

## IMPLEMENTASI ALGORITMA *K-NEAREST NEIGHBOR* DENGAN NORMALISASI *Z-SCORE* DALAM KLASIFIKASI PENERIMA BANTUAN SOSIAL DESA SERUNAI

Novi Safitri, Dadan Kusnandar, Shantika Martha

### INTISARI

Beberapa program pemerintah dalam menanggulangi masalah kemiskinan, diantaranya Program Keluarga Harapan (PKH), Bantuan Pangan Non Tunai (BPNT), Bantuan Sosial Tunai (BST), dan Bantuan Langsung Tunai (BLT) dana desa. Desa Serunai merupakan desa yang berada di Kecamatan Salatiga Kabupaten Sambas, telah menerapkan pemberian bantuan sosial sejak tahun 2020. Penentuan penerima bantuan sosial yang dilakukan dengan cara manual dapat menjadi penyebab ketidaktepatan sasaran penerima bantuan. Oleh karenanya penelitian ini bertujuan mengklasifikasikan penerima bantuan sosial dan mencari akurasi terbaik menggunakan algoritma *K-Nearest Neighbor* dengan normalisasi *z-score*. Penerima bantuan sosial diklasifikasikan menjadi tiga, yaitu penerima PKH, penerima BLT dana desa dan tidak menerima bantuan. Klasifikasi penerima bantuan menggunakan perhitungan jarak Euclidean dengan sembilan parameter  $K$ , yaitu  $3 \leq K \leq 19$ ,  $K \in \text{bilangan ganjil}$ . Perhitungan akurasi dilakukan dengan confusion matrix. Data yang digunakan adalah data Pemutakhiran Keluarga Desa Serunai Tahun 2022 yang berdomisili di Dusun Kemboja, diperoleh dari website BKKBN dan kantor Desa Serunai, sebanyak 205 data dengan delapan variabel atribut dan satu variabel kelas. Data dibagi menjadi data latih (training) dan data uji (testing), dengan proporsi 90:10, 80:20, dan 70:30. Hasil penelitian menunjukkan bahwa klasifikasi terbaik diperoleh pada proporsi 90:10 dengan parameter  $K = 11$ , dimana dari 20 individu yang diprediksi, terdapat sepuluh individu diklasifikasikan sebagai penerima PKH (kelas 1), satu individu diklasifikasikan sebagai penerima BLT dana desa (kelas 2) dan sembilan individu diklasifikasikan sebagai bukan penerima bantuan (kelas 3). Jika disandingkan dengan data aktual terdapat dua individu yang salah diklasifikasikan sehingga didapatkan akurasi sebesar 90%, dengan recall sebesar 100% dan presisi sebesar 80%.

**Kata Kunci:** klasifikasi, normalisasi *z-score*, jarak Euclidean.

### PENDAHULUAN

Kemiskinan tidak hanya kecenderungan seseorang yang tidak mampu memenuhi kebutuhan hidupnya, tetapi juga merupakan suatu kegagalan negara dalam memenuhi hak masyarakatnya untuk hidup sejahtera [1]. Berbagai upaya yang dilakukan pemerintah dalam menanggulangi kemiskinan, diantaranya dengan memberikan program bantuan sosial seperti Program Keluarga Harapan (PKH), Bantuan Pangan Non Tunai (BPNT), Bantuan Sosial Tunai (BST) dan Bantuan Langsung Tunai (BLT) dana desa [2]. Bantuan ini bersifat selektif dan bertujuan untuk menjangkau pihak yang benar-benar layak menerima bantuan.

Desa Serunai merupakan desa yang berada di Kecamatan Salatiga, Kabupaten Sambas. Desa Serunai telah menerapkan program bantuan sosial sejak tahun 2020 sampai saat ini. Secara umum bantuan ini berhasil diterapkan, namun penetapan calon penerima bantuan yang dilakukan secara manual menjadi kekhawatiran sebagai pemicu kasus tidak tepat sasaran. Data *mining* sebagai bentuk kemajuan teknologi saat ini dapat dimanfaatkan petugas desa untuk menetapkan status calon penerima bantuan dengan cepat dan akurat. Salah satu fungsi data *mining* adalah untuk melakukan klasifikasi [3]. Klasifikasi adalah proses memilah label/kelas data yang bertujuan untuk memprediksi label/kelas dari data baru yang belum diketahui [4]. Pada penelitian ini dilakukan klasifikasi status penerima bantuan sosial menggunakan algoritma *K-Nearest Neighbor*. Algoritma *K-Nearest Neighbor* dihitung berdasarkan kedekatan atau kemiripan kasus satu dengan kasus lainnya [5]. Algoritma *K-Nearest Neighbor* memiliki

keunggulan yaitu proses yang cepat, sederhana, mudah dipelajari, efektif pada data latih dengan jumlah yang besar dan menghasilkan performa yang cukup baik [6]. Data yang digunakan pada penelitian ini perlu dilakukan normalisasi karena memiliki atribut dengan rentang data yang berbeda-beda dan memiliki nilai *outlier*. Normalisasi yang digunakan adalah normalisasi *z-score* karena lebih stabil terhadap *outlier* [7]. Tujuan penelitian ini yaitu untuk mengklasifikasikan penerima bantuan sosial menggunakan algoritma *K-Nearest Neighbor* dengan normalisasi *z-score* dan mencari akurasi terbaik yang dihasilkan oleh algoritma *K-Nearest Neighbor* dengan normalisasi *z-score*. Batasan masalah pada penelitian ini menggunakan perhitungan jarak *Euclidean* dan menggunakan sembilan parameter *K* yaitu  $3 \leq K \leq 19$ ,  $K \in \text{bilangan ganjil}$ . Data penelitian menggunakan data warga berdomisili di Dusun Kemboja pada Pemutakhiran Keluarga Desa Serunai Tahun 2022, data penelitian terbagi menjadi data latih (*training*) dan data uji (*testing*) dengan proporsi perbandingan 90:10, 80:20, dan 70:30, dan pengklasifikasian terbagi menjadi tiga yaitu penerima PKH, penerima BLT dana desa dan tidak menerima bantuan.

Langkah pertama pada penelitian ini adalah melakukan input data penelitian. Data penelitian terdiri dari data numerik dan kategorik. Pada data numerik dilakukan uji *outlier* dan data kategorik dikonversi menjadi numerik. Setelah dilakukan uji *outlier* pada data numerik, terdapat nilai *outlier* sehingga dilakukan normalisasi *z-score*. Data penelitian dibagi menjadi data latih (*training*) dan data uji (*testing*) dengan proporsi perbandingan 90:10, 80:20, dan 70:30, selanjutnya dilakukan tahapan algoritma *K-Nearest Neighbor* hingga didapatkan hasil klasifikasi dan kemudian dievaluasi menggunakan *confusion matrix*.

### Data Mining

Data *mining* adalah proses yang menggunakan pembelajaran mesin, kecerdasan buatan, statistik dan matematika dalam analisis dan ekstraksi pengetahuan serta ekstraksi informasi dari *database* besar dan dapat digunakan sebagai alat pengambilan keputusan di masa akan datang [5]. Adapun enam fungsi data *mining* yaitu deskripsi, estimasi, prediksi, klasifikasi, *cluster* dan asosiasi [3].

Data *mining* meliputi pengumpulan data, ekstraksi data, analisis data dan statistik data. Berkaitan dengan penggunaan data yang ada dalam *database*, data *mining* sering disinonimkan oleh banyak peneliti sebagai proses KDD (*Knowledge Discovery in Database*) [5]. Langkah-langkah proses KDD meliputi pembersihan data, integrasi data, transformasi data, data *mining*, evaluasi pola dan presentasi pengetahuan [8].

### Normalisasi Z-Score

Umumnya nilai-nilai atribut data memiliki rentang yang berbeda akan menyebabkan bias pada proses data *mining*. Untuk menghindari bias, maka data dapat dinormalisasikan atau distandarisasikan, sehingga dengan demikian data memiliki bobot yang sama. Ada banyak teknik yang digunakan dalam normalisasi salah satunya normalisasi *z-score*.

Normalisasi *z-score* dengan memanfaatkan *mean* dan standar deviasi akan lebih stabil terhadap nilai pencilan/*outlier* maupun nilai-nilai baru yang lebih kecil dari *min* dan yang lebih besar dari *max* [7]. Normalisasi *z-score* dapat dihitung dengan Persamaan (1) [9]:

$$z_{il} = \frac{(x_{il} - \mu_l)}{\sigma_l} \quad (1)$$

Keterangan:

- $z_{il}$  : Hasil normalisasi *z-score* objek ke-*i* pada variabel atribut ke-*l*
- $x_{il}$  : Nilai data asli objek ke-*i* pada variabel atribut ke-*l*
- $\mu_l$  : Nilai rata-rata variabel atribut ke-*l*
- $\sigma_l$  : Standar deviasi variabel atribut ke-*l*

## Klasifikasi

Klasifikasi merupakan fungsi dari data *mining* yang berguna mendeskripsikan, mengidentifikasi dan memilah label/kelas data atau konsep data, bertujuan menggunakan objek yang label/kelasnya telah diketahui sebelumnya untuk dapat memprediksi label/kelas objek baru yang belum diketahui [4]. Proses klasifikasi melewati dua tahapan, yaitu pembelajaran dan klasifikasi [4]. Klasifikasi memiliki empat komponen dasar yaitu kelas, atribut, dataset latih, dan dataset uji [5].

## *K-Nearest Neighbor*

*K-Nearest Neighbor* merupakan teknik klasifikasi yang sederhana, efektif pada data latih dengan jumlah yang besar dan menghasilkan performa yang cukup baik [6]. Algoritma *K-Nearest Neighbor* bertujuan mengklasifikasikan objek baru berdasarkan jumlah mayoritas *K*-tetangga terdekat dari perhitungan jarak atribut pada data uji dan data latih. Salah satu kelemahan *K-Nearest Neighbor* adalah penentuan parameter *K* yang bias. *K* ditentukan secara bebas, tergantung pada peneliti itu sendiri, umumnya *K* diambil dalam bilangan ganjil. *K* adalah parameter yang mengacu kepada jumlah tetangga terdekat dari suatu objek yang diprediksi kelasnya untuk dapat menentukan mayoritas tetangga bagi suatu objek. Algoritma *K-Nearest Neighbor* memiliki dua komponen penting yaitu jumlah tetangga terdekat (*K*) dan jarak [5]. Adapun langkah-langkah algoritma *K-Nearest Neighbor* adalah sebagai berikut [5]:

1. Menentukan parameter *K*. Parameter *K* sebaiknya diambil bilangan ganjil untuk memudahkan penentuan (*voting*) mayoritas dan minoritas *K*. Parameter *K* bilangan ganjil juga berguna menghindari terjadinya lebih dari satu kelas (kelas imbang) yang memenangkan *voting* untuk ditetapkan sebagai hasil klasifikasi [7].
2. Menghitung jarak data uji yang akan diprediksi oleh semua jarak data latih yang ada menggunakan metode jarak *Euclidean*.
3. Mengurutkan hasil jarak *Euclidean* dan menentukan tetangga terdekat sesuai dengan *K* yang dipilih (diurutkan dari nilai terkecil ke nilai terbesar).
4. Pilih kelas mayoritas dari parameter *K* yang telah ditentukan sebagai kelas klasifikasi data baru yang diprediksi.

## Jarak *Euclidean*

Jarak *Euclidean* adalah rumus pengukuran jarak terpopuler. Semakin kecil jarak *Euclidean*, semakin mirip kedua atribut tersebut [5]. Jarak harus selalu bernilai positif. Jarak *Euclidean* dapat dihitung menggunakan Persamaan (2) [10]:

$$d(x_h, x_j) = |x_h, x_j| = \sqrt{\sum_{l=1}^p (x_{hl} - x_{jl})^2} \quad (2)$$

Keterangan:

- $d(x_h, x_j)$  : Jarak *Euclidean* antara data latih (*training*) ke-*h* dengan data uji (*testing*) ke-*j*  
 $|x_h, x_j|$  : Jarak *Euclidean* antara data latih (*training*) ke-*h* dengan data uji (*testing*) ke-*j* selalu bernilai positif  
 $x_{hl}$  : Data latih (*training*) ke-*h* pada variabel atribut ke-*l*  
 $x_{jl}$  : Data uji (*testing*) ke-*j* pada variabel atribut ke-*l*  
 $p$  : Jumlah variabel atribut yang diamati

### Confusion Matrix

*Confusion matrix* merupakan alat informasi kelas aktual dan prediksi yang berguna untuk evaluasi hasil klasifikasi dalam prediksi objek “True” atau “False” [11]. *Confusion matrix* adalah sebuah matriks yang menunjukkan hasil dari klasifikasi aktual dan prediksi dengan ukuran  $M \times M$ , dimana  $M$  adalah banyaknya label/kelas klasifikasi. Pada penelitian ini menggunakan *confusion matrix* berukuran  $3 \times 3$ , karena mempunyai tiga label/kelas. *Confusion matrix* berukuran  $3 \times 3$  ditampilkan pada Gambar 1 [12]:

		Predicted			
		A	B	C	
Actual	A	AA	AB	AC	→ TP → FN
	B	BA	BB	BC	→ TN
	C	CA	CB	CC	→ FP

**Gambar 1.** *Confusion Matrix*  $3 \times 3$

Pada Gambar 3.1 *confusion matrix* berukuran  $3 \times 3$ , AA adalah data aktual positif benar diprediksi sebagai positif atau *True Positive* (TP), AB dan AC adalah data aktual positif salah diprediksi sebagai negatif atau *False Negative* (FN), BA dan CA adalah data aktual negatif salah diprediksi sebagai positif atau *False Positive* (FP) dan BB, BC, CB dan CC adalah data aktual negatif benar diprediksi sebagai negatif atau *True Negative* (TN).

Keterangan:

*True Positive* (TP) : Jumlah data aktual dengan kelas positif dan model memprediksi positif

*False Negative* (FN) : Jumlah data aktual dengan kelas positif dan model memprediksi negatif

*False Positive* (FP) : Jumlah data aktual dengan kelas negatif dan model memprediksi positif

*True Negative* (TN) : Jumlah data aktual dengan kelas negatif dan model memprediksi negatif

Berikut adalah *performance metrics* yang digunakan untuk mengukur kinerja model [13]:

1. Akurasi (*accuracy*) adalah proporsi jumlah data yang diprediksi benar setelah dilakukan uji hasil klasifikasi. Akurasi merupakan nilai kedekatan antara hasil prediksi dengan data aktual. Akurasi dapat dihitung menggunakan Persamaan (3):

$$accuracy = \frac{TP+TN}{\text{total data}} \times 100\% \quad (3)$$

2. *Sensitivity* atau *recall* adalah proporsi kasus pada aktual positif diprediksi positif secara benar. *Recall* merupakan kemampuan algoritma untuk menemukan kembali informasi dari data aktual berdasarkan hasil prediksi. *Recall* dapat dihitung menggunakan Persamaan (4):

$$recall = \frac{TP}{TP+FN} \times 100\% \quad (4)$$

3. *Precision* adalah proporsi kasus pada prediksi positif yang juga sebenarnya kasus positif. *Precision* merupakan konsistensi dari kemampuan algoritma dalam memberikan prediksi hasil sesuai dengan data aktual. *Precision* dapat dihitung menggunakan Persamaan (5):

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

## PEMBAHASAN

Data pada penelitian ini adalah data warga pada Pemutakhiran Keluarga Desa Serunai Tahun 2022, yang berdomisili di Dusun Kemboja Desa Serunai Kecamatan Salatiga Kabupaten Sambas dan berjumlah 205 individu. Variabel penelitian ditampilkan pada Tabel 1.

**Tabel 1.** Variabel Penelitian

Variabel	Satuan	Keterangan
Jenis Kelamin ( $X_1$ )	-	1. Laki-Laki 2. Perempuan
Pekerjaan ( $X_2$ )	-	1. Tidak Bekerja 2. Petani 3. Wiraswasta 4. PNS/TNI/POLRI 5. Pegawai Swasta 6. Pensiunan 7. Pekerja Lepas
Pendidikan ( $X_3$ )	-	1. Tidak Sekolah 2. Tidak Tamat SD 3. Tamat SD 4. Tamat SLTP 5. Tamat SLTA 6. Tamat PT/Akademi
Jenis Dinding Rumah ( $X_4$ )	-	1. Kayu/Papan 2. Tembok
Jenis Lantai Rumah ( $X_5$ )	-	1. Kayu/Papan 2. Semen 3. Keramik/Granit/Marmer/Ubin/Tegel/Teraso
Luas Rumah ( $X_6$ )	$m^2$	-
Jumlah Tanggungan ( $X_7$ )	Orang	-
Usia ( $X_8$ )	Tahun	-
Penerima Bantuan ( $Y$ )	-	1. PKH 2. BLT 3. Tidak menerima bantuan

Variabel data penelitian kategorik dikonversikan menjadi numerik seperti pada Tabel 1 dan variabel data numerik dilakukan uji *outlier*, setelah dilakukan uji *outlier* terdapat data yang mengandung *outlier* sehingga perlu dilakukan normalisasi *z-score* dengan perhitungan pada Persamaan (1).

Data dibagi menjadi data latih (*training*) dan data uji (*testing*) dengan proporsi 90:10, 80:20, dan 70:30. Pada pembagian data terlebih dahulu dilakukan pengacakan data menggunakan *software* Microsoft Office Excel pada fungsi “*Rand*” kemudian dilakukan “*Sort*”. Dari proporsi 90:10 didapatkan 185 data sebagai data latih (*training*) dan 20 data sebagai data uji (*testing*), proporsi 80:20 didapatkan 164 data sebagai data latih (*training*) dan 41 data sebagai data uji (*testing*), dan proporsi 70:30 didapatkan 144 data sebagai data latih (*training*) dan 61 data sebagai data uji (*testing*). Kemudian dilakukan perhitungan algoritma *K-Nearest Neighbor*. Misal dilakukan analisis algoritma *K-Nearest Neighbor* pada proporsi 90:10. Berikut adalah langkah-langkah algoritma *K-Nearest Neighbor*:

1. Menentukan parameter  $K$ . Pada penelitian ini menggunakan sembilan parameter  $K$  yaitu  $3 \leq K \leq 19$ ,  $K \in$  bilangan ganjil
2. Melakukan perhitungan jarak *Euclidean* menggunakan perhitungan pada Persamaan (2).
3. Mengurutkan hasil jarak *Euclidean* dari nilai terkecil sampai nilai terbesar. Hasil pengurutan jarak *Euclidean* diberi peringkat sesuai dengan jumlah tetangga terdekatnya (parameter  $K$ ).
4. Menentukan hasil klasifikasi berdasarkan kelas mayoritasnya.

Dilakukan langkah-langkah algoritma *K-Nearest Neighbor* diatas untuk setiap individu data uji (*testing*) dan didapatkan hasil klasifikasi yang ditampilkan pada Tabel 2.

**Tabel 2.** Hasil Klasifikasi Data Uji (*Testing*) dengan Proporsi 90:10

Individu ke	$\hat{Y}$									Y
	K3	K5	K7	K9	K11	K13	K15	K17	K19	
186	3	3	3	3	3	3	3	3	3	3
187	1	1	1	1	1	1	1	1	1	1
188	1	1	1	1	1	1	1	1	1	1
189	2	2	2	2	2	1	1	1	1	2
190	3	3	3	3	3	3	3	3	3	3
191	3	3	3	3	3	3	3	3	3	3
192	3	3	3	3	3	3	3	3	3	3
193	3	3	3	3	3	3	3	3	3	3
194	3	1	1	1	1	1	1	1	1	1
195	1	1	1	1	1	1	1	1	1	3
196	1	1	1	1	1	1	1	1	1	1
197	1	1	1	1	1	1	3	1	3	3
198	3	3	1	1	1	1	1	1	1	1
199	1	1	1	1	1	1	1	1	1	1
200	3	3	3	3	3	3	3	3	3	3
201	1	1	1	1	1	1	1	1	1	1
202	1	1	1	1	3	3	1	3	3	3
203	3	3	3	3	3	3	3	3	3	3
204	1	1	1	1	1	1	1	1	1	1
205	3	3	3	3	3	3	3	3	3	3

Berdasarkan Tabel 2, jika disandingkan hasil klasifikasi yang diperoleh dengan data aktualnya, pada parameter  $K = 3$  terdapat lima individu yang salah diklasifikasikan yaitu individu ke 194, 195, 197, 198 dan 202,  $K = 5$  terdapat empat individu yang salah diklasifikasikan yaitu individu ke 195, 197, 198 dan 202,  $K = 7$  dan  $K = 9$  terdapat tiga individu yang salah diklasifikasikan yaitu individu ke 195, 197 dan 202,  $K = 11$  terdapat dua individu yang salah diklasifikasikan yaitu 195 dan 197,  $K = 13$  terdapat tiga individu yang salah diklasifikasikan yaitu 189, 195 dan 197,  $K = 15$  terdapat tiga individu yang salah diklasifikasikan yaitu 189, 195 dan 202,  $K = 17$  terdapat tiga individu yang salah diklasifikasikan yaitu 189, 195 dan 197 dan  $K = 19$  terdapat dua individu yang salah diklasifikasikan yaitu 189 dan 195. Kemudian misal pada parameter  $K = 11$  dilakukan *confusion matrix* yang ditampilkan pada Tabel 3.

**Tabel 3.** *Confusion Matrix* dengan Parameter  $K = 11$  pada Proporsi 90:10

Aktual	Prediksi			
	Kelas 1	Kelas 2	Kelas 3	
Kelas 1	8	0	0	TP
Kelas 2	0	1	0	TN
Kelas 3	2	0	9	FP

Berdasarkan Tabel 3, pada parameter  $K = 11$  terdapat delapan individu pada aktual memiliki kelas “1” benar diprediksi memiliki kelas “1”, sembilan individu pada data aktual memiliki kelas “3” benar diprediksi memiliki kelas “3”, satu individu pada data aktual memiliki kelas “2” benar diprediksi memiliki kelas “2”, dan dua individu pada data aktual memiliki kelas “3” salah diprediksi memiliki kelas “1”.

Dilakukan analisis algoritma *K-Nearest Neighbor* pada semua proporsi data yang telah ditentukan yaitu 90:10, 80:20, dan 70:30 sehingga didapatkan hasil klasifikasi. Hasil klasifikasi dilakukan perhitungan akurasi, *recall* dan presisi dengan Persamaan (3), (4), dan (5) sehingga diperoleh hasil seperti yang ditampilkan pada Tabel 4.

**Tabel 4.** Perbandingan Hasil Akurasi dengan Berbagai Proporsi pembagian Data

K	Proporsi								
	70% data latih ( <i>training</i> ) 30% data uji ( <i>testing</i> )			80% data latih ( <i>training</i> ) 20% data uji ( <i>testing</i> )			90% data latih ( <i>training</i> ) 10% data uji ( <i>testing</i> )		
	<i>Accuracy</i>	<i>Recall</i>	<i>Precision</i>	<i>Accuracy</i>	<i>Recall</i>	<i>Precision</i>	<i>Accuracy</i>	<i>Recall</i>	<i>Precision</i>
3	70,49%	75,86%	66,67%	60,97%	70,58%	52,17%	75,00%	75,00%	66,67%
5	67,21%	79,31%	62,16%	63,41%	76,47%	54,17%	80,00%	87,50%	70,00%
7	75,41%	<b>82,76%</b>	70,59%	68,29%	<b>82,35%</b>	58,33%	85,00%	<b>100,00%</b>	72,73%
9	77,05%	<b>82,76%</b>	72,73%	<b>73,17%</b>	<b>82,35%</b>	<b>63,64%</b>	85,00%	<b>100,00%</b>	72,73%
11	<b>81,97%</b>	<b>82,76%</b>	<b>80,00%</b>	<b>73,17%</b>	<b>82,35%</b>	<b>63,64%</b>	<b>90,00%</b>	<b>100,00%</b>	<b>80,00%</b>
13	77,05%	75,86%	75,86%	70,73%	<b>82,35%</b>	60,87%	85,00%	<b>100,00%</b>	72,73%
15	75,41%	75,86%	73,33%	70,73%	<b>82,35%</b>	60,87%	85,00%	<b>100,00%</b>	72,73%
17	75,41%	75,86%	73,33%	<b>73,17%</b>	76,47%	60,50%	85,00%	<b>100,00%</b>	72,73%
19	75,41%	75,86%	73,33%	<b>73,17%</b>	76,47%	60,50%	<b>90,00%</b>	<b>100,00%</b>	<b>80,00%</b>

Berdasarkan Tabel 4 pada proporsi pembagian data 90:10 memperoleh akurasi terbaik pada parameter  $K = 11$  dan  $K = 19$  yaitu sebesar 90%, dengan *recall* sebesar 100% dan presisi sebesar 80%. Pada proporsi 80:20 menghasilkan akurasi terbaik pada parameter  $K = 9$  dan  $K = 11$  yaitu sebesar 73,17%, dengan *recall* sebesar 82,35% dan presisi sebesar 63,64%. Pada proporsi 70:30 menghasilkan akurasi terbaik pada parameter  $K = 11$  yaitu sebesar 81,97% dengan *recall* sebesar 82,76% dan presisi sebesar 80%. Pada penelitian ini dipilih parameter  $K = 11$  sebagai  $K$  terbaik pada klasifikasi *K-Nearest Neighbor* karena ditinjau dari masing-masing proporsi yang telah ditentukan dominan menghasilkan akurasi terbaik pada parameter  $K = 11$  dan karena penelitian ini berdasarkan pada kedekatan ketetanggaannya, maka jika dibandingkan antara  $K = 11$  dan  $K = 19$ ,  $K = 11$  lebih dekat ketetanggaannya.

## PENUTUP

Berdasarkan analisis yang telah dilakukan diperoleh hasil pada proporsi data 90:10 memperoleh akurasi terbaik pada parameter  $K = 11$  dan  $K = 19$  yaitu sebesar 90%, dengan *recall* sebesar 100% dan presisi sebesar 80%. Pada proporsi data 80:20 menghasilkan akurasi terbaik pada parameter  $K = 9$  dan  $K = 11$  yaitu sebesar 73,17%, dengan *recall* sebesar 82,35% dan presisi sebesar 63,64%. Pada proporsi data 70:30 menghasilkan akurasi terbaik pada parameter  $K = 11$  yaitu sebesar 81,97% dengan *recall* sebesar 82,76% dan presisi sebesar 80%. Kemudian dapat ditarik kesimpulan bahwa klasifikasi algoritma *K-Nearest Neighbor* dengan menggunakan normalisasi *z-score* berhasil diimplementasikan dengan baik untuk permasalahan seleksi penerima bantuan sosial di Desa Serunai pada proporsi data 90:10 dengan parameter  $K = 11$ . Adapun hasil klasifikasi dari algoritma *K-Nearest Neighbor* dengan normalisasi *z-score* adalah pada proporsi data 90:10 dengan parameter  $K = 11$  yaitu dari 20 individu

terdapat sepuluh individu diklasifikasikan sebagai penerima PKH (kelas 1), satu individu diklasifikasikan penerima BLT dana desa (kelas 2), dan sembilan individu diklasifikasikan sebagai bukan penerima bantuan (kelas 3). Akurasi terbaik yang dihasilkan sebesar 90%, dengan *recall* sebesar 100% dan presisi sebesar 80%.

## DAFTAR PUSTAKA

- [1]. Nurjanah SL, Yovita I. Analisis Pengaruh Jumlah Penduduk, Indeks Pembangunan Manusia dan Tingkat Pengaruh terhadap Kemiskinan di Provinsi Riau Tahun 2002-2021. *Jurnal Ekonomi dan Bisnis*. 2022; 11(1):567-574.
- [2]. Hasanah RL, Hasan M, Pangesti WE, Wati FF, Gata W. Klasifikasi Penerima Dana Bantuan Desa Menggunakan Metode KNN (K-Nearest Neighbor). *Jurnal Techno Nusa Mandiri*. 2019; 16(1):1-6.
- [3]. Mustafa MS, Ramadhan MR, Thenata AP. Implementasi Data Mining untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier. *Creative Information Technology Journal*, 2017;4(2):151-162.
- [4]. Han J, Kamber M, Pei J. *Data Mining Concepts and Techniques Third Edition*. United States of America: Morgan Kaufman; 2012.
- [5]. Arhami M, Nasir M. *Data Mining Algoritma dan Implementasi*. Yogyakarta: Penerbit ANDI; 2020.
- [6]. Bhatia N. Survey of Nearest Neighbor Techniques. *International Journal of Computer Science and Information Security*. 2010; 8(2):302-305.
- [7]. Suyanto. *Data Mining untuk Klasifikasi dan Klasterisasi Data*. Bandung: Informatika; 2019.
- [8]. Vulandari RT. *Data Mining Teori dan Aplikasi Rapidminer*. Yogyakarta: Penerbit Gava Media; 2017.
- [9]. Kusnadi A, Pane IZ, Khaeruzzaman Y, Clara VC. *Ekstraksi Fitur dan Pengenalan Wajah Konsep dan Aplikasinya*. Malang: CV Literasi Nusantara Abadi; 2022.
- [10]. Anton H, Rorres C. *Aljabar Linear Elementer Versi Aplikasi Edisi Delapan/Jilid 1*. Jakarta: Erlangga; 2004.
- [11]. Zulaikhah S, Aziz A, Harianto W. Optimasi Algoritma K-Nearest Neighbor dengan Normalisasi dan Seleksi Fitur untuk Klasifikasi Penyakit Liver. *Jurnal Mahasiswa Teknik Informatika*. 2021; 6(2):439-445.
- [12]. Wabang K, Nurhayati OD, Farikhin. Application of The Naive Bayes Classifier Algorithm to Classify Community Complaints. *Jurnal Rekayasa Sistem dan Teknologi Informasi*. 2022; 6(10):872-876.
- [13]. Siswa TA, Prihandoko. Penerapan Optimasi Berbasis Particle Swarm Optimization (PSO) Algoritma Naive Bayes dan K-Nearest Neighbor sebagai Perbandingan untuk Mencari Kinerja Terbaik dalam Mendeteksi Kanker Payudara. *Jurnal Media Teknologi Informasi dan Komputer*. 2018; 2(2):38-45.