

Penerapan Particle Swarm Optimization Untuk Meningkatkan Kinerja Algoritma K-Nearest Neighbor Dalam Klasifikasi Penyakit Diabetes

Dyah susilowati^{1*}, Sutrisno², Muhammad Yunus³

¹Program Studi Pendidikan Teknologi Informasi, Universitas Bumigora
dyah.bumigora@gmail.com

²Program Studi Ilmu Komputer, Universitas Bumigora
.sutrisno@gmail.com

³Program Studi Manajemen Informasi Kesehatan, Politeknik Negeri Jember
m.yunus@poliije.ac.id

Keywords:

*Classification of Diabetes,
Particle Swarm
Optimization,
K-Nearest Neighbor*

ABSTRACT

The study aims to apply the KNN algorithm to classify diabetes by combining the PSO algorithm as a selection so as to obtain the best accuracy from the KNN algorithm in classifying, so that it can be applied to diagnose diabetes. This research consists of several stages of research including dataset collection, data pre-processing, data sharing, finding the optimal k-value to the classification process and performance or accuracy testing. From this research, the accuracy of the KNN algorithm before feature selection using PSO is 75% at k-optimal 19 and after feature selection using PSO is obtained an increase in accuracy to 77.213% at the same k value with features that affect the pima dataset, namely Glucose, Blood Pressure, Skin Thickness, Insulin, BMI, Diabetes Pedigree Function, and Age. Thus the use of KNN with PSO feature selection can be used to identify diabetes because it has a much better level of accuracy.

Kata Kunci

*Klasifikasi Diabetes
Particle Swarm
Optimization
K-Nearest Neighbor,*

ABSTRAK

Penelitian ini bertujuan untuk menerapkan algoritma KNN untuk melakukan klasifikasi penyakit diabetes dengan mengkombinasikan algoritma PSO sebagai seleksi sehingga memperoleh akurasi yang terbaik dari algoritma KNN dalam melakukan klasifikasi, sehingga dapat digunakan untuk melakukan identifikasi terhadap penyakit diabetes. Penelitian ini menggunakan tahapan data mining antara lain pengumpulan dataset, pre-processing data, klasifikasi Penyakit Diabetes pembagian data, pencarian nilai k-optimal hingga proses klasifikasi dan pengujian kinerja atau Particle Swarm Optimization akurasi. Dari penelitian ini diperoleh hasil akurasi algoritma KNN sebelum dilakukan seleksi fitur menggunakan PSO sebesar 75% pada k-optimal 19 dan setelah dilakukan seleksi fitur menggunakan PSO diperoleh peningkatan akurasi menjadi 77,213% pada nilai k=19 dan ditemukan atribut atau fitur yang berpengaruh pada dataset pima yaitu Glucose, Blood Pressure, Skin Thickness, Insulin, BMI, Diabetes Pedigree Function, dan age. Dengan demikian penggunaan KNN dengan seleksi fitur PSO dapat digunakan untuk melakukan identifikasi penyakit diabetes yang memiliki tingkat akurasi yang jauh lebih baik.

Korespondensi Penulis:

Dyah Susilowati,
Universitas Bumigora Mataram,
Jl. Ismail Marzuki No.22, Cilinaya, Kec. Cakranegara, Kota
Mataram, Nusa Tenggara Barat, 83127
Email: dyah.bumigora@gmail.com

**Submitted : 14-05-2023; Accepted : 25-06-2023;
Published : 30-06-2023**

Copyright (c) 2023 The Author (s)



This article is distributed under a Creative Commons Attribution-ShareAlike 4.0 International License (CC BY-SA 4.0)

1. PENDAHULUAN

Diabetes merupakan masalah yang menyebabkan kadar gula darah dalam tubuh manusia meningkat pada tingkat yang lebih tinggi secara berkelanjutan [1]. Selain itu diabetes juga menjadi penyebab terjadinya penyakit komplikasi, seperti penyakit stroke, ginjal, jantung koroner, gangguan pada mata dan saraf, bahkan hingga mengakibatkan amputasi dan kematian [2]. Menurut organisasi kesehatan dunia atau yang dikenal sebagai *World Health Organization* (WHO) diabetes membunuh 1,5 juta jiwa di tahun 2012. WHO juga memprediksi bahwa di tahun 2030 penyakit diabetes akan menjadi salah satu dari 7 faktor penyebab utama terjadinya kematian didunia [3]. Maka untuk mengurangi angka kematian tersebut, yaitu dengan melakukan deteksi secara dini terhadap penderita diabetes untuk mengantisipasi terjadinya komplikasi yang disebabkan diabetes yang berujung pada kematian.

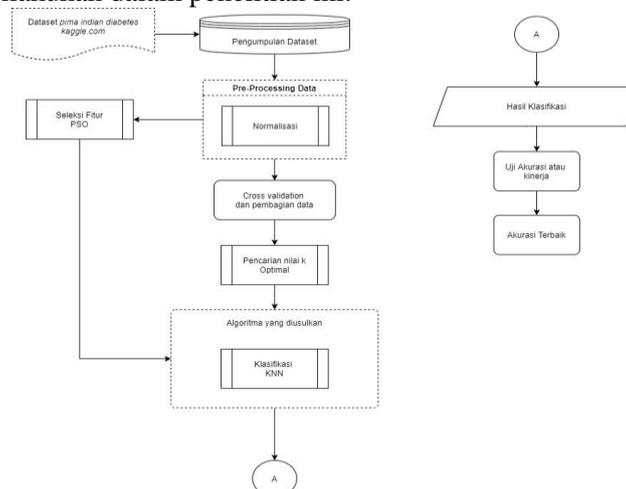
Data mining atau penambangan data bisa menjadi salah satu solusi atau metode yang bisa digunakan untuk melakukan prediksi atau identifikasi terhadap suatu masalah salah satunya adalah identifikasi penyakit diabetes [4]. Beberapa penelitian dibidang *data mining* juga sudah dilakukan untuk mengatasi permasalahan penyakit diabetes. Penelitian yang dilakukan [2], mengukur kinerja algoritma *k-Nearest Neighbor* (KNN) dalam menangani *dataset* penderita diabetes. Dari beberapa penelitian yang sudah dilakukan khususnya yang menerapkan algoritma KNN, masih memiliki kekurangan yaitu tidak menerapkan metode seleksi fitur untuk memilih atribut yang relevan yang dapat mengurangi waktu komputasi pada metode klasifikasi dan juga bisa mengoptimalkan nilai akurasi klasifikasi yang didapatkan. Dari uraian di atas, maka dalam penelitian ini penulis bertujuan untuk menerapkan *Particle Swarm Optimization* (PSO) untuk seleksi fitur guna meningkatkan kinerja algoritma KNN dalam klasifikasi penyakit diabetes. Dalam penelitian ini PSO yang digunakan dalam seleksi fitur merupakan pengembangan dari PSO yaitu *Binary Particle Swarm Optimization* (BPSO) untuk memperoleh kinerja yang jauh lebih baik dari algoritma KNN [5].

Algoritma KNN merupakan salah satu algoritma untuk melakukan klasifikasi dari suatu kumpulan data. Dimana algoritma KNN melakukan klasifikasi objek berdasarkan nilai k atau tetangga terdekat [6][7]. Sedangkan PSO merupakan salah satu algoritma kecerdasan *swarm* yang di kembangkan oleh Kennedy dan Eberhart pada tahun 1995. Proses pemikiran algoritma ini didorong oleh perilaku sosial dari kawanan burung atau kawanan ikan yang sejenis. Algoritma ini adalah salah satu algoritma evolusioner yang mirip dengan algoritma genetika dan menggunakan fungsi *fitness* untuk mengevaluasi kualitas dari solusi [8].

Pengkombinasian antara PSO dengan KNN dalam penelitian ini didasarkan pada beberapa penelitian yang menerapkan PSO dengan beberapa algoritma klasifikasi yang memperoleh peningkatan akurasi yang baik pada algoritma klasifikasi yang digunakan [9]. Penelitian bertujuan untuk mengetahui tingkat akurasi algoritma KNN dalam melakukan klasifikasi, dan untuk mengetahui tingkat akurasi algoritma KNN setelah diterapkan algoritma PSO sebagai seleksi fitur serta mengetahui fitur-fitur apa saja yang mempengaruhi hasil klasifikasi algoritma KNN.

2. METODE PENELITIAN

Penelitian ini terdiri dari beberapa tahapan penelitian yang dilakukan, berikut adalah tahapan-tahapan penelitian yang dilakukan dalam penelitian ini:



Gambar 1. Tahapan Penelitian

2.1 Pengumpulan Dataset

Data yang digunakan dalam penelitian ini adalah dataset *pima Indian diabetes* yang penulis peroleh dari situs *kagle.com* dimana data ini terdiri dari 768 data dan 8 atribut atau fitur. Atribut dataset pima ditunjukkan oleh Tabel 1 berikut : [10]

Tabel 1. Atribut dataset pima

Atribut	Keterangan	Label
<i>Pregnancies</i>	Angka kehamilan	X1
<i>Glucose</i>	Kadar glukosa 2 jam setelah makan. Menurut WHO salah satu yang menjadi kriteria penyakit diabetes, yaitu kadar glukosa minimal 200 mg/dl.	X2
<i>Blood Pressure</i>	Tekanan Darah	X3
<i>Skin Thickness</i>	Ketebalan Kulit	X4
Insulin	Insulin	X5
BMI	Berat Badan	X6
<i>Diabetes Pedigree Function</i>	Riwayat diabetes dalam keluarga	X7
<i>Age</i>	Umur	X8
<i>Outcome</i>	Status deabetes (1 = mengidap penyakit diabetes, 0 = tidak mengidap penyakit diabetes) .	Y

2.2 Pre-processing data

Tahapan pre-processing data yang dilakukan terdiri dari dua tahapan dimana pada tahapan pertama akan dilakukan tahapan normalisasi pada dataset pima yang digunakan, kemudian tahapan kedua akan dilakukan seleksi fitur menggunakan algoritma PSO. Namun tahapan seleksi fitur akan dilakukan setelah didapatkan nilai akurasi algoritma KNN sebelum dilakukan seleksi fitur sehingga akan diketahui perbedaan akurasi antara algoritma KNN sebelum dilakukan seleksi fitur menggunakan PSO dengan algoritma KNN setelah dilakukan seleksi fitur menggunakan PSO.

2.3 Pembagian Dataset

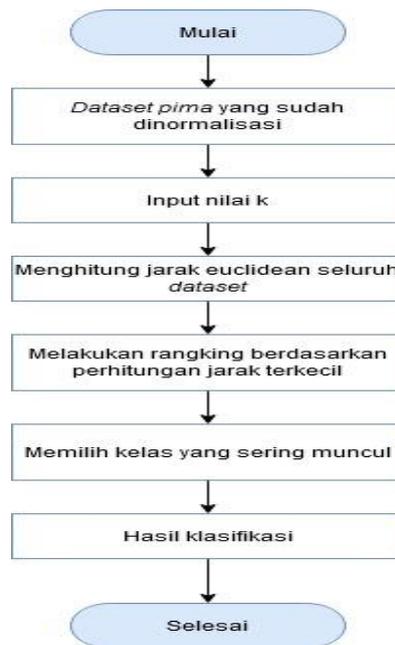
Akan dilakukan pembagian data menggunakan metode 10 fold cross validation untuk memperoleh hasil yang fit. Penggunaan $k\text{-fold}=10$ digunakan untuk menghilangkan bias pada data. Dalam $k\text{-fold}=10$, data akan dibagi menjadi 10 bagian atau fold atau 10 iterasi pengujian yang berukuran sama pada masing-masing fold, dimana semua data akan mendapat bagian menjadi data training dan data testing, sehingga memiliki 10 subset data untuk mengevaluasi kinerja algoritma klasifikasi yang digunakan.

2.4 Pengujian Nilai K Optimal

Setelah pembagian data dilakukan akan dilakukan tahapan pencarian nilai k optimal, karena dalam algoritma KNN penentuan nilai k sangat berpengaruh terhadap klasifikasi yang diperoleh, hal ini dikarenakan nilai k sangat sensitif terhadap hasil klasifikasi [11]. Oleh karena itu dalam penelitian ini akan dilakukan pengujian k dengan menggunakan range nilai 5 sampai dengan 30 untuk mendapatkan nilai k yang optimal.

2.5 Klasifikasi KNN

Pada tahap ini akan dilakukan proses klasifikasi menggunakan algoritma KNN dengan menggunakan nilai k optimal yang sudah diperoleh pada tahapan sebelumnya. Berikut adalah proses klasifikasi menggunakan algoritma KNN:



Gambar 2. Tahapan KNN

Berikut adalah contoh klasifikasi menggunakan algoritma KNN secara manual yang dilakukan, dimana data testing yang digunakan adalah dataset pima pada baris ke 15 dan akan dilakukan training terhadap 10 data pertama pada dataset pima, sebagai berikut:

Tabel 2. Dataset testing

No	X1	X2	X3	X4	X5	X6	X7	X8	Y
15	0.294	0.834	0.590	0.192	0.207	0.385	0.217	0.500	?

Saat ini nilai k yang akan digunakan sebagai uji coba adalah nilai k = 3, berikut adalah perhitungan euclidean pada setiap data antara data testing dengan data training.

$$\begin{aligned}
 (1,15) &= \sqrt{(0.353 - 0.294)^2 + (0.744 - 0.834)^2 + (0.590 - 0.590)^2 + (0.354 - 0.192)^2 + (0.000 - 0.207)^2 + (0.501 - 0.385)^2 + (0.234 - 0.217)^2 + (0.483 - 0.500)^2} \\
 &= \sqrt{0.09464} \\
 d &= 0.30763
 \end{aligned}$$

$$\begin{aligned}
 (2,15) &= \sqrt{(0.059 - 0.294)^2 + (0.427 - 0.834)^2 + (0.541 - 0.590)^2 + (0.293 - 0.192)^2 + (0.000 - 0.207)^2 + (0.501 - 0.385)^2 + (0.234 - 0.217)^2 + (0.167 - 0.500)^2} \\
 &= \sqrt{0.39786} \\
 d &= 0.63076
 \end{aligned}$$

Dari hasil perhitungan diperoleh jarak euclidean, dan dilakukan pranking jarak Euclidean dari yang terkecil ke terbesar dimana jarak Euclidean terkecil diberikan nilai pranking 1.

Tabel 3. Hasil perankingan nilai euclidean

Data	Jarak Euclidean Distance	Ranking	Y
1	0.30763	1	1
2	0.63076	6	0
3	0.47471	2	1
4	0.70826	8	0
5	0.94165	10	1
6	0.54102	3	0
7	0.68578	7	1
8	0.87864	9	0
9	0.59450	4	1

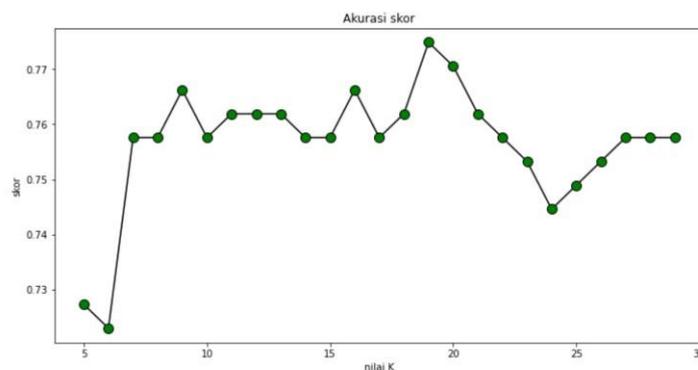
Berdasarkan pada Tabel 3 diatas maka diperoleh 3 jarak Euclidean terkecil atau terbaik yaitu pada data ke 1 dengan jarak Euclidean 0,30763, data ke 3 dengan jarak Euclidean 0,47471, dan data ke 6 dengan jarak Euclidean yaitu 0,54102. Dari ketiga jarak Euclidean terbaik maka selanjutnya akan dipilih kelas mayoritas, dengan kata lain yaitu kelas yang paling banyak muncul atau keluar. Kelas mayoritas yang muncul dan menjadi hasil klasifikasi terhadap dataset ke-15 adalah 1 yang merupakan kelas positif diabetes, sehingga dapat dinyatakan hasil klasifikasi terhadap data ke-15 adalah pasien positif diabetes dan kelas mayoritas yang keluar berada pada data ke-1 dan ke-3.

2.6 Pengujian Kinerja

Pada tahap ini akan dilakukan proses pengujian kinerja untuk mengetahui kinerja yang diperoleh pada algoritma KNN dalam melakukan klasifikasi penyakit diabetes berdasarkan pada dataset yang diperoleh. Selain itu pada tahap ini juga akan dilakukan pengujian kinerja untuk mengetahui peningkatan kinerja yang diperoleh algoritma KNN setelah diterapkan algoritma PSO sebagai seleksi fitur. Pengujian dilakukan dengan mengevaluasi perbandingan hasil Accuracy, diamana semakin tinggi nilai Accuracy yang diperoleh maka semakin baik pula metode atau algoritma yang digunakan.

3. HASIL DAN PEBAHASAN

Dari hasil pengujian pada nilai k optimal dengan range 5 sampai dengan 30 diperoleh nilai k optimal yaitu k=19 dengan perolehan skor akurasi sebesar 77,489%. Berikut adalah grafik perbandingan nilai k yang diperoleh :



Gambar 3. Grafik pengujian nilai k

Setelah nilai k optimal diperoleh, maka akan dilakukan tahapan pengujian klasifikasi menggunakan algoritma KNN terhadap dataset pima yang belum dilakukan normalisasi data. Dataset pima sebelum dilakukan normalisasi ditunjukkan oleh tabel 4 berikut :

Tabel 4. Dataset pima sebelum normalisasi

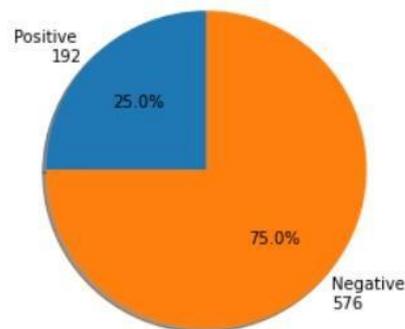
No	X1	X2	X3	X4	X5	X6	X7	X8	Y
1	6	148	72	35	0	33.6	0.627	50	1
2	1	85	66	29	0	26.6	0.351	31	0
3	8	183	64	0	0	23.3	0.672	32	1
4	1	89	66	23	94	28.1	0.167	21	0
5	0	137	40	35	168	43.1	2.288	33	1
...
767	1	126	60	0	0	30.1	0.349	47	1
768	1	93	70	31	0	30.4	0.315	23	0

Dari dataset pada tabel 3.1 tersebut dilakukan pengujian klasifikasi menggunakan algoritma KNN untuk mengetahui pengaruh normalisasi data pada hasil klasifikasi. Dari hasil pengujian yang dilakukan diperoleh akurasi klasifikasi pada algoritma KNN sebesar 74,359%. Pada tahapan berikutnya akan dilakukan tahapan pengujian yang sama namun dengan dataset yang sudah dilakukan pengolahan data menggunakan metode min-max normalization untuk mengetahui pengaruh min-max normalization terhadap hasil akurasi klasifikasi pada algoritma KNN. Berikut ini adalah dataset pima Indian diabetes setelah dilakukan tahapan normalisasi menggunakan metode min-max normalization.

Tabel 5. Dataset pima setelah normalisasi

No	X1	X2	X3	X4	X5	X6	X7	X8	Y
1	0.352941	0.743719	0.590164	0.353535	0.000000	0.500745	0.234415	0.483333	1
2	0.058824	0.427136	0.540984	0.292929	0.000000	0.396423	0.116567	0.166667	2
3	0.470588	0.919598	0.524590	0.000000	0.000000	0.347243	0.253629	0.183333	3
4	0.058824	0.447236	0.540984	0.232323	0.111111	0.418778	0.038002	0.000000	4
5	0.000000	0.688442	0.327869	0.353535	0.198582	0.642325	0.943638	0.200000	5
...
7	0.117647	0.613065	0.573770	0.272727	0.000000	0.548435	0.111870	0.100000	1
8	0.294118	0.608040	0.590164	0.232323	0.132388	0.390462	0.071307	0.150000	0

Dari dataset yang sudah dilakukan tahapan normalisasi menggunakan metode min-max normalization kemudian dilakukan klasifikasi menggunakan algoritma KNN dengan menggunakan nilai k optimal yang sudah diperoleh pada pengujian sebelumnya maka diperoleh hasil klasifikasi sebagai berikut:



768

Gambar 4. Hasil klasifikasi pada dataset pima

Berdasarkan dari Gambar 4 diperoleh hasil klasifikasi dari 768 data pada dataset pima terdapat 192 data yang terklasifikasi sebagai data positif diabetes dan 576 data terklasifikasi sebagai negatif diabetes. Berdasarkan klasifikasi yang dilakukan diperoleh confusion matrix untuk mengetahui nilai rata-rata akurasi algoritma KNN tanpa BPSO sebagai seleksi fitur dengan menggunakan pengujian model 10-fold cross validation sebagai berikut.

Tabel 6. Confusion matrix setelah normalisasi

Kelas	Klasifikasi Positif	Klasifikasi Negatif
Positif	442	58
Negatif	134	134

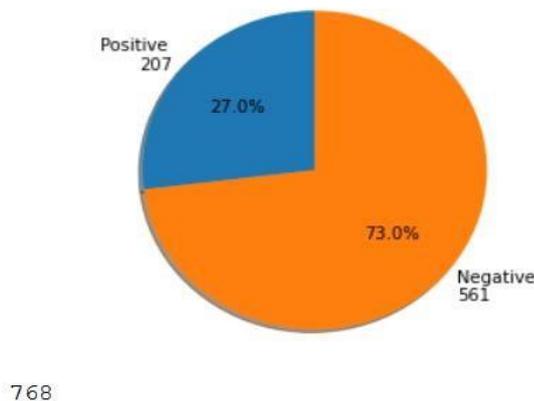
Berdasarkan Tabel 6 menunjukkan hasil algoritma KNN tanpa BPSO yang telah dikelompokkan menggunakan tabel confusion matrix. Dari tabel confusion matrix tersebut terlihat bahwa nilai True Positif (TP) dan True Negatif (TN) memiliki nilai yang tinggi hal ini bisa disebabkan karena adanya ketidak seimbangan pada dataset pima indian diabetes, dimana diketahui dari 768 data pada dataset pima indian diabetes terdapat 258 data yang dinyatakan positif diabetes dan 500 data dinyatakan negatif diabetes, dimana data negatif diabetes jauh lebih tinggi dibandingkan dengan data positif diabetes. Adapun perhitungan hasil confusion matrix manual dapat dirincikan sebagai berikut:

$$\begin{aligned}
 \text{Akurasi} &= \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \\
 &= \frac{442 + 134}{442 + 134 + 134 + 58} \times 100\% \\
 &= \frac{576}{768} \times 100\% \\
 &= 0.75 \times 100\% \\
 &= 75\%
 \end{aligned}$$

Setelah dilakukan normalisasi data menggunakan metode min-max normalization diperoleh peningkatan akurasi yang sebelumnya akurasi yang diperoleh algoritma KNN sebesar 74,739% kemudian menjadi 75%. Hal ini dikarenakan dataset pima setelah dilakukan pengolahan data menggunakan min-max normalization memiliki rentang nilai yang seimbang sehingga menghasilkan peningkatan akurasi yang diperoleh.

Tahapan pengujian BPSO-KNN merupakan tahapan untuk mengukur akurasi yang diperoleh algoritma KNN dalam melakukan klasifikasi pada dataset pima setelah dilakukan seleksi fitur menggunakan BPSO. Pada tahapan pengujian BPSO-KNN nilai k yang digunakan merupakan nilai k optimum yang sudah diperoleh dari pengujian sebelumnya yaitu k=19. Pada tahap ini dilakukan pengujian yang sama dengan tahapan pengujian sebelumnya yaitu dengan menggunakan metode 10-fold cross validation untuk memperoleh hasil yang fit.

Berikut hasil klasifikasi pada dataset pima indian diabetes setelah dilakukan seleksi fitur menggunakan BPSO.



Gambar 5. Hasil Klasifikasi BPSO-KNN

Gambar 5 merupakan hasil klasifikasi dengan algoritma KNN pada dataset pima setelah dilakukan seleksi fitur. Hasil klasifikasi menunjukkan terdapat 207 data terklasifikasi positif diabetes dan 561 data terklasifikasi negatif diabetes dari total keseluruhan dataset pima indian diabetes. Hasil ini memiliki perbedaan dengan hasil klasifikasi pada algoritma KNN sebelum dilakukan seleksi fitur.

Dari hasil klasifikasi tersebut diperoleh confusion matrix untuk mengetahui nilai rata-rata akurasi algoritma KNN setelah diterapkan BPSO sebagai seleksi fitur dengan menggunakan pengujian model 10-fold cross validation sebagai berikut.

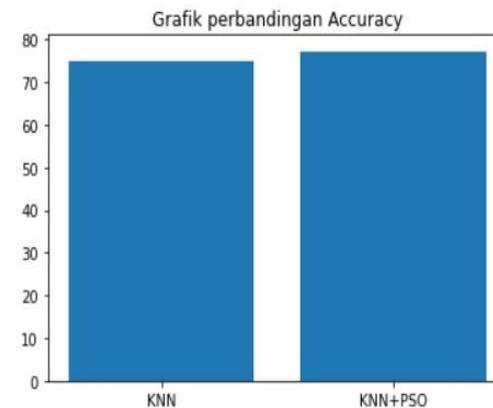
Tabel 7. Confusion Matrix BPSO-KNN

Kelas	Klasifikasi Positif	Klasifikasi Negatif
Positif	442	58
Negatif	134	134

Tabel 7 menunjukkan hasil algoritma KNN tanpa BPSO yang telah dikelompokkan menggunakan tabel confusion matrix. Dari tabel confusion matrix tersebut terlihat bahwa nilai True Positif (TP) dan True Negatif (TN) memiliki nilai yang tinggi yaitu 443 dan 150 hal ini bisa disebabkan karena adanya ketidak seimbangan pada dataset pima indian diabetes. Dari Tabel 4.10 diperoleh nilai akurasi sebagai berikut:

$$\begin{aligned}
 \text{Akurasi} &= \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \\
 &= \frac{443 + 150}{443 + 150 + 118 + 57} \times 100\% \\
 &= \frac{593}{768} \times 100\% \\
 &= 0.77213 \times 100\% \\
 &= 77,213 \%
 \end{aligned}$$

Dan Berdasarkan pengujian yang sudah dilakukan antara KNN tanpa BPSO dan KNN dengan BPSO sebagai seleksi fitur diperoleh perbandingan akurasi yang cukup signifikan. Berikut adalah grafik perbandingan perolehan akurasi antara algoritma KNN dengan BPSO-KNN dalam klasifikasi terhadap dataset pima indian diabetes yang ditunjukkan pada Gambar 6 :



Akurasi Score Knn = 75.000 %
Akurasi Score Knn + PSO : 77.214 %

Gambar 6. Grafik Perbandingan Akurasi antara KNN dengan BPSO-KNN

Berdasarkan Gambar 6 diatas, perolehan akurasi tertinggi diperoleh algoritma KNN setelah dilakukan seleksi fitur menggunakan BPSO dengan perolehan nilai akurasi sebesar 77,214% dibandingkan dengan menggunakan algoritma KNN saja tanpa seleksi fitur menggunakan BPSO. Hasil tertinggi yang diperoleh algoritma KNN setelah dilakukan seleksi fitur menggunakan BPSO membuktikan bahwa dengan adanya seleksi fitur BPSO untuk mengurangi dimensi data pada dataset yang digunakan dapat meningkatkan perolehan akurasi pada algoritma KNN dalam melakukan klasifikasi. Penelitian lain juga memperoleh hasil yang sama diantaranya adalah penelitian yang dilakukan oleh (Pradana & Aditsania, 2018) dan (Hasanuddin, 2016) yaitu dengan menerapkan seleksi fitur menggunakan BPSO atau PSO dapat meningkatkan akurasi pada algoritma klasifikasi yang diperoleh.

Seleksi fitur menggunakan PSO maupun BPSO, semata-mata hanya akan memilih atribut atau fitur-fitur yang memiliki tingkat akurasi atau nilai fitness yang terbaik atau posisi-posisi terbaik dari partikel, dengan kata lain atribut atau fitur yang dipilih mengacu pada nilai akurasi terbaik dari fitur [12]. Kemudian fitur tersebut digunakan kembali untuk menjadi acuan untuk mencari atau melakukan seleksi fitur Kembali. Sehingga bisa saja atribut atau fitur yang memiliki nilai korelasi rendah yang dapat terseleksi lebih banyak dibandingkan dengan atribut atau fitur dengan nilai korelasi yang tinggi, sehingga dapat meningkatkan nilai akurasi yang diperoleh algoritma KNN. Hal ini juga selaras dengan penelitian yang dilakukan oleh [7].

4. KESIMPULAN

Dari hasil pengujian sistem maka dapat disimpulkan bahwa:

- Berdasarkan hasil dan pembahasan pada proses klasifikasi pada dataset pima indian diabetes menggunakan algoritma KNN diperoleh hasil klasifikasi 192 data terklasifikasi positif diabetes dan 576 data terklasifikasi negatif diabetes dan diperoleh tingkat akurasi atau tingkat korelasi antara nilai prediksi dari sistem dengan nilai sesungguhnya yang mencapai 75%.
- Diperoleh fitur-fitur pada dataset pima indian diabetes yang berpengaruh terhadap hasil klasifikasi pada algoritma KNN yaitu Glucose (X2), Blood Pressure (X3), Skin Thickness (X4), Insulin (X5), BMI (X6), Diabetes Pedigree Function (X7), dan Age (X8).
- Dari hasil pengujian menggunakan confusion matrix diperoleh peningkatan akurasi klasifikasi pada algoritma KNN setelah diterapkan algoritma BPSO untuk seleksi fitur menjadi 77,214% sehingga tampak selisih akurasi yang diperoleh yaitu sebesar 2,214%. Hal tersebut dikarenakan

BPSO hanya memilih fitur-fitur yang dapat menghasilkan akurasi klasifikasi data yang tinggi. Hal ini menunjukkan bahwa dengan adanya BPSO sebagai seleksi fitur, dapat meningkatkan perolehan akurasi pada algoritma KNN dalam melakukan klasifikasi. Dengan demikian penggunaan KNN dengan seleksi fitur BPSO dapat digunakan untuk melakukan identifikasi penyakit diabetes karena memiliki tingkat akurasi yang jauh lebih baik dibandingkan dengan menggunakan algoritma KNN saja.

UCAPAN TERIMA KASIH

Penulis menyampaikan penghargaan dan ucapan terima kasih yang sebesar-besarnya kepada pihak yang telah memberikan kontribusi didalam penelitian ini, khususnya Universitas Bumigora Mataram.

REFERENSI

- [1] D. K. Choubey, P. Kumar, S. Tripathi, and S. Kumar, "Performance evaluation of classification methods with PCA and PSO for diabetes," *Netw. Model. Anal. Heal. Informatics Bioinforma.*, vol. 9, no. 1, pp. 1–30, Dec. 2020, doi: 10.1007/S13721-019-0210-8/METRICS.
- [2] A. M. Argina, "Penerapan Metode Klasifikasi K-Nearest Neighbor pada Dataset Penderita Penyakit Diabetes," *Indones. J. Data Sci.*, vol. 1, no. 2, pp. 29–33, Jul. 2020, doi: 10.33096/IJODAS.V1I2.11.
- [3] Kementerian Kesehatan RI., "Infodatin tetap produktif, cegah, dan atasi Diabetes Melitus 2020," *Pusat Data dan Informasi Kementerian Kesehatan RI*. pp. 1–10, 2020.
- [4] H. Hairani, G. S. Nugraha, M. N. Abdillah, and M. Innuddin, "Komparasi Akurasi Metode Correlated Naive Bayes Classifier dan Naive Bayes Classifier untuk Diagnosis Penyakit Diabetes," *InfoTekJar J. Nas. Inform. dan Teknol. Jar.*, vol. 3, no. 1, pp. 6–11, Sep. 2018, doi: 10.30743/INFOTEKJAR.V3I1.558.
- [5] D. Y. Shailendra, "Seleksi Fitur Menggunakan Metode Hybrid Particle Swarm Optimization dengan Operasi Local Search (HPSO-LS) untuk Klasifikasi Data," 2017.
- [6] M. Yunus, N. Kadek, and A. Pratiwi, "Prediksi Status Gizi Balita Dengan Algoritma K-Nearest Neighbor (KNN) di Puskesmas Cakranegara," *JTIM J. Teknol. Inf. dan Multimed.*, vol. 4, no. 4, pp. 221–231, Feb. 2023, doi: 10.35746/JTIM.V4I4.328.
- [7] K. Saxena, Z. Khan, S. Singh, M.-T. Research, and S. 1&3, "Diagnosis of Diabetes Mellitus using K Nearest Neighbor Algorithm," *Int. J. Comput. Sci. Trends Technol.*, vol. 2, Accessed: Jun. 30, 2023. [Online]. Available: www.ijcstjournal.org
- [8] I. Lina, P. Politeknik, and B. Jepara, "Implementasi Algoritma Particle Swarm Optimization (PSO) dan K-Nearest Neighbor (K-NN) Dalam Memprediksi Keberhasilan Anak SMK Mendapatkan Kerja," *Technol. J. Ilm.*, vol. 13, no. 4, pp. 339–350, Oct. 2022, Accessed: Jun. 30, 2023. [Online]. Available: <https://ojs.uniska-bjm.ac.id/index.php/JIT/article/view/8167>
- [9] O. Findik, I. Babaoğlu, and E. Ülker, "A color image watermarking scheme based on hybrid classification method: Particle swarm optimization and k-nearest neighbor algorithm," *Opt. Commun.*, vol. 283, no. 24, pp. 4916–4922, Dec. 2010, doi: 10.1016/J.OPTCOM.2010.07.020.
- [10] U. M. LEARNING, "Pima Indians Diabetes Database | Kaggle." <https://www.kaggle.com/datasets/uciml/pima-indians-diabetes-database> (accessed Jun. 30, 2023).
- [11] S. Ulya, M. Arief Soeleman, F. Budiman, and M. Teknik Informatika, "Optimasi Parameter K Pada Algoritma K-NN Untuk Klasifikasi Prioritas Bantuan Pembangunan Desa," *Techno.Com*, vol. 20, no. 1, pp. 83–96, Feb. 2021, doi: 10.33633/tc.v20i1.4215.
- [12] Y. Xue, T. Tang, W. Pang, and A. X. Liu, "Self-adaptive parameter and strategy based particle swarm optimization for large-scale feature selection problems with multiple classifiers," *Appl. Soft Comput.*, vol. 88, p. 106031, Mar. 2020, doi: 10.1016/J.ASOC.2019.106031.