



Comparative Analysis of C4.5 and CART Algorithms for Classification of Stroke

Analisis Perbandingan Algoritma C4.5 dan CART untuk Klasifikasi Penyakit Stroke

**Suryani^{1*}, Desvita Rahmadani², Ali Alamuddin Muzafar³,
Abdul Hamid⁴, Rahmatul Annisa⁵, Mustakim⁶**

^{1,2,3,4,5,6} Program Studi Sistem Informasi, Fakultas Sains dan Teknologi,
UIN Sultan Syarif Kasim Riau

^{1,3,5,6} Puzzle Research Data Technology (Predatech), Fakultas Sains dan Teknologi,
UIN Sultan Syarif Kasim Riau

E-Mail: ¹12050320389@students.uin-suska.ac.id, ²12050323639@students.uin-suska.ac.id,
³12050310450@students.uin-suska.ac.id, ⁴12050312122@students.uin-suska.ac.id,
⁵11753202097@students.uin-suska.ac.id, ⁶mustakim@uin-suska.ac.id

Corresponding Author: Suryani

Abstract

Cerebrovascular disease or stroke occurs when the blood supply to the brain is cutoff due to blockage (ischemic) or ruptured blood vessels (hemorrhagic) that disrupt the central nervous system. According to the World Health Organization (WHO), 70% of deaths occur due to stroke. Based on the 2018 Riskesdas Results Report the national stroke prevalence reached 10.9%. Stroke is the number 3 killer in Indonesia. The high mortality rate due to stroke is due to the difficulty of making accurate predictions. This disease strikes suddenly and death can occur in less than 24 hours. Based on these problems, a machine learning algorithm is needed to accurately classify stroke to carry out prevention and early treatment to reduce mortality. This study compares the C4.5 and CART decision tree algorithms to get the best classification results. The best results are obtained by the C4.5 algorithm with an accuracy of 95.76%, while the accuracy of the CART algorithm is only 95.11% with the distribution of training data and test data are equally significant, which is 70:30.

Keyword: CART, Classification, C4.5, Stroke

Abstrak

Penyakit *cerebrovascular* atau stroke terjadi akibat terputusnya suplai darah menuju otak akibat penyumbatan (*ischemic*) atau pembuluh darah pecah (*hemorrhagic*) sehingga mengganggu sistem syaraf pusat. Menurut *World Health Organization* (WHO) 70% kematian dunia terjadi akibat stroke. Berdasarkan Laporan Hasil Riskesdas tahun 2018, prevalensi stroke nasional menyentuh persentase 10,9%. Stroke adalah pembunuh nomor 3 di Indonesia. Tingginya kematian akibat stroke dikarenakan sulitnya melakukan prediksi yang akurat. Penyakit ini menyerang tiba-tiba, kematian dapat terjadi kurang dari 24 jam. Berdasarkan permasalahan tersebut, maka dibutuhkan algoritma *machine learning* yang mampu melakukan klasifikasi stroke dengan akurat, guna melakukan pencegahan dan perawatan dini sehingga dapat mengurangi angka kematian. Penelitian ini membandingkan algoritma *decision tree* C4.5 dan CART untuk mendapatkan hasil klasifikasi yang terbaik. Hasil terbaik didapatkan oleh algoritma C4.5 dengan akurasi sebesar 95,76%, sedangkan akurasi algoritma CART hanya 95,11% dengan pembagian data latih dan data uji sama besar, yakni 70:30.

Kata Kunci: CART, C4.5, Kalsifikasi, Stroke

1. PENDAHULUAN

Stroke atau *cerebrovascular* termasuk Penyakit Tidak Menular (PTM) yang ditandai dengan penyumbatan (*ischemic*) atau pecahnya pembuluh darah (*hemorrhagic*) akibat aliran darah menuju otak mengalami gangguan [1]. Penyumbatan pada pembuluh darah menuju otak menyebabkan matinya sel-sel otak karena kekurangan asupan oksigen dan nutrisi [2][3]. Resiko terbesar yang dialami penderita stroke akibat rusaknya pembuluh darah adalah kematian [4]. Menurut World Health Organization (WHO) angka kematian akibat stroke mencapai 70% dari total kematian di dunia. Lebih dari 36 juta orang kehilangan nyawa, dan 9 juta darinya terjadi sebelum usia 60 tahun [5]. Kematian akibat stroke banyak terjadi di negara dengan penghasilan rendah dan menengah, salah satunya adalah Indonesia [6][5][7].

Stroke menjadi penyebab kematian tertinggi nomor tiga di Indonesia setelah jantung dan kanker [5]. Berdasarkan Laporan Hasil Riskesdas tahun 2018, prevalensi stroke nasional menyentuh persentase 10,9%, dan provinsi Kalimantan Timur memiliki catatan prevalensi tertinggi, yakni 14,7 per mil [8]. Kasus kematian stroke di Indonesia didominasi oleh stroke hemoragik. *Intracerebral Hemorrhage* (ICH) atau stroke hemoragik adalah subtype stroke kedua yang menyebabkan cacat berat dan kematian [9]. ICH adalah disfungsi neurologis akut yang ditandai dengan pecahnya pembuluh darah arteri, vena dan kapiler sehingga memicu terjadinya pendarahan primer pada otak [4]. Pendarahan ini kemudian menyebabkan penderita stroke hilang kesadaran, mengalami kelumpuhan akibat tidak berfungsinya panca indera, dan kematian [4].

Kematian akibat stroke sulit diperkirakan karena gejala klinisnya tidak terduga dan berkembang sangat cepat [9]. Maka dari itu, dibutuhkan keterlibatan teknologi yang dinamakan *machine learning* untuk melakukan klasifikasi terhadap penyakit stroke, dalam hal ini digunakan sebuah metode pengolahan data yaitu data mining. Data mining adalah cabang ilmu yang menggabungkan bidang-bidang ilmu komputer yang dimanfaatkan guna mencari pola dan informasi menarik dari sekumpulan data, atau disebut sebagai proses penguraian pengetahuan dalam database menggunakan metode tertentu, seperti kecerdasan buatan, *machine learning* dan statistik. [10][11].

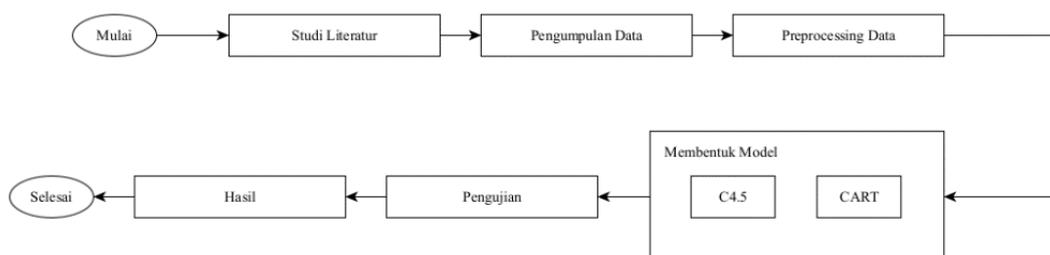
Salah satu metode yang dapat diterapkan dalam data mining adalah klasifikasi [12]. Klasifikasi adalah metode sederhana yang digunakan untuk mengenali kelas atau model data yang kemudian digunakan sebagai pendekatan dalam prediksi suatu permasalahan [13]. Metode klasifikasi yang diterapkan pada penelitian ini adalah algoritma C4.5 dan CART, keduanya adalah algoritma untuk membangun pohon keputusan atau *decision tree*. Berdasarkan pada penelitian sebelumnya oleh Gangavarapu, dkk (2021) terkait *Analyzing the Performance of Stroke Prediction using ML Classification Algorithms*, diperoleh hasil algoritma Naive Bayes Classifiers (NBC) unggul dengan nilai akurasi sebesar 82% [14].

Dengan data stroke yang sama [15] telah dilakukan penelitian berikutnya oleh Kelvin Leonardi dan Zakarias Situmorang (2022) dengan topik perbandingan algoritma C4.5 dan Naive Bayes Classifiers (NBC) untuk prediksi stroke, hasil penelitian ini algoritma C4.5 lebih baik dengan nilai akurasi 95% [16]. Penelitian selanjutnya dilakukan oleh Jody Alwin Irawadi dan Siti Sunendiari (2021) yang melakukan perbandingan algoritma CHAID, *Exhaustive CHAID* dan CART untuk klasifikasi kanker payudara, didapatkan hasil algoritma CART lebih baik dengan nilai akurasi sebesar 84,9 % [17]. Selanjutnya, berdasarkan penelitian oleh Anna Hendri Soleliza Jones dan Muhchromin Sucron Makmun yang membahas implementasi metode CART untuk klasifikasi penyakit hepatitis pada anak, diketahui algoritma CART memiliki tingkat akurasi yang baik, yakni 94% [18].

Berdasarkan rujukan tersebut, maka penelitian ini melakukan penerapan algoritma CART untuk melakukan klasifikasi penyakit stroke. Karena data yang digunakan sama dengan penelitian terdahulu oleh Kelvin Leonardi dan Zakarias Situmorang (2022) maka akan dilakukan perbandingan hasil akurasi algoritma C4.5 dengan algoritma CART. Berdasarkan perbandingan tersebut, akan ditentukan algoritma mana yang memiliki performa terbaik untuk melakukan klasifikasi penyakit stroke.

2. BAHAN DAN METODE

Metodologi penelitian adalah penjabaran alur proses yang dilakukan pada penelitian ini. Metodologi penelitian divisualkan pada Gambar 1.



Gambar 1. Metodologi Penelitian

Studi Literatur

Studi literatur pada penelitian ini merupakan tahapan untuk mengumpulkan dasar teori yang dibutuhkan. Studi literature dimulai dengan mengumpulkan informasi yang sesuai dengan topik yang sedang diteliti. Informasi tersebut didapat melalui laporan penelitian terdahulu, tesis, maupun karya ilmiah lainnya.

Pengumpulan Data dan *Preprocessing* Data

Pengumpulan data merupakan proses mencari data yang dibutuhkan untuk penelitian. Data yang digunakan pada penelitian ini adalah dasaset stroke bersumber dari Kaggle. Data yang telah dikumpulkan kemudian dilakukan *preprocessing* yang terdiri dari *cleaning* dan tranformasi supaya data siap digunakan.

Pembentukan Model dan Pengujian

Pada tahapan ini dilakukan pembentukan model pengujian algoritma C4.5 dan CART menggunakan tool RapidMiner. Hasil pengujian tersebut akan dibandingkan sehingga diperoleh algoritma dengan akurasi yang paling tinggi untuk pengklasifikasian penyakit stroke atau *cerebrovascular*.

2.1 *Decision Tree*

Metode yang populer digunakan untuk klasifikasi dengan metode data mining adalah *decision tree* [11]. Pohon keputusan atau *decision tree* adalah model yang digunakan dalam mengklasifikasikan objek yang divisualisasikan melalui bambur daun dan cabang [17]. Daun (*leaf*) merepresentasikan kelas data, sementara cabangnya adalah representasi kondisi dari atribut objek yang terukur [17].

2.2 *Algoritma C4.5*

Algoritma C4.5 merupakan pengembangan *Iterative Dichotomiser 3* (ID3) oleh J. Ross Quinlan [20]. Algoritma ini telah digunakan untuk melakukan klasifikasi data yang basisnya adalah numerik dan kategorik [20]. C4.5 memiliki inputan berupa data *training* dan data *sample*. Data training adalah data contoh untuk membangun *decision tree* yang sebelumnya telah diuji kebenarannya. Sedangkan data *sample* adalah record data yang akan menjadi parameter dalam klasifikasi data [11].

Perhitungan algoritma C4.5 melalui beberapa tahapan, dimulai dengan pencarian nilai *entropy*, kemudian menghitung nilai *gain*. Berdasarkan hasil perhitungan tersebut barulah kemudian dibangun pohon keputusan atau *decision tree*. Adapun rumus untuk menghitung *entropi* dan *gain*, dapat dilihat pada persamaan (1) dan (2) berturut-turut [21].

$$Entropy (s) = \sum_{i=0}^n -p_i * \log_2 p_i \quad (1)$$

Keterangan :

- S : Himpunan kasus
- n : Jumlah partisi S
- p_i : Proporsi dan S_i terhadap S

Setelah dilakukan perhitungan *entropy* pada setiap kasus, kemudian dilakukan perhitungan *information gain* menggunakan persamaan 2.

$$Gain (S, A) = Entropy (S) - \sum_{i=0}^n \frac{|S_i|}{|S|} * Entropy (S_i) \quad (2)$$

Keterangan :

- S : Himpunan kasus
- n : Jumlah partisi atribut A
- |S| : Jumlah kasus dalam S
- | S_i | : Jumlah kasus pada partisi ke i

Kemudian, dilanjutkan dengan mencari nilai *gain ratio*. Untuk menghitung *gain ratio* dibutuhkan perhitungan *split information* dengan persamaan 3.

$$SplitInformation (S, A) = -\sum_{i=1}^n \frac{S_i}{S} \log_2 \frac{S_i}{S} \quad (3)$$

Setelah mendapatkan nilai *Split Information*, dilanjutkan dengan melakukan perhitungan *gain ratio* dengan persamaan 4.

$$GainRatio = -\frac{Gain (S,A)}{SplitInformation (S,A)} \quad (4)$$

2.3 Classification and Regression Trees (CART)

CART dikenal sebagai salah satu algoritma *machine learning* pembuat pohon keputusan yang memiliki kemampuan baik dalam melakukan klasifikasi dan analisis regresi [22]. Algoritma ini selalu membentuk cabang biner [18]. Prinsip kerjanya adalah memilah seluruh amatan menjadi dua gugus amatan, kemudian hasil gugus amatan tersebut akan dipilah kembali untuk membuat dua gugus berikutnya sehingga didapat jumlah amatan minimum dalam setiap gugus [18].

Pembentukan *decision tree* menggunakan algoritma CART dilakukan dengan cara menghitung indeks gini pada masing-masing kelas dengan menggunakan rumus pada persamaan 5. Setelahnya, ditentukan total indeks gini pembelahan menggunakan persamaan 6.

$$j(t) = 1 - \sum_{j=1}^m P^2(j|t) \quad (5)$$

Pada persamaan 5, node t dibagi menjadi dua subset D1 dan D2 sesuai dengan ukuran b1 dan b2 masing-masing. Berdasarkan pembagian subset tersebut akan didapatkan total indeks gini pembelahan yang terdapat pada persamaan 6.

$$Gini_{pembelahan}(t) = \frac{b_1}{b} gini(D_1) + \frac{b_2}{b} gini(D_2) \quad (6)$$

Keterangan:

- Ginipembelahan : Nilai indeks gini setiap variable
- Gini(D1) : Nilai indeks gini subset D1 pada setiap variable
- Gini(D2) : Nilai indeks gini subset D2 pada setiap variabel
- b : Jumlah data pada suatu variable
- b1 : Jumlah data pada subset d1
- b2 : Jumlah data pada subset d2

Node terminal pada suatu class ditentukan apabila jumlah sampel pada class sama atau kurang dari 5. Label node terminal ditentukan berdasarkan persamaan 7 dengan memperhatikan jumlah terbanyak sebagai berikut:

$$P(j_0|t) = \max_j P(j|t) = \max_j \frac{m_j(t)}{m(t)} \quad (7)$$

Pengukuran akurasi pada algoritma CART menggunakan *confusion matrix*, yang mana ini adalah metode populer untuk mengukur kinerja algoritma klasifikasi.

3. HASIL DAN PEMBAHASAN

3.1 Preprocessing Data

Penelitian ini menggunakan *dataset* penyakit stroke dengan 12 atribut dan 5110 *record* data yang bersumber dari situs Kaggle [15]. Supaya data siap digunakan, perlu dilakukan *preprocessing* yang terdiri dari *cleaning* dan transformasi data.

Pada proses *cleaning* data, terdapat 202 *record* yang tidak dapat digunakan dan dilakukan penghapusan. Sehingga data yang siap untuk digunakan berjumlah 4909 *record* sebagaimana ditunjukkan pada Tabel 1.

Tabel 1. Dataset Stroke

No	ID	Gender	Umur	Hipertensi	...	Kadar Gula	BMI	Merokok	Stroke
1	9046	Male	67	0	...	228.69	36.06.00	Sebelumnya Ya	Ya
...					...				
1045	1665	Female	79	1	...	174.12	24	Tidak Pernah	Ya
1046	56669	Male	81	0	...	186.21	29	Sebelumnya Ya	Ya
...					...				
4909	39708	Male	55	0	...	56.87	28.09.00	Sebelumnya Ya	Tidak

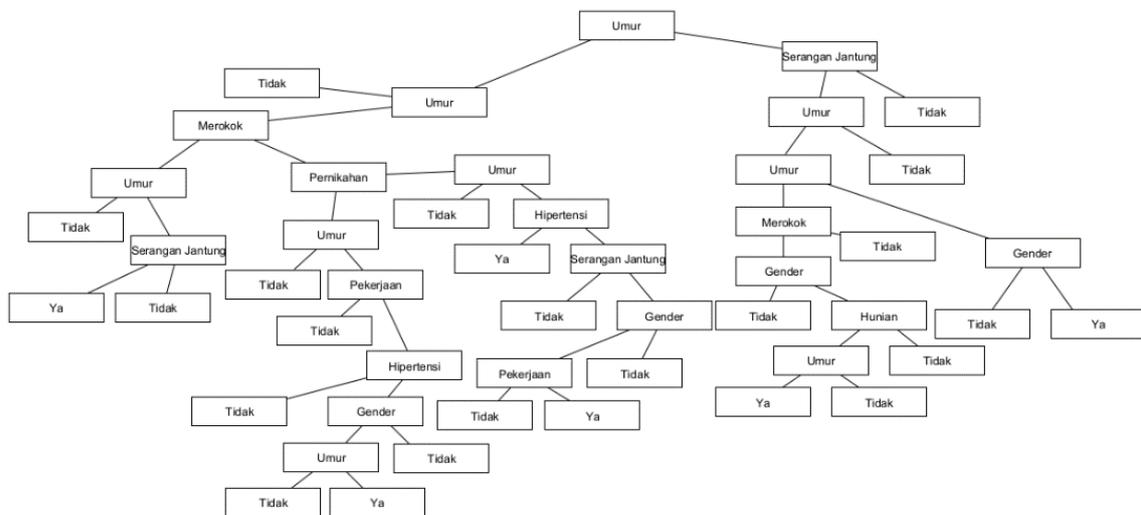
Transformasi data untuk merubah atribut bernilai teks menjadi numerik. Transformasi data ditunjukkan pada Tabel 2.

Tabel 2. Transformasi Data

No	ID	Gender	Umur	Hipertensi	...	Kadar Gula	BMI	Merokok	Stroke
1	9046	0	67	0	...	228.69	36.06.00	2	Ya
...					...				
1045	1665	1	79	1	...	174.12	24	0	Ya
1046	56669	0	81	0	...	186.21	29	2	Ya
...					...				
4909	39708	0	55	0	...	56.87	28.09.00	2	Tidak

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* 60:40, sehingga dihasilkan *decision tree* yang ditunjukkan pada Gambar 2.

**Gambar 1.** Decision Tree C4.5 Hold-Out 60:40

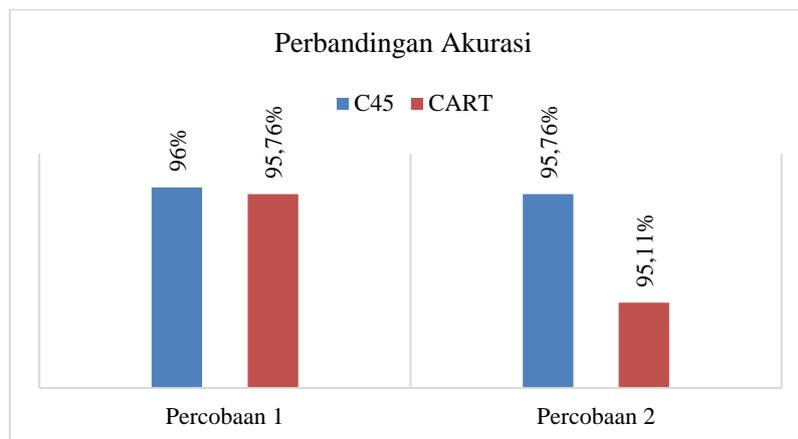
Berdasarkan *decision tree* perhitungan model C4.5 yang ditunjukkan pada Gambar 2, atribut paling berpengaruh adalah umur, serangan jantung, merokok dan status pernikahan. Berdasarkan hasil ini, kelompok usia penderita stroke adalah 68 sampai dengan 78 tahun dengan memperhatikan atribut data lainnya. Kelompok usia yang bukan stroke berdasarkan hasil ini adalah ≤ 58 tahun. Sebanyak 2328 data bukan stroke, dan 45 diantaranya terklasifikasi stroke dengan catatan memiliki riwayat serangan jantung. Setelah didapat *decision tree* dengan *hold-out* 60:40, kemudian dilakukan perhitungan untuk melihat tingkat akurasi dengan *confusion matrix*. Hasil akurasi ditunjukkan pada Tabel 3.

Tabel 3. Akurasi C4.5 Hold-Out 60:40

Accuracy : 95.80%			
	True Tidak	True Ya	Class Precision
Pred. Tidak	4682	188	96.14%
Pred. Ya	18	21	53.85%
Class Recall	99.62%	10.05%	

Tabel 4 merupakan hasil akurasi algoritma C4.5 dengan *hold-out* 60:40. Berdasarkan hasil tersebut, diketahui ada 4682 *record* data yang terklasifikasi benar sebagai bukan penderita stroke, dan ada 188 data yang terklasifikasi sebagai benar penderita stroke. Selanjutnya dilakukan percobaan kedua dengan *hold-out* 70:30, sehingga didapat *decision tree* yang terbentuk adalah pada Gambar 3.

kemudian membuat cairan masuk ke ruang intestium di seluruh susunan saraf pusat akibat perubahan pada tekanan kapiler. Kondisi ini akan membuat neuron dan di sekitarnya mengalami koma bahkan kematian [24]. Berdasarkan percobaan yang telah dilakukan pada penelitian ini menggunakan algoritma C4.5 dan CART dapat dilihat pada Bagan 1.



Gambar 6. Perbandingan Akurasi C4.5 dan CART

Perbandingan hasil akurasi pada Bagan 1 menggunakan pembagian data latih data uji 60:40 dan 70:30. Pengamatan ini bertujuan menentukan algoritma mana yang memiliki nilai akurasi atau performa terbaik dalam klasifikasi penyakit stroke. Berdasarkan pengamatan ini, dapat disimpulkan bahwa algoritma C4.5 memiliki performa lebih baik dengan nilai akurasi mencapai 96% pada percobaan pertama, dan 95,76% pada percobaan kedua.

4. KESIMPULAN

Berdasarkan hasil pembahasan dan percobaan yang telah dilakukan, percobaan pertama dengan *hold-out* data 60:40 menunjukkan hasil algoritma C4.5 memiliki tingkat akurasi yang baik mencapai 96% daripada algoritma CART dengan akurasi hanya 95,76%. Kemudian, dilakukan percobaan kedua dengan *hold-out* data 70:30. Algoritma C4.5 menunjukkan akurasi sebesar 95,76% sedangkan algoritma CART hanya sebesar 95,11%. Secara keseluruhan, dapat disimpulkan bahwa algoritma C4.5 merupakan metode klasifikasi terbaik untuk penyakit stroke ketika dibandingkan dengan algoritma CART.

REFERENSI

- [1] A. Faisal and A. Subekti, "Deep Neural Network untuk Prediksi Stroke," *Jurnal Edukasi dan Penelitian Informatika*, pp. 443–449, 2021.
- [2] O. Almadani and R. Alshammari, "Prediction of Stroke using Data Mining Classification Techniques," *International Journal of Advanced Computer Science and Applications*, 2018. [Online]. Available: www.ijacsa.thesai.org
- [3] R. A. Zuama, S. Rahmatullah, and Y. Yuliani, "Analisis Performa Algoritma Machine Learning pada Prediksi Penyakit Cerebrovascular Accidents," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 6, no. 1, p. 531, Jan. 2022, doi: 10.30865/mib.v6i1.3488.
- [4] Dody and Z. Huzaiifah, "HUBUNGAN ANTARA KLASIFIKASI STROKE DENGAN GANGGUAN FUNGSI MOTORIK PADA PASIEN STROKE," *Journal of Nursing Invention*, pp. 94–97, 2021.
- [5] I. Lishania, R. Goejantoro, and Y. N. Nasution, "Perbandingan Klasifikasi Metode Naive Bayes dan Metode Decision Tree Algoritma (J48) pada Pasien Penderita Penyakit Stroke di RSUD Abdul Wahab Sjahranie Samarinda," *Jurnal EKSPONENSIAL*, vol. 10, no. 2, 2019.
- [6] A. Byna and M. Basit, "PENERAPAN METODE ADABOOST UNTUK MENGOPTIMASI PREDIKSI PENYAKIT STROKE DENGAN ALGORITMA NAÏVE BAYES," *Jurnal Sisfokom (Sistem Informasi dan Komputer)*, vol. 9, no. 3, pp. 407–411, Nov. 2020, doi: 10.32736/sisfokom.v9i3.1023.
- [7] H. K V, H. P, G. Gupta, V. P, and P. K B, "STROKE PREDICTION USING MACHINE LEARNING ALGORITHMS," *International Journal of Innovative Research in Engineering & Management*, vol. 8, no. 4, Jul. 2021, doi: 10.21276/ijirem.2021.8.4.2.
- [8] "Direktorat P2PTM." <http://p2ptm.kemkes.go.id/> (accessed Jun. 16, 2022).

- [9] E. Hartono, M. Puspitasari, and O. Adam, "GAMBARAN TEKANAN DARAH PADA PASIEN STROKE HEMORAGIK DENGAN DIABETES MELITUS DAN NON DIABETES MELITUS DI BAGIAN SARAF RUMKITAL DR.RAMELAN SURABAYA," *Jurnal Sinaps*, 2019.
- [10] I. Kamila, U. Khairunnisa, and Mustakim, "Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan," *Jurnal Ilmiah Rekayasa dan Manajemen Sistem Informasi*, vol. 5, no. 1, pp. 119–125, 2019.
- [11] F. Elfaladonna and A. Rahmadani, "ANALISAMETODE CLASSIFICATION-DECISSION TREE DAN ALGORITMA C.45 UNTUK MEMPREDIKSI PENYAKIT DIABETES DENGAN MENGGUNAKAN APLIKASI RAPID MINER," *Journal Science and Information Technology*, pp. 10–17, 2021.
- [12] M. Swathy and K. Saruladha, "A comparative study of classification and prediction of Cardio-Vascular Diseases (CVD) using Machine Learning and Deep Learning techniques," *ICT Express*, vol. 8, no. 1, pp. 109–116, Mar. 2022, doi: 10.1016/j.ict.2021.08.021.
- [13] D. Saputra, W. Irmayani, D. Purwaningtias, and J. Sidauruk, "A Comparative Analysis of C4.5 Classification Algorithm, Naïve Bayes and Support Vector Machine Based on Particle Swarm Optimization (PSO) for Heart Disease Prediction," *International Journal of Advances in Data and Information Systems*, vol. 2, no. 2, Nov. 2021, doi: 10.25008/ijadis.v2i2.1221.
- [14] G. Sailasya and G. L. Aruna Kumari, "Analyzing the Performance of Stroke Prediction using ML Classification Algorithms," 2021. [Online]. Available: www.ijacsa.thesai.org
- [15] "Stroke Prediction Dataset Kaggle." <https://www.kaggle.com/datasets/fedesoriano/stroke-prediction-dataset> (accessed Jun. 16, 2022).
- [16] K. L. Kohsasih and Z. Situmorang, "Analisis Perbandingan Algoritma C4.5 Dan Naïve Bayes Dalam Memprediksi Penyakit Cerebrovascular," *JURNAL INFORMATIKA*, vol. 9, no. 1, pp. 13–17, 2022, [Online]. Available: <http://ejournal.bsi.ac.id/ejurnal/index.php/ji>
- [17] J. Alwin Irawadi and S. Sunendiari, "Penerapan dan Perbandingan Tiga Metode Analisis Pohon Keputusan pada Klasifikasi Penderita Kanker Payudara," *Journal Riset Statistika*, vol. 1, pp. 19–27, 2021, doi: 10.29313/v1i01.7097.
- [18] A. Hendri Soleliza Jones and M. Sucron Makmun, "Implementasi Metode CART untuk Klasifikasi Diagnosis Penyakit Hepatitis Pada Anak," *Journal of Informatics, Information System, Software Engineering and Applications*, vol. 3, no. 2, pp. 61–70, 2021, doi: 10.20895/INISTA.V3I2.
- [19] W. D. Septiani, "Komparasi Metode Klasifikasi Data Mining Algoritma C4.5 Dan Naive Bayes Untuk Prediksi Penyakit Hepatitis," *Jurnal Pilar Nusa Mandiri*, vol. 13, no. 1, pp. 76–84, 2017.
- [20] I. Junaedi, N. Nuswantari, and V. Yasin, "UNTUK DATA MINING ANALISIS TINGKAT RISIKO KEMATIAN NEONATUM PADA BAYI," 2019. [Online]. Available: <http://journal.stmikjayakarta.ac.id/index.php/jisicomTelp.+62-21-3905050>,
- [21] D. P. Utomo and M. Mesran, "Analisis Komparasi Metode Klasifikasi Data Mining dan Reduksi Atribut Pada Data Set Penyakit Jantung," *Jurnal Media Informatika Budidarma*, vol. 4, no. 2, pp. 437–444, 2020.
- [22] M. M. Ghiasi, S. Zendejboudi, and A. A. Mohsenipour, "Decision tree-based diagnosis of coronary artery disease: CART model," *Computer Methods and Programs in Biomedicine*, vol. 192, Aug. 2020, doi: 10.1016/j.cmpb.2020.105400.
- [23] D. Touriano, E. Fernando, P. Siagian, and H. A. Rohayani, "Sistem Pakar Mendiagnosis Penyakit Jantung dengan Metode Fuzzy Set," 2014.
- [24] F. P. Prasticha, D. M. Rezha, A. S. Setyoningrum, and I. Setiawan, "HUBUNGAN ANTARA USIA, HIPERTENSI, KEBIASAAN MEROKOK DENGAN MORTALITAS STROKE ISKEMIK," *Proceeding Book National Symposium and Workshop Continuing Medical Education XIV*, pp. 429–439, 2021.