

Klasifikasi Masyarakat Penerima BPNT Program Sembako 2021 di Kelurahan Tiakar dengan Menggunakan Metode KNN Classifier

Annisa Hidayatul Nufus¹, Helma²

^{1,2},Departemen Matematika,Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Negeri Padang (UNP)

Article Info

Article history:

Received January 24, 2023

Revised March 15, 2023

Accepted June 30, 2023

Keywords:

K-Nearest Neighbor
classification
confusion matrix

Kata Kunci:

K-Nearest Neighbor
Klasifikasi
confusion matrix

ABSTRACT

At the time of the implementation of the food program in the Tiakar Sub-District, there was no information available on the criteria that would make the head of a family eligible to be a recipient of assistance, even though this was very important so that the government's goal of fulfilling the need for nutritious food could be felt by those who really needed it. The purpose of this study was to classify the heads of families in Tiakar Subdistrict as eligible or not eligible to receive staple foods using the K-Nearest Neighbor method. This study uses interview data conducted with the heads of families in the Tiakar Village. The data analysis step is to divide the data into training data and test data by 80%:20%, determine the number of nearest neighbours, calculate the dissimilarity distance and choose a class, calculate the level of accuracy using the confusion matrix and then choose the optimal K. Based on the results of the study, it was found that the K value that was good to use in the classification of family heads in Tiakar Subdistrict was $K = 3$ because it had an accuracy percentage of 95%.

ABSTRAK

Pada saat pelaksanaan program pangan di Kelurahan Tiakar belum tersedia informasi kriteria yang menjadikan kepala keluarga layak menjadi penerima bantuan padahal hal tersebut sangat penting agar tujuan pemerintah untuk memenuhi kebutuhan pangan bergizi dapat dirasakan oleh pihak yang memang membutuhkan. Tujuan penelitian ini untuk mengklasifikasikan kepala keluarga di Kelurahan Tiakar menjadi layak atau tidak layak menerima sembako menggunakan metode *K-Nearest Neighbor*. Penelitian ini menggunakan data interview yang dilakukan kepada kepala keluarga di Kelurahan Tiakar. Langkah analisis data yang dilakukan yaitu membagi data menjadi data latih dan data uji sebesar 80%:20%, menentukan jumlah tetangga terdekat, menghitung jarak ketakmiripan dan memilih kelas, menghitung tingkat akurasi menggunakan *confusion matrix* kemudian memilih K yang optimal. Berdasarkan hasil penelitian diperoleh nilai K yang baik digunakan dalam klasifikasi kepala keluarga di Kelurahan Tiakar adalah $K=3$ karena memiliki persentase akurasi 95%.

This is an open access article under the [CC BY-SA](#) license.



Penulis pertama/sesuai:

(Annisa Hidayatul Nufus)

Departemen Matematika, Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam,

Universitas Negeri Padang, Jl.Prof.Dr. Hamka, Air Tawar barat, Padang Utara, Padang, 25171

Padang, Sumatera Barat

Email: annisahidayatul55@gmail.com

1. PENDAHULUAN

Salah satu program yang dibentuk oleh pemerintah untuk mengurangi beban masyarakat dalam pemenuhan kebutuhan pokok adalah program Bantuan Pangan Non Tunai (BPNT) yang telah dikembangkan menjadi Program Sembako. Sasaran program sembako adalah keluarga dengan kondisi sosial ekonomi terendah yang terdaftar dalam Keluarga Penerima Manfaat (KPM) yang telah ditetapkan oleh Kementerian Sosial dan bersumber dari Data Terpadu Kesejahteraan Sosial (DTKS) yang telah diverifikasi oleh pemerintah daerah. Dengan Program Sembako, indeks bantuan yang semula Rp.110.000/KPM/bulan naik menjadi Rp.150.000/KPM/bulan. Selain itu Program Sembako memperluas komoditas yang dapat dibeli sehingga tidak hanya berupa beras dan telur seperti pada program BPNT. Melalui mekanisme akun elektronik berupa Kartu Keluarga Sejahtera (KKS) yang digunakan hanya untuk kebutuhan bahan pangan di tempat yang telah bekerjasama dengan bank Himpunan Bank Negara (Himbara) yang kemudian disebut E-warong. Program ini diselenggarakan sesuai dengan PERPRES RI No.63 Tahun 2017 tentang penyaluran Bantuan Sosial secara Non Tunai [1].

Pada tahun 2021 pelaksanaan penyaluran bantuan Program Sembako di Kelurahan Tiakar dengan jumlah penerima manfaat bantuan Program Sembako sebanyak 177 kepala keluarga. Berdasarkan informasi yang dimiliki pihak Kelurahan Tiakar, cukup banyak jumlah masyarakat yang berada di bawah Garis Kemiskinan Nasional (GKN) sehingga dikatakan layak menjadi penerima bantuan pangan. Dalam pelaksanaan program bantuan pangan belum tersedia informasi pasti kriteria yang menjadikan kepala keluarga layak menjadi penerima bantuan tersebut. Pada tahun 2021 tercatat bahwa dari 238 kepala keluarga yang diajukan pihak kelurahan kepada kepala Dinas Sosial sebagai masyarakat yang berada dibawah garis kemiskinan sehingga mereka termasuk kedalam penerima program bantuan pangan, namun dilapangan hanya 177 kepala keluarga yang dinyatakan layak menerima program bantuan tersebut. Rendahnya jumlah kepala keluarga yang dinyatakan layak menerima bantuan bila dibandingkan dengan semua kepala keluarga yang diajukan pihak kelurahan sehingga perlu dilakukan klasifikasi kepala keluarga mana yang berhak mendapatkan bantuan dan kepala keluarga mana yang tidak berhak mendapatkan bantuan.

Data mining adalah suatu proses untuk mengidentifikasi informasi penting didalam berbagai database besar. Secara umum, dalam pengerjaannya data mining terdapat dua pendekatan yaitu *unsupervised learning* dan *supervised learning* [2]. Pada pendekatan *supervised learning* terdapat dua kelompok data yaitu data latih (*training*) dan data uji (*testing*). Memisahkan data menjadi data latih dan data uji dimaksudkan agar model yang diperoleh nantinya memiliki kemampuan yang lebih dalam melakukan pendugaan klasifikasi data. Contoh tekniknya adalah klasifikasi (*classifier*) [2]. Klasifikasi adalah proses untuk menemukan model dengan tujuan memprediksi kategori untuk data yang tidak diketahui kategorinya [3]. Teknik klasifikasi bekerja dengan mengelompokkan data berdasarkan data latih dan nilai klasifikasi [4]. Beberapa klasifikasi tipe *supervised learning* yaitu *K-Nearest Neighbor*, *Naive Bayes*, *Support Vector Machine*, *Decision Trees*, *Linear Regression*, dan *Neural Network* [5].

Metode *K-Nearest Neighbor classifier* adalah proses klasifikasi yang berbasis jarak dalam menentukan kedekatan antar data yang akan menjadi tetangga (*neighbor*) terdekatnya dan memilih kelas atau kategori berdasarkan kategori *K* tetangga. Parameter *K* adalah banyaknya tetangga terdekat yang berfungsi penting dalam penentuan hasil prediksi. Metode KNN *classifier* memiliki kelebihan, antara lain ketahanan terhadap data latih yang banyak *noise* (gangguan) dan efektif terhadap data latih yang berukuran besar [6].

Algoritma KNN menggunakan informasi mengenai variabel dari kumpulan data latih agar dalam proses klasifikasi terhadap data uji dapat dihitung kedekatan data baru tersebut dengan semua data latih.



Berdasarkan hal tersebut, diperoleh tetangga terdekat dari data baru yang selanjutnya digunakan untuk memprediksi kategori dari data baru. Variabel yang digunakan sangat mempengaruhi hasil klasifikasi sehingga perlu dilakukan analisa mendalam terhadap variabel yang dipilih untuk mengetahui variabel yang penting dan variabel yang tidak relevan [7]. Klasifikasi KNN memiliki dua tahap. Pertama adalah penentuan tetangga terdekat dan yang kedua adalah penentuan kelas menggunakan tetangga tersebut[8].

Salah satu masalah dalam pengklasifikasian adalah penentuan nilai K . Nilai K memiliki dampak yang besar dalam penentuan tingkat akurasi. Jumlah data yang banyak tidak selalu harus menggunakan nilai K yang besar begitu juga dengan data dengan jumlah sedikit tidak harus menggunakan nilai K yang kecil. Dalam beberapa studi menunjukkan bahwa metode KNN optimal ketika menggunakan nilai K yang relatif besar. Pada studi lainnya jika nilai K terlalu besar maka akan banyak tetangga terdekat yang tidak relevan. Artinya, nilai K yang besar tidak memungkinkan hasil yang baik dibandingkan dengan nilai K yang kecil. Pada kasus data yang akan dikelompokkan menjadi dua kelas, tidak hanya variabel numerik saja yang baik menggunakan nilai K yang kecil tetapi apabila juga menggunakan variabel kategorik [9]. Nilai K yang dipilih akan menentukan tingkat akurasi dalam proses klasifikasi. Nilai K yang optimal adalah nilai K yang memberikan akurasi yang tinggi.

Pada hal klasifikasi, akurasi yang baik adalah hal yang sangat diperhatikan [10]. Akurasi adalah persentase ketepatan suatu data yang telah diklasifikasikan dengan benar yang digambarkan melalui *confusion matrix*[11]. *Confusion matrix* digunakan untuk menghitung tingkat akurasi pada data mining dengan ukuran $n \times n$, dimana n merupakan banyaknya kategori (kelas) yang berbeda *confusion matrix* mengelompokkan data menjadi dua klasifikasi yaitu klasifikasi hasil prediksi dan klasifikasi yang sebenarnya. Dari *confusion matrix* terdapat tiga parameter yang akan dihitung yaitu akurasi, presisi, dan *recall* [12].

2. METODE

Jenis penelitian ini merupakan penelitian terapan. Data yang digunakan merupakan data primer yang diperoleh dari observasi dan wawancara pada *survey* yang dilakukan kepada kepala keluarga di Kelurahan Tiakar. Populasi dalam penelitian ini adalah semua kepala keluarga yang berada di Kelurahan Tiakar. Dalam Penelitian ini teknik pengambilan data kepala keluarga di Kelurahan Tiakar menggunakan metode *purposive sampling* melalui *survey* secara wawancara dan observasi. Artinya, sampel akan diambil di Kelurahan Tiakar berdasarkan kriteria dan kebutuhan pada penelitian ini. Untuk mendapatkan banyak sampel yang akan mewakili semua kepala keluarga digunakan Rumus Slovin.

$$n = \frac{N}{1 + Ne^2}$$

Dimana:

n : jumlah sampel

N : jumlah populasi

e : Presentase kelonggaran ketelitian kesalahan pengambilan sampel yang masih bisa ditolerir
($e = 0.1$)

Pada penelitian ini, ada beberapa langkah dalam mengklasifikasikan data, yaitu:

- a. Mengumpulkan data
- b. Membagi data menjadi data latih dan data uji

Pada penelitian ini data yang diperoleh dibagi menjadi dua kategori, yaitu 80% data latih dan 20% data uji dari seluruh data yang dimiliki. Perbandingan tersebut merupakan salah satu perbandingan data yang menghasilkan nilai akurasi yang tinggi dan sering digunakan untuk pengklasifikasian data

menjadi dua kategori [13]. Para peneliti biasanya menggunakan proporsi data latih dan data uji sebesar 80% : 20% [14].

c. Menentukan jumlah tetangga terdekat K

Pada penelitian ini nilai K yang digunakan merupakan nilai K yang relatif kecil untuk mengurangi gangguan (*noise*) pada proses pengklasifikasian. Pada penelitian ini digunakan $K = 3, K = 5, K = 7, K = 9$.

d. Untuk setiap data uji

1) Menghitung jarak ketakmiripan

Pada penelitian ini, untuk mengetahui kategori satu data uji maka dilakukan penghitungan satu data baru tersebut terhadap semua data latih yang dimiliki menggunakan koefisien jarak ketakmiripan Gower. Dalam menghitung koefisien Gower, terlebih dahulu dilakukan perhitungan jarak ketakmiripan antara pasangan pengamatan berdasarkan variabel secara terpisah. Berikut beberapa rumus yang digunakan berdasarkan jenis variabel:

a) Variabel ke-C adalah nominal (kategorik)

Jarak ketakmiripan antara dua objek pada variabel ke-C dapat dihitung melalui cara berikut :

$$d_{ijc} = \begin{cases} 0, & \text{jika } x_i = x_j . \\ 1, & \text{jika } x_i \neq x_j . \end{cases} \quad (1)$$

b) Variabel ke-C adalah rasio (numerik)

Penghitungan jarak ketakmiripan antara dua objek data menggunakan variabel rasio dinyatakan dengan rumus :

$$d_{ijc} = \frac{|x_{ic} - x_{jc}|}{\max(x_c) - \min(x_c)} \quad (2)$$

dimana $\max(x_c)$ adalah nilai tertinggi dari seluruh data variabel ke C dan $\min(x_c)$ adalah nilai terendah dari seluruh data variabel ke-C.

c) Variabel ke-C adalah ordinal

Untuk mendapatkan ukuran jarak untuk variabel ordinal, data terlebih dahulu ditransformasi menggunakan rumus sebagai berikut:

$$x_{ic} = \frac{r_{ic} - 1}{R_c - 1} \quad (3)$$

dimana r_{ic} adalah nomor urutan kategori ordinal ke- i ($r = 1, \dots, R_c$), dan R_c adalah jumlah urutan maksimal dari variabel ke-C. Setelah dilakukan transformasi dengan cara tersebut, hasil yang diperoleh dapat digunakan dalam persamaan 2 untuk variabel numerik.

Selanjutnya koefisien jarak ketakmiripan Gower dapat dihitung sebagai berikut:

$$d_G(\mathbf{X}_i, \mathbf{X}_j) = \frac{\sum(w_{ijc} \times d_{ijc})}{\sum w_{ijc}} \quad (4)$$



dimana d_{ijc} adalah ukuran ketakmiripan antara objek ke- i dan ke- j dengan variabel ke- C ($C = 1, \dots, m$), dan w_{ijc} bernilai nol jika nilai variabel C objek ke- i atau ke- j hilang (*missing*), bernilai satu jika tidak [14].

2) Memilih kelas

Setelah jarak yang diperoleh diurutkan dari terkecil ke terbesar, kelas terbanyak dari K tetangga terdekat dari data yang akan dievaluasi (data uji) merupakan kelasnya.

e. Menghitung tingkat akurasi pengklasifikasian

Tingkat keakuratan dalam melakukan klasifikasi dapat diketahui dengan melihat keakuratan prediksi kategori suatu data uji KNN *classifier*. Menghitung tingkat akurasi dengan menggunakan *confusion matrix* yang terdiri dari nilai akurasi, presisi dan *recall*.

1) Akurasi

Akurasi dari suatu pengukuran adalah tingkat kedekatan pengukuran terhadap nilai yang sebenarnya [15]. Nilai akurasi dapat diketahui menggunakan persamaan berikut:

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

2) Presisi

Presisi dari suatu sistem pengukuran, disebut juga reproduktifitas atau pengulangan adalah sejauh mana pengulangan pengukuran dalam kondisi yang tidak berubah mendapatkan hasil yang sama [15]. Nilai presisi dapat diketahui menggunakan persamaan berikut :

$$Presisi = \frac{TP}{TP + FP} \times 100\% \quad (6)$$

3) Recall

Recall merupakan proporsi kasus positif yang diprediksi secara benar [15]. Nilai recall dapat diketahui menggunakan persamaan berikut:

$$Recall = \frac{TP}{TP + FN} \times 100\% \quad (7)$$

f. Memilih K yang optimal

Nilai tetangga terdekat (nilai K) memiliki peran penting dalam metode KNN *classifier*. Nilai K yang dipilih akan menentukan tingkat akurasi dalam proses klasifikasi. Nilai K yang optimal adalah nilai K yang memberikan akurasi yang tinggi.

3. HASIL DAN PEMBAHASAN

Data yang telah diperoleh dari hasil *survey* akan dianalisis menggunakan metode KNN *classifier*. terdapat 100 data sampel dengan 3 kategori. Berikut akan dijelaskan hasil dari analisis klasifikasi kepala keluarga di Kelurahan Tiakar menjadi layak atau tidak layak.

3.1. Menghitung Koefisien Gower

Pada penelitian ini koefisien Gower digunakan untuk mencari ukuran ketakmiripan agar diketahui K tetangga terdekat suatu data. Semakin kecil ukuran ketakmiripan maka semakin besar ukuran kemiripan dan tingkat kedekatan antar dua data. Ukuran ketakmiripan diperoleh dengan menghitung jarak masing-masing variabel antar dua data terlebih dahulu.

Berikut akan diilustrasikan perhitungan jarak Gower antara objek pengamatan R57 yang ada pada data uji dengan tujuh data latih pertama. Pada Tabel 1 disajikan data objek R57 dan tujuh data latih pertama dengan dimisalkan simbol "L" untuk kategori Layak dan simbol "TL" untuk kategori Tidak Layak seperti berikut

Tabel 1. Data Latih dan Data uji

No	Nama	Tanggungan	Penghasilan	Pekerjaan	Kategori
1	R1	5	500.000-1.000.000	Buruh	L
2	R2	4	500.000-1.000.000	Buruh	L
3	R3	4	500.000-1.000.000	Buruh	L
4	R4	7	1.000.001-3.000.000	Wiraswasta	L
5	R5	3	3.000.001-5.000.000	PNS	TL
...	⋮
...	500.000-1.000.000
76	R95	6	3.000.001-5.000.000	Buruh	L
77	R96	5	>5.000.000	Buruh	TL
78	R97	6	>5.000.000	PNS	TL
79	R98	4	500.000-1.000.000	Wiraswasta	TL
80	R100	3	500.000-1.000.000	Buruh	L
81	R57	4	500.000-1.000.000	Buruh	-(data uji)

Pada Tabel 1 dapat dilihat terdapat objek pengamatan yaitu R57 yang ingin diprediksi layak atau tidak layaknya dalam menerima bantuan pangan. Status objek data R57 dapat diketahui dengan melihat koefisien jarak ketakmiripan Gower dengan masing - masing data latih melalui langkah - langkah sebagai berikut :

a. Menghitung jarak ketakmiripan masing-masing variabel R57 dengan semua data latih untuk semua variabel.

1) Jumlah Tanggungan Kepala Keluarga (X_1)

Variabel jumlah tanggungan merupakan variabel rasio (numerik). Berikut ini merupakan contoh perhitungan antara data baru (R57) ke data pertama (R1).

$$d_{R57,R1} = \frac{|x_{ic} - x_{jc}|}{\max(x_c) - \min(x_c)}$$

$$= \frac{|4 - 5|}{7 - 1}$$

$$= 0,1667$$

2) Penghasilan Kepala Keluarga (X_2)

Variabel penghasilan merupakan variabel ordinal sehingga harus ditransformasikan terlebih dahulu menggunakan rumus sebagai berikut:

$$x_{ic} = \frac{r_{ic} - 1}{R_c - 1}$$



Setelah dilakukan transformasi dengan cara tersebut, hasil yang diperoleh dapat digunakan dalam persamaan 1 untuk variabel numerik.

Tabel 2. Transformasi Data Penghasilan

	Penghasilan	Hasil Transformasi (X_2)
1	500.000-1.000.000	$x_{12} = \frac{1-1}{4-1} = 0$
2	1.000.001-3.000.000	$x_{12} = \frac{2-1}{4-1} = \frac{1}{3}$
3	3.000.001-5.000.000	$x_{12} = \frac{3-1}{4-1} = \frac{2}{3}$
4	>5.000.000	$x_{12} = \frac{4-1}{4-1} = 1$

Selanjutnya, akan dihitung jarak dengan data hasil transformasi. Diketahui dari data hasil transformasi bahwa nilai $\max(X_2) = 1$ dan $\min(X_2) = 0$, sehingga perhitungan jarak dilakukan seperti berikut:

$$d_{R57,R1} = \left| \frac{0-0}{1-0} \right|$$

$$= 0$$

3) Pekerjaan Kepala Keluarga (X_3)

Pada penelitian ini kriteria pekerjaan dikategorikan ke dalam variabel nominal sehingga jaraknya dapat diukur sebagai berikut:

$$d_{ijc} = \begin{cases} 0, & \text{jika } x_i = x_j . \\ 1, & \text{jika } x_i \neq x_j . \end{cases}$$

Untuk objek R57 memiliki pekerjaan sebagai buruh dan untuk objek R1 juga memiliki pekerjaan sebagai buruh, karena pekerjaan R57 = pekerjaan R1 maka

$$d_{R57,R1} = 0$$

b. Menghitung ukuran ketakmiripan Gower

Setelah diperoleh jarak masing-masing variabel antara objek pengamatan R57 dan tujuh data latih pertama yang dapat disajikan dalam tabel berikut:

Tabel 3. Jarak Masing-Masing Variabel R57 dengan Data Latih

No	Nama	Jarak		
		Tanggungan (X_1)	Penghasilan (X_2)	Pekerjaan (X_3)
1	R1	0,1667	0	0
2	R2	0	0	0
3	R3	0	0	0
4	R4	0,5	0,3333	1
5	R5	0,1667	0,6667	1
⋮	⋮	⋮	⋮	⋮

No	Nama	Jarak		
		Tanggungans (X_1)	Penghasilan (X_2)	Pekerjaan (X_3)
76	R95	0,3333	0	0,1111
77	R96	0,1667	0,6667	0,2778
78	R97	0,3333	1	0,7778
79	R98	0	1	0,6667
80	R100	0,1667	0	0

Selanjutnya, akan dicari koefisien jarak ketakmiripan Gower antara R57 dengan tujuh data latih pertama. Rumus perhitungan jarak Gower sebagai berikut:

$$d_G(\mathbf{X}_i, \mathbf{X}_j) = \frac{\sum(w_{ijc} \times d_{ijc})}{\sum w_{ijc}}$$

dimana d_{ijc} adalah ukuran ketakmiripan antara objek ke- i dan ke- j dengan variabel ke- C ($C = 1, \dots, m$), dan w_{ijc} bernilai nol jika nilai variabel C objek ke- i atau ke- j hilang (*missing*), bernilai satu jika tidak. Berikut ini merupakan contoh perhitungan manual untuk mencari nilai jarak gower antara data baru (R57) ke data pertama (R1).

$$\begin{aligned} d_G(R57, R1) &= \frac{\sum(w_{ijc} \times d_{ijc})}{\sum w_{ijc}} \\ &= \frac{(1 \times x_1) + (1 \times x_2) + (1 \times x_3)}{1 + 1 + 1} \\ &= 0,05 \end{aligned}$$

Menghitung jarak antara data uji (R57) dengan latih lainnya dihitung menggunakan cara yang sama. Hasil dari perhitungan jarak gower dapat dilihat pada tabel 3 berikut ini:

Tabel 4. Hasil Perhitungan Gower dengan Data Uji 1 (R57)

No	Nama	Jarak			Koefisien Gower
		(X_1)	(X_2)	(X_3)	
1	R1	0,1667	0	0	0,055556
2	R2	0	0	0	0
3	R3	0	0	0	0
4	R4	0,5	0,3333	1	0,611111
5	R5	0,1667	0,6667	1	0,611111
⋮	⋮	⋮	⋮	⋮	⋮
76	R95	0,3333	0	0	0,1111
77	R96	0,1667	0,6667	0	0,2778
78	R97	0,3333	1	1	0,7778
79	R98	0	1	1	0,6667
80	R100	0,1667	0	0	0,055556

3.2. Penduga Kategori

Setelah diperoleh hasil perhitungan jarak Gower antara objek data R57 dengan semua data latih, langkah selanjutnya adalah menduga kategori data tersebut. Untuk mengetahui kategori objek data R57



hal yang harus dilakukan adalah mencari kategori terbanyak dari data yang menjadi tetangga (*neighbor*) terdekat. Tetangga terdekat diperoleh dengan cara mengurutkan jarak Gower terkecil ke terbesar. Semakin kecil jarak gower berarti jarak antar data semakin dekat. Berdasarkan hasil tersebut, ditentukan penduga kategori bagi data uji tersebut, diperoleh tetangga terdekat objek R57 berdasarkan masing-masing nilai K sebagai berikut :

a. Untuk $K = 3$

Tabel 5. Peringkat Data dengan $K=3$

Nama	(X_1)	(X_2)	(X_3)	Koefisien Gower	Kategori	Peringkat
R2	0	0	0	0	L	1
R3	0	0	0	0	L	2
R11	0	0	0	0	L	3

Nilai $K = 3$ artinya dibutuhkan tiga data tetangga terdekat yang telah diurutkan dari terkecil ke terbesar. Berdasarkan peringkat penghitungan koefisien jarak Gower antara objek data R57 dengan semua data latih. Dari Tabel 5 diatas diketahui bahwa ketiga data memiliki kategori “Layak”, maka dapat disimpulkan bahwa data uji (R57) dengan $K = 3$ memiliki kategori “Layak”.

b. Untuk $K = 5$

Tabel 6. Peringkat Data dengan $K=5$

Nama	(X_1)	(X_2)	(X_3)	Koefisien Gower	Kategori	Peringkat
R2	0	0	0	0	L	1
R3	0	0	0	0	L	2
R11	0	0	0	0	L	3
R25	0	0	0	0	L	4
R59	0	0	0	0	L	5

Nilai $K = 5$ artinya dibutuhkan lima data tetangga terdekat yang telah diurutkan dari terkecil ke terbesar. Berdasarkan peringkat penghitungan koefisien jarak Gower antara objek data R57 dengan semua data latih. Dari Tabel 6 diatas diketahui bahwa kelima data memiliki kategori “Layak”, maka dapat disimpulkan bahwa data uji (R57) dengan $K = 5$ memiliki kategori “Layak”.

c. Untuk $K = 7$

Tabel 7. Peringkat Data dengan $K=7$

Nama	(X_1)	(X_2)	(X_3)	Koefisien Gower	Kategori	Peringkat
R2	0	0	0	0	L	1
R3	0	0	0	0	L	2
R11	0	0	0	0	L	3
R25	0	0	0	0	L	4
R59	0	0	0	0	L	5
R74	0	0	0	0	L	6
R88	0	0	0	0	L	7

Nilai $K = 7$ artinya dibutuhkan tujuh data tetangga terdekat yang telah diurutkan dari terkecil ke terbesar. Berdasarkan peringkat penghitungan koefisien jarak Gower antara objek data R57 dengan

semua data latih. Dari Tabel 7 diatas diketahui bahwa ketujuh data memiliki kategori “Layak”, maka dapat disimpulkan bahwa data uji (R57) dengan $K = 7$ memiliki kategori “Layak”.

3.3. Menghitung Tingkat Akurasi

Dari hasil ukuran ketakmiripan yang menghasilkan prediksi kategori kelayakan seorang kepala keluarga dalam menerima bantuan program pangan akan dicari tingkat keakuratan hasil prediksi tersebut. Tingkat keakuratan diukur dengan perhitungan akurasi, *recall*, dan presisi menggunakan informasi yang diperoleh dari *confusion matrix*. Semakin tinggi nilai akurasi, presisi, dan, *recall* maka semakin tinggi tingkat keakuratan hasil prediksi begitu juga sebaliknya. Adapun cara memperoleh perhitungan akurasi, presisi, dan *recall* melalui informasi *confusion matrix* yang berisikan data yang dibagi menjadi dua kategori yaitu kategori prediksi dan kategori sebenarnya.

Tabel 8. Confusion Matrix

Kategori Prediksi	Kategori Sebenarnya	
	Layak	Tidak Layak
Layak	TP = 11	FP = 1
Tidak Layak	FN = 0	TN = 8

Dari Tabel 8, sebagai contoh digunakan $K = 3$ diperoleh data bahwa nilai akurasi adalah sebesar 95%, presisi 91,66%, dan *recall* 100%. Perhitungan selanjutnya, yaitu membandingkan beberapa hasil tingkat akurasi dari algoritma KNN dengan nilai $K = 3, K = 5, K = 7$ dan $K = 9$. Hasil ini akan dicatat lalu dibandingkan satu sama lain untuk menetapkan nilai K yang paling optimal yaitu penggunaan KNN dengan tingkat akurasi tertinggi.

Tabel 10. Akurasi KNN dari Berbagai Nilai K

	K=3	K=5	K=7	K=9
<i>Accuracy</i>	95%	90%	95%	90%
<i>Precision</i>	91,66%	90,9%	91,66%	90,9%
<i>Recall</i>	100%	90,9%	100%	90,9%

Pada Tabel 10 dapat dilihat bahwa nilai akurasi, presisi, dan *recall* yang cukup tinggi yaitu mendekati 100%, sehingga dapat disimpulkan bahwa KNN *classifier* yang digunakan dalam penelitian ini cocok untuk menentukan klasifikasi kepala keluarga di Kelurahan Tiakar menjadi layak dan tidak layak menerima program bantuan pangan. Hasil penelitian menunjukkan bahwa nilai $K = 3$ dan $K = 7$ merupakan nilai K yang paling optimal dengan tingkat akurasi sebesar 95%. Sedangkan untuk $K = 5$ dan $K = 9$ memiliki tingkat akurasi sebesar 90%.

Berdasarkan hasil penelitian ini, nilai K yang baik digunakan dalam klasifikasi kepala keluarga di Kelurahan Tiakar adalah $K = 3$ dan $K = 7$. Pada penelitian ini, nilai $K = 3$ umumnya memiliki prediksi kategori yang tepat untuk semua data yang menjadi tetangga terdekat dan nilai K yang kecil mengurangi tingkat gangguan (*noise*) dalam proses klasifikasi.



4. KESIMPULAN

Berdasarkan pada hasil penelitian dapat dilihat bahwa nilai akurasi tertinggi yaitu 95% sehingga dapat disimpulkan bahwa KNN *classifier* yang digunakan dalam penelitian ini cocok untuk menentukan klasifikasi kepala keluarga di Kelurahan Tiakar menjadi layak dan tidak layak menerima program bantuan pangan. Nilai K yang baik digunakan dalam klasifikasi kepala keluarga di Kelurahan Tiakar adalah $K = 3$ dimana memiliki tingkat akurasi paling tinggi dan nilai K yang kecil mengurangi tingkat gangguan (*noise*) dalam proses klasifikasi.

REFERENSI

- [1] Kemenkeu RI Kanwil DJPb Prov Kaltim. 2020. Indeks Manfaat Program Sembako. <http://djp.kemenkeu.go.id/indeks-manfaat-program-sembako>.
- [2] Prasetyowati, E. 2017. Data Mining Pengelompokan Data untuk Informasi dan Evaluasi. Duta Media Publishing, Jawa Timur.
- [3] Budiman, I. M. dan R. Ramadina. 2015. Penerapan Fungsi Data Mining Klasifikasi untuk Prediksi Masa Studi Mahasiswa Tepat Waktu pada Sistem Informasi Akademik Perguruan Tinggi. JUPITER (Jurnal Penelitian Ilmu dan Teknologi Komputer).7(1):39-50.
- [4] Sari, Y.A. dan S. Adinugroho. 2018. Implementasi Data Mining Menggunakan Weka. UB Press, Malang.
- [5] Rismiyati, F.A.D. dan M. Ary. 2019. Ukuran Akurasi Klasifikasi Penyakit Mesothelioma Menggunakan Algoritma K-Nearest Neighbor dan Backward Elimination. Jurnal Sains dan Teknologi Informasi.5(1): 2527-9114.
- [6] Indriyanto, A. 2021. Algoritma K-Nearest Neighbor untuk Prediksi Nasabah Asuransi. NEM, Google LLC.
- [7] Rahmadani, E.V., F. P. Syafriah dan H. H. Nisa. 2020. Algoritma C4.5 dan K-Nearest Neighbors (KNN) untuk Memetakan Matakuliah dan Keterlambatan Kelulusan Mahasiswa. Kreatif Industri Nusantara, Bandung.
- [8] Lestari, M.E.I. 2015. Penerapan algoritma klasifikasi Nearest Neighbor (KNN) untuk mendeteksi penyakit jantung. Faktor Exacta. 7(4): 366-371
- [9] Paryudi, I. 2019. What Affects K Value Selection In K-Nearest Neighbor. International Journal Of Scientific and Technology Research. 8(7).
- [10] Visa, S., dkk. 2011. Confusion matrix-based feature selection. MAICS.710:120-127
- [11] Devita, R. N., H. W. Herwanto dan A. P. Wibawa. 2018. Perbandingan Kinerja Metode Naive Bayes Dan K-Nearest Neighbor Untuk Klasifikasi Artikel Berbahasa Indonesia. JTIK (Jurnal Teknologi Informasi dan Ilmu Komputer). 5(4):427-434.
- [12] Ramadhani, M. dan D. H. Murti. 2018. Klasifikasi Ikan Menggunakan Oriented Fast and Rotated Brief (Orb) Dan K-Nearest Neighbor (Knn). JUTI J. Ilm. Teknol. Inf. 16(2): 8-115.
- [13] Indriati dan R. Achmad. 2016. Sentiment Analysis For Review Mobile Applications Using Neighbor Method Weighted K-Nearest Neighbor (Nwknn). Journal of Environmental Engineering and Sustainable Technology. 3(1): 23-32.
- [14] Sulizial, L.M. dan A. Fauzan. 2020. Comparing Naive Bayes, K-Nearest Neighbor, And Neural Network Classification Methods Of Seat Load
- [15] ulc, Z., M. Matejka dan J. Prochzka. 2016. Modifications of the Gower similarity coefficient. In 19th Conference of Applications of Mathematics and Statistics in Economics-(AMSE 2016).