unggul diujikan daripada *K-Nearest Neighbor*. Referensi tersebut akan digunakan sebagai petunjuk untuk melakukan penelitian mengenai klasifikasi terhadap penyakit diare yang akan diterapkan pada data penyakit diare Puskesmas Srondol, Kecamatan Banyumanik, Kota Semarang. Aktualisasi penelitian ini bertujuan untuk mengidentifikasi hasil klasifikasi yang terbentuk dan mengevaluasi ketepatan klasifikasi diare yang didapatkan.

2. TINJAUAN PUSTAKA

Diare merupakan penyakit yang disinyalir dengan suatu peralihan wujud dan konsistensi pada feses yaitu lebih lunak atau encer serta kekerapan defekasi lebih dari 3 kali sehari (Setiadi, 2013). Penyakit diare digolongkan menjadi dua kategori, yakni diare akut dan kronik. Diare akut merupakan diare yang masih termasuk kategori ringan dan terjadi kurang dari 14 hari. Diare akut diklasifikasikan lagi menjadi koleriform yaitu jenis diare yang ringan dengan gejala diare pada umumnya dan disentriform yaitu diare yang kondisi fesesnya terdapat darah dan lendir. Diare kronik merupakan diare yang terjadi dengan waktu lebih dari 14 hari dan termasuk kategori berat (Irwan, 2017).

Klasifikasi didefinisikan sebagai mekanisme penemuan model yang bertujuan untuk memperkirakan kelas data uji berdasarkan data latih yang telah diketahui (Han *et al.*, 2011). *Fuzzy K-Nearest Neighbor* merupakan algoritma dalam klasifikasi berdasarkan logika *fuzzy*, tetangga terdekat, dan kedekatan jarak pada data uji terhadap setiap data latih. Konsep *Fuzzy K-Nearest Neighbor* adalah memprediksi data uji dengan mengaplikasikan nilai keanggotaan terhadap masing-masing kelas data dengan nilai keanggotaan berkisar antara 0 sampai 1 (Prasetyo, 2014).

Pengukuran jarak tetangga terdekat pada metode *Fuzzy K-Nearest Neighbor* dapat menerapkan jarak *euclidean*. Jarak *euclidean* memiliki *similarity* yang tinggi pada jarak dua titik dan terdapat formula yang berbeda untuk masing-masing tipe data sehingga lebih sesuai untuk diaplikasikan pada jenis data penelitian ini. Formula jarak *euclidean* dapat dihitung sesuai dengan Persamaan (1) (Han *et al.*, 2011).

$$d(x_i, x_h) = \sqrt{\sum_{l=1}^{P} (diff(x_{il}, x_{hl}))^2} \quad (1)$$

dengan keterangan:

$d(x_i, x_h)$: jarak *euclidean* dari data x_i terhadap data x_h
P : banyakya variabel bebas
x_{il} : data uji ke-i dalam variabel ke-l
x_{hl} : data latih ke-h dalam variabel ke-l
$diff(x_{il}, x_{hl})$: selisih data uji ke-i dengan data latih ke-h pada variabel ke-l.

Kalkulasi nilai selisih jarak *euclidea*n bergantung pada tipe data yang digunakan. Formula perhitungan nilai selisih menurut tipe data dijabarkan pada Tabel 1 (Prasetyo, 2014).

Tabel 1. Formula Selisih pada Jarak *Euclidean* Berdasarkan Tipe Data

Tipe Atribut	Formula Jarak
Nominal	$diff(x_{il}, x_{hl}) = \begin{cases} 0 & jika\ x_{il} = x_{hl} \\ 1 & jika\ x_{il} \neq x_{hl} \end{cases}$
Ordinal	$diff(x_{il}, x_{hl}) = \frac{\lvert x_{il} - x_{hl} \rvert}{(q-1)}$
Rasio atau Interval	$diff(x_{il}, x_{hl}) = \lvert x_{il} - x_{hl} \rvert$

dengan keterangan:

$diff(x_{il}, x_{hl})$: nilai selisih data uji ke-i dengan data latih ke-h pada variabel ke-l
q : jumlah pengkategorian dalam x
x_{il} : data uji ke-i pada variabel prediktor ke-l
x_{hl} : data latih ke-h pada variabel prediktor ke-l

Hasil penentuan jarak *euclidean* dari data uji terhadap data latih disusun secara urut mulai data yang paling kecil ke yang paling besar, kemudian dipilih sebanyak K tetangga terdekat dari pengurutan jarak *euclidean* yaitu sebanyak K jarak terkecil sehingga membentuk data tetangga dan dinamakan x_g, dengan $g = 1,2, \dots, K$. Hasil pengurutan jarak *euclidean* dan penentuan nilai K selanjutnya digunakan untuk menentukan nilai keanggotaan pada data tetangga terdekat. Nilai keanggotaan dari data tetangga x_n ke dalam kelas c_j dapat diformulasikan sesuai dengan Persamaan (2).

$$\mu(x_g, c_j) = \begin{cases} 0 & jika\ x_g\ bukan\ milik\ kelas\ c_j \\ 1 & jika\ x_g\ milik\ kelas\ c_j \end{cases} \quad (2)$$

Nilai keanggotaan data tetangga x_g kemudian digunakan untuk menghitung nilai keanggotaan masing-masing kelas data uji x_i. Li et al. (2007) menyatakan bahwa nilai keanggotaan data uji x_i ke dalam kelas c_j dapat dihitung menggunakan formula pada Persamaan (3).

$$\mu(x_i, c_j) = \frac{\sum_{n=1}^{K}[\mu(x_g, c_j)] \times [d(x_i, x_g)^{\frac{-2}{(m-1)}}]}{\sum_{n=1}^{K}[d(x_i, x_g)^{\frac{-2}{(m-1)}}]} \quad (3)$$

dengan keterangan:

c_j : kelas ke- j dengan $j = 1,2, \dots, C$
C : banyaknya kelas pada variabel respon
x_i : data uji ke-i
x_g : yaitu data tetangga ke-g pada K tetangga terdekat
$\mu(x_i, c_j)$: nilai keanggotaan data uji x_i ke kelas c_j
K : banyaknya tetangga terdekat yang diterapkan
$\mu(x_g, c_j)$: nilai keanggotaan dari data tetangga x_g pada K tetangga terdekat di kelas c_j,
$d(x_i, x_g)$: jarak *euclidean* data uji x_i terhadap data tetangga x_g
m : banyaknya pembobot pangkat dengan m > 1, m ditetapkan dengan melihat banyaknya kategori variabel respon yang ditetapkan.

Penjabaran langkah-langkah dari konsep *Fuzzy K-Nearest Neighbor* (FK-NN) adalah sebagai berikut:

1. Menetapkan nilai K tetangga terdekat. Nilai K menggunakan K ganjil karena jumlah kategori kelas data bernilai genap. Contoh K yang diterapkan pada penelitian ini yakni K=3.
2. Menghitung jarak *euclidean* menggunakan Persamaan (1).

3. Melakukan kalkulasi nilai keanggotaan pada data uji dalam masing-masing kelas menggunakan Persamaan (3).
4. Memilih nilai keanggotaan terbesar untuk ditentukan menjadi hasil akhir kelas data uji dengan formula pada Persamaan (4) (Prasetyo, 2012).

$$y' = ArgMax_{j=1}^{c}(\mu(x_i, c_j)) \tag{4}$$

Evaluasi kinerja klasifikasi dapat diukur dengan menggunakan *confusion matrix* yang dapat didefinisikan sebagai alat pengujian yang mampu diterapkan untuk menentukan seberapa baik hasil kategorisasi dilakukan. *Confusion matrix* diuraikan dalam tabel yang berisi klasifikasi jumlah data yang tepat dan dan tidak tepat sesuai dengan Tabel 2 (Indriani, 2014).

Tabel 2. Pengukuran *Confusion Matrix Classification*

		Kelas Prediksi	
		Positif	Negatif
Kelas Aktual	Positif	*True Positive* (TP)	*False Negative* (FN)
	Negatif	*False Positive* (FP)	*True Negative* (TN)

Ukuran ketepatan klasifikasi yang dipertimbangkan adalah akurasi dan spesifisitas. Akurasi merupakan penghitungan banyaknya data yang dikategorikan dengan benar (Qadrini *et al*., 2021). Formula untuk menghitung akurasi ditulis dalam Persamaan (5).

$$Akurasi = \frac{TP + TN}{TP + FP + TN + FN} \times 100\% \tag{5}$$

Spesifisitas merupakan perbandingan antara banyaknya *True Negative* (TN) dengan banyaknya nilai aktual negatif (Hidayat *et al*., 2021). Spesifisitas dapat dihitung menggunakan formula pada Persamaan (6).

$$Spesifisitas = \frac{TN}{TN + FP} \times 100\% \tag{6}$$

3. METODE PENELITIAN

Penelitian ini menerapkan data sekunder yang didapatkan dari Puskesmas Srondol, Kecamatan Banyumanik, Kota Semarang. Data tersebut merupakan data penyakit diare tahun 2022 yang terdiri atas 528 data. Variabel penelitian terdiri atas 2 variabel yakni variabel respon dan variabel prediktor. Variabel respon yaitu diagnosis penyakit diare (Y) dibagi menjadi dua kategori yaitu koleriform dan disentriform. Variabel prediktor terdiri dari 13 variabel yaitu umur (X_1), kondisi feses (X_2), frekuensi buang air besar (X_3), pusing (X_4), mual (X_5), muntah (X_6), demam (X_7), perut melilit (X_8), kram perut (X_9), kembung (X_{10}), lemas (X_{11}), lendir (X_{12}), dan darah (X_{13}).

Struktur data pada penelitian ini terdiri atas n objek data, P variabel prediktor, dan variabel respon (y). Struktur data penelitian ditampilkan dalam Tabel 3.

Tabel 3. Struktur Data Penelitian

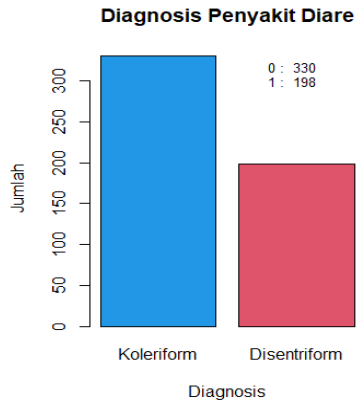
	x_1	x_2	...	x_P	y
1	x_{11}	x_{12}	...	x_{1P}	y_1
2	x_{21}	x_{22}	...	x_{2P}	y_2
3	x_{31}	x_{32}	...	x_{3P}	y_3
4	x_{41}	x_{42}	...	x_{4P}	y_4
⋮	⋮	⋮	⋱	⋮	⋮
n	x_{n1}	x_{n2}	...	x_{nP}	y_n

Pengolahan data dalam penelitian ini mengimplementasikan algoritma *Fuzzy K-Nearest Neighbor* dan diterapkan menggunakan *software Rstudio 4.3.0*. Tahap-tahap yang digunakan untuk melakukan pengolahan dan analisis data di penelitian ini diuraikan melalui 6 tahap, yaitu:

1. Mengumpulkan data dan memasukkan data ke dalam sebuah *file* menggunakan *Microsoft Excel 2016* dari *database* Puskesmas Srondol.
2. Memasukkan data ke dalam *database* program *Rstudio 4.3.0.*
3. Melakukan *preprocessing* data berupa melengkapi *missing values*.
4. Melakukan klasifikasi dengan *Fuzzy K-Nearest Neighbor.*
   a. Memisahkan data untuk dijadikan data uji dan data latih dengan perbandingan data latih dengan data uji yakni 80% : 20%.
   b. Menetapkan K tetangga terdekat.
   c. Melakukan kalkulasi jarak *euclidean* pada semua data latih dengan data uji memakai formulasi sesuai Persamaan (1).
   d. Mengurutkan data yang memiliki jarak *euclidean* dari yang terkecil hingga terbesar.
   e. Menghitung nilai keanggotaan $\mu(x_g, c_j)$ dari data tetangga x_g dalam kelas c_j dengan menggunakan formula pada Persamaan (2).
   f. Mengukur nilai keanggotaan data uji x_i atau nilai $\mu(x_i, c_j)$ ke dalam kelas data c_j, menggunakan Persamaan (3).
   g. Menetapkan nilai keanggotaan yang paling besar dari hasil penghitungan nilai $\mu(x_i, c_j)$ untuk dijadikan kelas data menggunakan persamaan (4). Kelas yang memiliki nilai keanggotaan terbesar akan dipilih sebagai kelas prediksi data uji.
5. Melakukan evaluasi model dengan melihat *Confusion Matrix.*
   a. Membentuk *Confusion Matrix* dengan dua kelas.
   b. Menghitung nilai *accuracy* dan *specifity*.
   c. Menentukan model terbaik.
6. Melakukan analisis dari hasil klasifikasi dan menarik kesimpulan.

4. HASIL DAN PEMBAHASAN

Data diagnosis penyakit diare terbagi menjadi kategori koleriform dan disentriform dengan jumlah seperti dalam Gambar 1.



Gambar 1. Jumlah Kategori Diagnosis Diare

Gambar 1 menunjukkan bahwa jumlah kategori koleriform yaitu sebanyak 330 data atau sebesar 62,5%, sedangkan jumlah kategori disentriform yaitu sebanyak 198 data atau sebesar 37,5%. Jumlah pasien diare koleriform lebih banyak dari diare disentriform karena puskesmas lebih banyak menangani kasus diare yang ringan.

Data penyakit diare perlu disortir untuk dicek kelengkapan data dan dilakukan perbaikan melalui tahap *preprocessing* data. Data mempunyai *missing value* pada variabel perut melilit (X_8) dan akan ditangani dengan mengisi variabel menggunakan nilai modus. Hasil penanganan *missing value* disajikan dalam Tabel 4.

Tabel 4. Penanganan *Missing Value* pada Variabel X_8

X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	X_{13}	Y
2,4	0	3	0	0	0	0	1	1	0	0	0	0	0
2,1	0	5	0	0	0	1	1	0	0	0	0	0	0

Data penyakit diare dibagi menjadi 2 bagian yaitu data latih dan data uji dengan perbandingan 80:20. *Dataset* yang sudah dipisahkan kemudian ditentukan nilai K tetangga terdekatnya yang dicontohkan menggunakan K = 3 karena memiliki ketepatan klasifikasi tertinggi dibanding nilai K lainnya. Nilai K digunakan untuk menentukan data tetangga terdekat pada hasil penghitungan dan pengurutan jarak *euclidean* berdasarkan Persamaan (1). Contoh penerapan nilai K = 3 pada penghitungan jarak *euclidean* pada data uji ke-1 terhadap 422 data latih disajikan sesuai dengan Tabel 5.

Tabel 5. Urutan Jarak *Euclidean* Data Uji ke-1 terhadap Data Latih

No	Data Latih ke-	X_1	X_2	X_3	X_4	X_5	…	X_{13}	Y	Jarak
1	2	19	1	5	1	1	…	1	1	1,32
2	402	18,9	1	4	0	1	…	0	0	2,57
3	10	20,1	1	5	1	1	…	1	1	2,63
4	311	18	0	3	1	1	…	0	0	2,65
5	291	19,8	1	5	1	1	…	1	1	2,74
6	191	19,1	1	3	0	1	…	0	0	2,81
7	365	19,4	1	7	1	1	…	1	1	2,91
8	178	19,1	1	8	1	1	…	1	1	2,98
9	364	18,5	1	5	1	1	…	1	1	3,02
10	35	18,7	1	3	1	1	…	0	0	3,05
…	…	…	…	…	…	…	…	…	…	…
422	87	89,9	0	3	0	0		1	1	71,89

Nilai K sebanyak 3 menunjukkan banyaknya 3 tetangga terdekat untuk dilihat kelas datanya pada variabel diagnosis sehingga dapat digunakan untuk menghitung nilai keanggotaan. Nilai yang dihasilkan dari jarak *euclidean* selanjutnya digunakan untuk menghitung nilai keanggotaan masing-masing kelas data tetangga terdekat pada variabel diagnosis menggunakan Persamaan (2). Nilai keanggotaan yang dihasilkan dari proses klasifikasi pada K = 3 ditampilkan pada Tabel 6.

Tabel 6. Nilai Keanggotaan pada Data Tetangga Terdekat Sebanyak 3

No	Jarak	Kelas Data	Nilai Keanggotaan Kelas Koleriform (0)	Nilai Keanggotaan Kelas Disentriform (1)
1	1,32	1	0	1
2	2,57	1	0	1
3	2,63	0	1	0

Tabel 6 menunjukkan bahwa data ke-1 memiliki nilai keanggotaan sebesar 0 pada kelas koleriform karena memiliki kelas data yang berbeda, sedangkan pada pada kelas data disentriform memiliki nilai keanggotaan sebesar 1 karena memiliki kelas data yang sama. Data ke-2 juga memiliki nilai keanggotaan sebesar 0 pada kelas koleriform karena memiliki kelas data yang berbeda, sedangkan pada pada kelas data disentriform memiliki nilai keanggotaan sebesar 1 karena memiliki kelas data yang sama. Data ke-3 memiliki nilai keanggotaan sebesar 1 pada kelas koleriform karena memiliki kelas data yang sama, sedangkan pada pada kelas data disentriform memiliki nilai keanggotaan sebesar 1 karena memiliki kelas data yang berbeda.

Hasil nilai keanggotaan pada masing-masing tetangga terdekat digunakan untuk menghitung nilai keanggotaan pada data uji untuk menetapkan hasil akhir kategori kelas data melalui penerapan *Fuzzy K-Nearest Neighbor*. Penentuan nilai keanggotaan masing-masing kelas data uji dihitung menggunakan formula pada Persamaan (3) dengan nilai K = 3 pada data uji ke-1. Penghitungan nilai keanggotaan untuk kelas koleriform dan disentriform menggunakan Persamaan (3) adalah sebagai berikut:

$$(x_1, c_0) = \frac{\left(0 \times \left(1{,}32^{\frac{-2}{2-1}}\right)\right) + \left(0 \times \left(2{,}57^{\frac{-2}{2-1}}\right)\right) + \left(1 \times \left(2{,}63^{\frac{-2}{2-1}}\right)\right)}{\left(1{,}32^{\frac{-2}{2-1}}\right) + \left(2{,}57^{\frac{-2}{2-1}}\right) + \left(2{,}63^{\frac{-2}{2-1}}\right)}$$

$$(x_1, c_0) = \frac{0{,}1447}{0{,}8741} = 0{,}1655$$

$$(x_1, c_1) = \frac{\left(1 \times \left(1{,}32^{\frac{-2}{2-1}}\right)\right) + \left(1 \times \left(2{,}57^{\frac{-2}{2-1}}\right)\right) + \left(0 \times \left(2{,}63^{\frac{-2}{2-1}}\right)\right)}{\left(1{,}32^{\frac{-2}{2-1}}\right) + \left(2{,}57^{\frac{-2}{2-1}}\right) + \left(2{,}63^{\frac{-2}{2-1}}\right)}$$

$$(x_1, c_1) == \frac{0{,}7294}{0{,}8741} = 0{,}8345$$

Hasil penghitungan nilai keanggotaan masing-masing kelas data uji menunjukkan bahwa nilai kenggotaan untuk kelas koleriform adalah sebesar 0,1655, sedangkan nilai keanggotaan untuk kelas disentriform adalah sebesar 0,8345. Penghitungan tersebut menunjukkan bahwa nilai keanggotaan terbesar adalah nilai keanggotaan kelas disentriform sehingga data uji ke-1 termasuk dalam kelas disentriform atau kelas 1. Contoh salah satu hasil klasifikasi menggunakan K tetangga terdekat sebanyak 3 dengan tampilan dalam bentuk kategorik 0 untuk kelas koleriform dan 1 untuk kelas disentriform diuraikan pada Tabel 7.

Tabel 7. Hasil Klasifikasi pada K = 3

Data Uji	Diagnosis Aktual	Diagnosis Prediksi
1	1	1
2	1	1
3	1	0
4	0	1
5	1	1
6	1	1
7	0	1
8	0	0
9	1	0
10	0	0
…	…	…
106	0	0

Tabel 7 menunjukkan bahwa sebanyak 85 data diklasifikasikan dengan tepat dan sebanyak 21 data salah diklasifikasikan. Sebanyak 51 data dikelompokkan untuk masuk ke kelas koleriform, sedangkan 55 data dikelompokkan ke kelas disentriform.

Evaluasi dari proses klasifikasi dapat diukur berdasarkan *confusion matrix* untuk melihat ketepatan algoritma *Fuzzy K-Nearest Neighbor* dalam menggolongkan data pada setiap kelas data uji. Perhitungan *confusion matrix* menerapkan nilai K sebanyak 3 karena hasil evaluasi mendapatkan ketepatan klasifikasi tertinggi pada K = 3 dibandingkan dengan nilai K tetangga terdekat yang lain. Berdasarkan hasil confusion matrix pada nilai K = 3, diperoleh bahwa dari 51 data aktual koleriform, sebanyak 38 data berhasil diklasifikasikan

dengan benar sebagai koleriform, sedangkan 13 data salah diklasifikasikan sebagai disentriform. Sementara itu, dari 55 data aktual disentriform, sebanyak 47 data berhasil diklasifikasikan dengan benar sebagai disentriform, sedangkan 8 data salah diklasifikasikan sebagai koleriform. Dengan demikian, secara keseluruhan terdapat 85 data yang terklasifikasi dengan tepat dan 21 data yang mengalami kesalahan klasifikasi dari total 106 data uji. Hasil ini menunjukkan bahwa algoritma Fuzzy K-Nearest Neighbor dengan K = 3 mampu memberikan performa klasifikasi yang cukup baik dalam membedakan diagnosis diare koleriform dan disentriform. Ukuran ketepatan klasifikasi, yaitu akurasi dan spesifisitas, selanjutnya dihitung berdasarkan jumlah klasifikasi benar dan salah tersebut menggunakan Persamaan (5) dan (6).

$$Akurasi = \frac{38 + 47}{38 + 8 + 47 + 13} \times 100 = 80{,}19\%$$

$$Spesifisitas = \frac{47}{47 + 8} \times 100\% = 85{,}45\%$$

Penghitungan nilai akurasi dengan K = 3 mendapatkan hasil sebesar 80,19%. Besarnya nilai akurasi tersebut menunjukkan bahwa data uji berjumlah 106 data terklasifikasikan secara benar sebesar 80,19% dan terklasifikasikan secara salah sebesar 19,81%. Hasil penghitungan spesifisitas dengan K = 3 memperoleh nilai sebesar 85,45%. Besarnya nilai spesifisitas tersebut menunjukkan bahwa data uji sebanyak 106 data mampu mengklasifikasikan kelas negatif atau kelas disentriform secara benar sebesar 85,45% dan terklasifikasikan secara salah sebesar 14,55%. Nilai-nilai ketepatan klasifikasi tersebut menunjukkan bahwa algoritma *Fuzzy K-Nearest Neighbor* mampu mengklasifikasikan data penyakit diare secara baik.

5. KESIMPULAN

Kesimpulan yang didapatkan dari pengujian *Fuzzy K-Nearest Neighbor* dalam melakukan klasifikasi penyakit diare yaitu hasil klasifikasi penyakit diare dapat teridentifikasi dan diklasifikasikan dengan baik pada masing-masing nilai K tetangga terdekat dan memperoleh ketepatan klasifikasi tertinggi untuk nilai K = 3 yang memiliki nilai akurasi sebesar 80,19% dan spesifisitas sebesar 85,45%. Hasil klasifikasi diagnosis diare dapat digunakan menjadi model untuk memprediksi data dengan variabel yang sama sehingga GUI R mampu diterapkan sebagai alat untuk mengklasifikasikan data dengan lebih cepat dan efisien.

DAFTAR PUSTAKA

Han, J., Kamber, M., dan Pei, J. (2011). *Data Mining: Concepts and Techniques 3rd Edition*. Waltham: Morgan Kaufmann.

Hidayat, W., Ardiansyah, M., dan Setyanto, A. (2021). Pengaruh Algoritma ADASYN dan SMOTE terhadap Performa Support Vector Machine pada Ketidakseimbangan Dataset Airbnb. *Jurnal Pendidikan Informatika,* Vol. 5, No. 1: 11-20.

Indriani, A. (2014). Klasifikasi Data Forum dengan Menggnakan Metode Naive Bayes Classifier. *Seminar Nasional Aplikasi Teknologi Informasi,* Vol. 1, No. 1: 5-10.

Irwan. (2017). *Epidemiologi Penyakit Menular.* Yogyakarta: CV. Absolute Media.

Li, D., Deogun, J. S., dan Wang, K. (2007). *Gene Function Classification Using Fuzzy K-Nearest Neighbor Approach*. 644-647.

Nugraha, S. D., Putri, R. R., dan Wihandika, R. C. (2017). Penerapan Fuzzy K-Nearest Neighbor (FK-NN) Dalam Menentukan Status Gizi Balita. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer,* Vol. 1, No. 9: 925-932.

Prasetyo, E. (2012). *Data Mining: Konsep dan Aplikasi Menggunakan Matlab.* Yogyakarta: Penerbit Andi.

Prasetyo, E. (2014). *Data Mining: Mengolah Data menjadi Informasi Menggunakan MATLAB.* Yogyakarta: Penerbit Andi.

Qadrini, L., Seppewali, A., & Aina, A. (2021). Decision Tree dan Adaboost pada Klasifikasi Penerima Program Bantuan Sosial. *Jurnal Inovasi Penelitian,* Vol. 2, No. 7: 1959-1966.

Setiadi. (2013). *Konsep dan Praktik Penulisan Keperawatan.* Yogyakarta: Graha Ilmu.