

Perbandingan Optimalisasi Hasil Klasifikasi Menggunakan Pso Pada Algoritma C.45 Dan Cart (Studikasuk Prediksi Penyakit Stroke)

Muhammad Gus Choyyin¹, Handoyo Widi Nugroho²

^{1,2}Fakultas Komputer Informatics & Business Institute Darmajaya, Jalan Z.A. Pagar Alam, No.93 Gedong Meneng ,
Bandar Lampung, Lampung, Indonesia
E-mail: 2221210028.mail@darmajaya.ac.id¹, handoyo.wn@darmajaya.ac.id²

Abstract — This study aims to compare the performance of four different classification models, namely CART, Decision Tree C4.5, and their respective versions with Particle Swarm Optimization (PSO). Stroke can cause damage to the brain and significantly affect bodily functions. Therefore, early detection and management of stroke are crucial to prevent worse outcomes. In this context, the use of technology and algorithms in decision support systems becomes relevant to aid in stroke disease diagnosis in a classification context. Evaluation is performed using four main performance metrics: accuracy, precision, recall, and Area Under Curve (AUC). Data is tested using these four models, and the results are compared to identify the most effective model in predicting the target class. The results show that the Decision Tree C4.5 model, especially when enhanced with PSO, provides the same accuracy of 96.27%. Both models also have the same precision rate, which is 98.43%. Decision Tree C4.5 with PSO also has the same recall rate, which is 96.90%, and the same AUC, which is 0.995. It consistently delivers better performance than the CART model in terms of accuracy, precision, recall, and AUC.

Keywords: Classification, CART, Decision Tree C4.5, Particle Swarm Optimization (PSO), Accuracy, Precision, Recall, Area Under Curve (AUC).

Abstrak— Penelitian ini bertujuan untuk membandingkan kinerja empat model klasifikasi berbeda, yaitu CART, Decision Tree C4.5, dan versi masing-masing dengan Penyesuaian Swarm Partikel (PSO), Stroke dapat menyebabkan kerusakan pada otak dan mempengaruhi fungsi tubuh secara signifikan. Oleh karena itu, deteksi dini dan pengelolaan penyakit stroke sangat penting untuk mencegah dampak yang lebih buruk. Dalam konteks ini, penggunaan teknologi dan algoritma dalam sistem pendukung keputusan menjadi relevan untuk membantu diagnosis penyakit stroke.dalam konteks klasifikasi. Evaluasi dilakukan menggunakan empat metrik kinerja utama: akurasi, presisi, recall, dan Area Under Curve (AUC). Data diuji menggunakan empat model tersebut, dan hasilnya dibandingkan untuk mengidentifikasi model yang paling efektif dalam memprediksi kelas target. Hasil menunjukkan bahwa model Decision Tree C4.5, terutama saat ditingkatkan dengan PSO, Decision Tree C4.5 dengan PSO memberikan akurasi yang sama yaitu 96,27%. Kedua model juga memiliki tingkat presisi yang sama, yaitu 98,43%. DecesionTree C4.5 dengan PSO juga memiliki tingkat recall yang sama, yaitu 96,90%. dan AUC yang sama, yaitu 0,995. secara konsisten memberikan kinerja yang lebih baik daripada model CART dalam hal akurasi, presisi, recall, dan AUC.

Kata Kunci: Klasifikasi, CART, Decision Tree C4.5, Penyesuaian Swarm Partikel (PSO), Akurasi, Presisi, Recall, Area Under Curve (AUC).

I. PENDAHULUAN

Penyakit stroke merupakan salah satu penyakit yang sering menimbulkan dampak yang serius terhadap kesehatan manusia. Stroke dapat menyebabkan kerusakan pada otak dan mempengaruhi fungsi tubuh secara signifikan. Oleh karena itu, deteksi dini dan pengelolaan penyakit stroke sangat penting untuk mencegah dampak yang lebih buruk. Dalam konteks ini, penggunaan teknologi dan algoritma dalam sistem pendukung keputusan menjadi relevan untuk membantu diagnosis penyakit stroke.[1][2]

Perkembangan teknologi dan kecerdasan buatan membuka peluang untuk mengimplementasikan algoritma pemodelan data guna mendukung deteksi dini penyakit stroke. Dua algoritma yang umum digunakan dalam pemodelan data adalah Classification and Regression Trees (CART) dan C4.5 (C45). Kedua algoritma ini memiliki kelebihan dan kekurangan masing-masing. Oleh karena itu,

perbandingan antara algoritma CART dan C45 menjadi subjek penelitian yang menarik, terutama ketika diterapkan berbasis Particle Swarm Optimization (PSO).[2][3]

CART dan C45 adalah dua algoritma pohon keputusan yang populer dalam dunia kecerdasan buatan. Pohon keputusan merupakan model prediktif yang dapat menggambarkan hubungan antara berbagai variabel input dan output dalam bentuk struktur pohon. Keduanya sering digunakan dalam klasifikasi data, termasuk untuk diagnosis penyakit. Namun, pemilihan algoritma yang tepat dapat sangat mempengaruhi akurasi prediksi.[4][5]

penelitian ini, kami akan membahas perbandingan antara algoritma CART dan C45 dalam konteks deteksi penyakit stroke. Kami juga akan mengeksplorasi pemanfaatan Particle Swarm Optimization (PSO) sebagai pendekatan berbasis swarm untuk meningkatkan performa kedua algoritma tersebut. PSO telah terbukti efektif dalam mencari solusi optimal dalam ruang pencarian kompleks, dan penggabungan PSO dengan algoritma pohon keputusan diharapkan dapat meningkatkan akurasi prediksi serta efisiensi waktu.[6]

Melalui penelitian ini, diharapkan dapat ditemukan solusi yang lebih optimal dalam deteksi dini penyakit stroke, yang dapat memberikan kontribusi positif terhadap pengembangan sistem pendukung keputusan di bidang kesehatan. Selain itu, perbandingan antara algoritma CART dan C45 berbasis PSO akan memberikan pemahaman yang lebih baik tentang kelebihan dan kekurangan masing-masing algoritma, sehingga dapat membantu pemilihan algoritma yang sesuai dengan kebutuhan dan karakteristik data pada kasus penyakit stroke[7]

Metode pemodelan data yang digunakan dalam penelitian ini bertujuan untuk memberikan kontribusi dalam meningkatkan akurasi dan efisiensi deteksi penyakit stroke. Penelitian terkait deteksi penyakit dengan menggunakan algoritma pohon keputusan telah menjadi fokus utama dalam dunia medis dan kecerdasan buatan. Dengan membandingkan algoritma CART dan C45 berbasis PSO, diharapkan dapat diidentifikasi metode yang paling optimal untuk aplikasi pada dataset spesifik penyakit stroke.[8]

Pentingnya penelitian ini terletak pada dampak positif yang dapat dihasilkan, yaitu penyediaan solusi yang lebih efisien dan akurat dalam mendeteksi gejala penyakit stroke. Deteksi dini melalui sistem pendukung keputusan dapat memungkinkan intervensi lebih cepat dan pengelolaan yang lebih efektif terhadap pasien. Oleh karena itu, penyelidikan mengenai perbandingan algoritma berbasis PSO menjadi relevan untuk merumuskan rekomendasi terbaik dalam aplikasi di lapangan kesehatan.[7]

Selain itu, penelitian ini dapat memberikan wawasan tentang keunggulan dan kelemahan masing-masing algoritma, serta dampak integrasi Particle Swarm Optimization pada kinerja keduanya. PSO sebagai metode optimasi dapat membantu algoritma pohon keputusan mencapai solusi yang lebih baik dalam menangani kompleksitas dataset penyakit stroke.pada konteks implementasi, penelitian ini akan membahas metodologi eksperimental yang melibatkan pengumpulan data, preprocessing, pemodelan menggunakan algoritma CART dan C45 berbasis PSO, serta evaluasi performa keduanya. Evaluasi tersebut akan melibatkan parameter-parameter kritis seperti akurasi, presisi, recall, dan F1-score, sehingga dapat memberikan gambaran yang komprehensif tentang keunggulan relatif dari masing-masing algoritma.

Penting untuk mencatat bahwa penelitian ini bukan hanya berfokus pada aspek teoretis, tetapi juga memiliki dampak praktis dalam konteks penerapan di bidang kesehatan. Hasil dari penelitian ini diharapkan dapat memberikan panduan yang berharga bagi para profesional kesehatan, peneliti, dan pengembang sistem informasi kesehatan untuk memilih metode deteksi penyakit stroke yang paling sesuai dengan kebutuhan spesifik mereka.

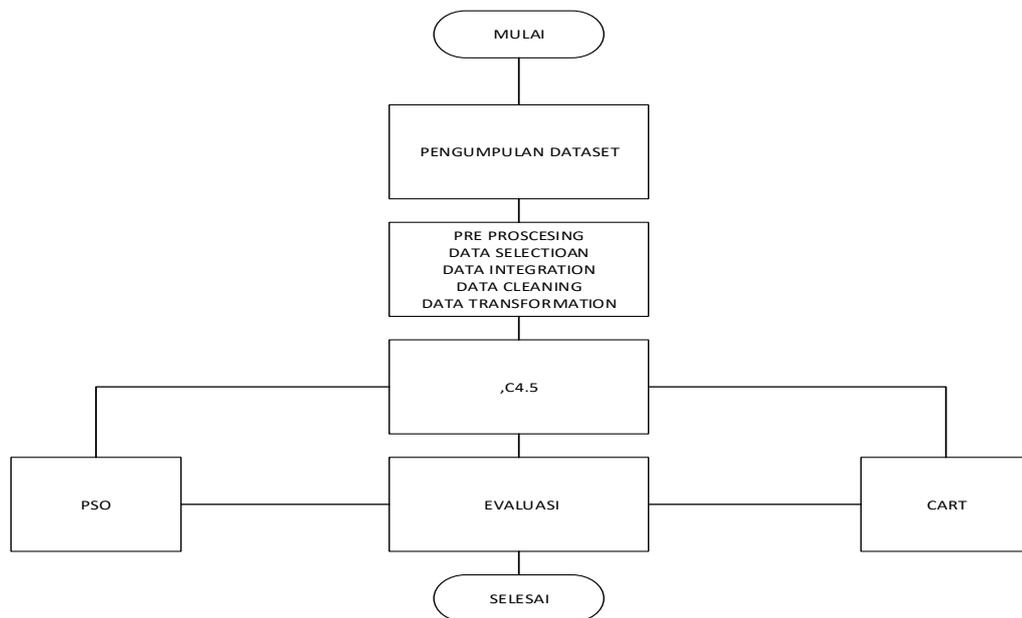
demikian, penelitian ini diharapkan dapat memberikan kontribusi positif dalam menghadapi tantangan kompleks deteksi penyakit stroke dan meningkatkan kualitas layanan kesehatan melalui penerapan teknologi kecerdasan buatan.

Datamining data adalah suatu proses ekstraksi informasi dari database yang besar, dan dalam pelaksanaannya, melibatkan berbagai teknik seperti statistik, matematika, kecerdasan buatan, dan pembelajaran mesin[9]. Salah satu metode yang digunakan dalam proses penambangan data adalah

klasifikasi, yang merupakan landasan dalam analisis data[5]. Menurut penelitian oleh [4], klasifikasi adalah suatu teknik dalam penambangan data yang terkait dengan pembelajaran mesin, dan berfungsi sebagai metode untuk memprediksi penentuan keanggotaan dalam kelompok berdasarkan data yang telah ada.

II. METODOLOGI PENELITIAN

Pada penelitian ini akan membahas langkah-langkah dari proses penelitian yang akan dilaksanakan, dalam melakukan analisa dan mencari pola data untuk dijadikan sebuah dataset untuk memudahkan penelitian dan dapat berjalan dengan sistematis sehingga memenuhi tujuan yang diinginkan maka dibuat alur tahapan penelitian secara detail pada gambar 1



Gambar 1. Tahapan Penelitian

2.1 Teknik Pengumpulan Data

Teknik pengumpulan data pada penelitian ini adalah mengambil dataset dari <https://www.kaggle.com/datasets/fedesoriano/stroke-prediction-dataset>.

Data sebanyak 5.110 record yang terdiri dari 12 atribut 1 label atau kelas, atribut dataset dan diskripsinyaseperti tabel 1 dibawah ini:

Tabel 1 Atribut Dataset dan diskripsinya

Atribut	Deskripsi	Kriteria
Id	Id data	84-72951
Gender	Jenis kelamin -Pria -Wanita	Male
		Female
Age	Usia	10-82
Hypertension	Riwayat hipertensi	1=Yes
		0=No
Heart_disease	Riwayat penyakit jantung	1=Yes
		0=No
Ever_married	Pernah menikah	Yes
		No
		Private
		Self-

Work_type	Tipe pekerjaan	employed
		Govt_job
		Never_work ed
		Children
Residence_ty pe	Tipe tempat tinggal	Urban
		Rural
Avg_glucose_ level	Level kadar gula	55-271
Bmi	Indeks masa tubuh	11-92
Smoking_stat us	Status merokok	Formerly smoked
		Never smoked
		smoked
Stroke	Label	1=Yes
		0=No

2.2 Preprocessing

Sebelum menerapkan tahap implementasi, langkah pra-pemrosesan harus dilakukan terlebih dahulu. Meskipun jumlah data awal dapat dikumpulkan, tidak semua data akan dimanfaatkan, dan tidak semua atribut akan digunakan karena perlu melalui proses awal pra-pemrosesan data atau yang dikenal sebagai persiapan data..[10]

2.3 Data Selection

Dalam proses pengolahan data dari dataset yang terkumpul, dilakukan seleksi data dengan cara memilih serta memisahkan data berdasarkan kriteria yang telah ditetapkan. Langkah selanjutnya adalah mengurangi jumlah atribut dan catatan yang ada untuk mendapatkan dataset yang masih informatif..[11]

2.4 Data Integration

Penyatuan data dari berbagai sumber merupakan tahapan yang penting. Proses tersebut tidak hanya berguna untuk mengurangi duplikasi dan inkonsistensi data yang timbul akibat pengambilan dari multiple sumber data. [10]

2.5 Data Cleaning

Di tahap ini, proses penyingkiran data yang tidak relevan pada karakteristik kualitas udara akan dilakukan untuk memastikan bahwa hasil klasifikasi tidak terpengaruh. Langkah-langkah melibatkan pembersihan manual terhadap data yang kosong (null) dan pengisian nilai yang hilang pada data yang tidak lengkap (missing value), serta mengevaluasi untuk penghapusan atribut atau penggantian data, serta proses identifikasi atau eliminasi yang diperlukan.[12]

2.6 Data Transformation

Setelah melalui tahap pembersihan data, integrasi data, dan pemilihan data, langkah berikutnya adalah melakukan proses Transformasi Data. Awalnya, data mengenai nilai sikap tercatat dalam format non-numerik seperti yang terlihat dalam tabel pada bagian kolom Pembersihan Data. Data kemudian diubah ke dalam format numerik agar dapat diolah dalam proses eksplorasi data..[5]

III. HASIL DAN PEMBAHASAN

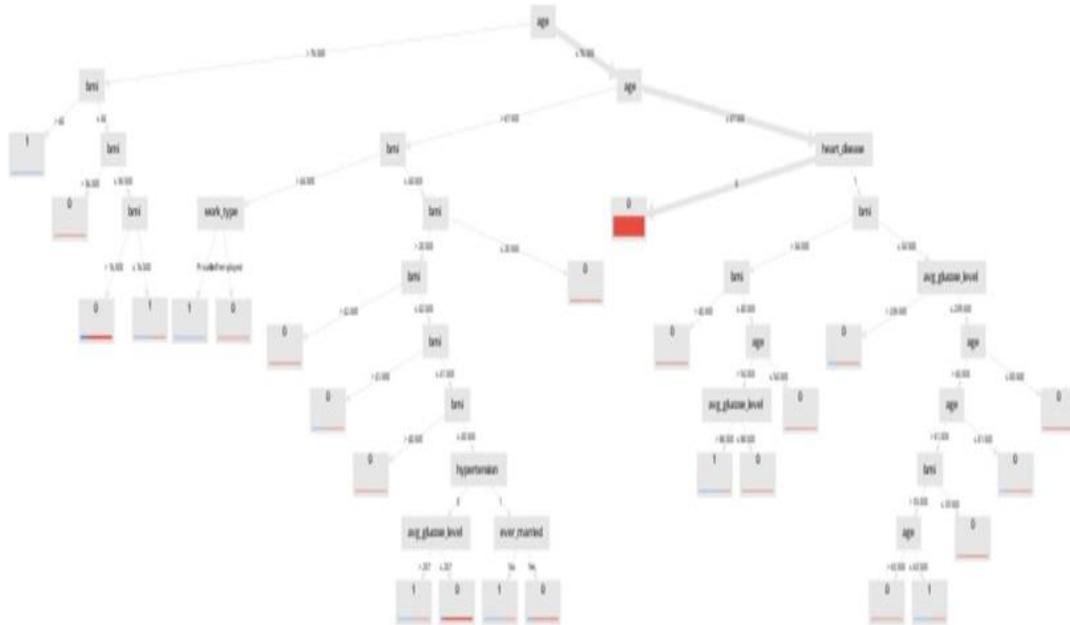
Hasil, dan pembahasan penelitian. Software RapidMiner digunakan sebagai alat bantu untuk memproses data. Proses, hasil, dan pembahasan terbagi ke dalam beberapa subbab, termasuk Pengolahan dan distribusi Data serta Implementasi algoritma pada RapidMiner.

3.1 Algoritma Cart

Penerapan algoritma Cart pada rapidminer dengan menggunakan split validation dengan nilai akurasi, precision, confusion matrix atau nilai recall dan nilai AUC dengan pembagian data training dan data testing sebesar 80:20, dan Patricle Swarm Optimization (PSO) dan Cart Split Data (80:20)

3.2 Algoritma C4.5

Setelah melakukan preprocessing data, kemudian dilakukan pemodelan algoritma C4.5 menggunakan tools RapidMiner untuk mengetahui bagaimana decision tree yang terbentuk. Pada percobaan pertama menggunakan pembagian data dengan hold-out 80:20, sehingga dihasilkan decision tree yang ditunjukkan pada Gambar 2.



Gambar 2 Tampilan Pohon Keputusan *Decision Tree* C4.5

Penerapan algoritma decision Tree pada rapidminer dengan menggunakan split validation dengan nilai akurasi, precision, confusion matrix atau nilai recall dan nilai AUC dengan pembagian data training dan data testing sebesar 80:20, dan Patricle Swarm Optimization (PSO) dan Decision Tree C4.5 Split Data (80:20)

Berikut hasil perbandingan dari dua algoritma tersebut dapat dilihat pada tabel 2 dibawah ini

Tabel 2 Hsil nilai perbandingan dua algoritma

No	Keterangan	Algoritma Cart	Cart dan PSO	Decision Tree	Decision Tree dan PSO
1	Akurasi	81,82%	85,12%	96,27%	96,27%
2	Precisiin	82,61%	86,24%	98,43%	98,43%
3	Recall	97,94%	96,91%	96,90%	96,90%
4	Auc	0,823	0,847	0,995	0,995

Berdasarkan tabel di atas, kita dapat membuat beberapa kesimpulan:

1. Algoritma Decision Tree, baik dengan maupun tanpa PSO, memberikan performa yang lebih tinggi dibandingkan dengan algoritma CART dalam hal akurasi, presisi, recall, dan AUC. Dengan PSO, kedua algoritma tersebut memiliki tingkat akurasi yang sama, yaitu 96,27%, presisi yang sama, yaitu 98,43%, recall yang sama, yaitu 96,90%, dan AUC yang sama, yaitu 0,995. Namun, pada algoritma Decision Tree, PSO tidak menghasilkan peningkatan yang signifikan dalam performa dibandingkan dengan Decision Tree tanpa PSO.
2. Algoritma CART, meskipun memiliki performa yang lebih rendah dibandingkan dengan Decision Tree, tetap memberikan hasil yang cukup baik dalam hal akurasi, presisi, recall, dan AUC. Namun, perlu diperhatikan bahwa peningkatan performa dari CART ke Decision Tree (baik dengan maupun tanpa PSO) cukup signifikan.
3. Peningkatan performa dari CART ke Decision Tree menunjukkan bahwa penggunaan algoritma Decision Tree, terutama dengan PSO, dapat menjadi pilihan yang lebih baik dalam konteks klasifikasi data.
4. Tingkat recall yang tinggi pada semua algoritma menunjukkan bahwa model-model tersebut mampu mengidentifikasi sebagian besar kasus positif dengan benar.

Dengan demikian, dalam konteks klasifikasi data, algoritma Decision Tree, terutama yang ditingkatkan dengan PSO, mungkin menjadi pilihan yang lebih baik untuk digunakan berdasarkan performa yang dihasilkannya.

IV. SIMPULAN

1. Performa Decision Tree: Dari tabel, terlihat bahwa algoritma Decision Tree, baik dengan maupun tanpa PSO, memberikan hasil yang jauh lebih baik daripada CART dalam semua metrik evaluasi yang digunakan, yaitu akurasi, presisi, recall, dan AUC. Hal ini menunjukkan bahwa algoritma Decision Tree lebih mampu mengklasifikasikan data dengan lebih baik daripada CART dalam konteks penelitian ini.
2. Pengaruh PSO Meskipun menggunakan PSO tidak menghasilkan peningkatan yang signifikan dalam performa Decision Tree dalam hal akurasi, presisi, recall, dan AUC, namun penggunaan PSO masih memberikan hasil yang konsisten dengan versi non-PSO. Ini menunjukkan bahwa PSO mungkin tidak diperlukan atau tidak memberikan keuntungan tambahan yang signifikan dalam konteks penggunaan algoritma Decision Tree pada dataset yang digunakan.
3. Performa yang Konsisten: Perlu diperhatikan bahwa performa Decision Tree, baik dengan maupun tanpa PSO, cukup konsisten dalam semua metrik evaluasi yang digunakan. Hal ini menunjukkan bahwa algoritma tersebut dapat diandalkan dalam konteks klasifikasi data, terutama dalam deteksi penyakit stroke yang menjadi fokus penelitian ini.
4. Tingkat Recall yang Tinggi: Tingkat recall yang tinggi pada semua model menunjukkan bahwa model-model tersebut efektif dalam mengidentifikasi kasus positif (dalam konteks ini, kemungkinan adanya stroke). Hal ini penting untuk mencegah adanya kasus yang terlewatkan atau tidak terdeteksi.

Dengan demikian, berdasarkan analisis ini, dapat disimpulkan bahwa algoritma Decision Tree, baik dengan maupun tanpa PSO, merupakan pilihan yang lebih baik daripada CART dalam konteks klasifikasi data pada studi ini. Meskipun penggunaan PSO tidak memberikan peningkatan yang signifikan dalam performa, algoritma Decision Tree tetap menunjukkan performa yang konsisten dan efektif dalam deteksi penyakit stroke.

UCAPAN TERIMA KASIH

Terima kasih kami ucapkan kepada semua pihak atas dukungan dan fasilitas yang diberikan selama proses penelitian. Ucapan terima kasih selanjutnya kami ucapkan kepada rekan-rekan di Institut Informatika dan Bisnis Darmajaya yang selalu memberikan bantuan moral sepanjang proses penelitian ini .

DAFTAR PUSTAKA

- [1] K. L. Kohsasih and Z. Situmorang, “Comparative Analysis of C4.5 and Naïve Bayes Algorithms in Predicting Cerebrovascular Disease,” *J. Inform.*, vol. 9, no. 1, pp. 13–17, 2022.
- [2] E. R. Pambudi, Sriyanto, and Firmansyah, “Teknika 16 (02): 221-226,” *Ijccs*, vol. 16, No.02, no. x, pp. 221–226, 2022.
- [3] A. F. Riany and G. Testiana, “Penerapan Data Mining untuk Klasifikasi Penyakit Stroke Menggunakan Algoritma Naïve Bayes,” *J. SAINTEKOM*, vol. 13, no. 1, pp. 42–54, 2023, doi: 10.33020/saintekom.v13i1.352.
- [4] Y. Azhar, A. K. Firdausy, and P. J. Amelia, “Perbandingan Algoritma Klasifikasi Data Mining Untuk Prediksi Penyakit Stroke,” *SINTECH (Science Inf. Technol. J.*, vol. 5, no. 2, pp. 191–197, 2022, doi: 10.31598/sintechjournal.v5i2.1222.
- [5] Y. N. Paramitha, A. Nuryaman, A. Faisol, E. Setiawan, and D. E. Nurvazly, “Klasifikasi Penyakit Stroke Menggunakan Metode Naïve Bayes,” *J. Siger Mat.*, vol. 04, no. 01, pp. 11–16, 2023.
- [6] C. Lintang Bhuana and H. W. Nugroho, “Perbandingan Decision Tree C4.5 Dan Support Vector Machine (Svm) Dalam Klasifikasi Penderita Stroke Berbasis Pso,” vol. 6, no. 1, pp. 24–34, 2023.
- [7] R. S. Rohman, R. A. Saputra, and D. A. Firmansaha, “Komparasi Algoritma C4.5 Berbasis PSO Dan GA Untuk Diagnosa Penyakit Stroke,” *CESS (Journal Comput. Eng. Syst. Sci.*, vol. 5, no. 1, p. 155, 2020, doi: 10.24114/cess.v5i1.15225.
- [8] Suryani, D. Rahmadani, A. A. Muzafar, A. Hamid, R. Annisa, and Mustakim, “Analisis Perbandingan Algoritma C4.5 dan CART untuk Klasifikasi Penyakit Stroke,” *SENTIMAS Semin. Nas. Penelit. dan Pengabd. Masy.*, pp. 197–206, 2022, [Online]. Available: <https://journal.irpi.or.id/index.php/sentimas>
- [9] F. N. Ikhromr, I. Sugiyarto, U. Faddillah, and B. Sudarsono, “Implementasi Data Mining Untuk Memprediksi Penyakit Diabetes Menggunakan Algoritma Naives Bayes dan K-Nearest Neighbor,” *INTECOMS J. Inf. Technol. Comput. Sci.*, vol. 6, no. 1, pp. 416–428, 2023.
- [10] F. Meila Azzahra Sofyan, A. Putri Riyandoro, D. Fitriani Maulana, J. Haerul Jaman, S. Informasi, and U. Singaperbangsa Karawang, “Penerapan Data Mining dengan Algoritma C5.0 Untuk Prediksi Penyakit Stroke,” *Juli*, vol. 6, pp. 619–625, 2023, [Online]. Available: <https://ojs.trigunadharma.ac.id/index.php/jsk/index>
- [11] N. A. Iskandar, I. Ernawati, and Y. Widiastiwi, “Klasifikasi Diagnosis Penyakit Stroke Dengan Menggunakan Metode Random Forest,” *Semin. Nas. Mhs. Ilmu Komput. dan Apl.*, pp. 432–441, 2022, [Online]. Available: <https://conference.upnvj.ac.id/index.php/senamika/article/view/2190>
- [12] K. Akmal, A. Faqih, and F. Dikananda, “Perbandingan Metode Algoritma Naïve Bayes Dan K-Nearest Neighbors Untuk Klasifikasi Penyakit Stroke,” *JATI (Jurnal Mhs. Tek. Inform.*, vol. 7, no. 1, pp. 470–477, 2023, doi: 10.36040/jati.v7i1.6367.