

KOMPARASI ALGORITMA K-NEAREST NEIGHBOR (K-NN), SUPPORT VECTOR MACHINE (SVM), DAN DECISION TREE DALAM KLASIFIKASI PENYAKIT STROKE

Andri Setiawan¹, Rangga Febrio Waleska², Muhammad Adji Purnama³,
Rahmaddeni⁴, Lusiana Efrizoni⁵

^{1,2,3,4,5}Program Studi Teknik Informatika, STMIK AMIK RIAU

Jln. Purwodadi indah KM.10, Sidomulyo barat, Sidomulyo Bar., Kec. Tampan, Kota Pekanbaru, Riau 28294

¹ 2110031802092@sar.ac.id, ² 2110031802117@sar.ac.id, ³ 2110031802125@sar.ac.id,

⁴ rahmaddeni@sar.ac.id, ⁵ lusiana@stmik-amik-riau.ac.id

Abstract

Stroke is a leading cause of death and disability in humans. It can disrupt blood flow to the brain, impairing organ function, depriving certain organs of elasticity, oxygen, or nutrients, and ultimately causing them to die rapidly. The aim of this study is to classify individuals who have experienced a stroke. In this research, the K-NN method has the ability to handle complex problems unaffected by various factors and its strong, intensive, and non-assumptive nature. The SVM method has the capability to handle complex problems unaffected by various factors and effectively perform training. Meanwhile, Decision Tree develops knowledge based on training data and its labels, making predictions related to categories or class labels. These methods will classify stroke data from 4981 records. The test results with an 80:20 data split show that the K-NN method achieves 94% accuracy, SVM achieves 95% accuracy, while Decision Tree achieves 92% accuracy. From these results, the SVM method performs better compared to the two methods, K-NN and Decision Tree. This study utilizes Streamlite to create visually appealing data visualizations.

Keywords : The disease of stroke, Decision Tree, K-Nearst Neighbors (KNN), Support Vector Machine (SVM), and Streamlite

Abstrak

Penyakit stroke adalah penyebab kematian dan kecacatan pada manusia. Stroke dapat menyebabkan aliran darah otak terganggu. Ini menghambat semua fungsi organ, menyebabkan organ-organ tertentu kehilangan elastisitas, oksigen, atau nutrisi, dan akhirnya menyebabkan mereka mati dengan cepat. Tujuan dari penelitian ini adalah untuk mengklasifikasi seseorang yang mengalami stroke. Dalam penelitian ini, metode K-NN mempunyai kemampuan untuk menangani masalah yang kompleks tanpa terpengaruh oleh berbagai faktor dan sifatnya yang kuat, intensif, dan tidak asumsif, metode SVM memiliki kemampuan untuk menangani masalah yang kompleks tanpa terpengaruh oleh berbagai factor dan melakukan pelatihan secara efektif, sedangkan Decision Tree mengembangkan pengetahuan berdasarkan data pelatihan dan labelnya, melakukan prediksi terkait kategori atau label kelas. Dari metode-metode tersebut akan melakukan klasifikasi pada data penyakit stroke dari 4981 record. Hasil pengujian metode dengan *splitting* data 80:20 menunjukkan bahwa metode K-NN mendapatkan hasil akurasi 94%, dan SVM mendapatkan hasil akurasi 95%, sedangkan Decision Tree mendapatkan hasil 92%. Dari hasil tersebut metode SVM lebih baik dibandingkan dua metode K-NN dan Decision Tree. Studi ini menggunakan Streamlite untuk membuat visualisasi data menjadi lebih menarik.

Kata kunci : Penyakit Stroke, Decision Tree, K-Nearst Neighbors (KNN), Support Vector Machine (SVM), dan Streamlite

1. Pendahuluan

Salah satu masalah kesehatan yang paling umum di negara berkembang dan dewasa adalah stroke. Stroke adalah penyebab utama cerebral palsy bilateral dan triple di seluruh dunia. 70% korban stroke di negara-negara tersebut memiliki tekanan darah tinggi dan masalah jantung, 87% korban stroke di negara-negara yang sama mengalami kecemasan dan depresi[1]. Di seluruh dunia, stroke adalah penyebab kematian dan kecacatan paling umum. Untuk membedakan antara stroke iskemik dan perdarahan otak, diagnosis bergantung pada gambaran klinis dan pencitraan otak [2]. Menurut peneliti [3], Sekitar 15 juta orang menderita stroke setiap tahun, atau 1 dari 6 orang di seluruh dunia, menurut statistik stroke global. Di Indonesia, 43,1% orang berusia 75 tahun ke atas dan 0,2% orang berusia 15 hingga 24 tahun menderita stroke. Stroke yang tidak diobati dapat menyebabkan kematian yang lebih tinggi.

Stroke terjadi ketika satu atau lebih pembuluh darah otak tersumbat atau pecah, mengganggu sebagian atau seluruh fungsi otak. Akibatnya, otak kekurangan darah, oksigen, dan nutrisi yang diperlukan untuk berkembang, yang menyebabkan kematian cepat[2]. Sedangkan, menurut dari [4], Stroke juga dikenal sebagai stroke iskemik akut, yang disebabkan oleh stroke yang terjadi secara perlahan (selama detik) atau cepat (selama kemacetan) sebelum gejala seperti nyeri dan pembengkakan muncul yang berhubungan dengan lokasi stroke yang paling parah.

Algoritma *K-Nearest Neighbors*, sering dikenal sebagai metode *K-NN*, banyak digunakan untuk klasifikasi dan prediksi karena pengajarannya yang menyeluruh. Kekuatan metode *K-NN* meliputi sifatnya yang kuat, intensif, dan tidak asumsif terhadap pencilaan [5]. Menurut [6], Teknik *K-NN* digunakan untuk membandingkan beberapa studi kasus baru dengan studi kasus lama yang didasarkan pada analisis bobot dari berbagai kasus. *Support Vector Machine (SVM)* adalah teknik seleksi algoritmik yang mencapai ambang akurasi klasifikasi tinggi dengan membandingkan seperangkat parameter standar dengan diskerta kandidat, juga dikenal sebagai kualitas yang dirasakannya. Keuntungan dari *Support Vector Machines (SVM)* adalah kemampuannya untuk menangani masalah yang kompleks tanpa terpengaruh oleh berbagai faktor. *SVM* adalah kemampuan untuk melakukan pelatihan secara efektif, and harus berbgaiian dari pembelajaran teknik[7]. *Decision Tree* atau juga dikenal sebagai "pohon keputusan", adalah

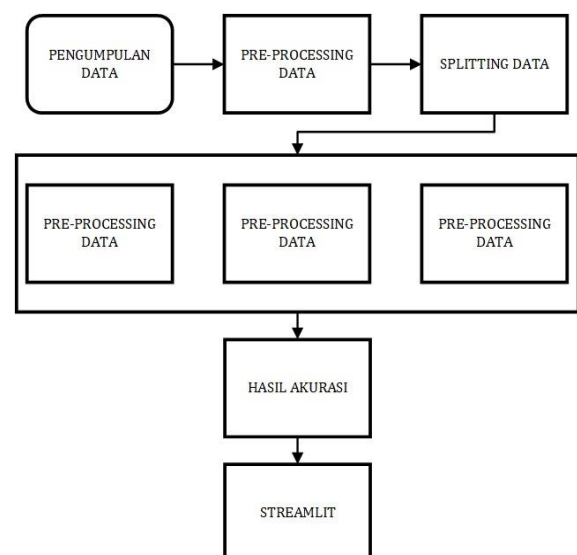
struktur data yang terdiri dari simpul (node) dan rusuk (edge). Pada sebuah pohon, simpul terdiri dari tiga jenis: simpul akar (root node), simpul percabangan (branch node), dan simpul daun (leaf node)[8]. Pendekatan ini bermanfaat untuk mengembangkan pengetahuan berdasarkan data pelatihan dan labelnya, melakukan prediksi terkait kategori atau label kelas, dan melakukan klasifikasi pada data yang baru diperoleh[9].

Berdasarkan penelitian terdahulu yang telah meneliti terkait tentang penyakit stroke, peneliti [10] merangkum hasil pasien stroke di RSUD Abdul Wahab Sjahranie pada bulan November dan Desember 2017 dengan tingkat akurasi 81,25% dengan metode *Naïve bayes* dan tingkat akurasi 87,5% dengan metode *Decision Tree* algoritma [J48]. Algoritma *K-Nearest Neighbors* memiliki nilai ketepatan *Performance Vector* terbaik sebesar 94.36%, sedangkan algoritma *Naïve bayes* memiliki nilai ketepatan terendah sebesar 90.10%, menurut peneliti[3].

Berdasarkan uraian diatas, Dalam penelitian ini, algoritma *K-Nearest Neighbors (KNN)*, *Support Vector Machine (SVM)*, dan *Decision Tree* akan digunakan. Ketiga algoritma tersebut digunakan untuk mengklasifikasikan data tentang penyakit stoke.

2. Metode Penelitian

Pada bagian ini menjelaskan tahap dari penelitian yang dilakukan seperti yang di sajikan pada gambar 1.



Gambar 1. Tahap Penelitian

2.1 Pengumpulan data

Tahap pertama pada penelitian ini adalah pengumpulan data dari situs Kaggle. Dataset yang digunakan merupakan data public. Data penyakit stroke didapatkan pada dataset dari kaggle berjumlah 4982 record.

2.2 Pre-procesing

Pre-Procesing data adalah proses pemilihan data sehingga dapat digunakan dengan cara yang lebih terstruktur dengan menghilangkan atribut yang tidak perlu, membuat data lebih sistematis dan menghilangkan noise[11]. Pembersihan data adalah tahap preprocessing data di mana data hilang dan duplikat diperiksa. Jika ada data hilang, mereka dapat diimputasi dengan mean dan median atau dihapus. Jika ada data duplikat, mereka dapat dihapus[12]. Langkah pertama Sebelum proses klasifikasi dilakukan transformasi data, Transformasi data merubah data category menjadi data numeric.

2.3 Spliting Data

Langkah selanjutnya adalah memisahkan data. Pemisahan data adalah proses membagi data menjadi dua bagian, yaitu data uji dan latih[13]. Menurut peneliti [1], salah satu yang mempengaruhi seberapa baik model klasifikasi berfungsi pada algoritma pembelajaran mesin adalah model *splitting data*, yang di gunakan untuk mempartisi dataset. Sharing data bertujuan untuk memastikan bahwa model yang dibangun dapat diterapkan pada data baru. Dalam proses penelitian ini, algoritma K-NN, SVM, dan Tree of Choice akan digunakan untuk menguji data dengan rasio 80:20.

2.4 K-Nearest Neighbour (K-NN)

Metode *K-NN* merupakan teknik untuk mengelompokan data yang sudah diklasifikasikan. Jika nilai fitur tidak sesuai dengan probabilitas, kinerja algoritma sangat dipengaruhi oleh karakteristik. [13]. Menurut peneliti [14], *K-Nearest Neighbor (KNN)* adalah metode untuk mengurutkan perangkat berdasarkan pengukuran dominasi yang paling dekat dengan objek; ini membutuhkan perangkat kategori sebagai sistem yang mampu menemukan statistik. Untuk menghitung kesamaan yang ada antara kasus baru dan kasus lama, algoritma *K-Nearest Neighbor (K-NN)* digunakan bersama dengan pendekatan untuk menghitung kesamaan yang ada antara kasus

baru dan kasus lama. Prinsip kerja *K-Nearest Neighbor (K-NN)* adalah pendekatan yang didasarkan pada perhitungan bobot dan dalam proses pembelajarannya menggunakan analogi/belajar melalui analogi. Dalam proses klasifikasi, *K-NN* memiliki tabel yang paling baru dibuat dengan algoritma yang memberikan nilai kepada populasi berdasarkan jumlah titik data pengamatan dari variabel k-tailed[15]. Dalam penerapan algoritma *K-NN*, digunakan persamaan 1[16].

$$d_i = \sqrt{\sum_{i=1}^n (X_i - Y_i)^2} \quad (1)$$

Keterangan :

d = jarak kedekatan
i = data variabel
y = data testing
x = data training
n = Dimensi Data

2.5 Support Vector Machine (SVM)

Support Vector Machine (SVM) merupakan algoritma pembelajaran yang diawasi untuk analisis data dan pengenalan pola dalam penambangan data. *Support Vector Machine (SVM)* mengubah dokumen teks menjadi vektor sebelum menggunakannya untuk klasifikasi[17]. Untuk menjadi pemisah kedua kelas data dengan garis batas pemisah yang maksimum, *support vector machine (SVM)* menarik garis yang ideal[18]. Menurut [7], *SVM* memiliki tujuan untuk menemukan hiperrata terbaik yang memungkinkan klasifikasi dan prediksi yang akurat dengan memisahkan secara maksimal kelas-kelas dari titik data. Menurut [19], Metode pengklasifikasian *Support Vector Machine (SVM)* dapat digunakan secara linier atau non-linier. Tujuan dari metode *Support Vector Machine (SVM)* adalah untuk menghitung hasil dari support vector sehingga peneliti hanya perlu mengetahui fungsi kernelnya dan fungsi non-liniernya. Metode ini memungkinkan peneliti untuk menemukan hyperplane yang terbaik dengan menemukan titik yang maksimum dan menghitung margin hyperplane.

Persamaan *Support vector machine (SVM)* :

$$f(x) = w^T \Phi(x) + b \quad (4)$$

Keterangan :

b = Bias
 $x = (x_1, x_2, \dots, x_D)^T$ = Variabel Input

$w = (w_0, w_2, \dots, w_D)^T$ = Parameter Bobot
 $\Phi(x)$ = Fungsi Transformasi Fitur

2.6 Decision Tree

Salah satu pendekatan yang paling umum untuk klasifikasi dan prediksi adalah *Decision Tree*. Struktur pohon digunakan untuk menentukan konsekuensi dan urutan keputusan. Hasilnya adalah variabel yang memprediksi respons atau output Y [20]. Menurut [9], pendekatan ini bermanfaat untuk mengembangkan pengetahuan berdasarkan data latih dan labelnya, melakukan prediksi terkait kategori atau label kelas, dan melakukan klasifikasi pada data yang baru diperoleh. Ada banyak algoritma klasifikasi yang tersedia di domain pembelajaran mesin. Konsep umum algoritma pohon keputusan dibahas dalam skrip ini. Persamaan 2 berikut menggunakan rumus untuk menghitung gain[21].

$$gain(S, A) = Entropy(s) - (x + a)^n - \sum_1^n = \frac{1}{|S|} \sum_i |S_i| \log_2 \frac{|S_i|}{|S|} \quad (2)$$

Keterangan :

S = kumpulan Kasus
 n = Jumlah Kasus Partisi atribut A
 A = Atribut
 $|S_i|$ = Jumlah Kasus Partisi ke i
 $|S|$ = Jumlah Kasus Dalam S

Sementara perhitungan nilai entropy persamaan 3.

$$Entropy(s) = \sum_1^n - p_i \log_2 p_i \quad (3)$$

Keterangan :

S = kumpulan Kasus
 n = Jumlah Partisi S
 A = Fitur
 P_i = Perbandingan dari S_i Terhadap S

2.7 Streamlite

Streamlit adalah kerangka kerja web yang ditujukan untuk menyebarkan model dan

visualisasi dengan mudah menggunakan bahasa Python. Ini cepat dan minimalis, tetapi memiliki tampilan yang cukup baik dan ramah pengguna. Widget bawaan yang dapat digunakan untuk input pengguna termasuk pengunggahan gambar, penggeser, input teks, dan elemen hypertext markup language (HTML) lain yang sudah dikenal, seperti tombol checkbox dan tombol radio[22]. Streamlit menjadi pilihan yang bagus untuk mereka yang baru mengenal dunia data dan ingin membuat proyek yang dapat digunakan kapan saja. Paket ini membuat pembelajaran mesin dan ilmu data menjadi lebih mudah diterapkan[23].

3. Analisa Hasil dan Pembahasan

3.1 Pengumpulan Data

Studi ini menggunakan data dari tahun 2021. Data ini berasal dari <https://www.kaggle.com/datasets/zzettrkalpakbal/full-filled-brain-stroke-dataset>, dengan 11 atribut, dan totalnya 4981 data. Dataset penyakit stroke yang disajikan pada Tabel 1.

Keterangan :

- Gender : male atau female.
- Age : umur.
- Hypertension : 0 pasien tidak menderita hipertensi, 1 pasien menderita hipertensi.
- heart_disease : 0 pasien menderita penyakit jantung.
- ever_married : No atau Yes.
- work_type : children, govt, private, never worked, private, atau self employed.
- Residence_type : Rural atau Urban.
- avg_glucose_level : rata-rata kadar glukosa.
- Bmi : indeks masa tubuh.
- Smoking_status : formerly smoked, never smoked, smoke atau unknown.
- Stroke : satu pasien mengalami stroke atau nol jika tidak.

TABEL I. DATASET PENYAKIT STROKE

gender	Age	hypertension	heart_disease	ever_married	work_type	Residence_type	avg_glucose_level	Bmi	Smoking_status	Stroke
Male	67	0	1	Yes	Private	Urban	228.69	36.6	Formerly Smoked	1
Male	80	0	1	Yes	Private	Rural	105.92	32.5	Never Smoked	1
Female	49	0	1	Yes	Private	Urban	171.23	34.4	Smoke	1
Female	79	1	0	Yes	Self Employed	Rural	174.12	24	Never Smoke	1
Male	81	0	0	Yes	Private	Urban	186.21	29	Formerly Smoke	1

3.2 Preprocessing Data

Sebelum proses klasifikasi dilakukan transformasi data, Transformasi data merubah data category menjadi data numeric. Berdasarkan

Hasilnya, nilai atribut *ever_married*, *work_type*, *residence_type*, dan *smoking_status* diubah menjadi "0" dan "1". Tabel 2 menunjukkan transformasi data.

TABEL II. DATASET PENYAKIT STROKE

gender	Age	hypertension	heart_disease	ever_married	work_type	Residence_type	avg_glucose_level	Bmi	Smoking_status	Stroke
1	67.0	0	1	1	2	1	228.69	36.6	1	1
1	80.0	0	1	1	2	0	105.92	32.5	2	1
1	49.0	0	0	1	2	1	171.23	34.4	3	1
0	79.0	1	0	1	3	0	174.12	24.0	2	1
0	81.1	0	0	1	2	1	186.21	29.0	1	1

3.3 Klasifikasi

setelah melakukan tahapan preprocessing data dengan menggunakan transformasi data, Selanjutnya kita melakukan tahapan klasifikasi dengan menggunakan spliting data 80:20 pada algoritma KNN, SVM Dan Decision Tree.

	precision	recall	f1-score	support
0	0.95	0.99	0.97	943
1	0.00	0.00	0.00	54
accuracy			0.94	997
macro avg	0.47	0.50	0.48	997
weighted avg	0.89	0.94	0.92	997

Gambar 2. Hasil Menggunakan Metode K-NN

Berdasarkan Gambar 2, hasil yang diperoleh selama tahap penerapan metode K-NN menunjukkan tingkat akurasi sebesar 94%.

	precision	recall	f1-score	support
0	0.95	1.00	0.97	929
1	0.00	0.00	0.00	53
accuracy			0.95	982
macro avg	0.47	0.50	0.49	982
weighted avg	0.89	0.95	0.92	982

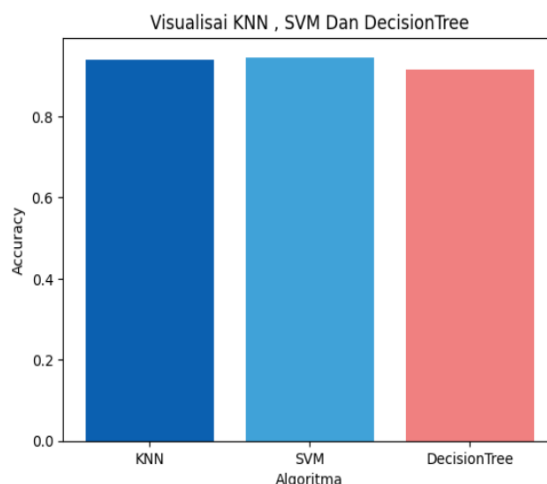
Gambar 3. Hasil Menggunakan Metode SVM

Berdasarkan Gambar 3, adapun nilai akurasi yang dihitung menggunakan metode SVM mendapatkan hasil akurasi sebesar 95%.

	precision	recall	f1-score	support
0	0.95	0.96	0.96	943
1	0.20	0.19	0.19	54
accuracy			0.92	997
macro avg	0.58	0.57	0.58	997
weighted avg	0.91	0.92	0.91	997

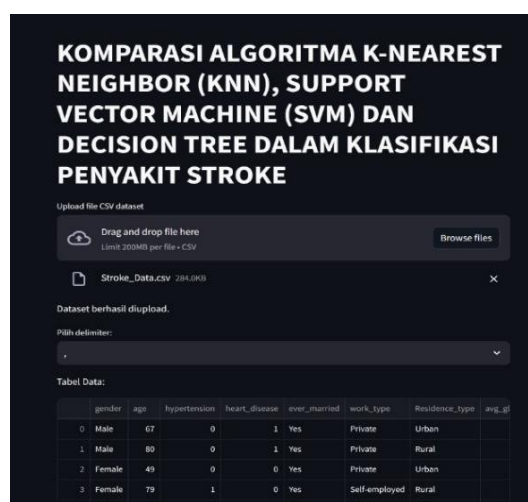
Gambar 4. Hasil Menggunakan Metode Decision Tree

Sedangkan berdasarkan Gambar 4, nilai akurasi menggunakan metode Decision Tree mendapatkan hasil sebesar 92%.



Gambar 5. Visualisasi Hasil Perbandingan

Pada Gambar 5, merupakan hasil perbandingan algoritma machine learning, accuracy dari KNN 94% dan SVM mendapatkan hasil yaitu 95% sedangkan accuracy dari metode Decision Tree mendapatkan hasil rendah yaitu 92%.



Gambar 6. User interface dengan streamlit

Pada gambar 6, merupakan antar muka pengguna (user interface) untuk aplikasi web. Di mana menampilkan hasil visualisasi dari akurasi metode K-NN, SVM dan Decision tree.

4. Hasil

TABEL III. HASIL AKURASI

Method	Acuracy	Precision	Recall
KNN	94%	0,95%	0,99%
SVM	95%	0,95%	1,00%

Decision Tree	92%	0,95%	0,96%
---------------	-----	-------	-------

Pada tabel 3, setelah melakukan evaluasi terhadap beberapa *splitting data* penelitian ini mengambil nilai *splitting data* terbaik yaitu 80:20. hasil akurasi dari tiga metode *K-NN*, *SVM* Dan *Decision Tree* mendapatkan hasil *accuracy* 0.94, *precision* 0.95, dan *recall* 0.99 pada algoritma *K-NN*, Sedangkan *SVM* mendapatkan *accuracy* 0.95, *precision* 0.95, *recall* 1.00 Dan *Decision Tree* hasil *accuracy* 0.92, *precision* 0.95, *recall* 0.96.

5. Kesimpulan dan Saran

Menurut hasil penelitian, tiga algoritma *Decision Tree*, *K-NN*, dan *SVM* digunakan. dari tiga algoritma menggunakan *splitting data* 80:20 dapat disimpulkan tingkat hasil akurasi terbaik terdapat pada algoritma *KNN* mendapatkan akurasi 94% dan *SVM* dengan tingkat akurasi 95% sedangkan algoritma *Decision Tree* memperoleh tingkat akurasi 92%. Pada penelitian ini, metode *SVM* menghasilkan hasil klasifikasi yang lebih baik diantara tiga metode.

Saran untuk peneliti selanjutnya, agar mencoba melakukan perbandingan klasifikasi dengan algoritma yang lain. Supaya dapat dilihat algoritma mana yang lebih bagus dan akurat.

Daftar Pustaka:

- [1] N. Aliffiyanti Iskandar, I. Ernawati, and Y. Widiastiw, "Klasifikasi Diagnosis Penyakit Stroke Dengan Menggunakan Metode Random Forest," 2022. [Online]. Available: <https://www.kaggle.com/fedesoriano/stroke>
- [2] A. Byna and M. Basit, "PENERAPAN METODE ADABOOST UNTUK MENOPTIMASI 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.
- [3] K. Akmal, A. Faqih, and F. Dikananda, "PERBANDINGAN METODE ALGORITMA NAÏVE BAYES DAN K-NEAREST NEIGHBORS UNTUK KLASIFIKASI PENYAKIT STROKE," 2023. [Online]. Available: www.researchgate.net
- [4] N. D. Saputri, K. Khalid, and D. Rolliawati, "Comparison of Bagging and Adaboost Methods on C4.5 Algorithm for Stroke Prediction," *SISTEMASI*, vol. 11, no. 3, p. 567, Sep. 2022, doi: 10.32520/stmsi.v11i3.1684.
- [5] Y. Pratama, A. Prayitno, D. Nazrian, N. Aini, Y. Rizki, and E. Rasywir, "BULLETIN OF COMPUTER SCIENCE RESEARCH Klasifikasi Penyakit Gagal Jantung Menggunakan Algoritma K-Nearest Neighbor," 2022, doi: 10.47065/bulletincsr.v3i1.203.
- [6] D. Efriadi, R. Rahmadden, A. Agustin, and J. Junadhi, "Prediksi Penambahan Piutang Iuran Jaminan Sosial Ketenagakerjaan menggunakan Algoritma K-Nearest Neighbor," *Edumatic: Jurnal Pendidikan Informatika*, vol. 6, no. 1, pp. 49–57, Jun. 2022, doi: 10.29408/edumatic.v6i1.5255.
- [7] R. abibah br Lumbantobing, "Prediksi Harga Cryptocurrency Menggunakan Algoritma Support Vector Machine," 2023.
- [8] R. Ridho, "KLASIFIKASI DIAGNOSIS PENYAKIT COVID-19 MENGGUNAKAN METODE DECISION TREE," 2021. [Online]. Available: <https://jurnal.umj.ac.id/index.php/just-it/index>
- [9] D. P. Sinambela, H. Naparin, M. Zulfadhilah, and N. Hidayah, "Implementasi Algoritma Decision Tree dan Random Forest dalam Prediksi Perdarahan Pascasalin," *Jurnal Informasi dan Teknologi*, vol. 5, no. 3, pp. 58–64, Sep. 2023, doi: 10.60083/jidt.v5i3.393.
- [10] A. Wahab, S. Samarinda, 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 Comparison of the Classification for Naive Bayes Method and the Decision Tree Algorithm (J48) for Stroke Patients in Abdul Wahab Sjahranie Samarinda Hospital," *Jurnal EKSPONENSIAL*, vol. 10, no. 2, 2019.
- [11] M. Nurkholifah, Jasmarizal, Y. Umar, and Rahmadden, "ANALISA PERFORMA ALGORITMA MACHINE LEARNING DALAM PREDIKSI PENYAKIT LIVER," *Jurnal Indonesia : Manajemen Informatika dan Komunikasi*, vol. 4, no. 1, pp. 164–172, Jan. 2023, doi: 10.35870/jimik.v4i1.149.
- [12] N. Yolanda Paramitha *et al.*, "Klasifikasi Penyakit Stroke Menggunakan Metode Naive Bayes," 2023. [Online]. Available: <https://www.kaggle.com/datasets/zzettrkalpakbal/full-filled->
- [13] A. Putri, C. Syaficha Hardiana, E. Novfuja, F. Try Puspa Siregar, Y. Fatma, and R.

- Wahyuni, "Comparison of K-NN, Naive Bayes and SVM Algorithms for Final-Year Student Graduation Prediction Komparasi Algoritma K-NN, Naive Bayes dan SVM untuk Prediksi Kelulusan Mahasiswa Tingkat Akhir," *Institut Riset dan Publikasi Indonesia (IRPI) MALCOM: Indonesian Journal of Machine Learning and Computer Science Journal Homepage*, vol. 3, no. 1, pp. 20–26, 2023.
- [14] M. Ikhsan and R. Syair Habibi, "PENERAPAN SISTEM CERDAS BERBASIS CASE BASE BASED REASONING (CBR) DAN METODE K-NEAREST NEIGHBOR UNTUK IDENTIFIKASI MASALAH DATA CENTER," 2023. [Online]. Available: <http://e-journal.stmiklombok.ac.id/index.php/jire> ISSN.2620-6900
- [15] S. N. Luqman *et al.*, "Komparasi Algoritma Klasifikasi Genre Musik pada Spotify Menggunakan CRISP-DM," 2021.
- [16] N. Indriyani, E. Ali STMIK Amik Riau, and U. Rio STMIK Amik Riau, "SATIN-Sains dan Teknologi Informasi Menentukan Kualitas Pelayanan Maskapai Penerbangan Domestik Dengan Menggunakan Metode Naïve Bayes," vol. 6, no. 1, 2020, [Online]. Available: <http://hubud.dephub.go.id/website/BandaraDet>
- [17] D. Septhya, K. Rahayu, S. Rabbani, V. Fitria, Y. Irawan, and R. Hayami, "MALCOM: Indonesian Journal of Machine Learning and Computer Science Implementation of Decision Tree Algorithm and Support Vector Machine for Lung Cancer Classification Implementasi Algoritma Decision Tree dan Support Vector Machine untuk Klasifikasi Penyakit Kanker Paru," vol. 3, pp. 15–19, 2023.
- [18] S. S. T. E. A. R. , A. N. Ade Silvia Handayani1, "KLASIFIKASI KUALITAS UDARA DENGAN METODE SUPPORT VECTOR MACHINE," *JIRE (Jurnal Informatika & Rekayasa Elektronik)* , 2020.
- [19] J. I. Matematika and S. Adi, "MATHunesa Tahun 2022 KOMPARASI METODE SUPPORT VECTOR MACHINE (SVM), K-NEAREST NEIGHBORS (KNN), DAN RANDOM FOREST (RF) UNTUK PREDIKSI PENYAKIT GAGAL JANTUNG Atik Wintarti".
- [20] A. Irzky Shafarindu, E. Patimah, Y. Marintan Siahaan, A. Wisnu Wardhana, B. Vicky Haekal, and D. Sandya Prasvita, *Klasifikasi Data Penjualan pada Supermarket dengan Metode Decision Tree*. 2021. [Online]. Available: https://www.kaggle.com/aungpyaeap/supermarket-sales?select=supermarket_sales+-+Sheet1.csv.
- [21] F. Akbar, H. Wira Saputra, A. Karel Maulaya, and M. Fikri Hidayat, "MALCOM: Indonesian Journal of Machine Learning and Computer Science Implementation of Decision Tree Algorithm C4.5 and Support Vector Regression for Stroke Disease Prediction Implementasi Algoritma Decision Tree C4.5 dan Support Vector Regression untuk Prediksi Penyakit Stroke," vol. 2, pp. 61–67, 2022.
- [22] W. Hastomo, N. Aini, A. Satyo, B. Karno, and L. M. R. Rere, "Metode Pembelajaran Mesin untuk Memprediksi Emisi Manure Management," 2022.
- [23] R. Safri Irawansyah, L. A. Syamsul Irfan A, and G. Wahyu Wiriasto, "ANALISIS SENTIMEN TERHADAP PROGRAM MERDEKA BELAJAR-KAMPUS MERDEKA (MBKM) PADA TWITTER MENGGUNAKAN ALGORITMA NAIVE BAYES CLASSIFIER (NBC)," 2010.