

## Optimasi K-Nearest Neighbor Dengan Particle Swarm Optimization Untuk Klasifikasi *Idiopathic Thrombocytopenic Purpura*

Roudlotul Jannah Alfirdausy<sup>1\*</sup>, Izzatul Aliyyah<sup>2</sup>, Aris Fanani<sup>3</sup>

<sup>1,2,3</sup>Program Studi Matematika, Fakultas Sains dan Teknologi, UIN Sunan Ampel Surabaya  
Gn. Anyar, Kec. Gn. Anyar, Surabaya, Jawa Timur 60294

<sup>1\*</sup>09030220048@student.uinsby.ac.id

<sup>2</sup>09040220057@student.uinsby.ac.id, <sup>3</sup>arisfanani@uinsby.ac.id

(Naskah masuk: 23 Juli 2023; diterima untuk diterbitkan: 31 Oktober 2023)

**ABSTRAK** – *Immune Thrombocytopenic Purpura (ITP)* dianggap sebagai gangguan hematologi yang disebabkan oleh kerusakan pada trombosit akibat respons autoimun tubuh, yang mengakibatkan kemudahan terjadinya memar atau pendarahan yang berlebihan pada individu yang terkena. Pentingnya deteksi dini penyakit ITP tidak dapat diabaikan, karena kelainan ini dapat berdampak kronis atau jangka panjang. Oleh karena itu, penelitian ini bertujuan untuk mengklasifikasikan penyakit ITP dengan akurasi yang lebih baik guna menghindari kesalahan dalam diagnosis pasien, serta memungkinkan penanganan dan pengobatan yang tepat dan segera. Dalam penelitian ini, digunakan metode kombinasi PSO-KNN. Hasil yang diperoleh menunjukkan peningkatan yang signifikan dalam akurasi, sensitivitas, dan spesifisitas dibandingkan dengan metode KNN standar. Akurasi mencapai 91.8%, meningkat sebesar 4.9% dari KNN standar, sensitivitas mencapai 91.2%, meningkat sebesar 11.8% dari KNN standar, dan spesifisitas mencapai 92.6%, walaupun mengalami penurunan sebesar 3.7% dari KNN standar. Waktu pelatihan dan pengujian dengan metode PSO-KNN juga lebih efisien dibandingkan dengan KNN standar. Hal ini menunjukkan bahwa penggunaan algoritma PSO mampu mengoptimalkan hasil klasifikasi dari KNN, sehingga dapat menjadi alat yang lebih andal dalam diagnosis dini penyakit ITP.

**Kata Kunci** – *Immune Thrombocytopenic Purpura; Klasifikasi; KNN; PSO.*

## K-Nearest Neighbor Optimization With Particle Swarm Optimization For *Idiopathic Thrombocytopenic Purpura* Classification

**ABSTRACT** – *Immune Thrombocytopenic Purpura (ITP)* is a hematological disorder caused by damage to platelets due to the body's autoimmune response, which results in easy bruising or excessive bleeding in affected individuals. The importance of early detection of ITP cannot be ignored, because this disorder can have chronic or long-term impacts. Therefore, this study aims to classify ITP disease with better accuracy in order to avoid errors in patient diagnosis, and enable appropriate and immediate treatment and treatment. In this research, the PSO-KNN combination method was used. The results obtained show significant improvements in accuracy, sensitivity, and specificity compared to standard KNN methods. Accuracy reached 91.8%, an increase of 4.9% from standard KNN, sensitivity reached 91.2%, an increase of 11.8% from standard KNN, and specificity reached 92.6%, although it decreased by 3.7% from standard KNN. Training and testing time with the PSO-KNN method is also more efficient compared to standard KNN. This shows that the use of the PSO algorithm is able to optimize the classification results from KNN, so that it can become a more reliable tool in early diagnosis of ITP disease.

**Keywords** – *Immune Thrombocytopenic Purpura; Classification; KNN; PSO.*

## 1. PENDAHULUAN

ITP (*Immune Thrombocytopenic Purpura*) merupakan kondisi yang terjadi pada sistem darah dimana trombosit mengalami kerusakan yang disebabkan oleh respon autoimun. Salah satu ciri utama ITP adalah penurunan kadar trombosit dalam darah kurang dari  $100 \times 10^9$  per liter tanpa adanya penyebab lain sehingga menyebabkan trombositopenia. ITP dapat terjadi akibat penghancuran kompleks trombosit-antibodi dalam sistem retikuloendotelial yang melebihi kapasitas produksi sumsum tulang [1]. Kondisi ini bisa menyebabkan seseorang mudah memar atau berdarah secara berlebihan dikarenakan tingkat trombosit yang rendah [2].

ITP dapat menyerang anak-anak ataupun orang dewasa. Kasus ITP pada anak-anak, seringkali terjadi setelah mereka mengalami infeksi virus dan dapat sembuh sepenuhnya tanpa pengobatan khusus [3]. Pada orang dewasa, angka kejadian ITP berkisar antara 1.6 hingga 6.6 kasus per 100.000 orang tiap tahun. Pada populasi dewasa, ITP cenderung menjadi kondisi kronis atau jangka panjang. Sebagai contoh, insiden tahunan ITP di Eropa Utara mencapai 2.68 kasus per 100.000 orang [4].

Beberapa penelitian yang terkait dengan penyakit ITP seperti penelitian oleh [5] memperoleh hasil penyakit ITP pada ibu hamil sangat berisiko melahirkan bayi dengan perdarahan di otak. Kemudian penelitian oleh [2] mendapatkan hasil diagnosa penyakit ITP dapat menggunakan metode sistem pakar yaitu *Variable Centered Intelligent Rule System (VCIRS)*. Selanjutnya penelitian oleh [3] dengan hasil terapi untuk penyakit ITP terdiri dari berbagai tahap pengobatan yang disesuaikan dengan kondisi klinis pasien, dengan tujuan meningkatkan kualitas hidup mereka.

Dalam penelitian ini, dilakukan klasifikasi penyakit ITP dari dua kelas antara kelas normal dan kelas penyakit ITP. Metode *self-classification* digunakan untuk mengorganisasi data dengan membagi objek-objek ke dalam kelas-kelas yang berbeda sesuai dengan jumlah yang diinginkan. Aturan-aturan yang telah ditetapkan dibuat untuk menjadi pembeda satu kelas dari yang lainnya dan menentukan objek mana yang termasuk dalam kelas tertentu berdasarkan perilaku dan atribut yang telah ditentukan [6]. Tujuan dari penelitian ini untuk mengetahui jumlah penderita kasus ITP dan tidak serta dapat mengenali orang yang menderita ITP agar dapat segera dilakukan penanganan melalui anamnesis terkait riwayat penyakit dan pengobatannya dengan tenaga kesehatan sehingga dapat menghindari kesalahan dalam diagnosis maupun penanganan terhadap pasien penderita ITP.

Penelitian ini menggunakan algoritma *K-*

*Nearest Neighbor (KNN)* untuk mengklasifikasikan data. KNN adalah metode klasifikasi yang fokus pada pencarian jarak terpendek dari data yang sedang dianalisis dan K data terdekat dari data *training*. [7]. Metode ini didasarkan pada konsep jarak terdekat di mana jarak antara data *training* dan data *testing* dihitung dengan metode *Euclidean* [8]. Selain itu, untuk meningkatkan hasil penelitian agar lebih optimal, digunakan pula metode *Particle Swarm Optimization (PSO)*. PSO adalah teknik pengoptimasi yang sederhana dan bisa diterapkan dengan mengubah beberapa parameter. [9]. Dalam PSO, dilakukan iterasi sampel untuk mencari posisi yang menghasilkan nilai fungsi minimal. Teknik optimasi PSO ini menggunakan titik acuan dari kualitas sebagai panduan dalam mencari solusi yang optimal [10]. PSO adalah suatu teknik yang digunakan untuk memaksimalkan hasil dengan kemampuan untuk mengatur pemilihan subset guna mencapai tingkat akurasi yang optimal [11].

Penelitian yang menggunakan kombinasi dua metode, seperti *K-Nearest Neighbor (KNN)* dan *Particle Swarm Optimization (PSO)*, dalam konteks pengenalan citra batik dengan ragam hias geometris, telah membuktikan potensi besar dalam meningkatkan akurasi hasil. Hasil dari penelitian tersebut menunjukkan bahwa penggabungan metode PSO ke dalam algoritma KNN dapat menghasilkan peningkatan signifikan dalam akurasi klasifikasi. Dalam kasus ini, terlihat bahwa penggunaan metode KNN-PSO berhasil meningkatkan akurasi sebesar 6% dibandingkan dengan penggunaan algoritma KNN standar. Akurasi pada KNN standar sebesar 83%, sedangkan setelah mengintegrasikan metode PSO, akurasi meningkat menjadi 89% [12].

Penelitian lainnya terkait upaya untuk meningkatkan akurasi algoritma *K-Nearest Neighbor (KNN)* dalam memprediksi potensi pelanggan yang akan berhenti menggunakan layanan atau *customer churn*. Dalam rangka mencapai hal ini, dua metode utama digunakan, yaitu normalisasi data dengan *Z-score* dan *Particle Swarm Optimization (PSO)* untuk mengoptimalkan nilai K yang paling sesuai. Normalisasi data dengan *Z-score* adalah langkah penting dalam memastikan setiap fitur dalam dataset memiliki skala yang seragam, sehingga mencegah pengaruh outlier atau perbedaan skala pada hasil klasifikasi. Sementara itu, PSO digunakan untuk mengatur parameter K dalam algoritma KNN, yang sangat penting dalam menentukan jumlah tetangga terdekat yang digunakan dalam klasifikasi. Hasil dari penelitian ini menunjukkan peningkatan akurasi yang signifikan sebesar 14% setelah menerapkan normalisasi *Z-Score* dan PSO. Dari akurasi awal sebesar 68,5%, akurasi model meningkat menjadi 82,5% setelah proses normalisasi dan optimasi. Hal

ini menegaskan bahwa pendekatan ini berhasil dalam meningkatkan kinerja algoritma KNN dalam memprediksi *customer churn*[13].

Peningkatan sebesar 6% ini menunjukkan bahwa pendekatan KNN-PSO mampu mengoptimalkan proses klasifikasi citra batik dengan lebih baik. Hal ini mengindikasikan bahwa penggunaan PSO membantu dalam mengatur parameter-parameter dalam KNN secara lebih efisien, sehingga memungkinkan model untuk lebih baik dalam mengenali dan mengklasifikasikan citra batik dengan ragam hias geometris. Hasil ini memberikan kontribusi penting dalam pengembangan teknik pengenalan citra dan membuktikan bahwa pendekatan kombinasi antara KNN dan PSO dapat meningkatkan kinerja algoritma klasifikasi secara signifikan.

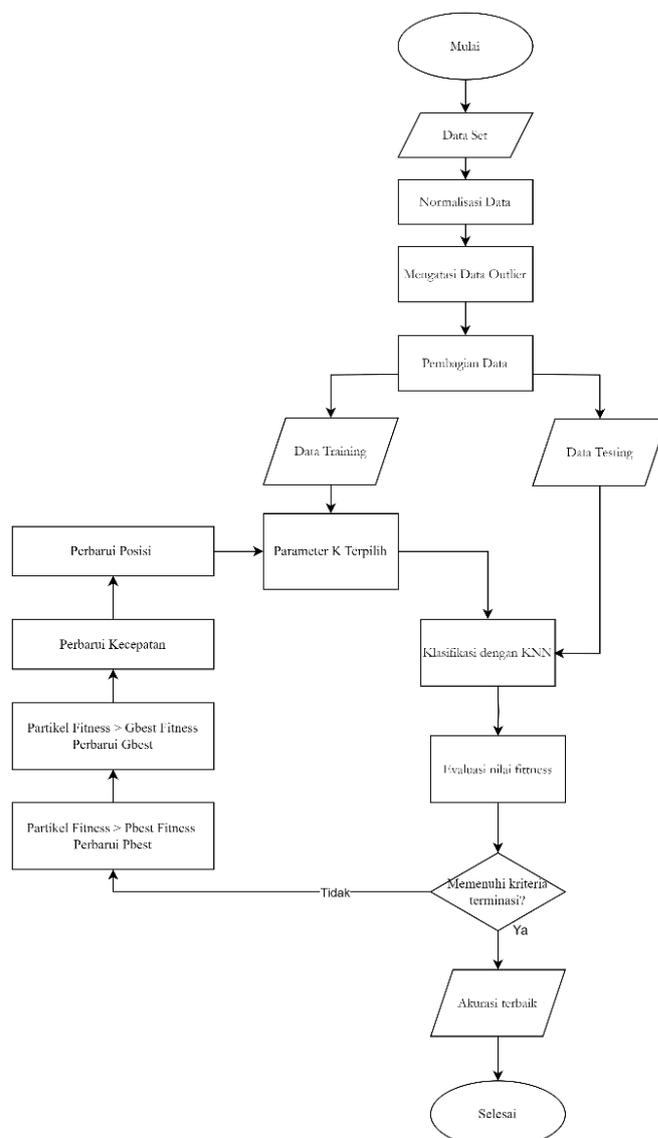
Penelitian yang dilakukan pada *K-Nearest Neighbor* (KNN) Berbasis *Particle Swarm Optimization* (PSO) untuk menganalisis sentimen terhadap Tokopedia menunjukkan hasil yang cukup baik. Dalam penelitian ini, akurasi mencapai 97.9%, *precision* (presisi) mencapai 96.17%, *recall* mencapai 96.62%, dan *f-measure* mencapai 96.39%. Hasil ini menggambarkan bahwa integrasi PSO dalam algoritma KNN telah berhasil meningkatkan performa klasifikasi secara signifikan. Tingginya akurasi dan metrik-metrik evaluasi lainnya menunjukkan bahwa model ini mampu mengenali dan mengklasifikasikan sentimen terhadap Tokopedia dengan tingkat ketepatan yang sangat tinggi. Selain itu, dalam penelitian ini juga terdapat bukti konkrit tentang peningkatan kinerja metode KNN setelah diberikan optimasi metode PSO. Hal ini dibuktikan dengan perubahan nilai kesalahan (*Root Mean Square Error*/RMSE) dari 0.093 menjadi 0.072 setelah proses optimasi dilakukan. Penurunan RMSE ini mengindikasikan bahwa hasil prediksi menjadi lebih mendekati nilai sebenarnya setelah penggunaan PSO. Ini menegaskan bahwa pendekatan PSO tidak hanya meningkatkan akurasi dalam analisis sentimen terhadap Tokopedia tetapi juga membuat model menjadi lebih presisi dalam memprediksi sentimen pelanggan [14].

Berdasarkan dari penelitian di atas dan penjelasan yang ada, maka penelitian ini menggunakan algoritma *K-Nearest Neighbor* (KNN) dengan *Particle Swarm Optimization* (PSO) untuk mengklasifikasi penyakit *Immune Thrombocytopenic Purpura* (ITP) agar menghindari kesalahan dalam diagnosis dan penderita dapat segera melakukan anamnesis serta mendapat penanganan dan pengobatan dari tenaga kesehatan.

## 2. METODE DAN BAHAN

Metode penelitian ini dibuat sesuai dengan

proses dari metode KNN berbasis PSO dengan tahap awal adalah mengumpulkan dataset penelitian. Tahap selanjutnya adalah klasifikasi yang dilakukan penggabungan metode antara KNN dengan PSO, dan tahap terakhir adalah mengukur akurasi hasil penelitian menggunakan *confusion matrix*. Diagram penelitian dapat dilihat pada Gambar 1.



Gambar 1. Diagram Alur PSO-KNN

Diagram alur tersebut menjelaskan tahapan dalam penelitian dengan deskripsi sebagai berikut.

**Tahap 1** : Subjek atau bahan penelitian dalam studi ini adalah kumpulan data baru tentang *Immune Thrombocytopenic Purpura* (ITP), yang diperoleh dari website Kaggle. Dataset ini terdiri dari total 357 data pasien, yang telah dikumpulkan dengan informasi yang beragam. Dari seluruh dataset tersebut, sebanyak 179 data teridentifikasi sebagai pasien yang menderita penyakit ITP, sementara 178 data lainnya teridentifikasi sebagai pasien dengan kondisi normal. Data-data ini mencakup beberapa parameter

penting, seperti jenis kelamin (Gender) terdiri dari 0 = perempuan dan 1 = laki-laki, usia (Umur dalam tahun), kadar hemoglobin (HB dalam bentuk g/dl), hasil dari *Thin Layer Chromatography* (TLC), *Mean Corpuscular Volume* (MCV) yaitu ukuran rata-rata sel darah merah, *Mean Corpuscular Hemoglobin* (MCH) dalam satu eritrosit, status vaksin *Pneumococcal Conjugate Vaccine* (PCV), dan jumlah trombosit atau platelet dalam darah. Klasifikasi pasien dalam dataset ini mengacu pada dua kategori utama, yaitu Kelas 0 yang mengindikasikan kondisi normal dan Kelas 1 yang mengindikasikan pasien positif ITP.

**Tahap 2:** Proses *preprocessing* atau praproses, adalah tahap kunci dalam data mining yang memiliki peran yang sangat penting dalam membantu mempersiapkan data sebelum dilakukan analisis lebih lanjut. Salah satu teknik *preprocessing* yang sering digunakan adalah normalisasi data. Normalisasi adalah proses yang bertujuan untuk mengubah skala atau rentang nilai dari setiap fitur dalam dataset sehingga memiliki skala yang seragam. Hal ini penting karena dalam dataset yang memiliki fitur-fitur dengan skala yang berbeda, algoritma data mining seperti clustering, klasifikasi, atau regresi dapat dipengaruhi oleh perbedaan skala ini. Normalisasi membantu menghindari bias yang tidak diinginkan yang mungkin muncul akibat perbedaan skala, sehingga memungkinkan algoritma untuk memberikan hasil yang lebih akurat. Proses normalisasi dapat dilakukan dengan *Z-Score* atau biasa disebut *Standard Scalar* [15].

$$X_{new} = \frac{X_{old} - \mu}{\sigma} \quad (1)$$

**Tahap 3:** Setelah proses normalisasi data selesai, tahap selanjutnya dalam analisis data adalah mendeteksi adanya outlier dalam data univariat. Outlier adalah nilai-nilai yang signifikan berbeda dari sebagian besar data dalam dataset dan dapat mempengaruhi hasil analisis statistik dan pemodelan. Deteksi outlier dapat dilakukan dengan mendefinisikan ambang batas yang dapat mengklasifikasikan data sebagai outlier [16]. Salah satu metode umum yang digunakan adalah dengan memanfaatkan konsep kuartil dan rentang interkuartil (IQR) dari data tersebut. Data dianggap sebagai outlier jika nilainya kurang dari Q1 dikurangi 1.5 kali IQR atau lebih besar dari Q3 ditambah 1.5 kali IQR. Nilai IQR sendiri adalah selisih antara kuartil ketiga (Q3) dengan kuartil pertama (Q1). Pendekatan ini membantu mengidentifikasi dan mengisolasi nilai-nilai yang berpotensi memengaruhi hasil analisis secara tidak wajar, sehingga memastikan bahwa data yang digunakan dalam analisis statistik dan pemodelan adalah data yang lebih konsisten [17].

**Tahap 4:** Metode klasifikasi yang digunakan pada penelitian ini adalah metode gabungan antara

KNN dan PSO yang dapat disebut sebagai metode PSO-KNN. *Particle Swarm Optimization* (PSO) adalah metode dalam sistem swarm intelligence yang digunakan dalam mendapatkan solusi optimasi dengan mencari penyelesaian di ruang pencarian. Dalam PSO, kita mengikuti konsep swarm, di mana sejumlah partikel (atau agen) bergerak dalam ruang pencarian, dan mereka berkomunikasi satu sama lain untuk mencari solusi yang paling optimal. Metode PSO-KNN menggabungkan keunggulan PSO dalam mencari solusi optimal dengan algoritma KNN, yang merupakan metode klasifikasi berdasarkan jarak antara data yang ingin diklasifikasikan dengan data pelatihan [18]. Dalam KNN, mencari sekelompok  $n$  objek dalam data training yang memiliki kemiripan atau kedekatan terbesar dengan data dalam data baru atau data testing [19]. Berikut adalah tahap-tahap klasifikasi menggunakan metode KNN [20]:

1. Tentukan nilai  $K$ .
2. Hitung jarak *Euclidean* dari data pengujian dan data pelatihan.  

$$\text{dist}(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (2)$$
3. Kelompokkan data menurut hasil jarak *Euclidean*.
4. Kelompokkan data menurut nilai terendah atau tetangga yang paling dekat.
5. Pilih kelas dengan probabilitas terbanyak dari nilai  $K$  yang digunakan untuk hasil prediksi.

**Tahap 5:** Evaluasi hasil klasifikasi menggunakan *confusion matrix* seperti yang ditunjukkan pada Tabel 1.

Tabel 1 Representasi Confusion Matrix [1]

Actual	Predicted	
	Positive	Negative
Positive	True Positive (TP)	False Negative (FN)
Negative	False Positive (FP)	True Negative (TN)

Evaluasi berdasarkan *Confusion Matrix* sebagai berikut:

1. *Accuracy* (akurasi) adalah jumlah hasil test terprediksi benar di antara total data. Akurasi dapat dihitung dengan persamaan (3).  

$$\text{Akurasi} = \frac{TP+TN}{TP+FN+FP+TN} \times 100 \quad (3)$$
2. *Sensitivity* (sensitifitas) adalah jumlah orang yang menunjukkan hasil positif menderita penyakit. Sensitifitas dapat dihitung dengan persamaan (4).  

$$\text{Recall} = \frac{TP}{TP+FN} \times 100 \quad (4)$$
3. *Specificity* (spesifisitas) adalah jumlah orang yang menunjukkan hasil negatif diantara orang yang tidak sakit. Spesifisitas dapat dihitung dengan

persamaan (5).  

$$\text{Specificity} = \frac{TN}{TN+FP} \times 100\% \quad (5)$$

### 3. HASIL DAN PEMBAHASAN

Sebelum mengaplikasikan metode klasifikasi PSO-KNN, tahap penting yang harus dilakukan adalah normalisasi menggunakan metode *Standard Scalar*. Dengan melakukan normalisasi, setiap fitur dalam dataset akan memiliki rentang nilai yang sama. Normalisasi membantu menghindari bias yang mungkin timbul ketika satu atau beberapa fitur memiliki nilai yang jauh lebih besar dibandingkan dengan yang lainnya, sehingga mencegah fitur-fitur tersebut mendominasi proses klasifikasi. Penerapan normalisasi dengan metode *Standard Scalar* memberikan dasar yang lebih baik, untuk proses pengambilan keputusan, membantu dalam mengidentifikasi fitur-fitur yang paling relevan dalam memprediksi hasil, dan secara keseluruhan meningkatkan kualitas analisis data. Data yang telah ternormalisasi ditunjukkan oleh Tabel 2.

Tabel 2 Data Hasil Normalisasi [2]

Sex	Age	HB	TLC	MCV	MCH	PCV	Plat-let
1.22	0.70	-1.22	13.25	1.19	1.57	-1.57	3.03
-0.82	0.60	-0.10	8.91	-0.28	-0.12	-0.20	3.86
-0.82	1.40	0.77	6.73	0.85	0.52	0.96	0.73
1.22	1.05	-2.10	3.35	-2.27	-2.42	-1.96	-0.66
1.22	-0.88	-1.39	3.25	-0.88	-1.81	-0.84	1.28
...	...	...	...	...	...	...	...
1.22	0.51	0.35	-0.55	0.10	-0.81	1.15	-0.87
1.22	-0.93	0.89	-0.57	0.56	0.54	0.89	0.05
-0.82	-0.43	-0.60	-0.57	0.50	0.42	-0.66	-0.55
1.22	-0.48	-1.60	-0.57	1.05	1.21	-1.87	-0.71
-0.82	0.65	-2.30	-0.59	0.29	-0.81	-2.30	-0.81

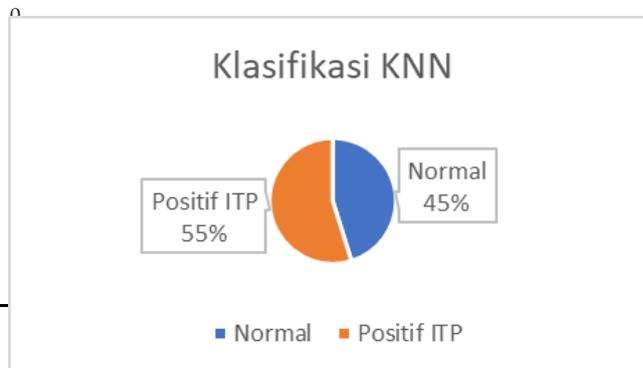
Setelah didapatkan hasil normalisasi, langkah selanjutnya mendeteksi adanya outlier dengan metode IQR (*Interquartile Range*). Metode IQR digunakan untuk mengidentifikasi adanya nilai-nilai ekstrim yang dapat mempengaruhi hasil analisis statistik. Dengan menghitung batas atas dan batas bawah berdasarkan IQ. Outlier merujuk pada nilai-nilai yang secara signifikan berbeda dari nilai-nilai lain dalam dataset dan dapat mempengaruhi analisis statistik dan model yang dibangun. Deteksi dan penanganan outlier merupakan langkah penting dalam analisis data, karena nilai-nilai yang tidak biasa ini dapat menyebabkan bias dalam hasil analisis statistik dan pemodelan. Nilai-nilai yang berada di luar batas-batas ini dianggap sebagai outlier. Setelah proses deteksi outlier menggunakan metode IQR, terdapat hasil bahwa dari total 357 data,

hanya tersisa 303 data yang tidak mengandung outlier, sementara 54 data lainnya diidentifikasi sebagai outlier.

Dalam kerangka penelitian ini, perbandingan data training dan testing yang digunakan adalah 80:20 dari total 303 data terdiri dari kelas 1 sebanyak 162 data dan kelas 0 sebanyak 141 data, yang telah melewati proses normalisasi dan deteksi outlier menggunakan metode IQR. Dengan perbandingan ini, 242 data terdiri dari 108 data kelas 0 dan 134 data kelas 1 dimasukkan ke dalam kelompok data training, sedangkan 61 data terdiri dari 27 data di kelas 0 dan 34 data di kelas 1 dijadikan sebagai kelompok data testing. Penggunaan perbandingan ini adalah suatu pendekatan yang umum dalam pembangunan model klasifikasi, di mana data training digunakan untuk melatih model klasifikasi, sedangkan data testing digunakan untuk menguji kinerja model yang telah dibangun.

#### Klasifikasi dengan Metode KNN

Dengan menggunakan jarak antara data yang di prediksi dengan tetangga terdekat dihitung dengan *Euclidean Distance*. Pada tahapan klasifikasi dengan metode KNN didapatkan hasil dari training untuk kelas 1 sejumlah 134 data dan untuk kelas 0 sejumlah 108 data. Sedangkan untuk jumlah hasil testing pada kelas 1 sejumlah 28 data dan pada kelas 0 sejumlah 33 data yang ditunjukkan pada Gambar 2.



Gambar 2 Diagram Hasil Klasifikasi KNN [2]

Berdasarkan hasil analisis grafik dari dataset yang terdiri dari 303 data pasien, dapat disimpulkan bahwa dari jumlah tersebut, sebanyak 162 pasien atau setara dengan 55% dari total data, dapat diklasifikasikan sebagai pasien yang tergolong dalam kelas 1. Ini menunjukkan bahwa lebih dari separuh pasien dalam dataset ini memiliki penyakit *Immune Thrombocytopenic Purpura* (ITP), yang dicirikan oleh gangguan autoimun yang mengakibatkan penurunan jumlah trombosit dalam darah. Sementara itu, sebanyak 141 pasien atau sekitar 45% dari total data dianggap dalam kelas 0. Ini menunjukkan bahwa ada sejumlah pasien dalam dataset yang tidak menunjukkan gejala atau kondisi

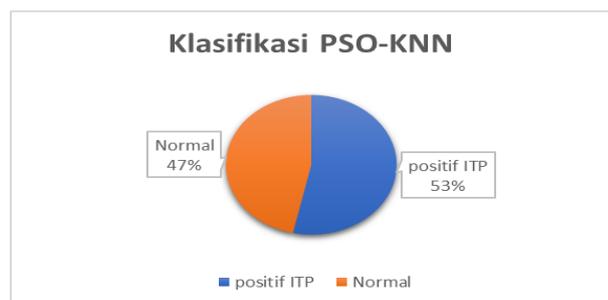
yang terkait dengan ITP.

### Klasifikasi dengan Metode PSO-KNN

Algoritma KNN membutuhkan penentuan jumlah tetangga terdekat ( $K$ ) yang merupakan salah satu parameter penting dalam proses klasifikasi. Dalam beberapa kasus, memilih nilai  $K$  yang tepat dapat berdampak signifikan pada kinerja dan akurasi model KNN. Dengan menggunakan optimasi PSO-KNN, kita dapat secara efektif menentukan nilai  $K$  yang optimal untuk algoritma KNN dalam analisis klasifikasi.

Uji coba performa optimasi PSO dilakukan dengan mengoperasikan sistem menggunakan sebanyak 100 iterasi. Pada tahap ini, parameter-parameter kunci dalam algoritma PSO telah ditentukan dengan cermat. Nilai dari parameter kognitif ( $c1$ ) dan parameter sosial ( $c2$ ) masing-masing diatur sebesar 0.5. Selain itu, nilai *inertia weight* ( $w$ ) diatur sebesar 0.9. Inertia weight memainkan peran penting dalam menyesuaikan seberapa besar langkah perubahan partikel dalam setiap iterasi. Dengan pengaturan parameter-parameter ini, uji coba PSO dilakukan untuk memaksimalkan kinerja algoritma dalam klasifikasi penyakit ITP. Hasil dari uji coba ini akan memberikan gambaran tentang sejauh mana pengaturan parameter-parameter PSO yang telah ditentukan dapat memberikan hasil yang optimal dalam memprediksi penyakit ITP pada data pasien.

Setelah melalui 100 iterasi dalam proses optimasi, didapatkan nilai  $K$  terbaik untuk metode KNN, yaitu sebesar 18. Nilai  $K$  ini memiliki peran penting dalam proses klasifikasi, karena menentukan jumlah tetangga terdekat yang akan digunakan untuk mengklasifikasikan setiap data. Selanjutnya, proses klasifikasi dengan metode KNN menggunakan nilai  $K$  yang telah dihasilkan dari hasil optimasi. Dalam tahap *training*, terdapat 108 data yang diklasifikasikan sebagai kelas 0 (tidak menderita penyakit ITP) dan 134 data yang diklasifikasikan sebagai kelas 1 (menderita penyakit ITP). Sedangkan pada tahap *testing*, terdapat 27 data yang termasuk ke dalam kelas 0 dan 34 data yang termasuk ke dalam kelas 1.



Gambar 3. Diagram Hasil Klasifikasi PSO-KNN [3]

Berdasarkan hasil analisis grafik pada Gambar 3, dari dataset yang terdiri dari 303 data pasien, dapat disimpulkan bahwa sebanyak 168 pasien atau sekitar 53% dari total data, dapat diklasifikasikan sebagai pasien yang tergolong dalam kelas 1. Ini mengindikasikan bahwa lebih dari separuh pasien dalam dataset ini mengalami penyakit *Immune Thrombocytopenic Purpura* (ITP), yang ditandai dengan gangguan autoimun yang menyebabkan penurunan jumlah trombosit dalam darah. Sementara itu, sebanyak 135 pasien atau sekitar 47% dari total data dianggap dalam kelas 0. Ini menunjukkan bahwa ada sejumlah pasien dalam dataset yang tidak menunjukkan gejala atau kondisi yang terkait dengan ITP.

### Analisis Hasil Perbandingan

Dalam klasifikasi KNN dengan menggunakan data *training* dan data *testing* yang selanjutnya dihasilkan *confusion matrix* untuk pengukuran performa KNN dengan PSO-KNN dapat dilihat pada Tabel 3.

Tabel 3. *Confusion Matrix* Klasifikasi KNN dan PSO-KNN [3]

Actual	Predicted			
	KNN		PSO-KNN	
	Kelas 1	Kelas 0	Kelas 1	Kelas 0
Kelas 1	26	1	25	2
Kelas 0	7	27	3	31

Perhitungan dilakukan dengan *confusion matrix* dan didapatkan hasil akurasi, sensitivitas, dan spesifisitas dari model KNN dan PSO-KNN pada Tabel 4.

Tabel 4. Hasil akurasi, sensitivitas, dan spesifisitas KNN dan PSO-KNN [4]

Hasil Klasifikasi	KNN	PSO-KNN
Akurasi	0.869	0.918
Sensitivitas	0.794	0.912
Spesifisitas	0.963	0.926
Waktu Training	0.0023 s	0.0022 s
Waktu Testing	0.0027 s	0.0030 s

Berdasarkan tabel diatas diketahui bahwa dengan penentuan nilai  $K$  dengan metode *Particle Swarm Optimization* untuk meningkatkan akurasi algoritma *K-Nearest Neighbor*. Waktu traing dan testing PSO-KNN lebih cepat daripada KNN standar dengan hasil akurasi yang lebih baik maka PSO mampu mengoptimalkan dan meningkatkan hasil klasifikasi. Pada penelitian ini akurasi KNN-PSO meningkat daripada menggunakan KNN standar dengan peningkatan hingga 4.9% dimana akurasi untuk KNN standar adalah 86.9% sedangkan PSO-

KNN meningkat hingga 91.8%. Hasil Sensitivitas meningkat hingga 11.8% dimana Sensitivitas KNN sebesar 79.4% dan PSO-KNN sebesar 91.2%. Sedangkan pada hasil Spesifisitas menurun hingga 3.7% dimana spesifisitas PSO-KNN 92.6% dan KNN sebesar 96.3%.

#### 4. KESIMPULAN

Penelitian ini menggunakan perbandingan 80:20 untuk data *training* dan *testing* sehingga terdapat 242 data yang termasuk kedalam data *training* dan 61 data yang termasuk kedalam data *testing*. Dengan menggunakan optimasi PSO-KNN, kita dapat secara efektif menentukan nilai K yang optimal untuk algoritma KNN dalam analisis klasifikasi untuk meningkatkan nilai akurasi. Hasil yang diperoleh dari klasifikasi dengan metode kombinasi PSO-KNN adalah nilai akurasi sebesar 91.8% dengan peningkatan 4.9% dari KNN standar, nilai sensitivitas sebesar 91.2% dengan peningkatan 11.8% dari KNN standar, dan nilai spesifitas sebesar 92.6% dengan penurunan 3.7% dari KNN standar. % Waktu pelatihan dan pengujian dari PSO-KNN juga lebih cepat daripada KNN standar sehingga PSO mampu mengoptimalkan dan meningkatkan hasil klasifikasi dari KNN. Dari hasil juga didapatkan 168 pasien masuk dalam kelas 1 atau sebanyak 53% pasien penderita penyakit ITP dan 47% pasien normal atau sebanyak 135 masuk dalam kelas 0.

#### DAFTAR PUSTAKA

- [1] L. D. Yen *et al.*, “Risiko Terjadinya Ekimosis Pada Penyakit Graves: Sebuah Kajian Literatur,” *J. Kedokt. dan Kesehat. Publ. Ilm. Fak. Kedokt. Univ. Sriwij.*, vol. 10, no. 1, pp. 61–66, 2023, doi: 10.32539/jkk.v10i1.19818.
- [2] M. I. R. Hasibuan, “Sistem Pakar Mendiagnosa Penyakit Itp (Idiopathic Thrombocytopenic Purpura) Menggunakan Metode Variable Centered Intelligent Rule System (VCIRS),” *J. Sist. Komput. dan Inform.*, vol. 1, no. 2, p. 94, 2020, doi: 10.30865/json.v1i2.1954.
- [3] I. Ma and A. T. Sandhu, “Immune Thrombocytopenia,” *Hosp. Med. Clin.*, vol. 6, no. 1, pp. 53–66, 2017, doi: 10.1016/j.ehmc.2016.07.005.
- [4] A. Haryanto, R. Hartono, and I. Isngadi, “Manajemen Anestesi pada Seksio Sesarea dengan Idiopathic Thrombocytopenic Purpura: Serial kasus,” *J. Anestesi Obstet. Indones.*, vol. 4, no. 2, pp. 100–109, 2021, doi: 10.47507/obstetri.v4i2.61.
- [5] M. C. Wenno, M. Nasir, and C. A. Rieuwpassa, “Immune Thrombositopenia Purpura in Pregnancy At the Undate General Hospital Palu,” *J. Med. Prof.*, vol. 3, no. 3, pp. 197–203, 2019.
- [6] I. Romli and A. T. Zy, “Penentuan Jadwal Overtime Dengan Klasifikasi Data Karyawan Menggunakan Algoritma C4.5,” *J. Sains Komput. Inform. (J-SAKTI)*, vol. 4, no. 2, pp. 694–702, 2020.
- [7] R. L. Hasanah, M. Hasan, W. E. Pangesti, F. F. Wati, and W. Gata, “Klasifikasi Penerima Dana Bantuan Desa Menggunakan Metode Knn (K-Nearest Neighbor),” *J. Techno Nusa Mandiri*, vol. 16, no. 1, pp. 1–6, 2019, doi: 10.33480/techno.v16i1.25.
- [8] M. M. Baharuddin, H. Azis, and T. Hasanuddin, “Analisis Performa Metode K-Nearest Neighbor Untuk Identifikasi Jenis Kaca,” *Ilk. J. Ilm.*, vol. 11, no. 3, pp. 269–274, 2019, doi: 10.33096/ilkom.v11i3.489.269-274.
- [9] A. P. Giovani, A. Ardiansyah, T. Haryanti, L. Kurniawati, and W. Gata, “Analisis Sentimen Aplikasi Ruang Guru Di Twitter Menggunakan Algoritma Klasifikasi,” *J. Teknoinfo*, vol. 14, no. 2, p. 115, 2020, doi: 10.33365/jti.v14i2.679.
- [10] N. F. Istighfarin, R. A. Rahmasti, and H. Nugroho, “Penerapan Metode Particle Swarm Optimization (PSO) Dan Genetic Algorithm (GA) Pada Sistem Optimasi Visible Light Communication (VLC) Untuk Menentukan Posisi Robot,” *Simetris J. Tek. Mesin, Elektro dan Ilmu Komput.*, vol. 11, no. 1, pp. 279–286, 2020, doi: 10.24176/simet.v11i1.4052.
- [11] S. A. Dainamang, N. Hayatin, and D. R. Chandranegara, “Analisis Sentimen Media Sosial Twitter terhadap RUU Omnibus Law dengan Metode Naive Bayes dan Particle Swarm Optimization,” *Komputika J. Sist. Komput.*, vol. 11, no. 2, pp. 211–218, 2022, doi: 10.34010/komputika.v11i2.6037.
- [12] K. Widyatmoko, E. Sugiarto, M. Muslih, and F. Budiman, “Optimasi Metode K-Nearest Neighbor Dengan Particle Swarm Optimization Untuk Pengenalan Citra Batik Dengan Ragam Hias Geometris,” *J. Inform. Upgris*, vol. 8, no. 1, 2022, doi: 10.26877/jiu.v8i1.11705.
- [13] M. A. Imron and B. Prasetyo, “Improving Algorithm Accuracy K-Nearest Neighbor Using Z-Score Normalization and Particle Swarm Optimization to Predict Customer Churn,” *J. Soft Comput. Explor.*, vol. 1, no. 1, pp. 56–62, 2020, doi: 10.52465/josce.v1i1.7.
- [14] D. Pajri, Y. Umaidah, and T. N. Padilah, “K-Nearest Neighbor Berbasis Particle Swarm Optimization untuk Analisis Sentimen Terhadap Tokopedia,” *J. Tek. Inform. dan Sist. Inf.*, vol. 6, no. 2, pp. 242–253, 2020, doi: 10.28932/jutisi.v6i2.2658.
- [15] Ahmad Harmain, P. Paiman, H. Kurniawan, K. Kusri, and Dina Maulina, “Normalisasi Data Untuk Efisiensi K-Means Pada Pengelompokan Wilayah Berpotensi Kebakaran Hutan Dan Lahan Berdasarkan Sebaran Titik Panas,” *Tek. Teknol. Inf. dan Multimed.*, vol. 2, no. 2, pp. 83–89, 2022, doi: 10.46764/teknimedia.v2i2.49.
- [16] E. A. Fadlilah, “Identifikasi Anomali Data Akademik Menggunakan Dbscan Outlier Detection,” *Pros. Sains Nas. dan Teknol.*, vol. 12, no. 1, p. 336, 2022, doi: 10.36499/psnst.v12i1.7012.

- [17] P. R. Sihombing, Suryadiningrat, D. A. Sunarjo, and Y. P. A. C. Yuda, "Identifikasi Data Outlier (Pencilan) dan Kenormalan Data Pada Data Univariat serta Alternatif Penyelesaiannya," *J. Ekon. dan Stat. Indones.*, vol. 2, no. 3, pp. 307–316, 2022, doi: 10.11594/jesi.02.03.07.
- [18] M. Mansur, T. Prahasto, and F. Farikhin, "Particle Swarm Optimization Untuk Sistem Informasi Penjadwalan Resource Di Perguruan Tinggi," *J. Sist. Inf. Bisnis*, vol. 4, no. 1, pp. 11–19, 2014, doi: 10.21456/vol4iss1pp11-19.
- [19] H. Leidiyana, "Penerapan Algoritma K-Nearest Neighbor Untuk Penentuan Resiko Kredit Kepemilikan Kendaraan Bermotor," *J. Penelit. Ilmu Komputer, Syst. Embed. Log.*, vol. 1, no. 1, pp. 65–76, 2013.
- [20] A. G. Karegowda *et al.*, "Cascading K-means Clustering and K-Nearest Neighbor Classifier for Categorization of Diabetic Patients 148," *Int. J. Eng. Adv. Technol.*, vol. 1, no. 3, pp. 147–151, 2012, [Online]. Available: <http://www.galaxy.gmu.edu/interface/I01/I2001Proceedings/Jbreault>