

IMPLEMENTASI METODE *NEIGHBOR WEIGHTED K-NEAREST NEIGHBOR* PADA PENGKLASIFIKASIAN STATUS GIZI BALITA DI WILAYAH KERJA PUSKESMAS WONOREJO KOTA SAMARINDA

Implementation of the Neighbor Weighted K-Nearest Neighbor Method in the Classification of Nutritional Status of Toddlers in the Work Area of Puskesmas Wonorejo, Samarinda City

Putri Aisha¹, M. Fathurahman^{2*}, Surya Prangga³

^{1,2,3}Program Studi Statistika, Jurusan Matematika FMIPA Universitas Mulawarman

Jl. Barong Tongkok No. 4 Kampus Gunung Kelua, Samarinda-Kalimantan Timur 75123 Indonesia

E-mail Corresponding Author: fathur@fmipa.unmul.ac.id

Abstrak: *Neighbor Weighted K-Nearest Neighbor* (NWKNN) merupakan salah satu metode klasifikasi dalam *machine learning* yang dikembangkan dari metode *K-Nearest Neighbor* (KNN). Metode NWKNN memiliki kelebihan dibanding KNN, terutama untuk menangani masalah jumlah kelas yang tidak seimbang dalam pengklasifikasian data *mining*. Konsep dari metode NWKNN adalah pemberian bobot yang besar untuk kelas minoritas dan bobot yang kecil untuk kelas mayoritas. Metode NWKNN dalam penelitian ini diterapkan pada pengklasifikasian status gizi balita di Kota Samarinda. Menurut *World Health Organization* (WHO), Indonesia merupakan salah satu negara dengan prevalensi *stunting* pada balita tertinggi di regional Asia Tenggara pada Tahun 2005 sampai dengan 2017, dengan rata-rata prevalensi *stunting* pada balita sebesar 36,4%. Status gizi adalah salah satu faktor penyebab terjadinya *stunting* pada balita. Penelitian ini bertujuan mendapatkan hasil pengklasifikasian status gizi balita menggunakan metode NWKNN dan nilai akurasi. Data yang digunakan adalah data balita di wilayah kerja Puskesmas Wonorejo Kota Samarinda Tahun 2022. Berdasarkan hasil klasifikasi status gizi balita menggunakan metode NWKNN, terdapat 93 balita yang diklasifikasikan secara tepat dari 128 balita dengan nilai akurasi sebesar 72,65%. Nilai akurasi ini termasuk cukup baik dan menunjukkan bahwa metode NWKNN layak digunakan untuk memprediksi ketepatan klasifikasi status gizi balita di wilayah kerja Puskesmas Wonorejo, Kota Samarinda.

Kata Kunci: Data Mining, Klasifikasi, *Machine Learning*, NWKNN, Status Gizi Balita

Abstract: *Neighbor-weighted K-nearest neighbor* (NWKNN) is a classification method in machine learning that was developed from the *K-nearest neighbor* (KNN) method. The NWKNN method has advantages over KNN, especially for dealing with the problem of an unbalanced number of classes in data mining classification. The NWKNN method gives large weights to the minority class and small weights to the majority class. The NWKNN method in this research was applied to classify the nutritional status of toddlers in Samarinda City. According to the *World Health Organization* (WHO), Indonesia is one of the countries with the highest prevalence of *stunting* in toddlers in the Southeast Asia region from 2005 to 2017, with an average prevalence of *stunting* in toddlers of 36.4%. Nutritional status is one of the factors causing *stunting* in toddlers. This research aims to obtain the results of classifying the nutritional status of toddlers using the NWKNN method and its accuracy values. The data used is data on toddlers in the work area of the Wonorejo Health Center, Samarinda City, in 2022. Based on the results of the classification of nutritional status of toddlers using the NWKNN method, 93 toddlers were classified correctly out of 128 toddlers, with an accuracy value of 72.65%. This accuracy value is quite good and shows that the NWKNN method is suitable for predicting the accuracy of the classification of nutritional status of toddlers in the working area of the Wonorejo Health Center, Samarinda City.

Keywords: Classification, Data Mining, Machine Learning, Nutritional Status of Toddlers, NWKNN

1. PENDAHULUAN

Data *mining* merupakan serangkaian proses dalam melakukan penambangan informasi atau pengetahuan yang ada pada basis data yang besar [1]. Berbagai metode telah dikembangkan dan diterapkan untuk pengklasifikasian data *mining* melalui *machine learning*. *K-Nearest Neighbor* (KNN) merupakan salah satu metode *supervised learning* yang merupakan bagian dari *machine learning* [2]. Konsep dari metode KNN adalah mengklasifikasikan data *mining* yang memiliki kemiripan ciri atau sifat berdasarkan jarak ketetanggaan. Pengklasifikasian data *mining* menggunakan metode KNN dapat dilakukan dengan mengidentifikasi sejumlah k tetangga terdekat kemudian mengambil kelas terbanyak dalam menentukan jenis kelas dari data. KNN termasuk metode *supervised learning* sederhana dibanding metode *supervised learning* lainnya karena hanya menggunakan memori dan tidak membentuk model pada kerangka kerjanya [3]. Adanya keterbatasan ini, mengakibatkan metode KNN sensitif terhadap data *mining* yang memiliki kelas dengan frekuensi datanya tidak seimbang. Pada kenyataannya, terdapat banyak data *mining* yang memiliki sifat kelas dengan frekuensi data yang tidak seimbang. Kinerja dari metode KNN memilih kelas mayoritas sebagai data *training* dalam jumlah yang besar, sedangkan untuk kelas minoritas dalam jumlah yang kecil [4]. Fenomena ini menyebabkan tingkat akurasi metode KNN menjadi rendah dan tidak layak digunakan untuk memprediksi ketepatan klasifikasi data *mining* yang memiliki sifat kelas dengan frekuensi data yang tidak seimbang. Oleh karena itu, dikembangkan metode *Neighbor Weighted K-Nearest Neighbor* (NWKNN) untuk mengatasi kelemahan dari metode KNN.

Metode NWKNN merupakan metode yang dikembangkan dari KNN dengan kerangka kerja menggunakan pembobotan, yakni pemberian bobot yang besar untuk kelas minoritas dan bobot yang kecil untuk kelas mayoritas pada variabel *output*. Berdasarkan prinsip pembobotan ini, metode NWKNN efektif untuk mengklasifikasikan data *mining* yang memiliki sifat kelas dengan frekuensi data yang tidak seimbang dan memiliki tingkat akurasi yang baik [5]. Penelitian yang mengkaji dan menerapkan metode NWKNN telah banyak dilakukan. [6] menerapkan metode NWKNN pada analisis *text mining*, khususnya pengklasifikasian dokumen *review* aplikasi *mobile*. Hasil dari penelitian ini menunjukkan bahwa metode NWKNN memiliki akurasi sebesar 79,7% pada komposisi data *training* dan *testing*, masing-masing sebesar 80% dan 20%. [7] mengklasifikasikan risiko hipertensi menggunakan NWKNN. Penelitian ini menghasilkan metode NWKNN mampu mengklasifikasikan risiko hipertensi dengan baik pada *training* sebanyak 100 data *training* dan data *testing* sebanyak 25 data. Nilai ketetanggaan (K) dan eksponen (E) yang optimal adalah $K = 10$, dan nilai $E = 4$ dengan akurasi sebesar 88%. [8] Mengklasifikasikan penyakit gagal ginjal menggunakan metode NWKNN. Hasil penelitian ini menunjukkan bahwa metode NWKNN dapat mengklasifikasikan penyakit gagal ginjal pada saat data *training* yang digunakan sebanyak 150 data dan data *testing* sebanyak 50 data dengan nilai $K = 2$ dan nilai $E = 2$ dengan hasil akurasi sebesar 88%. [9] menggunakan metode NWKNN untuk mengklasifikasikan status pembayaran premi pada perusahaan asuransi jiwa. Penelitian ini menghasilkan nilai K dan E yang optimal untuk klasifikasi status pembayaran premi menggunakan NWKNN sebesar $K = 3$ dan $E = 6$ dengan nilai akurasi sebesar 75%. [10] Mengklasifikasikan jenis *Attention Deficit Hyperactivity Disorder* (ADHD) pada anak usia dini menggunakan NWKNN. Hasil penelitian ini menunjukkan bahwa metode NWKNN dapat mengklasifikasikan jenis ADHD pada usia dini dengan baik pada data *training* yang digunakan sebanyak 80 data dan data *testing* sebanyak 20 data dengan nilai $K = 10$ dan nilai $E = 4$ dengan hasil akurasi sebesar 95%. [11] Mengklasifikasikan jenis ADHD pada anak usia dini menggunakan metode KNN, Fuzzy KNN (FKNN), dan NWKNN. Metode NWKNN memiliki rata-rata nilai akurasi tertinggi dibanding KNN dan FKNN. [12] mengklasifikasikan kinerja akademik siswa menggunakan NWKNN dengan seleksi fitur *information gain*. Hasil dari penelitian ini memberikan hasil akurasi sebesar 71,9% dengan nilai optimum dari $K = 45$ dan $E = 6$. [13] membandingkan metode KNN dan NWKNN pada pengklasifikasian risiko kredit perbankan. Penelitian ini menghasilkan metode NWKNN sebagai metode terbaik untuk memprediksi klasifikasi risiko kredit dengan nilai akurasi sebesar 95,73%. [14] mengkaji dan menerapkan metode NWKNN pada pengklasifikasian demam dengan hasil yang diperoleh adalah nilai K dan E yang berbeda-beda tidak berpengaruh terhadap akurasi, dengan nilai akurasi sebesar 100%.

Metode NWKNN pada penelitian ini diimplementasikan pada pengklasifikasian status gizi balita di wilayah kerja Puskesmas Wonorejo Kota Samarinda. Pada Tahun 2022 masih ditemukan balita gizi kurang dan gizi buruk di wilayah kerja Puskesmas Wonorejo Kota Samarinda. Balita merupakan anak yang berusia kurang dari 5 tahun, yakni berusia 1 sampai dengan 3 tahun dan anak usia prasekolah. Anak balita adalah salah satu kelompok rentan

gizi. Masa balita merupakan suatu masa dimana periode perkembangan fisik dan mental yang pesat, sehingga memerlukan gizi yang baik. Jika balita mengalami gizi buruk, maka perkembangan otaknya kurang dan akan berpengaruh pada kehidupan di usia sekolah dan prasekolah. Kesehatan balita sangat dipengaruhi oleh gizi yang terserap di dalam tubuh [15].

2. METODOLOGI

2.1. Data dan Variabel Penelitian

Data yang digunakan pada penelitian ini adalah data sekunder yang diperoleh dari Puskesmas Wonorejo Kota Samarinda Tahun 2022. Berdasarkan data penelitian terbentuk 2 variabel penelitian, yaitu variabel *input* dan *output*. Variabel *input* terdiri atas jenis kelamin (X_1), usia (X_2), berat badan (X_3), dan tinggi badan (X_4). Adapun variabel *output* adalah status gizi balita (Y). Variabel penelitian secara lengkap disajikan pada Tabel 1.

Tabel 1. Variabel Penelitian

Variabel	Nama Variabel	Skala Pengukuran Variabel
X_1	Jenis kelamin, dengan kategori: 1 = Laki-laki 2 = Perempuan	Nominal
X_2	Usia	Rasio
X_3	Berat badan	Rasio
X_4	Tinggi badan	Rasio
Y	Status gizi balita, dengan kategori: 1 = Gizi buruk 2 = Gizi kurang 3 = Gizi baik 4 = Gizi lebih.	Ordinal

2.2. Tahapan Analisis Data

Tahapan analisis data dalam penelitian ini adalah sebagai berikut:

1. Melakukan analisis statistika deskriptif terhadap data penelitian yang bertujuan untuk mengetahui gambaran awal dari data pada variabel input dan output.
2. Melakukan standarisasi data pada variabel usia (X_2), berat badan (X_3), dan tinggi badan (X_4) menggunakan Persamaan (1) [16].

$$\bar{x}_q = \frac{1}{n} \sum_{i=1}^n \bar{x}_{iq} \quad (1)$$

3. Melakukan pengacakan data dengan menggunakan *software R*, dengan tujuan agar semua data memiliki peluang yang sama untuk terpilih menjadi data *training* dan data *testing*.
4. Menentukan data *training* dan data *testing* dengan proporsi yang digunakan sebesar 70% data *training* dan 30% data *testing* (70:30), dan 90% data *training* dan 10% data *testing* (90:10). Data *training* digunakan untuk membentuk NWKNN, sedangkan data *testing* digunakan untuk memprediksi ketepatan klasifikasi variabel *output* [16]. Penentuan data *training* (tr) dan data *testing* (te) menggunakan formula pada Persamaan (2) dan (3).

$$tr = p \times n \quad (2)$$

$$te = n - tr \quad (3)$$

5. Menentukan nilai K , yaitu $K = 3, 5, 7$, dan 9 .

6. Menentukan banyaknya data dalam *subset* data *training* menggunakan metode *k-fold cross validation* [17]. Metode ini dapat menjelaskan rata-rata keberhasilan dari suatu sistem dengan cara melakukan pengulangan dari pengacakan dataset sehingga sistem tersebut teruji untuk beberapa variabel *input* yang acak. Setelah menentukan nilai $k = 10$ pada *k-fold cross validation*, menentukan banyaknya data dalam satu *subset* data menggunakan Persamaan (4) pada data *training*.

$$B = \frac{tr}{k} \quad (4)$$

7. Melakukan perhitungan jarak ketetangaan antara data *testing* dan data *training* pada setiap *subset* data. Perhitungan dilakukan dengan menghitung jarak antara data *testing* dan data *training* pada setiap variabel menggunakan Persamaan (5) dan (6) yang sesuai dengan skala pengukuran variabel. Selanjutnya, menghitung jarak Gower [18] dengan Persamaan (7) menggunakan hasil perhitungan jarak data *testing* terhadap data *training* pada setiap variabel. Berdasarkan metode *k-fold cross validation* dengan $k = 10$, maka 1 *subset* data berfungsi sebagai data *testing* dan 9 *subset* data lainnya berfungsi sebagai data *training*. Setiap *subset* data bergantian untuk dijadikan data *testing*, misalnya *subset* 1 menjadi data *testing*, maka *subset* 2 sampai dengan *subset* 10 sebagai data *training* dan seterusnya hingga *subset* 10 menjadi data *testing* dan *subset* 1 hingga *subset* 9 sebagai data *training*. Adapun perhitungan jarak Gower adalah sebagai berikut:

Jarak Gower pada skala data nominal dapat dihitung menggunakan Persamaan (5).

$$d_q(a, b) = \begin{cases} 0, & tr_{a,q} = te_{b,q} \\ 1, & tr_{a,q} \neq te_{b,q} \end{cases} \quad (5)$$

Jarak Gower pada skala data interval dan rasio dihitung menggunakan Persamaan (6).

$$d_q(a, b) = \frac{(tr_{a,q} - te_{b,q})}{\max(x_q) - \min(x_q)} \quad (6)$$

Sementara itu, jarak antara dua objek dihitung menggunakan Persamaan (7).

$$d(a, b) = \frac{1}{r} \sum_{q=1}^r d_q(a, b) \quad (7)$$

8. Mengklasifikasikan data berdasarkan nilai K dengan cara mengurutkan datanya menggunakan jarak ketetangaan dari terkecil hingga terbesar, kemudian menentukan hasil klasifikasi dengan *voting* kelas berdasarkan data *training* yang telah diurutkan.
9. Melakukan pemilihan nilai K optimum menggunakan kriteria nilai akurasi tertinggi. Perhitungan nilai akurasi ketepatan data *testing* untuk setiap *subset* data pada nilai K masing-masing proporsi dihitung terlebih dahulu menggunakan Persamaan (8), kemudian dilanjutkan dengan menghitung rata-rata akurasi masing-masing nilai K menggunakan Persamaan (9).

$$A_{(g,K)} = \frac{\text{Banyak data klasifikasi benar dalam satu subset data testing}}{\text{Banyak data di dalam satu subset data testing}} \quad (8)$$

$$\text{Akurasi}(K) = \frac{\text{Total akurasi pada ketetangaan ke } - K}{\text{Banyak subset}} \times 100\% \quad (9)$$

10. Melakukan perhitungan bobot untuk setiap kelas menggunakan nilai $E = 3, 5, 7, \text{ dan } 9$. Perhitungan nilai bobot setiap kelas menggunakan Persamaan (10).

$$\text{Weight}_{(g,h)} = \frac{1}{\left[\frac{\text{Num}(C_h^g)}{\text{Min}\{\text{Num}(C_h^g)\}} \right]^{1/E}} \quad (10)$$

11. Melakukan perhitungan skor pada setiap kelas menggunakan Persamaan (10), kemudian melihat skor terbesar untuk menentukan hasil klasifikasi. Setelah diperoleh hasil klasifikasi berdasarkan nilai skor data *testing* pada setiap kelas, kemudian dilakukan perhitungan akurasi hasil klasifikasi dari masing-masing nilai E pada data testing menggunakan Persamaan (11) dan menghitung rata-rata keseluruhan nilai akurasi berdasarkan nilai E menggunakan Persamaan (12).

$$Skor(b, C_h) = Weight_{(g,h)} \left(\sum_{T_a \in KNN(b)}^{a=1} (d(a, b) \delta(T_a, C_h)) \right) \quad (11)$$

$$Akurasi(E) = \frac{Total\ akurasi\ pada\ nilai\ E}{Banyaknya\ subset} \times 100\% \quad (12)$$

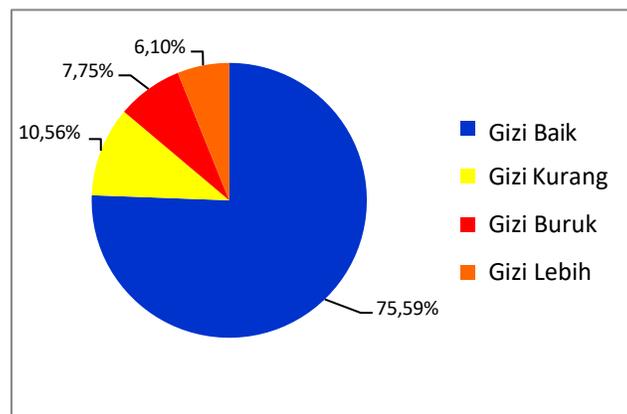
12. Melakukan perhitungan nilai akurasi klasifikasi NWKNN menggunakan Persamaan (13).

$$Akurasi = \frac{Banyak\ data\ uji\ benar\ klasifikasi}{Banyak\ data} \times 100\% \quad (13)$$

13. Melakukan interpretasi terhadap hasil klasifikasi terbaik.

3. HASIL DAN PEMBAHASAN

Pada bagian hasil dan pembahasan diawali dengan deskripsi data penelitian untuk variabel *input* dan *output*. Adapun deskripsi dari variabel *output*, yakni status gizi balita ditampilkan pada Gambar 1.



Gambar 1. Distribusi Status Gizi Balita di Wilayah Kerja Puskesmas Wonorejo Kota Samarinda Tahun 2022

Berdasarkan Gambar 1, tampak bahwa dari 426 balita di wilayah kerja Puskesmas Wonorejo Kota Samarinda Tahun 2022, balita dengan status gizi buruk sebesar 7,75%; gizi kurang sebesar 10,56%; gizi baik sebesar 75,59%; dan gizi lebih sebesar 6,10%. Hasil ini menunjukkan bahwa data status gizi balita memiliki frekuensi data pada masing-masing kategorinya tidak seimbang. Oleh karena itu, data status gizi balita di wilayah kerja Puskesmas Wonorejo Kota Samarinda Tahun 2022 layak digunakan untuk metode NWKNN.

Setelah dilakukan analisis statistik deskriptif terhadap data penelitian, dilanjutkan dengan melakukan standarisasi data dengan tujuan untuk mendapatkan data yang berada pada rentang yang sama dengan cara melakukan transformasi terhadap data asli menggunakan *z-score*. Standarisasi dilakukan pada data usia (X_2), berat badan (X_3), dan tinggi badan balita (X_4). Hasil yang diperoleh disajikan pada Tabel 2.

Tabel 2. Hasil Standardisasi Data

X_2	X_3	X_4
0,240	0,101	0,267
-0,148	0,549	0,425
1,145	0,549	0,883
⋮	⋮	⋮
1,534	0,101	0,820

Setelah diperoleh data yang telah distandarisasi, menentukan banyaknya data dalam *subset* data *training* menggunakan metode *Fold Cross Validation* dengan $k = 10$. Hasil yang diperoleh diberikan pada Tabel 3.

Tabel 3. Banyaknya Data dalam Subset Data Training

Subset	Nomor Sampel	Y	Jumlah Sampel
1	207	3	30
	158	3	
	404	2	
	⋮	⋮	
	108	3	
⋮	⋮	⋮	⋮
	146	3	
	60	3	
	154	3	
	407	2	
10	154	3	28

Berdasarkan data pada Tabel 3, dilakukan penghitungan jarak ketetanggaan antara data *testing* terhadap data *training* menggunakan jarak Gower. Perhitungan jarak Gower diawali dengan menghitung jarak data *testing* terhadap data *training* pada setiap variabel menggunakan Persamaan (5) dan (6), kemudian dilanjutkan dengan menghitung jarak Gower menggunakan Persamaan (7). Setelah didapatkan jarak Gower untuk setiap data *testing* dilanjutkan dengan mengurutkan jarak Gower untuk pengklasifikasian berdasarkan ketetanggaan terdekat. Pada Tabel 4 diberikan contoh hasil pengurutan jarak Gower dan pengklasifikasian data.

Tabel 4. Hasil Perhitungan Jarak Gower 268 Data Training terhadap Data Testing ke-1 dari Subset 1

Rank	Data Training		Data Testing ke-1 dari Subset 1 (Balita ke- 207)	Batas KNN	Klasifikasi Data Testing ke-1
	Balita	Klasifikasi	$d(a, b)$		
1	24	Gizi Baik	0,008974	3	Gizi Baik
2	29	Gizi Baik	0,010623		Gizi Baik
3	56	Gizi Baik	0,010714		Gizi Baik
4	281	Gizi Baik	0,013141		Gizi Baik
5	31	Gizi Baik	0,013828	5	Gizi Baik
6	74	Gizi Baik	0,020879		Gizi Baik
7	260	Gizi Baik	0,025046	7	Gizi Baik
8	133	Gizi Baik	0,025183		Gizi Baik
9	61	Gizi Baik	0,025687	9	Gizi Baik
⋮	⋮	⋮	⋮		⋮
268	412	Gizi Baik	0,617079		

Selanjutnya, dilakukan pemilihan nilai K optimum dan proporsi terbaik dapat dilakukan dengan melihat nilai akurasi yang dihasilkan, yakni diawali dengan menghitung akurasi masing-masing nilai K pada setiap *subset* kemudian menghitung rata-rata keseluruhan akurasi untuk setiap nilai K . Hasil yang diperoleh seperti pada Tabel 5.

Tabel 5. Hasil Pemilihan Nilai K Optimum

Proporsi Data <i>Training:Testing</i>	Nilai Akurasi dengan <i>Fold Cross Validation</i>			
	$K = 3$	$K = 5$	$K = 7$	$K = 9$
70:30	77,3	77,3	76,8	75,7
90:10	76,3	75,4	75,9	76,2

Pada Tabel 5 tampak bahwa pada proporsi data *training* dan *testing* sebesar 70:30, dengan batas 3 tetangga terdekat (3-NN) dan batas 5 tetangga terdekat (5-NN), diperoleh nilai akurasi tertinggi sebesar 77,3%. Hasil ini menunjukkan bahwa nilai K optimum yang digunakan pada pengklasifikasian status gizi balita dengan metode NWKNN adalah $K = 3$ dan $K = 5$.

Setelah didapat nilai K optimum, dihitung nilai bobot dan skor kelas pada *subset* data menggunakan proporsi data *training* dan *testing* sebesar 70:30, nilai $K = 3$ dan $K = 5$, dan nilai $E = 3, 5, 7, 9$. Pada Tabel 6 diberikan contoh hasil perhitungan nilai bobot dan skor kelas pada *subset* 1 dengan $E = 3$ pada $K = 3$.

Tabel 6. Nilai Bobot dan Skor Kelas pada *Subset* 1 dengan $E = 3$ pada $K = 3$

No. Sampel	Skor (b, C_h)				Klasifikasi Data Asli (Y)	Hasil Klasifikasi Data <i>Training</i>
	C_1	C_2	C_3	C_4		
207	0	0	0,013	0	Gizi Baik	Gizi Baik
158	0,021	0	0,008	0,013	Gizi Baik	Gizi Buruk*
404	0,022	0	0,016	0	Gizi Kurang	Gizi Buruk*
⋮	⋮	⋮	⋮	⋮	⋮	⋮
108	0,478	0	0,249	0	Gizi Baik	Gizi Buruk

*) Hasil klasifikasi tidak tepat.

Berdasarkan nilai bobot dan skor kelas pada semua *subset*, dapat diperoleh nilai akurasi dari masing-masing nilai E pada data *testing* seperti yang diberikan pada Tabel 7.

Tabel 7. Nilai Akurasi Prediksi Ketepatan Klasifikasi Berdasarkan Nilai E

<i>Subset</i>	$E = 3$	$E = 5$	$E = 7$	$E = 9$
1	63,33	73,33	66,67	73,33
2	76,67	73,33	76,67	76,67
3	73,33	76,67	83,33	83,33
⋮	⋮	⋮	⋮	⋮
10	64,29	67,86	78,57	75,00
Nilai Akurasi (%)	71,76	74,12	75,52	75,17

Pada Tabel 7, terlihat bahwa nilai $E = 7$ memiliki nilai akurasi terbesar. Hasil ini menunjukkan bahwa nilai bobot $E = 7$ yang digunakan pada metode NWKNN untuk memprediksi ketepatan klasifikasi status gizi balita di wilayah kerja Puskesmas Wonorejo Kota Samarinda. Setelah didapatkan proporsi data *training* dan data *testing*, masing-masing sebesar 70% dan 30%, yakni 298 data *training* dan 128 data *testing* dengan nilai $K = 3, K = 5$, dan $E = 7$ dapat diperoleh nilai akurasi metode NWKNN pada prediksi ketepatan klasifikasi status gizi balita di wilayah kerja Puskesmas Wonorejo Kota Samarinda. Adapun nilai akurasinya disajikan pada Tabel 8.

Tabel 8. Nilai Akurasi Prediksi Ketepatan Klasifikasi Metode NWKNN

Data <i>Testing</i>	$K = 3$	$K = 5$	Klasifikasi Data Asli (Y)
32	Gizi Baik	Gizi Baik	Gizi Baik
381	Gizi Baik*	Gizi Baik*	Gizi Lebih
201	Gizi Baik	Gizi Baik	Gizi Baik
292	Gizi Baik	Gizi Baik	Gizi Baik
352	Gizi Baik*	Gizi Baik*	Gizi Buruk
⋮	⋮	⋮	⋮
172	Gizi Baik	Gizi Baik	Gizi Baik
Prediksi Benar	84	93	
Nilai Akurasi (%)	65,63	72,65	

*) Hasil klasifikasi tidak tepat

Berdasarkan Tabel 8, dapat diketahui bahwa hasil perhitungan akurasi dari metode NWKNN dengan menggunakan nilai E optimum sebesar $E = 7$, dengan nilai K optimum sebesar $K = 3$ dan $K = 5$ diperoleh nilai akurasi prediksi ketepatan klasifikasi status gizi balita sebesar 65,63% pada nilai $K = 3$ dan nilai akurasi sebesar 72,65% pada nilai $K = 5$. Hasil ini menunjukkan bahwa metode NWKNN terbaik adalah metode NWKNN dengan proporsi data *training* dan *testing*, masing-masing sebesar 70% dan 30%; nilai ketetapan $K = 5$ dan nilai bobot $E = 7$. Nilai akurasi prediksi ketepatan klasifikasi sebesar 72,65% termasuk baik dan menunjukkan bahwa metode NWKNN layak digunakan untuk memprediksi ketepatan klasifikasi status gizi balita di wilayah kerja Puskesmas Wonorejo Kota Samarinda.

4. KESIMPULAN

Metode NWKNN dapat memprediksikan ketepatan klasifikasi data *mining* yang memiliki frekuensi data yang tidak seimbang pada kelasnya. Data status gizi balita di wilayah kerja Puskesmas Wonorejo Kota Samarinda Tahun 2022 yang digunakan sebagai variabel output dalam penelitian ini memiliki kelas dengan frekuensi data yang tidak seimbang. Metode NWKNN layak digunakan untuk memprediksi ketepatan klasifikasi status gizi balita di wilayah kerja Puskesmas Wonorejo Kota Samarinda karena memiliki nilai akurasi yang baik.

DAFTAR PUSTAKA

- [1] M. Kantardzic, *Data Mining: Concepts, Models, Methods, and Algorithms*, Hoboken, New Jersey: John Wiley & Sons, 2020.
- [2] P. Dangeti, *Statistics for Machine Learning*, Birmingham, UK: Packt Publishing, 2017.
- [3] E. Russano and E. F. Avelino, *Fundamentals of Machine Learning Using Python*, Oakville, Canada: Arcler Press, 2020.
- [4] Z. Haizadeh, M. Taheri and M. Z. Jahromi, "Nearest neighbor classification with locally weighted distance for imbalanced data," *International Journal of Computer and Communication Engineering*, vol. 3, no. 2, pp. 81-86, 2014.
- [5] S. Tan, "Neighbor-weighted k-nearest neighbor for unbalanced text corpus," *Expert Systems with Applications*, vol. 28, pp. 667-671, 2005.
- [6] Indriati and A. Ridok, "Sentiment analysis for review mobile applications using neighbor weighted k-nearest neighbor (NWKNN) method," *Journal of Environmental Engineering & Sustainable Technology*, vol. 3, no. 1, pp. 23-32, 2016.
- [7] B. L. Yudha, H. Muflikhah and R. C. Wihandika, "Klasifikasi risiko hipertensi menggunakan metode neighbor weighted k-nearest neighbor (NWKNN)," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 2, no. 2, pp. 897-904, 2018.
- [8] A. H. Hadi, D. E. Ratnawati and C. Dewi, "Identifikasi Penyakit Gagal Ginjal menggunakan metode neighbor weighted k-nearest neighbor (NWKNN)," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 2, no. 9, pp. 2562-2569, 2018.
- [9] Grasella, I. Purnamasari and F. D. T. Amijaya, "Klasifikasi status pembayaran premi menggunakan algoritma neighbor weighted k-nearest neighbor (NWKNN) (Studi kasus: PT. Bumiputera Kota Samarinda)," *Variance: Journal of Statistics and Its Applications*, vol. 1, no. 2, pp. 56-63, 2019.

- [10] P. N. Fadilla, Indriati and E. D. Ratnawati, "Identifikasi jenis attention deficit hyperactivity disorder (ADHD) pada anak usia dini menggunakan metode neighbor weighted k-nearest neighbor (NWKNN)," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 3, no. 3, pp. 194-200, 2016.
- [11] Indriati and A. Kusyanti, "Metode ensemble classifier untuk mendeteksi jenis attention deficit hyperactivity disorder pada anak usia dini," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 6, no. 3, pp. 301-308, 2019.
- [12] R. A. Azizah, F. A. Bachtiar and S. Adinugroho, "Klasifikasi kinerja akademik siswa menggunakan neighbor weighted k-nearest neighbor dengan seleksi fitur information gain," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 9, no. 3, pp. 605-614, 2022.
- [13] D. T. Wilujeng, M. Fatekurohman and I. M. Tirta, "Analisis risiko kredit perbankan menggunakan algoritma k-nearest neighbor dan nearest weighted k-nearest neighbor," *Indonesian Journal of Applied Statistics*, vol. 5, no. 2, pp. 142-148, 2023.
- [14] S. S. Milania, C. Suhery and T. Rismawan, "Klasifikasi demam menggunakan metode neighbor weighted k-nearest neighbor," *Journal of Computing Engineering, System and Science*, vol. 8, no. 2, pp. 250-261, 2023.
- [15] A. A. A. Hidayat, G. Marini and A. P. M. Tyas, "Factors affecting nutritional status in children aged 6–24 months," *Open Access Macedonian Journal of Medical Sciences*, vol. 8, pp. 291-295, 2020.
- [16] E. Prasetyo, *Data Mining: Mengolah Data menjadi Informasi Menggunakan MATLAB*, Yogyakarta: Andi , 2014.
- [17] T. T. Wong and N. Y. Yang, "Dependency analysis of accuracy estimates in k-Fold Cross Validation," *IEEE Transactions on Knowledge and Data Engineering*, vol. 29, no. 11, pp. 2417-2427, 2017.
- [18] A. Aditya, B. N. Sari and T. N. Padilah, "Perbandingan pengukuran jarak Euclidean dan Gower pada kluster k-medoids," *Jurnal Teknologi dan Sistem Komputer*, vol. 9, no. 1, pp. 1-7, 2021.

