

**KOMPARASI METODE KLASIFIKASI *DECISION TREE*
ALGORITMA C4.5 DAN *RANDOM FOREST*
UNTUK PREDIKSI PENYAKIT STROKE**

SKRIPSI

**Diajukan Sebagai Salah Satu Syarat untuk Memperoleh Gelar
Sarjana Matematika pada Fakultas MIPA**

Oleh :

**NUR AZIZAH
NIM 08011181722011**



**JURUSAN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS SRIWIJAYA
2021**

LEMBAR PENGESAHAN

KOMPARASI METODE KLASIFIKASI *DECISION TREE* ALGORITMA C4.5 DAN *RANDOM FOREST* UNTUK PREDIKSI PENYAKIT STROKE

SKRIPSI

Diajukan Sebagai Salah Satu Syarat untuk Memperoleh Gelar Sarjana
di Jurusan Matematika pada Fakultas MIPA

Oleh:

NUR AZIZAH
NIM. 08011181722011

Pembimbing Pembantu

Endang Sri Kresnawati, M.Si
NIP. 19770208 200212 2 003

Indralaya, November 2021

Pembimbing Utama

Dr. Yulia Resti, M.Si
NIP. 19730719 199702 2 001



HALAMAN PERSEMBAHAN

MOTTO

**“Kemarin adalah sejarah dan Esok adalah harapan.
Namun yang paling penting adalah
Hari ini sebuah kenyataan yang harus dijalani sebaik-baiknya”**

-Unknown-

Skripsi ini ku persembahkan kepada :

- 1. Allah SWT**
- 2. Mamak dan Abah**
- 3. Saudaraku**
- 4. Keluarga besar**
- 5. Dosen**
- 6. Sahabat dan Temanku**
- 7. Almamater**

KATA PENGANTAR

Assalamu'alaikum Warahmatullahi Wabarakatuh

Puji syukur atas kehadirat Allah SWT karena atas limpahan rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul **“Komparasi Metode Klasifikasi Decision Tree Algoritma C4.5 dan Random Forest untuk Prediksi Penyakit Stroke”**. Sholawat serta salam semoga selalu tercurahkan kepada junjungan kita Nabi Muhammad SAW beserta keluarga, sahabat, dan pengikutnya hingga akhir zaman. Skripsi ini sebagai salah satu syarat untuk memperoleh gelar Sarjana Sains jurusan Matematika di Fakultas Matematika dan Ilmu Pengetahuan Alam.

Terima kasih yang sebesar-besarnya penulis ucapkan kepada orang tua tercinta, yaitu **Bapak Rosichin Darmo** dan **Ibu Sopiah** serta **Ibu Tuliah** yang tak pernah lupa mendoakan yang terbaik untuk penulis, telah mendidik dan merawat dengan baik serta memberikan kasih sayang dan selalu memberikan dukungan apapun yang penulis pilih. Penulisan skripsi ini tidak terlepas dari bantuan berbagai pihak, baik secara langsung maupun tidak langsung. Penulis juga menyampaikan ucapan terima kasih kepada :

1. Bapak **Drs. Sugandi Yahdin, M.M** selaku Ketua Jurusan Matematika serta Ketua Seminar skripsi atas ilmu yang telah diberikan dan memberikan arahan, saran, serta mengatur jalannya seminar sehingga dapat berjalan dengan baik.

2. Ibu **Dr. Dian Cahyawati Sukanda, M.Si** selaku Sekretaris Jurusan Matematika yang telah membantu dalam proses penjadwalan seminar tugas akhir ini.
3. Ibu **Dra. Ning Eliyati, M. Pd** selaku Sekretaris Seminar skripsi yang telah membantu dan memberikan catatan masukan bagi penulis.
4. Ibu **Dr. Yulia Resti, M.Si** selaku Dosen Pembimbing Utama yang telah membimbing, meluangkan waktu serta pikiran dalam mengarahkan penulis agar dapat menyelesaikan skripsi dengan baik.
5. Ibu **Endang Sri Kresnawati, M.Si** selaku Dosen Pembimbing Pembantu yang telah membimbing, meluangkan waktu serta pikiran dalam mengarahkan penulis agar dapat menyelesaikan skripsi dengan baik.
6. Ibu **Des Alwine Zayanti, M.Si** selaku Dosen Pembahas yang telah memberikan tanggapan, saran, serta masukan agar skripsi ini dapat diselesaikan lebih baik.
7. Ibu **Novi Rustiana Dewi, M.Si** selaku Dosen Pembahas yang telah memberikan tanggapan, saran, serta masukan agar skripsi ini dapat diselesaikan lebih baik.
8. **Seluruh Dosen di Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Sriwijaya** atas ilmu yang telah diberikan.
9. Bapak **Irwansyah** selaku admin dan Ibu **Hamidah** selaku pegawai Tata Usaha Jurusan Matematika FMIPA.

10. Saudara laki-laki yang saya banggakan **Mas Agus, Mas Andi, Mas Ali**, dan saudara kembar saya **Azizi** yang sama-sama berjuang menyelesaikan studi, serta **keponakan saya**.
11. Sahabat-sahabat tim skripsi **Oliv, Mega, Kalim, Tesya, Mufhlika, Abu, Agung, Yuda, Rendy, Wawan**, dan **Shohif**. Sahabat-sahabatku di jurusan **Meta, Rina, Rizma**, dan **Nabilah** dan teman-teman **Angkatan 17** yang tidak bisa disebutkan satu persatu.
12. Para sahabat semasa sekolah **Nurul, Rye, Ermawati, Lili, Syarah, Vira, Adel**, dan **Riska** yang telah saling memberikan dukungan satu sama lain serta teman-teman satu daerah di **IKMS**.
13. Semua pihak yang telah membangun penulis dalam menyelesaikan skripsi ini yang tidak dapat disebutkan satu persatu.

Wassalamu'alaikum Warahmatullahi Wabarakatuh.

Indralaya, Oktober 2021

Nur Azizah
NIM.08011181722011

**CLASSIFICATION METHOD COMPARISON OF DECISION TREE
ALGORITHM C4.5 AND RANDOM FOREST
FOR STROKE PREDICTION**

By :

**Nur Azizah
08011181722011**

ABSTRACT

Stroke is a disease that disrupts the nervous system of the human brain. Stroke disease as one of the leading causes of death and serious disability with a high possibility of becoming an epidemic in the world in the next few decades. Therefore, it is necessary to predict as a first step in anticipating the occurrence of stroke in order to prevent or minimize the occurrence of disability. This research uses secondary data obtained from kaggle.com. This data has 11 variables and 5110 data. Prediction of occurrence using the Decision Tree and Random Forest methods. In the Random Forest method, 120 trees were built. The results of this research are the accuracy of the decision tree of 92.56% and the random forest of 93.80%. Precision decision tree is 95.45% and random forest is 98.76%. Recall decision tree is 97.09% and random forest is 94.91%.

Keywords : Stroke, Decision Tree, Random Forest

**KOMPARASI METODE KLASIFIKASI *DECISION TREE*
ALGORITMA C4.5 DAN *RANDOM FOREST*
UNTUK PREDIKSI PENYAKIT *STROKE***

Oleh :

**Nur Azizah
08011181722011**

ABSTRAK

Stroke merupakan penyakit yang menganggu sistem saraf otak manusia. Penyakit *stroke* tercatat sebagai salah satu penyebab utama kematian dan kecacatan serius dengan kemungkinan besar menjadi epidemi di dunia dalam beberapa dekade mendatang. Oleh karena itu perlu dilakukan prediksi sebagai langkah awal dalam mengantisipasi terjadinya *stroke* guna pencegahan atau meminimalisir terjadinya kecacatan. Penelitian ini menggunakan data sekunder yang diperoleh dari *kaggle.com* dimana memiliki 11 variabel dan 5110 data. Prediksi terjadinya *stroke* menggunakan metode *Decision Tree* dan *Random Forest*. Pada metode *Random Forest* pohon yang dibangun sebanyak 120 pohon. Hasil dari penelitian ini yaitu *accuracy decision tree* sebesar 92.56% dan *random forest* sebesar 93.80%. *Precision decision tree* sebesar 95.45% dan *random forest* sebesar 98.76%. *Recall decision tree* sebesar 97.09% dan *random forest* sebesar 94.91%.

Kata Kunci : *Stroke, Decision Tree, Random Forest.*

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PENGESAHAN.....	ii
HALAMAN PERSEMAHAN	iii
KATA PENGANTAR.....	iv
ABSTRACT	vii
ABSTRAK	viii
DAFTAR ISI.....	ix
DAFTAR TABEL	xi
DAFTAR GAMBAR.....	xii
DAFTAR LAMPIRAN	xiii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Perumusan Masalah	4
1.3 Batasan Masalah.....	5
1.4 Tujuan	5
1.5 Manfaat	6
BAB II TINJAUAN PUSTAKA.....	7
2.1 Penyakit <i>Stroke</i>	7
2.2 Data <i>Mining</i>	7
2.2.1 <i>Machine Learning</i>	8
2.3 <i>Preprocessing Data</i>	9
2.4 Klasifikasi	9
2.5 <i>Decision Tree</i>	9
2.5.1 Algoritma C4.5	10
2.6 <i>Random Forest</i>	11
2.7 <i>Confusion Matrix</i>	14
BAB III METODOLOGI PENELITIAN	16
3.1 Tempat.....	16
3.2 Waktu	16
3.3 Metode Penelitian.....	16
BAB IV HASIL DAN PEMBAHASAN	19

4.1.	Deskripsi Data.....	19
4.2.	Diskritisasi Data	20
4.3.	Penyajian Data	20
4.4.	Partisi Data.....	22
4.5.	Metode <i>Decision Tree</i>	23
4.6.	Metode <i>Random Forest</i>	37
4.7.	Analisis Hasil	49
BAB V	KESIMPULAN DAN SARAN	50
5.1	Kesimpulan	50
5.2	Saran.....	50
DAFTAR PUSTAKA	52	
LAMPIRAN	55	

DAFTAR TABEL

Tabel 2.1 <i>Confusion matrix</i>	14
Tabel 4.1 Deskripsi variabel	19
Tabel 4.2 Diskritisasi data.....	20
Tabel 4. 3 Penyajian variabel dependen.....	20
Tabel 4.4 Penyajian variabel independen.....	21
Tabel 4.5 Data <i>training</i>	22
Tabel 4.6 Data <i>testing</i>	22
Tabel 4.7 Perhitungan <i>entropy</i> dan <i>gain internal node</i> 1.1 metode <i>decision tree</i>	29
Tabel 4.8 Perhitungan <i>entropy</i> dan <i>gain internal node</i> 1.1.4 metode <i>decision tree</i>	31
Tabel 4.9 Perhitungan <i>entropy</i> dan <i>gain internal node</i> 1.1.4.1 metode <i>decision tree</i>	33
Tabel 4.10 Prediksi metode <i>decision tree</i>	35
<i>Tabel 4.11 Confusion matrix Decision Tree</i>	36
Tabel 4.12 Sampel pohon ke-1	37
Tabel 4.13 Perhitungan <i>entropy</i> dan <i>gain internal node</i> 1.1. metode <i>random forest</i>	41
Tabel 4.14 Perhitungan <i>entropy</i> dan <i>gain internal node</i> 1.1.4.....	43
Tabel 4.15 Perhitungan <i>entropy</i> dan <i>gain internal node</i> 1.1.4.1 metode <i>random forest</i>	44
Tabel 4.16 Prediksi metode <i>random forest</i>	47
Tabel 4.17 <i>Confusion matrix random forest</i>	47
Tabel 4.18 Perbandingan tingkat ketepatan klasifikasi.....	49

DAFTAR GAMBAR

Gambar 2.1 Metode <i>random forest</i>	12
Gambar 4.1 Pohon keputusan node pertama (<i>root node</i>).....	28
Gambar 4.2 Pohon keputusan <i>internal node</i> 1.1 metode <i>decision tree</i>	30
Gambar 4.3 Pohon keputusan <i>internal node</i> 1.1.4 metode <i>decision tree</i>	32
Gambar 4.4 Pohon keputusan <i>internal node</i> 1.1.4.1 metode <i>decision tree</i>	34
Gambar 4.5 Pohon keputusan <i>root node</i> metode <i>random forest</i>	41
Gambar 4.6 Pohon keputusan <i>internal node</i> 1.1 metode <i>random forest</i>	42
Gambar 4.7 Pohon keputusan <i>internal node</i> 1.1.4 metode <i>random forest</i>	43
Gambar 4.8 Pohon keputusan <i>terminal node</i> 1.1.4.1 metode <i>random forest</i>	45

DAFTAR LAMPIRAN

Lampiran 1	55
Lampiran 2	70

BAB I

PENDAHULUAN

1.1 Latar Belakang

Penyakit *stroke* merupakan penyakit yang mengganggu sistem saraf otak manusia. Terjadinya *stroke* dikarenakan suatu kondisi terganggunya pasokan darah ke otak yang mengakibatkan area otak yang rusak tidak dapat berfungsi dengan baik dalam mengendalikan bagian tubuh (Sakinah dkk., 2020). Penyakit *stroke* tercatat sebagai salah satu penyebab utama kematian dan kecacatan serius yang mempengaruhi jutaan nyawa manusia dengan kemungkinan besar menjadi epidemi di dunia dalam beberapa dekade mendatang. Pengambilan keputusan yang tepat terkait penyakit *stroke* menjadi peran utama yang dapat mengurangi kemungkinan kerusakan otak, kelumpuhan, dan lainnya (Thippa Reddy et al., 2020).

Pentingnya mengetahui gejala penyakit sebagai langkah awal dalam mengantisipasi adanya suatu penyakit yang membahayakan kesehatan atau bahkan menyebabkan kematian. Melakukan deteksi dini penyakit *stroke* berguna dalam pencegahan atau meminimalisir terjadinya kecacatan (Ahmed et al., 2019). Deteksi dini penyakit *stroke* dapat dilakukan melalui beberapa cara salah satunya data *mining*.

Data *mining* merupakan proses untuk memperoleh informasi pada suatu database besar yang mana dalam prosesnya menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* (Mardi, 2017). Salah satu teknik data *mining* yakni klasifikasi yang merupakan dasar dari analisis data

(Kurniawan, 2018). Menurut Bansal, Sharma and Goel (2017) klasifikasi merupakan teknik data *mining* yang berhubungan dengan *machine learning* dan merupakan suatu teknik untuk memprediksi keanggotaan kelompok berdasarkan data yang sudah ada.

Metode klasifikasi yang kerap kali digunakan dalam prediksi yakni *Decision Tree* dan *Random Forest*. *Decision Tree* merupakan metode klasifikasi dan prediksi yang cukup terkenal. *Decision tree* atau pohon keputusan merupakan sebuah diagram alir menyerupai pohon yang memiliki sebuah akar (*root node*) dan sebuah daun (*leaf node*). *Decision tree* digunakan untuk mengklasifikasikan suatu sampel data yang belum diketahui kelasnya ke dalam kelas-kelas yang ada dari sejumlah variabel bebas dengan sebuah variabel terikat, sehingga metode *decision tree* ini akan mengubah fakta yang besar menjadi sebuah pohon keputusan yang mempunyai aturan (Aziz dan Karpen, 2019).

Metode *random forest* merupakan metode yang digunakan pada klasifikasi data berdasarkan teknik pohon keputusan. Proses dari metode *random forest* merupakan adaptasi dari metode *decision tree* yang dikembangkan berdasarkan sampel *bootstrap* (Renata dan Ayub, 2020). Metode *random forest* menggunakan banyak pohon yang dihasilkan sehingga terbentuk kumpulan pohon atau hutan (*forest*) (Dewi dkk., 2011). Ketepatan klasifikasi menggunakan metode *random forest* akan lebih baik dibanding metode *decision tree* karena hasil dari klasifikasi tidak hanya menggunakan satu pohon keputusan melainkan banyak pohon (Nugroho dan Emiliyawati, 2017).

Manfaat utama menggunakan pohon keputusan yakni dapat menyelesaikan proses pengambilan keputusan yang kompleks menjadi lebih sederhana sehingga pengambilan keputusan akan lebih menjelaskan solusi masalah dengan mengubah bentuk data (tabel) menjadi sebuah model pohon keputusan (Damanik et al., 2019).

Beberapa algoritma yang dapat digunakan dalam membuat pohon keputusan antara lain ID3, C4.5, dan C5.0. Algoritma C4.5 merupakan pengembangan dari algoritma ID3 dimana kekurangan yang dimiliki oleh algoritma ID3 dapat ditutupi oleh algoritma C4.5 (Sulistiani and Aldino, 2020). Kelebihan algoritma C4.5 yakni memperoleh pohon keputusan yang mempunyai tingkat akurasi yang baik dan efisien dalam menangani variabel yang bertipe diskrit atau numerik (Pengaribuan et al., 2019).

Berdasarkan penelitian terdahulu yang menerapkan metode klasifikasi algoritma C4.5 yang dilakukan oleh (Rohman dkk., 2020) dalam penelitiannya terkait komparasi algoritma C4.5 berbasis PSO dan GA untuk diagnosa penyakit *stroke*. Data penelitian yang digunakan adalah dataset penyakit *stroke* yang diperoleh dari website *kaggle*. Hasil dari penelitian ini diperoleh *accuracy* dari algoritma C4.5 sebesar 89,93%, algoritma C4.5-PSO sebesar 91,63%, dan algoritma C4.5-GA sebesar 92,02%.

Penelitian lainnya yang dilakukan oleh (Ahmed et al., 2019) terkait prediksi *stroke* menggunakan distribusi *machine learning* berbasis *apache park*. Data penelitian yang digunakan adalah *healthcare dataset stroke* dari website *kaggle* dan dilakukan klasifikasi menggunakan metode *random forest*, *decision tree*, *support vector machine*, dan *logistic regression*. Diperoleh akurasi tertinggi yakni metode

random forest dengan nilai *accuracy* sebesar 90%, *decision tree* sebesar 79%, *support vector machine* dan *logistic regression* sebesar 77%. Oleh karena itu penulis mencoba melakukan penelitian untuk melihat perbandingan hasil klasifikasi kedua metode yakni *Decision Tree* Algoritma C4.5 dan *Random Forest* dalam pengklasifikasian penyakit *stroke* dengan menghitung nilai *accuracy*, *precision*, dan *recall* sebagai tingkat ketepatan klasifikasi dalam memprediksi penyakit *stroke*.

1.2 Perumusan Masalah

Adapun rumusan masalah pada penelitian ini adalah :

1. Bagaimana mengidentifikasi penyakit *stroke* dengan metode *decision tree* algoritma C4.5 dan *random forest* ?
2. Bagaimana ukuran ketepatan klasifikasi menggunakan metode *decision tree* algoritma C4.5 dan *random forest* ?

1.3 Batasan Masalah

Adapun batasan masalah pada penelitian ini adalah :

1. *Stroke Prediction Dataset* diperoleh dari website *kaggle.com* yang dipublish pada Februari 2021. Dataset terdiri dari 5110 sampel dengan variabel-variabelnya seperti jenis kelamin, usia, riwayat hipertensi, riwayat penyakit jantung, status menikah, tempat tinggal, jenis pekerjaan, kadar gula darah, BMI, status merokok, dan *stroke*.
2. Data dipartisi menjadi 70% data *training* dan 30% data *testing*.
3. Tingkat ketepatan klasifikasi dibatasi oleh nilai *accuracy*, *precision*, dan *recall*.

1.4 Tujuan

Adapun tujuan dari penelitian ini adalah :

1. Memprediksi penyakit *stroke* menggunakan metode *decision tree* algoritma C4.5 dan *random forest*.
2. Mengetahui perbandingan ukuran ketepatan klasifikasi dalam memprediksi penyakit *stroke* menggunakan metode *decision tree* algoritma C4.5 dan *random forest*.

1.5 Manfaat

Adapun manfaat dari penelitian ini adalah :

1. Sebagai bahan referensi bagi penelitian selanjutnya mengenai prediksi penyakit *stroke*.
2. Sebagai media pembelajaran terkait penggunaan metode *decision tree* algoritma C4.5 dan *random forest* dalam melakukan prediksi.

DAFTAR PUSTAKA

- Ahmad, A. (2017). Mengenal Artificial Intelligence, Machine Learning, Neural Network, dan Deep Learning. *Jurnal Teknologi Indonesia*.
- Ahmed, H., (2019). Stroke Prediction Using Distributed Machine Learning Based on Apache Spark. *International Journal of Advanced Science and Technology*, 28(15), 89–97.
- Aziz, A., dan Karpen. (2019). Diagnosa Penyakit Kulit Wajah Menggunakan Metode Decession Tree Dan Algoritma C4.5. *Jurnal Teknologi Dan Open Source*, 2(1), 74–86.
- Bansal, A., Sharma, M., and Goel, S. (2017). Improved K-mean Clustering Algorithm for Prediction Analysis using Classification Technique in Data Mining. *International Journal of Computer Applications*, 157(6), 35–40.
- Damanik, et al., (2019). Decision Tree Optimization in C4.5 Algorithm Using Genetic Algorithm. *Journal of Physics: Conference Series*, 1255(1).
- Darmawan, E. (2018). C4.5 Algorithm Application for Prediction of Self Candidate New Students in Higher Education. *Jurnal Online Informatika*, 3(1), 22.
- Dewi dkk., (2011). Penerapan Metode Random Forest dalam Driver Analysis (The Application of Random Forest in Driver Analysis). *Forum Statistika Dan Komputasi*, 16(1), 35–43.
- Eska, J. (2018). Penerapan Data Mining Untuk Prediksi Penjualan Wallpaper Menggunakan Algoritma C4.5. *Jurnal Teknologi Dan Sistem Informasi*, 2, 9–13.
- Ghorbani, R., and Ghousi, R. (2019). Predictive Data Mining Approaches in Medical Diagnosis: A review of Some Diseases Prediction. *International Journal of Data and Network Science*, 3(2), 47–70.
- Harryanto, F. F., dan Hansun, S. (2017). Penerapan Algoritma C4.5 untuk Memprediksi Penerimaan Calon Pegawai Baru di PT WISE. *Teknik Informatika Dan Sistem Informasi*, 3(2), 95–103.
- Hartono dkk., (2019). Gambaran Tekanan Darah Pada Pasien Stroke Hemoragik Dengan Diabetes Melitus Dan Non Diabetes Melitus Di Bagian Saraf Rumkital Dr.Ramelan Surabaya. *Journal of Chemical Information and Modeling*, 53(9), 1689–1699.

- Hastie, et al., (2001). *The Elements of Statistical Learning*. Springer.
- Indahsari, R. D., dan Sindunata, Y. (2014). Penerapan Data Mining Untuk Analisa Pola Perilaku Nasabah Dalam Pengkreditan Menggunakan Metode C.45 Studi Kasus Pada KSU Insan Kamil Demak. *Jurnal Ilmiah Teknologi Informasi Asia*, 8(2), 10–16.
- Kulkarni, V. Y., and Sinha, P. K. (2014). Effective Learning and Classification using Random Forest Algorithm. *International Journal of Engineering and Innovative Technology*, 3(11), 267–273.
- Kurniawan, Y. I. (2018). Perbandingan Algoritma Naive Bayes dan C.45 dalam Klasifikasi Data Mining. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 5(4), 455.
- Lim Jet al., (2017). Comparison of World Health Organization and Asia-Pacific body mass index classifications in COPD patients. *International Journal of COPD*, 12, 2465–2475.
- Mardi, Y. (2017). Data Mining : Klasifikasi Menggunakan Algoritma C4.5. *Jurnal Edik Informatika*, 2(2), 213–219.
- Nikooghadam et al., (2021). COVID-19 Prediction Classifier Model Using Hybrid Algorithms in Data Mining. *Original Article*, 9(88), 12723–12737. <https://doi.org/10.22038/ijp.2020.54272.4290>
- Nugroho, Y. S., dan Emiliyawati, N. (2017). Sistem Klasifikasi Variabel Tingkat Penerimaan Konsumen Terhadap Mobil Menggunakan Metode Random Forest. *Jurnal Teknik Elektro*, 9(1).
- Pengaribuan dkk., T (2019). Perbandingan Metode Algoritma C4.5 dan Extreme Learning Machine Untuk Mendiagnosis Penyakit Jantung Koroner. *Informatics Engineering Research and Technology*, 1(1), 1–7.
- Primajaya, A., dan Sari, B. N. (2018). Random Forest Algorithm for Prediction of Precipitation. *Indonesian Journal of Artificial Intelligence and Data Mining*, 1(1), 27.
- Rahman dkk, (2017). Klasifikasi Untuk Diagnosa Diabetes Menggunakan Metode Bayesian Regularization Neural Network (RBNN). *Jurnal Informatika*, 11(1), 36.
- Renata, E., dan Ayub, M. (2020). Penerapan Metode Random forest untuk Analisis Risiko pada dataset Peer to peer lending. *Jurnal Teknik Informatika Dan Sistem Informasi*, 6(3), 462–474.

- Richman et al., (2020). Nagging predictors. *Risks*, 8(3), 1–26.
- Rohman dkk, (2020). Komparasi Algoritma C4.5 Berbasis PSO Dan GA Untuk Diagnosa Penyakit Stroke. *CESS (Journal of Computer Engineering, System and Science)*, 5(1), 155.
- Sakinah dkk., (2020). Analisis Kinerja Algoritma Mesin Pembelajaran untuk Klarifikasi Penyakit Stroke Menggunakan Citra CT Scan. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 7(4), 833.
- Schapire, R. E. (2020). RFRSF: Employee Turnover Prediction Based on Random Forests and Survival Analysis. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 12343 LNCS, 503–515. https://doi.org/10.1007/978-3-030-62008-0_35
- Sihombing, P. R., and Arsani, A. M. (2021). Comparison of Machine Learning Methods in Classifying Poverty in Indonesia in 2018 Perbandingan Metode Machine Learning Dalam Klasifikasi Kemiskinan Di Indonesia Tahun 2018. *Jurnal Teknik Informatika (JUTIF)*, 2(1), 51–56.
- Somvanshi et al., (2017). A review of machine learning techniques using decision tree and support vector machine. *Proceedings - 2nd International Conference on Computing, Communication, Control and Automation, ICCUBEA 2016*.
- Sulistiani, H., and Aldino, A. A. (2020). Decision Tree C4.5 Algorithm for Tuition Aid Grant Program Classification (Case Study: Department of Information System, Universitas Teknokrat Indonesia). *Edutic - Scientific Journal of Informatics Education*, 7(1), 40–50.
- Sumpena et al., (2019). Comparison of C4 . 5 Algorithm and Naïve Bayes for Last Information on ICU Patients. *Journal Publications & Informatics Engineering Research*, 4(1), 88–94.
- Thippa Reddy et al., (2020). Antlion re-sampling based deep neural network model for classification of imbalanced multimodal stroke dataset. *Multimedia Tools and Applications*.
- Thupae et al., I (2018). Machine learning techniques for traffic identification and classifiacation in SDWSN: A survey. *Proceedings: IECON 2018 - 44th Annual Conference of the IEEE Industrial Electronics Society, November*, 4645–4650.