



**PENERAPAN ALGORITMA *PARTICLE SWARM OPTIMIZATION* UNTUK
OPTIMALISASI *FUZZY K-NEAREST NEIGHBOR***

***APPLICATION OF PARTICLE SWARM OPTIMIZATION ALGORITHM FOR FUZZY K-
NEAREST NEIGHBOR OPTIMIZATION***

Disna Putri Sari*, Prodi Matematika FMIPA UNY

Indira Ihnu Brilliant, Prodi Matematika FMIPA UNY

*e-mail: disnaputri.2020@student.uny.ac.id

Abstrak. Salah satu hal penting yang memengaruhi akurasi *fuzzy k-nearest neighbor* adalah penentuan parameter k dan m . Pada penelitian ini digunakan algoritma *particle swarm optimization* untuk menentukan kombinasi parameter k , m yang dapat menghasilkan akurasi paling baik. Penelitian ini dilakukan dengan tujuan untuk mengoptimalkan akurasi hasil klasifikasi algoritma *Fuzzy K-Nearest Neighbor* (FKNN) dengan penerapan *Particle Swarm Optimization* (PSO). Tahapan dari metode penelitian ini antara lain *preprocessing* data dengan normalisasi *min-max*, *processing* data dengan algoritma FKNN-PSO, dan evaluasi model dengan *confusion matrix*. Penelitian ini diimplementasikan pada *software python* dengan *wisconsin breast cancer dataset* yang diperoleh dari *scikitlearn python*. Data yang digunakan merupakan klasifikasi kanker payudara ke dalam dua kelas yaitu kanker ganas (*malignant*) dan kanker jinak (*benign*). Data tersebut dibagi ke dalam data latih dan data uji dengan perbandingan 70:30, sehingga diperoleh data latih sebesar 398 dan data uji sebesar 171. Hasil penelitian menunjukkan akurasi algoritma FKNN-PSO adalah 96,5% dengan $k=63$ dan $m=49$.

Kata kunci: *Fuzzy K-Nearest Neighbor*, Klasifikasi, *Particle Swarm Optimization*

Abstract. One of the important things that affects the accuracy of *fuzzy k-nearest neighbor* is the determination of the k and m parameters. In this research, the *particle swarm optimization* algorithm is used to determine the combination of k and m that can produce the best accuracy. This research was conducted with the aim of optimizing the accuracy of the classification results of the *Fuzzy K-Nearest Neighbor* (FKNN) Algorithm with the application of *Particle Swarm Optimization* (PSO). The stages of this research method include data preprocessing with *min-max* normalization, data processing with FKNN-PSO algorithm, and model evaluation with *confusion matrix*. This research is implemented on *python* software with *wisconsin breast cancer dataset* obtained from *scikitlearn python*. The data used is the classification of breast cancer in two classes, namely malignant cancer and benign cancer. The data is divided into training data and test data with a ratio of 70:30, resulting in training data of 398 and test data of 171. The results showed the accuracy of the FKNN-PSO algorithm was 96,5% with $k=63$ and $m=49$.

Keywords: *Fuzzy K-Nearest Neighbor*, Classification, *Particle Swarm Optimization*.

PENDAHULUAN

Klasifikasi merupakan analisis data dengan menemukan model dalam mendeskripsikan kelas data untuk memprediksi label kelas yang belum diketahui (Han et al., 2012). Salah satu metode klasifikasi adalah *k-nearest neighbor*. Algoritma *k-nearest neighbor* merupakan klasifikasi data berdasarkan mayoritas kelas dalam *k* tetangga terdekat. Algoritma ini termasuk dalam metode klasifikasi yang mudah diterapkan karena prosesnya yang sangat sederhana. Namun, pada metode ini vektor sampel yang ditetapkan ke suatu kelas tidak memiliki indikasi kekuatan keanggotaan dalam kelas tersebut, sehingga setiap vektor sampel dianggap sama pentingnya pada pemberian label kelas (Keller et al., 1985). Dengan demikian, diperkenalkan *fuzzy k-nearest neighbor* yang merupakan modifikasi dari algoritma *k-nearest neighbor*. Algoritma ini memanfaatkan teori *fuzzy* untuk membantu penetapan label kelas berdasarkan fungsi keanggotaan.

Salah satu proses klasifikasi menggunakan *fuzzy k-nearest neighbor* (FKNN) adalah menghitung jarak data uji ke setiap data latih. Beberapa jarak yang biasa digunakan antara lain *Euclidean*, *Manhattan*, *Canberra*, *Chebyshev*, *Bray-Curtis*, *Clark*, *Hamming*, *Bhattacharyya* (Ehsani & Drablos, 2020). Metode untuk mengukur jarak antara dua titik yang paling umum digunakan karena mewakili cara manusia memikirkan jarak di dunia nyata adalah *Euclidean distance* (Larose, 2005). Beberapa penelitian algoritma FKNN yang menggunakan *Euclidean distance* antara lain FKNN untuk prediksi kelulusan mahasiswa oleh Ahmad et al. (2018), FKNN untuk diagnosis penyakit Parkinson oleh Desinaini et al. (2019), FKNN untuk identifikasi status gizi balita oleh Nurida et al. (2019), FKNN untuk klasifikasi kelompok keahlian di Fakultas Ilmu Komputer Universitas Telkom oleh Kusuma et al. (2022), FKNN untuk klasifikasi curah hujan oleh Hidayatullaily & Zulhadi (2023), dan masih banyak lagi.

Fuzzy k-nearest neighbor merupakan algoritma klasifikasi yang menggunakan ukuran jarak tetangga terdekat untuk memperoleh solusi optimal. Menurut Han et al. (2012), normalisasi sangat berguna dalam algoritma klasifikasi yang berkaitan dengan pengukuran jarak. Beberapa metode normalisasi antara lain, *min-max*, *z-score*, dan *decimal scaling*. Penelitian oleh Pandey & Jain (2017) mengenai klasifikasi IRIS data dengan KNN menunjukkan hasil akurasi sebesar 100% untuk normalisasi *min-max* dan 85,71% untuk normalisasi *z-score*. Penelitian lain oleh Nasution et al. (2019) mengenai klasifikasi *wine* dengan metode KNN menghasilkan akurasi sebesar 65,92% untuk normalisasi *min-max* dan 65,85% untuk normalisasi *z-score*. Adapun penelitian oleh Soleh et al., (2022) mengenai klasifikasi penyakit diabetes dengan metode KNN menghasilkan akurasi sebesar 74% untuk normalisasi *min-max* dan 72% untuk normalisasi *zscore*. Normalisasi *min-max* juga banyak digunakan dalam penelitian terkait klasifikasi dengan metode FKNN. Penelitian-penelitian tersebut antara lain, FKNN untuk diagnosis penyakit anjing oleh Shanti et al. (2018), FKNN untuk identifikasi cedera pada pemain futsal oleh Sodik et al. (2019), FKNN untuk klasifikasi penyakit demam berdarah oleh Ramadhani et al. (2023), dan masih banyak lagi. Normalisasi ini juga dicontohkan untuk transformasi nilai atribut sebelum penghitungan jarak *Euclidean* pada metode KNN dalam buku yang ditulis oleh Han et al. (2012).

Setelah proses normalisasi, pada algoritma FKNN perlu dilakukan pembagian data latih dan data uji. Penelitian oleh Gholamy et al. (2018) membuktikan secara empiris bahwa perbandingan 70:30 adalah pembagian data uji dan data latih yang terbaik. Penelitian oleh Nguyen et al. (2021) tentang pengaruh *split* data pada *machine learning*, menunjukkan bahwa perbandingan 70:30 untuk data latih dan data uji menghasilkan akurasi terbaik. Penelitian lain

oleh Muraina (2022) menyatakan bahwa sebagian besar ahli menyarankan perbandingan 70:30 untuk *dataset* dengan ukuran kecil dari 100 sampai 1.000.000. Penelitian oleh Priyambodo et al. (2022) tentang prediksi waktu studi mahasiswa bidik misi dengan KNN, menunjukkan bahwa dengan beberapa persentase *split* data diperoleh akurasi terbaik dengan perbandingan *split* data 70:30.

Pada algoritma *fuzzy k-nearest neighbor* terdapat parameter k dan m . Parameter k adalah jumlah tetangga terdekat (data latih) dari data yang belum diketahui kelasnya (Han et al., 2012). Parameter m adalah nilai yang menentukan besarnya bobot ketika kontribusi setiap tetangga terhadap nilai keanggotaan dihitung (Keller et al., 1985). Pemilihan k dan m berpengaruh terhadap akurasi hasil klasifikasi. Penentuan k dan m dalam artikel ini dilakukan dengan algoritma *particle swarm optimization*. *Particle swarm optimization* (PSO) merupakan algoritma yang menjadikan populasi sebagai kandidat solusi (Engelbrecht, 2007). Algoritma ini dikembangkan berdasarkan perilaku sosial kawanan burung dengan konsep yang sangat sederhana dan dapat diterapkan melalui kode komputer (Kennedy & Eberhart, 1995). Beberapa penelitian yang telah dilakukan peneliti sebelumnya berkaitan dengan metode FKNN-PSO. Penelitian oleh Jamaluddin & Siringoringo (2017) menunjukkan bahwa akurasi hasil klasifikasi FKNN-Modified PSO lebih tinggi daripada FKNN dengan masing-masing akurasi sebesar 81% dan 71%. Penelitian lain oleh Ginting et al. (2020) dalam klasifikasi resiko kredit menunjukkan bahwa Modified PSO dapat mengoptimalkan FKNN sehingga akurasi hasil klasifikasinya sebesar 92,4%. Penelitian lain yang dilakukan oleh Kadry & Ismael (2020) menunjukkan akurasi hasil klasifikasi metode KNN, Modified KNN, dan Particle Optimized Score (POS) KNN dengan masing-masing nilai akurasinya yaitu, dan 83,94%, 85,56%, dan 88,59%. Penelitian lain yang dilakukan oleh Wijaya et al. (2023) dalam klasifikasi penyakit hati menunjukkan bahwa rata-rata akurasi hasil klasifikasi PSO-FKNN lebih tinggi daripada FKNN Hybrid Method yaitu sebesar 71,2% dan 66,8%.

Artikel ini akan membahas mengenai optimalisasi hasil klasifikasi *Fuzzy K-Nearest Neighbor* dengan *Particle Swarm Optimization*. PSO akan membantu pemilihan k dan m optimal dalam algoritma FKNN. Simulasi alur kerja algoritma akan ditunjukkan menggunakan *breast cancer dataset* dari *software python* (Rossum, 1980).

METODE

Teknik pengumpulan data yang digunakan adalah studi pustaka dengan mencari referensi dari buku, jurnal, skripsi yang berkaitan dengan topik penelitian. Data *wisconsin breast cancer* (kanker payudara) diperoleh dengan *load sklearn.dataset* di *software python*. Analisis data pada penelitian ini menggunakan algoritma FKNN-PSO dengan dimulai dari tahap *preprocessing* data, *processing* data, hingga evaluasi model. Berikut ini diuraikan keseluruhan langkah-langkah analisis data.

1. Melakukan normalisasi data dengan metode *min-max normalization* (Han et al., 2012).

$$v' = \frac{v - \min_A}{\max_A - \min_A} \quad (1)$$

Keterangan

- v : nilai data yang akan dinormalisasi
- v' : nilai data hasil normalisasi
- \min_A : nilai minimum atribut A
- \max_A : nilai maksimum atribut A

2. Memulai iterasi dari $t = 0$ dan melakukan inialisasi parameter awal antara lain jumlah partikel, swarm (S), posisi terbaik (P), dan kecepatan (v) (Parsopoulos & Vrahatis, 2010).

$$S = \{x_1, x_2, \dots, x_N\}; x_i = (x_{i1}, x_{i2}, \dots, x_{in})^T \in A, i = 1, 2, \dots, N \quad (2)$$

$$P = \{p_1, p_2, \dots, p_N\}; p_i = (p_{i1}, p_{i2}, \dots, p_{in})^T \in A, i = 1, 2, \dots, N \quad (3)$$

$$v_i = (v_{i1}, v_{i2}, \dots, v_{in})^T, i = 1, 2, \dots, N \quad (4)$$

3. Mensubstitusikan setiap nilai di S ke dalam algoritma *fuzzy k-nearest neighbor* untuk mencari *fitness value* yang berupa *error rate* hasil klasifikasi dengan langkah-langkah sebagai berikut (Keller et al., 1985).

- a. Membagi data latih dan data uji secara acak dengan perbandingan 70:30.

- b. Menghitung jarak *Euclidean* data uji ke setiap data latih.

$$d(x, y) = \sqrt{\sum_{k=1}^n (x_k - y_k)^2} \quad (5)$$

Keterangan:

$d(x, y)$: jarak *Euclidean* antara titik x dan y

n : jumlah dimensi

x, y : titik yang akan dihitung ukuran jaraknya

x_k, y_k : atribut (komponen) ke- k dari x dan y

- c. Mengurutkan data latih berdasarkan jarak terdekat, lalu menentukan k tetangga terdekat.

- d. Melakukan fuzzifikasi nilai keanggotaan k tetangga terdekat.

$$u_{ij}(x) = \begin{cases} 1, & x \in i \\ 0, & x \notin i \end{cases} \quad (6)$$

Keterangan:

i : kelas target

j : kelas data

$u_{ij}(x)$: nilai keanggotaan kelas i pada vektor j

- e. Menghitung derajat keanggotaan data uji terhadap masing-masing kelas.

$$u_i(x) = \frac{\sum_{j=1}^K u_{ij} \left(1 / \|x - x_j\|^{\frac{2}{(m-1)}} \right)}{\sum_{j=1}^K \left(1 / \|x - x_j\|^{\frac{2}{(m-1)}} \right)} \quad (7)$$

Keterangan:

x : data uji

x_j : data latih ke- j

$u_i(x)$: nilai keanggotaan data uji x terhadap kelas i

$x - x_j$: jarak data x ke data x_j dalam k tetangga terdekat

K : jumlah tetangga terdekat

m : bobot pangkat ($m > 1$).

- f. Melakukan klasifikasi data uji ke dalam kelas dengan nilai keanggotaan terbesar.

- g. Membuat *confusion matrix* untuk menghitung *error rate* hasil klasifikasi.

4. Menentukan $pbest$ dan $gbest$ (Parsopoulos & Vrahatis, 2010).

$$p_i(t) = \arg \min_t f_i(t) \tag{8}$$

$$p_g(t) = \arg \min_t p_i(t) \tag{9}$$

$$p_i(t+1) = \begin{cases} x_i(t+1), f(x_i(t+1)) \leq f(p_i(t)) \\ p_i(t+1), \text{ untuk yang lainnya} \end{cases} \tag{10}$$

Keterangan:

$p_i(t)$: $pbest$ pada iterasi ke- t

$p_g(t)$: $gbest$ pada iterasi ke- t

5. Memperbarui V dan S (Parsopoulos & Vrahatis, 2010).

$$v_{ij}(t+1) = wv_{ij}(t) + c_1R_1(p_{ij}(t) - x_{ij}(t)) + c_2R_2(p_{gj}(t) - x_{ij}(t)) \tag{11}$$

$$x_{ij}(t+1) = x_{ij}(t) + v_{ij}(t+1) \tag{12}$$

$$i = 1, 2, \dots, N; j = 1, 2, \dots, n; 0, 1 < w < 1, 2;$$

Keterangan:

$v_{ij}(t)$: kecepatan partikel i , dimensi j pada iterasi ke- t

$v_{ij}(t+1)$: kecepatan partikel i , dimensi j pada iterasi ke- $(t+1)$

$p_{ij}(t)$: *personal best* partikel i , dimensi j pada iterasi ke- t

$p_{gj}(t)$: *global best* dimensi j pada iterasi ke- t

$x_{ij}(t)$: posisi partikel i , dimensi j pada iterasi ke- t

$x_{ij}(t+1)$: posisi partikel i , dimensi j pada iterasi ke- $(t+1)$

w : bobot inersia

R_1, R_2 : bilangan *random* dari 0 sampai 1

c_1, c_2 : konstanta akselerasi

Pada proses iterasinya seringkali posisi partikel berada di luar ruang pencarian. Oleh karena itu dilakukan pembatasan kecepatan dan posisi sebagai berikut (Jain et al., 2022).

$$V_{max} = \delta(X_{max} - X_{min}) \tag{13}$$

$$X_{min} \leq S \leq X_{max}; -V_{max} \leq V \leq V_{max} \tag{14}$$

$$\delta \in (0, 1]$$

6. Mengulangi langkah 3, 4, 5, 6 sampai iterasi maksimal.

7. Melakukan evaluasi model dengan *confusion matrix* (Han et al., 2012).

Tabel 1. *Confusion Matrix*

Aktual	Prediksi		Total
	Ya	Tidak	
Ya	True Positive (TP)	False Negative (FN)	Positive
Tidak	False Positive (FP)	True Negative (TN)	Negative
Total	Positive	Negative	Positive +Negative

HASIL DAN PEMBAHASAN

Hasil

Sebelum memasuki proses analisis dan pengolahan data, berikut dijelaskan secara singkat mengenai data yang digunakan. Pada penelitian ini digunakan data yaitu *wisconsin breast cancer*. Banyak data adalah 569 dan banyak atribut adalah 30. Data terdiri dari dua kelas klasifikasi kanker yaitu kanker ganas (*malignant*) sebanyak 212 data dan kanker jinak (*benign*) sebanyak 357 data. Sebelum dilakukan perhitungan akurasi dengan algoritma FKNN-PSO, berikut diuraikan deskripsi data yang digunakan pada penelitian.

Tabel 2. Deskripsi *Wisconsin Breast Cancer Dataset*

Kelas	2
Banyak sampel setiap kelas	0 : 212 (M), 1: 357 (B)
Total sampel	569
Banyak atribut	30
Fitur	riil, positif
Tipe data	<i>Float, integer</i>

Tahapan algoritma ini secara garis besar sama seperti algoritma PSO dengan *fitness function* yang digunakan berupa algoritma FKNN. Dimulai dari *preprocessing* data berupa normalisasi *min-max*, *processing* data berupa algoritma PSO dengan nilai *fitness*-nya adalah *error rate* hasil klasifikasi FKNN, dan evaluasi model berupa *confusion matrix*. Tahapan selengkapnya diuraikan sebagai berikut.

1. *Preprocessing* data

Preprocessing data merupakan pengolahan data mentah sebelum diproses dalam tahapan selanjutnya. *Wisconsin breast cancer dataset* merupakan *dataset* siap pakai yang tidak memiliki *missing value*. Oleh karena itu, dalam tahapan ini tidak perlu dilakukan *cleaning* data. Pada penelitian ini, *preprocessing* data hanya berupa proses normalisasi data menggunakan metode *min-max*.

2. *Processing* data

Processing data merupakan serangkaian proses pengolahan data menggunakan algoritma FKNN-PSO. *Fitness function* yang digunakan dalam algoritma PSO berupa *error rate* hasil klasifikasi FKNN. Pada penelitian ini digunakan jumlah partikel sebanyak 10 dan iterasi maksimal sebanyak 50. *Swarm* (S) terdiri dari 10 pasangan k dan m yang diinisialisasi secara acak. Pada iterasi awal diasumsikan $S = P$, sedangkan kecepatan (v) ditetapkan bernilai 0.

$$S = \left\{ (3,2), (95,89), (85,87), (92,27), (86,44), \right. \\ \left. (25,15), (89,25), (91,80), (19,20), (25,21) \right\}$$

$$v = \{(0,0), (0,0), (0,0), (0,0), (0,0), (0,0), (0,0), (0,0), (0,0), (0,0)\}$$

Swarm (S) di atas disubstitusikan ke dalam algoritma FKNN untuk dicari *error rate* hasil klasifikasinya. Berdasarkan *error rate* tersebut akan ditentukan *pbest* dan *gbest* pada setiap iterasi. Adapun posisi partikel dan kecepatannya dibatasi dengan ketentuan sebagai berikut.

$$V_{max} = \delta(X_{max} - X_{min}) = 0,1(398 - 2) = 39,6$$

$$2 \leq S \leq 398 ; -39,6 \leq V \leq 39,6$$

Pada artikel ini dijalankan algoritma PSO dengan $w = 0,72$; $c_1 = c_2 = 2$. Berdasarkan percobaan tersebut diperoleh akurasi optimal sebesar 96,5% dengan $k=63$ dan $m=49$.

3. Evaluasi model

Setelah solusi optimal berhasil didapatkan, langkah selanjutnya adalah evaluasi model dengan *confusion matrix*. Hal ini bertujuan untuk membandingkan hasil kinerja model klasifikasi dengan data aktual yang ada. Solusi yang dihasilkan dari algoritma FKNN-PSO adalah $k = 63, m = 49$. *Confusion matrix* dari solusi tersebut disajikan sebagai berikut.

Tabel 3. *Confusion Matrix* Solusi Optimal

Aktual	Prediksi		Total
	0	1	
0	57	6	63
1	0	108	108
Total	57	114	171

Berdasarkan *confusion matrix* di atas diketahui bahwa terdapat 57 data kanker jinak (*benign*) diklasifikasikan dengan benar pada kelas kanker jinak (*benign*). Terdapat 108 data kanker ganas (*malignant*) diklasifikasikan dengan benar pada kelas kanker ganas (*malignant*). Terdapat 6 data kanker jinak (*benign*) diklasifikasikan dengan salah pada kelas kanker ganas (*malignant*). Tidak terdapat data kanker ganas (*malignant*) yang diklasifikasikan dengan salah pada kelas kanker jinak (*benign*).

Pembahasan

1. Penjelasan Evaluasi Model

Berdasarkan *confusion matrix* yang diperoleh pada hasil penelitian, maka diuraikan evaluasi model sebagai berikut.

<i>Error rate</i>	3,5%
Akurasi	96,5%
Presisi	M=100%
	B=94,7%
<i>Recall</i>	M=90%
	B=100%
<i>F1 Score</i>	M=95%
	B=97%

a. Akurasi (*Accuracy*)

Akurasi hasil klasifikasi *wisconsin breast cancer dataset* menggunakan algoritma FKNN-PSO adalah 96,5 %. Hal ini menunjukkan bahwa 96,5% data uji dapat diklasifikasikan dengan benar oleh model.

b. *Error Rate*

Error rate hasil klasifikasi *wisconsin breast cancer dataset* menggunakan algoritma FKNN-PSO adalah 3,5%. Hal ini menunjukkan bahwa 3,5% data uji tidak dapat diklasifikasikan dengan benar oleh model.

c. Presisi (*Precision*)

1) *Malignant* (kanker ganas)

Presisi kelas kanker ganas dari hasil klasifikasi *wisconsin breast cancer dataset* menggunakan algoritma FKNN-PSO adalah 1. Hal ini menunjukkan bahwa 100% data yang diprediksi sebagai kelas *malignant* merupakan hasil klasifikasi yang benar.

2) *Benign* (kanker jinak)

Presisi kelas kanker jinak dari hasil klasifikasi *wisconsin breast cancer dataset* menggunakan algoritma FKNN-PSO adalah 0,947368421. Hal ini menunjukkan bahwa 94,7% data yang diprediksi sebagai kelas *benign* merupakan hasil klasifikasi yang benar.

d. *Recall*

1) *Malignant* (kanker ganas)

Recall kelas kanker ganas dari hasil klasifikasi *wisconsin breast cancer dataset* menggunakan algoritma FKNN-PSO adalah 0,904761905. Hal ini menunjukkan bahwa 90% kelas aktual kanker ganas dapat diklasifikasikan dengan benar.

2) *Benign* (kanker jinak)

Recall kelas kanker jinak dari hasil klasifikasi *wisconsin breast cancer dataset* menggunakan algoritma FKNN-PSO adalah 1. Hal ini menunjukkan bahwa 100% kelas aktual kanker jinak dapat diklasifikasikan dengan benar.

e. *F1 Score*

1) *Malignant* (kanker ganas)

F1 Score kanker ganas dari hasil klasifikasi *wisconsin breast cancer dataset* menggunakan algoritma FKNN-PSO adalah 0,95. Hal ini menunjukkan bahwa nilai rata-rata harmonik presisi dan *recall* pada kelas kanker ganas sebesar 95%.

2) *Benign* (kanker jinak)

F1 Score kanker jinak dari hasil klasifikasi *wisconsin breast cancer dataset* menggunakan algoritma FKNN-PSO adalah 0,97. Hal ini menunjukkan bahwa nilai rata-rata harmonik presisi dan *recall* pada kelas kanker ganas sebesar 97%.

2. Kombinasi Parameter Algoritma PSO

Tabel 5. Hasil Akurasi FKNN-PSO

	w	c_1	c_2	k, m	Akurasi
Pengujian 1	0,72	2	2	63,49	96,5%
Pengujian 2	1	2	2	45,47	96,5%
Pengujian 3	0,5	1	1	46,29	96,5%
Pengujian 4	1	1	1	45,47	96,5%

Pada keempat pengujian yang telah dilakukan, diperoleh hasil akurasi optimal yang sama yaitu 96,5%. Namun, nilai k dan m berbeda-beda. Pengujian 1 diperoleh solusi $k = 63, m = 49$. Pengujian 2 diperoleh solusi $k = 45, m = 47$. Pengujian 3 diperoleh solusi $k = 46, m = 29$. Pengujian 4 diperoleh solusi $k = 45, m = 47$. Keempat pengujian tersebut menghasilkan kekonvergenan pengujian yang sama cepatnya.

3. Perbandingan Akurasi Algoritma FKNN dan KNN

Pada penelitian ini juga dilakukan perbandingan rata-rata akurasi antara metode FKNN dan KNN. Percobaan tersebut dilakukan dengan normalisasi maupun tanpa normalisasi. Hasil dari akurasi kedua metode tersebut disajikan sebagai berikut.

Tabel 6. Hasil Akurasi FKNN dan KNN

Normalisasi	FKNN	96%
	KNN	95,6%
Tanpa Normalisasi	FKNN	92,5%
	KNN	90,6%

SIMPULAN

FKNN-PSO merupakan algoritma PSO yang digunakan untuk mengoptimalkan akurasi hasil klasifikasi FKNN. Secara garis besar, tahapan dalam penelitian ini dibagi menjadi tiga yaitu *preprocessing* data, *processing* data, dan evaluasi model. Pada tahapan *preprocessing* data dilakukan normalisasi *min-max*. Selanjutnya, pada tahapan *processing* data dilakukan proses klasifikasi data sekaligus optimalisasi parameter k, m dengan algoritma FKNN-PSO. Terakhir, pada tahapan evaluasi model dilakukan perhitungan akurasi, *error rate*, presisi, *recall*, dan *F1 Score* dengan *confusion matrix*. Pada penelitian ini digunakan data berupa *wisconsin breast cancer*. Pada data tersebut dihasilkan akurasi sebesar 96,5% dengan $k = 63, m = 49$.

UCAPAN TERIMA KASIH

Terimakasih kepada koor Prodi Matematika dan seluruh Dosen Matematika yang telah memberikan ilmu hingga artikel ini dapat terselesaikan dengan baik.

DAFTAR PUSTAKA

- Ahmad, I., Sulistiyani, H., & Saputra, H. (2018). *Using Fuzzy K-Nearest Neighbor for Predicting University Students Graduation In Teknokrat*. 1(1), 47–52. <http://ejournal.uin-suska.ac.id/index.php/IJAIDM/index>
- Desinaini, L. N., Muallimah, A., Novitasari, D. C. R., & Hafiyusholeh, M. (2019). *Application of Fuzzy K-Nearest Neighbor (FKNN) To Detect the Parkinson's Disease*. 1(1), 8–16. <https://doi.org/10.15408/inprime.v1i1.12827>
- Ehsani, R., & Drablos, F. (2020). Robust Distance Measures for kNN Classification of Cancer Data. *Sage Journals*, 19, 1–9. <https://doi.org/10.1177/1176935120965542>
- Engelbrecht, A. P. (2007). *Computational Intelligence: An Introduction*. John Wiley & Sons, Ltd.
- Gholamy, A., Kreinovich, V., & Kosheleva, O. (2018). Why 70/30 or 80/20 Relation Between Training and Testing Sets: A Pedagogical Explanation. *Computer Science*, 1209, 1–6. https://scholarworks.utep.edu/cs_techrep/1209
- Ginting, W. C., Purba, R., & Arwin. (2020). *Optimization Performance of Fuzzy K-Nn with Modified Particle Swarm Optimization in Credit Risk Classification*. 4(2), 1417–1423.
- Han, J., Kamber, M., & Pei, J. (2012). *Data Mining Concepts and Techniques* (Third Edition). Morgan Kaufmann Publishers.
- Hidayatullaily, S., & Zulhadi. (2023). *Rainfall Classification Using Fuzzy K-Nearest Neighbor Method: Sumbawa Case Study*. 1(1), 01–10.
- Jain, M., Saihjal, V., Singh, N., & Singh, S. B. (2022). *An Overview of Variants and Advancements of PSO Algorithm*. 12(8392). <https://doi.org/10.3390/app12178392>
- Jamaluddin, & Siringoringo, R. (2017). Improved Fuzzy K-Nearest Neighbor Using Modified Particle Swarm Optimization. *IOP Publishing Ltd*. <https://doi.org/10.1088/1742-6596/930/1/012024>
- Kadry, R., & Ismael, O. (2020). *A New Hybrid KNN Classification Approach based on Particle Swarm Optimization*. 11(11), 291–296.
- Keller, J. M., Gray, M. R., & Givens, JR., J. A. (1985). A Fuzzy K-Nearest Neighbor Algorithm. *IEEE, SMC-15*(4).

- Kennedy, J., & Eberhart, R. C. (1995). Particle Swarm Optimization. *IEEE*, 4, 1942–1948. <https://doi.org/doi: 10.1109/ICNN.1995.488968>
- Kusuma, J., Kurniati, A. P., & Karo, I. M. K. (2022). *Analysis of Expertise Group Using The Fuzzy K-NN Classification Algorithm (Case Study: School of Computing Telkom University)*. 9(3), 564–572. <https://doi.org/10.30865/jurikom.v9i3.4215>
- Larose, D. T. (2005). *Discovering Knowledge in Data: An Introduction to Data Mining*. John Wiley & Sons, Inc.
- Muraina, I. O. (2022). *Ideal Dataset Splitting Ratios In Machine Learning Algorithms: General Concerns For Data Scientist And Data Analysts*. 496–504.
- Nasution, D. A., Khotimah, H. H., & Chamidah, N. (2019). *Perbandingan Normalisasi Data untuk Klasifikasi Wine Menggunakan Algoritma K-NN*. 4(1), 78. <https://doi.org/10.24114/cess.v4i1.11458>
- Nguyen, Q. H., Ly, H.-B., Ho, L. S., Al-Ansari, N., Le, H. V., Tran, V. Q., Prakash, I., & Pham, B. T. (2021). *Influence of Data Splitting on Performance of Machine Learning Models in Prediction of Shear Strength of Soil*. 1–15. <https://doi.org/10.1155/2021/4832864>
- Nurida, R., Sugiharti, E., & Alamsyah. (2019). *Implementation of Fuzzy K-Nearest Neighbor Method in Decision Support System for Identification of Under-five Children Nutritional Status Based on Anthropometry Index*. 1(1), 83–89. <https://journal.unnes.ac.id/sju/index.php/jaist>
- Pandey, A., & Jain, A. (2017). *Comparative Analysis of KNN Algorithm using Various Normalization Techniques*. 9(11), 36–42. <https://doi.org/10.5815/ijcnis.2017.11.04>
- Parsopoulos, K. E., & Vrahatis, M. N. (2010). *Particle Swarm Optimization and Intelligence: Advances and Applications*. Information Science Reference.
- Priyambodo, D., Nugroho, A., & Zaman, B. (2022). *Prediksi Ketepatan Waktu Studi Mahasiswa Bidik Misi Menggunakan K-Nearest Neighbour*. 5(2), 16–22.
- Ramadhani, F., Satria, A., & Sari, I. P. (2023). *Implementasi Metode Fuzzy K-Nearest Neighbor dalam Klasifikasi Penyakit Demam Berdarah*. 2(2), 58–62. <https://doi.org/10.56211/helloworld.v2i2.253>
- Rossum, G. van. (1980). *Python*.
- Shanti, D. M. F., Hidayat, N., & Wihandika, R. C. (2018). *Implementasi Metode F-KNN (Fuzzy K-Nearest Neighbor) Untuk Diagnosis Penyakit Anjing*. 2(12), 7401–7407.
- Sodiq, R. W. P., Hidayat, N., & Dewi, R. K. (2019). *Implementasi Metode Fuzzy K-Nearest Neighbor Untuk Identifikasi Cedera Pada Pemain Futsal*. 3(1), 1000–1004.
- Soleh, M., Andayati, D., & Rachmawati, Rr. Y. (2022). *Data Mining Model Klasifikasi Menggunakan Algoritma K-Nearest Neighbor dengan Normalisasi untuk Prediksi Penyakit Diabetes*. 18(2), 77–87. <https://doi.org/doi.org/10.36342/teika.v12i02.2911>
- Wijaya, I. G. B. S., Astuti, L. G., Suputra, I. P. G. H., Darmawan, I. D. M. B. A., Santiyasa, I. W., & Astawa, I. G. S. (2023). *Hybrid Implementation Fuzzy K-Nearest Neighbor (FK-NN) And Particle Swarm Optimization (PSO) Methods for Classification of Liver Disease*. 11(3), 635–644.