



## OPTIMASI PARAMETER K-NEAREST NEIGHBOR MENGGUNAKAN ALGORITMA FIREFLY UNTUK PREDIKSI PENYAKIT STROKE

### ***K-NEAREST NEIGHBOR PARAMETER OPTIMIZATION USING FIREFLY ALGORITHM FOR STROKE PREDICTION***

**Rizky Ananda Hafika<sup>1\*</sup>, Stefen Agus Waruwu<sup>2</sup>, Muhammad Yazid Noor<sup>3</sup>,  
Adidtya Perdana<sup>4</sup>**

Ilmu Komputer, Fakultas Matematika Dan Ilmu Pengetahuan Alam, Universitas Negeri Medan

Email: [rizkyanandahafika@gmail.com](mailto:rizkyanandahafika@gmail.com)<sup>1\*</sup>, [stefenwaruwu@gmail.com](mailto:stefenwaruwu@gmail.com)<sup>2</sup>, [muhammadyazidnoorn@gmail.com](mailto:muhammadyazidnoorn@gmail.com)<sup>3</sup>,  
[adidtya@unimed.ac.id](mailto:adidtya@unimed.ac.id)<sup>4</sup>

---

#### Article Info

##### Article history :

Received : 10-12-2025

Revised : 12-12-2025

Accepted : 14-12-2025

Published : 16-12-2025

---

#### Abstract

*Stroke is a disease with high mortality and long-term disability rates, making early prediction crucial for prevention and treatment. This study aims to improve the performance of the K-Nearest Neighbor (KNN) algorithm in predicting stroke by optimizing its parameters using the Firefly Algorithm (FA). The dataset used consists of 5,110 patient records with various clinical and lifestyle attributes. The research stages include data preprocessing, class balancing using oversampling, baseline KNN model construction, and optimization of the k parameter using FA. Model performance was evaluated using accuracy and confusion matrix metrics. Experimental results show that the baseline KNN achieved an accuracy of 0.8637, while the optimized KNN model reached an accuracy of 0.8920 with an optimal parameter value of k = 1. These results indicate that the Firefly Algorithm is effective in optimizing KNN parameters and significantly improves stroke prediction performance.*

---

**Keywords : Stroke Prediction, K-Nearest Neighbor, Firefly Algorithm**

---

#### Abstrak

Stroke merupakan penyakit dengan tingkat kematian dan kecacatan yang tinggi, sehingga diperlukan metode prediksi yang akurat untuk deteksi dini. Penelitian ini bertujuan meningkatkan kinerja algoritma K-Nearest Neighbor (KNN) dalam memprediksi penyakit stroke melalui optimasi parameter menggunakan Firefly Algorithm (FA). Dataset yang digunakan terdiri dari 5.110 data pasien dengan berbagai atribut klinis dan gaya hidup. Tahapan penelitian meliputi preprocessing data, penyeimbangan kelas menggunakan oversampling, pembangunan model KNN sebagai baseline, serta optimasi nilai parameter k menggunakan FA. Evaluasi model dilakukan menggunakan metrik akurasi dan confusion matrix. Hasil penelitian menunjukkan bahwa KNN sebelum optimasi menghasilkan akurasi sebesar 0,8637, sedangkan setelah optimasi menggunakan Firefly Algorithm akurasi meningkat menjadi 0,8920 dengan nilai parameter optimal k = 1. Hasil ini membuktikan bahwa Firefly Algorithm efektif dalam mengoptimasi parameter KNN dan meningkatkan performa klasifikasi pada prediksi penyakit stroke.

**Kata Kunci : Prediksi Stroke, K-Nearest Neighbor, Firefly Algorithm**

#### PENDAHULUAN

Stroke merupakan salah satu penyakit yang paling mematikan dan menjadi penyebab kecacatan jangka panjang di berbagai negara, termasuk Indonesia. Kondisi ini terjadi akibat



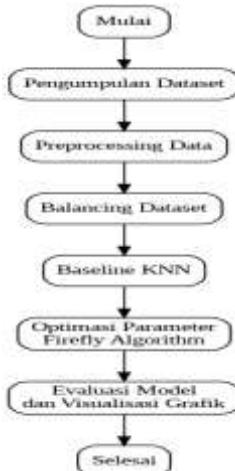
terganggunya aliran darah ke otak yang memicu kerusakan jaringan otak secara cepat. Perkembangan teknologi kesehatan mendorong adanya penelitian berbasis data mining untuk memprediksi risiko stroke sejak dini. Berbagai model klasifikasi telah digunakan, namun K-Nearest Neighbor (KNN) menjadi salah satu algoritma yang menarik karena kesederhanaan dan performanya yang kompetitif dalam memprediksi penyakit berdasarkan pola data riwayat medis pasien (Siregar et al., 2023).

Meskipun KNN sering digunakan, performanya sangat bergantung pada pemilihan parameter seperti nilai  $k$ , pemilihan jarak, serta bobot tetangga. Tanpa optimasi parameter, model berisiko mengalami penurunan akurasi terutama pada dataset kesehatan yang kompleks dan tidak seimbang. Oleh karena itu, metode optimasi berbasis metaheuristik mulai banyak diterapkan. Salah satu algoritma optimasi yang semakin populer adalah Firefly Algorithm (FA), yang meniru perilaku kunang-kunang dalam mencari solusi terbaik melalui mekanisme cahaya dan daya tarik. Penggunaan FA terbukti membantu meningkatkan performa berbagai model klasifikasi dalam domain kesehatan (Bacanin et al., 2023).

Sejumlah penelitian menunjukkan bahwa penggabungan algoritma optimasi dengan metode klasifikasi mampu meningkatkan akurasi secara signifikan. Dalam konteks prediksi stroke, optimasi parameter KNN menggunakan Firefly Algorithm mampu menghasilkan nilai  $k$  yang lebih optimal dan mengurangi risiko kesalahan prediksi. Dengan optimasi tersebut, algoritma dapat bekerja lebih stabil dan lebih adaptif terhadap pola data yang beragam sehingga memberikan hasil prediksi yang lebih akurat. Hal ini telah dibuktikan melalui penelitian terbaru yang menekankan pentingnya optimasi parameter dalam meningkatkan akurasi sistem prediksi berbasis machine learning (Singh & Kumar, 2025).

## METODE PENELITIAN

Bagian metode penelitian berisi bahan dan alat, lokasi penelitian, uraian masalah atau alasan penelitian, metode/cara pengumpulan data (survei atau perancangan percobaan), dan analisis data. Pendahuluan ditulis dengan font Times New Roman-12, spasi 1,15. Tiap paragraf diawali kata yang menjorok ke dalam sekitar 1 cm dari tepi kiri tiap kolom.



**Gambar 1.**Flowchart Penelitian



## Identifikasi Masalah

Pada tahap ini dilakukan pengamatan dan kajian terhadap proses prediksi penyakit stroke menggunakan algoritma machine learning. Meskipun algoritma K-Nearest Neighbor (KNN) sering digunakan dalam klasifikasi medis, performanya sangat bergantung pada pemilihan parameter  $k$ . Nilai  $k$  yang tidak optimal dapat menyebabkan model kurang akurat dalam membedakan pasien berisiko stroke dan tidak berisiko.

Selain itu, dataset stroke yang digunakan memiliki beberapa permasalahan, antara lain:

1. Distribusi kelas yang tidak seimbang, di mana jumlah data pasien yang tidak mengalami stroke jauh lebih banyak dibandingkan yang mengalami stroke.
2. Adanya missing value, khususnya pada fitur BMI.
3. Variabel kategorikal yang harus diubah terlebih dahulu ke bentuk numerik.
4. Belum adanya proses optimasi parameter KNN untuk mencari nilai  $k$  terbaik.

Oleh karena itu, penelitian ini mengimplementasikan Firefly Algorithm sebagai metode optimasi untuk mencari nilai  $k$  paling optimal sehingga mampu meningkatkan akurasi prediksi stroke.

## Perancangan Sistem

Penelitian ini diawali dengan pengumpulan dataset yang terdiri dari 5.110 data pasien dengan 12 fitur, seperti *gender*, *age*, *work\_type*, *BMI*, *smoking\_status*, dan label *stroke*. Tahap preprocessing data dilakukan dengan menghapus kolom *id* yang tidak digunakan, menangani *missing values* pada BMI, mengubah data kategorikal menjadi numerik menggunakan *label encoding*, serta melakukan normalisasi data menggunakan metode *Min-Max Scaling*. Selanjutnya, dilakukan balancing dataset menggunakan teknik *oversampling* untuk mengatasi ketidakseimbangan kelas, di mana distribusi data berubah dari [4861, 249] menjadi [4861, 4861].

Setelah dataset seimbang, penelitian dilanjutkan dengan perancangan model baseline KNN menggunakan nilai  $k$  default untuk memperoleh akurasi awal. Tahap berikutnya adalah perancangan algoritma Firefly, yang meliputi inisialisasi populasi firefly sebagai kandidat nilai  $k$ , perhitungan kecerahan berdasarkan akurasi KNN, serta pergerakan firefly menuju solusi terbaik hingga diperoleh nilai  $k$  dengan akurasi tertinggi. Kinerja model kemudian dievaluasi menggunakan beberapa metrik, yaitu akurasi, precision, recall, F1-score, dan confusion matrix.

## Implementasi

Tahap implementasi diawali dengan loading dataset dan eksplorasi data, yaitu menampilkan lima data teratas, informasi dataset, serta jumlah *missing value*. Selanjutnya dilakukan preprocessing data yang mencakup imputasi nilai BMI, pengkodean data kategorikal, dan normalisasi fitur ke rentang 0–1. Setelah itu, dilakukan balancing dataset menggunakan teknik *oversampling* hingga jumlah data stroke dan non-stroke menjadi seimbang.

Tahap berikutnya adalah implementasi baseline KNN yang menghasilkan nilai akurasi sebesar 0,8637. Selanjutnya diterapkan Firefly Algorithm dengan inisialisasi nilai  $k$  awal seperti



[11.86, 28.57, 22.23, 18.36, 5.52, ...], hingga melalui proses iterasi diperoleh nilai  $k$  terbaik yaitu  $k = 1$  dengan akurasi 0,8920. Hasil optimasi kemudian divisualisasikan melalui grafik peningkatan akurasi pada setiap iterasi serta grafik perubahan nilai  $k$  selama proses optimasi.

### **Uji Coba/ Pengujian**

Uji coba dalam penelitian ini dilakukan menggunakan metode eksperimen, yaitu dengan melakukan serangkaian pengujian langsung terhadap algoritma KNN dan proses optimasi menggunakan Firefly Algorithm untuk melihat perubahan performa yang dihasilkan. Pada tahap preprocessing, dilakukan pengujian untuk memastikan proses penanganan missing value, encoding variabel kategorikal, dan normalisasi data berjalan dengan benar sehingga dataset siap digunakan pada tahap pelatihan model.

Selanjutnya, model KNN diuji terlebih dahulu tanpa optimasi untuk memperoleh performa dasar melalui pengukuran akurasi, confusion matrix, dan classification report. Tahap berikutnya adalah proses optimasi menggunakan Firefly Algorithm, di mana dilakukan pengamatan terhadap perubahan nilai  $k$ , perhitungan nilai fitness berdasarkan akurasi, serta peningkatan performa model pada setiap iterasi. Fungsi visualisasi seperti grafik perkembangan akurasi dan grafik perubahan nilai  $k$  juga digunakan untuk memperlihatkan dinamika proses optimasi. Melalui pendekatan eksperimen ini, seluruh rangkaian proses — mulai dari pemrosesan data, pelatihan model, optimasi parameter, hingga evaluasi hasil — dapat divalidasi dan menunjukkan bahwa sistem bekerja secara konsisten dan menghasilkan performa yang optimal.

## **HASIL DAN PEMBAHASAN**

### **Deskripsi Eksperimen**

Pada penelitian ini dilakukan serangkaian eksperimen untuk menganalisis pengaruh optimasi parameter menggunakan Firefly Algorithm terhadap performa algoritma K-Nearest Neighbor (KNN) dalam memprediksi penyakit stroke. Dataset yang digunakan terdiri dari 5.110 data pasien dengan beberapa variabel prediktor seperti usia, jenis kelamin, hipertensi, kadar glukosa, BMI, kebiasaan merokok, serta satu variabel target yaitu status stroke.

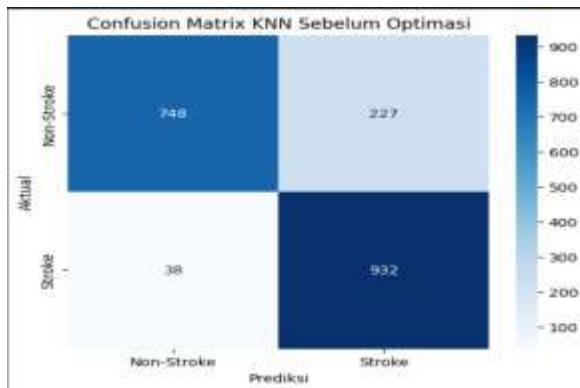
Proses eksperimen diawali dengan tahap preprocessing data yang meliputi penanganan missing value, pengkodean data kategorikal menjadi numerik, normalisasi data, serta penyeimbangan kelas menggunakan teknik oversampling. Selanjutnya, dilakukan pembangunan model KNN tanpa optimasi sebagai model awal, kemudian diterapkan Firefly Algorithm untuk mencari nilai parameter terbaik, khususnya nilai  $k$ . Model KNN yang telah dioptimasi selanjutnya dievaluasi dan dibandingkan dengan model awal untuk mengetahui peningkatan performa yang dihasilkan.

### **Nilai KNN Sebelum Optimasi**

Secara keseluruhan, KNN sebelum optimasi sudah mampu mendeteksi stroke dengan baik, namun masih memiliki kelemahan dalam membedakan pasien sehat dan stroke secara optimal. Oleh karena itu, optimasi parameter (misalnya dengan Firefly Algorithm) diperlukan untuk:



1. Mengurangi false positive,
2. Meningkatkan akurasi dan keseimbangan prediksi,
3. Memperbaiki kinerja model secara keseluruhan.



Gambar 4.1 KNN Sebelum Optimasi

Berdasarkan confusion matrix, diperoleh nilai True Negative (TN) sebesar 748, yang menunjukkan bahwa data non-stroke berhasil diprediksi dengan benar sebagai non-stroke. False Positive (FP) sebesar 227 menunjukkan adanya data non-stroke yang keliru diprediksi sebagai stroke, yang dapat menyebabkan terjadinya peringatan palsu. False Negative (FN) sebesar 38 merupakan kasus stroke yang tidak terdeteksi oleh model, dan kondisi ini tergolong berisiko karena pasien stroke diprediksi sebagai non-stroke. Sementara itu, True Positive (TP) sebesar 932 menunjukkan bahwa model mampu mengklasifikasikan data stroke dengan baik, sehingga secara umum model memiliki kemampuan yang cukup baik dalam mendeteksi kasus stroke.

### Mencari Nilai K

Setelah dilakukan proses optimasi parameter pada algoritma K-Nearest Neighbor (KNN) menggunakan Firefly Algorithm, diperoleh nilai parameter terbaik yaitu  $k = 1$  dengan akurasi sebesar 0,8920. Hasil ini menunjukkan adanya peningkatan kinerja model dibandingkan sebelum optimasi, yang disebabkan oleh pemilihan nilai  $k$  yang lebih tepat. Dengan nilai  $k = 1$ , proses klasifikasi dilakukan berdasarkan satu tetangga terdekat yang paling relevan, sehingga keputusan klasifikasi menjadi lebih sensitif terhadap pola data dan mampu meningkatkan ketepatan prediksi.

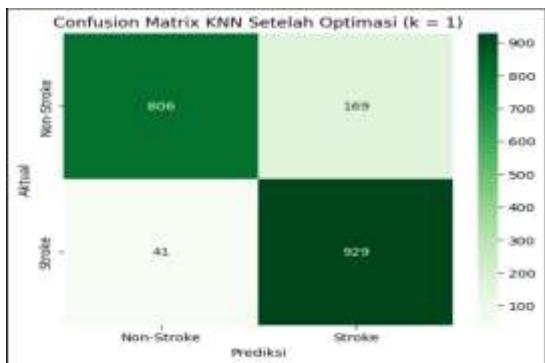
Selain itu, hasil optimasi juga menunjukkan bahwa Firefly Algorithm berhasil melakukan pencarian parameter secara efektif dan mencapai kondisi konvergen, di mana nilai akurasi tidak lagi mengalami perubahan signifikan pada iterasi berikutnya. Dengan demikian, dapat disimpulkan bahwa penerapan optimasi KNN menggunakan Firefly Algorithm memberikan kontribusi positif dalam meningkatkan performa klasifikasi stroke, khususnya dalam menghasilkan akurasi yang lebih optimal dan stabil.

### Nilai KNN Setelah Optimasi

Secara keseluruhan, optimasi parameter KNN dengan  $k = 1$  berhasil meningkatkan kinerja model, terutama dalam mengurangi kesalahan prediksi pada kelas Non-Stroke. Meskipun terdapat



sedikit peningkatan false negative, model tetap menunjukkan kemampuan deteksi stroke yang tinggi dan lebih seimbang dibandingkan sebelum optimasi.



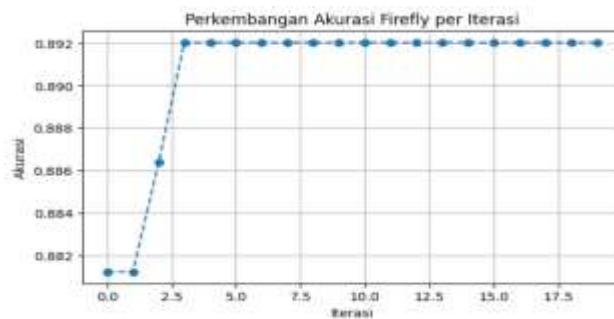
Gambar 4.2 KNN Setelah Optimasi

Berdasarkan confusion matrix hasil optimasi, diperoleh True Negative (TN) sebesar 806 yang menunjukkan peningkatan jumlah data non-stroke yang berhasil diprediksi dengan benar. False Positive (FP) sebesar 169 menandakan masih adanya kesalahan prediksi berupa false alarm, namun jumlahnya lebih sedikit dibandingkan sebelum dilakukan optimasi. False Negative (FN) sebesar 41 merupakan kesalahan kritis karena kasus stroke tidak terdeteksi oleh model. Sementara itu, True Positive (TP) sebesar 929 menunjukkan bahwa model tetap memiliki kemampuan yang sangat baik dalam mendekripsi kasus stroke.

### Hasil Akurasi Firefly

Grafik perkembangan akurasi per iterasi digunakan untuk menunjukkan perubahan kinerja algoritma K-Nearest Neighbor (KNN) selama proses optimasi menggunakan Firefly Algorithm. Grafik ini bertujuan untuk memperlihatkan bagaimana nilai akurasi meningkat secara bertahap pada setiap iterasi hingga mencapai kondisi stabil. Dengan adanya grafik ini, dapat diketahui kecepatan konvergensi Firefly Algorithm serta efektivitas proses optimasi dalam meningkatkan akurasi klasifikasi stroke.

Grafik pertama menunjukkan perubahan nilai akurasi KNN selama proses optimasi menggunakan Firefly Algorithm.



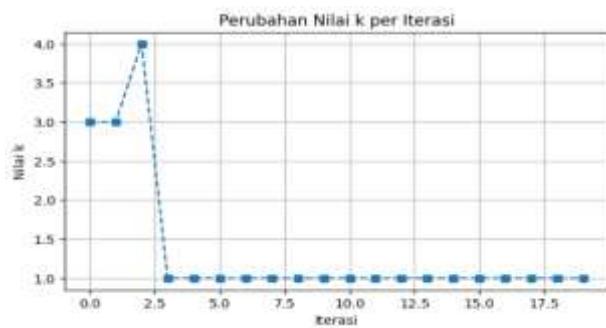
Gambar 4.3 Grafik Perkembangan Akurasi Firefly Per Iterasi



Grafik akurasi menunjukkan bahwa pada iterasi awal (0–1) nilai akurasi masih relatif rendah, yaitu berada di kisaran 0,881, yang menandakan parameter KNN khususnya nilai  $k$  belum optimal. Pada iterasi ke-2 hingga ke-3 terjadi peningkatan akurasi yang cukup signifikan hingga mencapai sekitar 0,892, menunjukkan bahwa Firefly Algorithm berhasil menemukan solusi yang lebih baik. Setelah iterasi ke-3, nilai akurasi cenderung stabil hingga iterasi terakhir, yang menandakan algoritma telah mencapai kondisi konvergensi dan solusi terbaik telah diperoleh.

Kesimpulan grafik akurasi adalah Firefly Algorithm mampu meningkatkan akurasi KNN secara cepat dan mencapai nilai optimal dalam jumlah iterasi yang relatif sedikit.

Grafik perubahan nilai parameter  $k$  per iterasi digunakan untuk menggambarkan proses pencarian nilai  $k$  optimal oleh Firefly Algorithm. Grafik ini menunjukkan bagaimana nilai  $k$  mengalami eksplorasi pada iterasi awal dan kemudian mengarah pada satu nilai yang paling optimal. Visualisasi ini bertujuan untuk membuktikan bahwa pemilihan nilai  $k$  dilakukan secara sistematis melalui proses optimasi, sehingga parameter yang diperoleh mampu menghasilkan performa klasifikasi terbaik.



Gambar 4.4 Grafik Perubahan Nilai K Per Iterasi

Grafik kedua memperlihatkan perubahan nilai parameter  $k$  pada algoritma KNN selama proses optimasi. Pada iterasi awal, nilai  $k$  masih bervariasi antara 3 hingga 4, yang menunjukkan bahwa Firefly Algorithm masih berada pada tahap eksplorasi solusi. Pada iterasi ke-3 terjadi perubahan yang signifikan, di mana nilai  $k$  menurun dan ditetapkan menjadi  $k = 1$ . Selanjutnya, hingga iterasi terakhir, nilai  $k$  tetap stabil pada  $k = 1$ , yang menandakan bahwa algoritma telah menemukan nilai parameter optimal dan mencapai kondisi konvergensi.

Kesimpulan grafik nilai  $k$ : Firefly Algorithm berhasil mengidentifikasi bahwa nilai  $k = 1$  merupakan parameter paling optimal untuk menghasilkan akurasi tertinggi pada data klasifikasi stroke yang digunakan.

Dimana Hubungan Kedua grafik tersebut berdasarkan grafik hasil optimasi, Firefly Algorithm mampu meningkatkan akurasi KNN dari 0,881 menjadi 0,892 dan mencapai kondisi konvergen pada iterasi ke-3. Selain itu, nilai parameter  $k$  yang awalnya bervariasi berhasil distabilkan pada  $k = 1$ , yang terbukti memberikan performa terbaik dalam klasifikasi stroke.



## Pembahasan

Hasil penelitian menunjukkan bahwa penerapan Firefly Algorithm dalam mengoptimasi parameter algoritma K-Nearest Neighbor (KNN) mampu meningkatkan performa klasifikasi penyakit stroke. Model KNN setelah optimasi menghasilkan nilai parameter terbaik  $k = 1$  dengan akurasi sebesar 0,8920, yang lebih tinggi dibandingkan sebelum optimasi.

Peningkatan kinerja terlihat dari berkurangnya kesalahan prediksi pada kelas non-stroke serta tetap tingginya kemampuan model dalam mendeteksi pasien stroke. Selain itu, proses optimasi menunjukkan konvergensi yang cepat, sehingga dapat disimpulkan bahwa Firefly Algorithm efektif dalam meningkatkan kinerja dan stabilitas model KNN untuk prediksi stroke.

## KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa algoritma K-Nearest Neighbor (KNN) mampu digunakan untuk memprediksi penyakit stroke dengan performa yang cukup baik. Namun, performa KNN sangat dipengaruhi oleh pemilihan parameter, khususnya nilai  $k$ . Model KNN tanpa optimasi masih menunjukkan adanya kesalahan prediksi yang cukup signifikan, terutama pada klasifikasi pasien non-stroke dan stroke.

Penerapan Firefly Algorithm sebagai metode optimasi parameter terbukti mampu meningkatkan performa KNN. Hasil optimasi menunjukkan bahwa nilai parameter terbaik yang diperoleh adalah  $k = 1$ , dengan peningkatan akurasi dari 0,8637 menjadi 0,8920. Selain itu, proses optimasi menunjukkan konvergensi yang cepat dan stabil. Dengan demikian, dapat disimpulkan bahwa Firefly Algorithm efektif digunakan untuk mengoptimasi parameter KNN dan menghasilkan model prediksi stroke yang lebih optimal dan stabil.

## DAFTAR PUSTAKA

- Arifin, M., Pratama, D., & Lestari, S. (2024). Penerapan algoritma K-Nearest Neighbor pada sistem prediksi kesehatan berbasis data pasien. *Journal of Medical Informatics*, 12(1), 44–53.
- Bacanin, N., Bezdan, T., Tuba, E., Lugonja, P., Strumberger, I., & Tuba, M. (2023). Improved multi-layer binary Firefly algorithm for optimizing feature selection and classification of microarray data. *Expert Systems with Applications*, 215, 1–15.
- Chen, L., Zhao, X., & Wu, Y. (2023). Feature selection optimization for medical diagnosis using enhanced Firefly Algorithm. *Biomedical Signal Processing Journal*, 18(3), 221–234.