

**KLASIFIKASI PENYAKIT DIABETES MENGGUNAKAN METODE
DECISION TREE DAN *RANDOM FOREST***

SKRIPSI

**Sebagai Salah Satu Syarat untuk Memperoleh Gelar
Sarjana Sains Bidang Matematika**



Oleh:

**SITI KALIMAH
NIM. 08011181722063**

**JURUSAN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS SRIWIJAYA
2022**

LEMBAR PENGESAHAN

KLASIFIKASI PENYAKIT DIABETES MENGGUNAKAN METODE
DECISION TREE DAN *RANDOM FOREST*

SKRIPSI

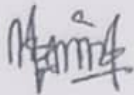
Sebagai Salah Satu Syarat untuk Memperoleh Gelar
Sarjana Sains Bidang Matematika

Oleh

SITI KALIMAH
NIM. 08011181722063

Indralaya, Juli 2022

Pembimbing Kedua



Novi Rustiana Dewi, M.Si
NIP.197011131996032002

Pembimbing Utama



Dr. Yulia Resti, M.Si
NIP.197307191997022001

Mengetahui,

Ketua Jurusan Matematika



Brs. Sugandi Yabidin, M.M
NIP. 19580727 198603 1003

PERNYATAAN INTEGRITAS

Yang bertanda tangan di bawah ini :

Nama : Siti Kalimah

NIM : 08011181722063

Judul : Klasifikasi Penyakit Diabetes Menggunakan Metode *Decision Tree* Dan *Random Forest*

Menyatakan dengan sesungguhnya bahwa seluruh data dan informasi yang dimuat dalam hasil penelitian ini, kecauli yang disebutkan dengan jelas sumbernya adalah hasil pengamatan dan investigasi saya sendiri dibawah supervisi pembimbing dan belum pernah atau tidak sedang diajukan sebagai syarat untuk memperoleh gelar kesarjanaan lain. Apabila dikemudian hari ditemukan adanya unsur plagiasi, maka saya bersedia menerima sanksi akademik dari Universitas Sriwijaya.

Demikian pernyataan ini saya buat dalam keadaan sadar dan tidak mendapat paksaan dari pihak manapun.



Indralaya, Juli 2022



Siti Kalimah

HALAMAN PERSEMBAHAN

MOTTO

“Berusaha lebih dari yang lain jika ingin mendapatkan hasil lebih dari yang lain dan ingat satu hal jangan pernah bandingkan dirimu dengan orang lain karena dirimu tetap dan akan selalu istimewa sebagaimana dirimu sendiri”

Skripsi ini ku persembahkan kepada:

- 1. Allah SWT**
- 2. Bapak dan Mamak**
- 3. Saudaraku**
- 4. Keluarga Besar**
- 5. Dosen**
- 6. Almamater**
- 7. Sahabat dan temanku**

KATA PENGANTAR

Assalamu'alaikum warrahmatullahi wabarakatuh

Puji syukur atas kehadiran Allah SWT yang telah memberikan rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul “**Klasifikasi Penyakit Diabetes Menggunakan Metode *Decision Tree* dan *Random Forest***”. Shalawat beserta salam semoga tetap tercurahkan kepada bimbingan kita, nabi agung nabi Muhammad SAW, beserta keluarga dan para pengikutnya.

Ucapan terima kasih yang sebesar-besarnya penulis sampaikan kepada orang tua tercinta, yaitu **bapak Sukarji** dan **ibu Jumaroh** yang selalu mendoakan dan selalu memberikan dukungan serta semangat untuk penulis sehingga penulis mampu menyelesaikan masa studi kuliah ini. Penulisan ini tidak lepas dari bantuan beberapa pihak, oleh Karena itu penulis ingin menyampaikan rasa terima kasih kepada :

1. **Kemenristekdikti** selaku pemberi beasiswa bidikmisi selama penulis menempuh masa perkuliahan hingga penulis mampu merasakan menjadi seorang mahasiswa dan dapat menyelesaikan jenjang Strata Satu di Jurusan Matematika Universitas Sriwijaya.
2. **Bapak Drs. Sugandi Yahdin, M.M** selaku ketua jurusan Matematika FMIPA universitas sriwijaya untuk ilmu yang telah diberikan.

3. **Ibu Dr. Dian Cahyawati Sukanda, M.Si** selaku sekretaris jurusan yang telah membantu proses administrasi pendaftaran seminar, dan juga atas ilmu yang diberikan.
4. **Ibu Dr. Yulia Resti, M.Si** selaku dosen pembimbing utama yang telah memberikan saran, kritikan, masukan, membimbing dan meluangkan waktu sehingga penulis mampu menyelesaikan skripsi dengan baik.
5. **Ibu Novi Rustiana Dewi, M.Si** selaku pembimbing kedua yang telah memberikan saran, masukan, dan membimbing penulis hingga mampu menyelesaikan skripsi dengan baik.
6. **Bapak Drs. Robinson Sitepu, M.Si** selaku ketua pelaksana seminar dan **Bapak Drs. Endro Setyo C, M.Si** selaku sekretaris pelaksana seminar yang telah membantu proses seminar sehingga seminar dapat dilaksanakan dengan baik.
7. **Ibu Dr.Ir. Herlina Hanum, M.Si** selaku pembahas 1 dan **Ibu Irmeilyana, M.Si** selaku pembahas 2 yang telah memberikan masukan, kritikan dan saran sehingga penulis mampu menyelesaikan skripsi dengan baik.
8. **Bapak Dr. Bambang Suprihatin M.Si** selaku pembimbing akademik yang telah memberikan arahan dan saran selama masa perkuliahan berlangsung.
9. **Seluruh Dosen di jurusan Matematika FMIPA UNSRI** yang telah memberikan ilmu nya kepada penulis selama perkuliahan di jurusan Matematika.

10. **Bapak Irwansyah dan ibu Hamida** yang telah membantu proses administrasi selama di jurusan Matematika FMIPA.
11. Tim hebat, tiada duanya **Abu, Agung, Yudha, Rendy, Wawan, Shohif, Oliv, Mega, Azizah, Tesya, Muflika** yang saling menguatkan satu sama lain, tidak pernah meninggalkan dan selalu siap untuk saling membantu satu sama lain.
12. **Teman-teman angkatan 17** jurusan Matematika FMIPA yang selalu memberikan tawa dan cerita yang tak pernah terlupakan dan akan menjadi sebuah memori kenangan abadi dimana kelak akan penulis ceritakan kepada semua orang bagaimana kisah kasih masa kuliah penulis dulu.
13. **Kakak tingkat 2015, 2016 serta adik tingkat 2018, 2019** atas semua cerita dan momen kebersamaanya.
14. **Semua pihak** yang terlibat dan terkait dalam hidup penulis yang tak dapat disebutkan satu persatu.

Indralaya, 2022

Siti kalimah
NIM. 08011181722063

CLASSIFICATION OF DIABETES USING DECISION TREE AND RANDOM FOREST METHOD

By:

Siti Kalimah

08011181722063

ABSTRACT

Diabetes is a metabolic disease characterized by hyperglycemia caused by defects in insulin secretion or reduced insulin production, slow insulin action or both. Diabetes is one of the deadliest diseases in the world. Accurate classification of people who have positive or negative laboratory test results have diabetes is important to get the right treatment. The purpose of this study is to classify the status of people who have positive or negative laboratory test results for diabetes using the Decision Tree C4.5 and Random Forest methods. In this study used data taken from kaggle.com. This data has a size of 520 and 17 variables. The variables are Age, Gender, Polyuria, Polydipsia, sudden weight loss, weakness, Polyphagia, Genital thrush, visual blurring, Itching, Irritability, delayed healing, partial paresis, muscle stiffness, Alopecia, Obesity, class. The results of this study indicate the level of accuracy, precision, recall, specificity, and F1 score on the Decision Tree C4.5 method respectively of 91.35%, 93.55%, 92.06%, 90.24%, and 92.80%. By using the Random Forest method, the accuracy, precision, recall, specificity, and F1 score levels respectively 98.08%, 100%, 96.88%, 100%, and 98.41%. Based on these measures, it is concluded that the Random Forest method is better than the Decision Tree C4.5 method in classifying the status of people who have positive or negative laboratory test results for diabetes.

Keywords: Diabetes, Decision Tree C4.5, Random Forest

KLASIFIKASI PENYAKIT DIABETES MENGGUNAKAN METODE *DECISION TREE* DAN *RANDOM FOREST*

By:

Siti Kalimah

08011181722063

ABSTRAK

Diabetes merupakan suatu penyakit metabolik dengan karakteristik hiperglikemia yang disebabkan oleh kelainan sekresi insulin atau berkurangnya produksi insulin, kerja insulin yang lambat atau bisa karena kedua-duanya. Penyakit diabetes menjadi salah satu jenis penyakit yang mematikan di dunia. Pengklasifikasian secara tepat orang-orang yang memiliki hasil tes laboratorium apakah positif atau negatif memiliki penyakit diabetes penting dilakukan untuk memperoleh penanganan yang tepat. Tujuan penelitian ini adalah mengklasifikasi status orang-orang yang memiliki hasil tes laboratorium apakah positif atau negatif memiliki penyakit diabetes menggunakan metode *Decision Tree C4.5* dan *Random Forest*. Pada penelitian ini digunakan data yang diambil dari *kaggle.com*. Data ini memiliki ukuran 520 dan 17 variabel. Variabel-variabel tersebut adalah *Age, Gender, Polyuria, Polydipsia, sudden weight loss, weakness, Polyphagia, Genital thrush, visual blurring, Itching, Irritability, delayed healing, partial paresis, muscle stiffness, Alopecia, Obesity, class*. Hasil penelitian ini menunjukkan tingkat akurasi, presisi, recall, specificity, dan F1 score pada metode *Decision Tree C4.5* secara berturut-turut sebesar 91.35%, 93.55%, 92.06%, 90.24%, dan 92.80%. Dengan menggunakan metode *Random Forest* diperoleh tingkat akurasi, presisi, recall, specificity, dan F1 score secara berturut-turut sebesar 98.08%, 100%, 96.88%, 100%, dan 98.41%. Berdasarkan ukuran-ukuran ini disimpulkan bahwa metode *Random Forest* lebih baik daripada metode *Decision Tree C4.5* dalam mengklasifikasi status orang-orang yang memiliki hasil tes laboratorium apakah positif atau negatif memiliki penyakit diabetes.

Kata kunci: Diabetes, *Decision Tree C4.5*, *Random Forest*

DAFTAR ISI

SKRIPSI.....	1
LEMBAR PENGESAHAN	ii
HALAMAN PERSEMBAHAN	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 Penelitian	5
1.5 Manfaat Penelitian	5
BAB II TINJAUAN PUSTAKA	5
2.1 Diabetes.....	5
2.2 Data Mining	7
2.3 Klasifikasi	8
2.4 <i>Decision Tree</i>	10
2.5 Algoritma C4.5.....	13
2.6 <i>Random Forest</i>	14
2.8 <i>Confusion Matrix</i>	16
BAB III METODOLOGI PENELITIAN	18
3.1 Tempat	18
3.2 Waktu	18

3.3	Data Penelitian	18
3.4	Metode Penelitian.....	19
DAFTAR ISI		
BAB IV	HASIL DAN PEMBAHASAN	21
4.1	Deskripsi Data, <i>Preprocessing</i> , Partisi Data.....	21
4.2	Mengklasifikasi Status Penyakit Diabetes dengan Metode <i>Decision Tree</i>	23
4.3	Mengklasifikasi Status Penyakit Diabetes dengan Metode <i>Random Forest</i>	33
4.4	Perbandingan Tingkat Ketepatan Klasifikasi	48
BAB V	KESIMPULAN DAN SARAN	50
5.1	Kesimpulan	50
5.2	Saran	51
DAFTAR PUSTAKA.....		52
LAMPIRAN.....		56

DAFTAR TABEL

Tabel 1.1 Tinjauan studi pustaka.....	2
Tabel 2.1 <i>Confusion Matrix</i>	16
Tabel 4.1 Deskripsi variabel.....	21
Tabel 4.2 Data <i>train</i> 80%	22
Tabel 4.3 Data <i>test</i> 20%	23
Tabel 4.3 Perhitungan <i>entropy</i> dan <i>gain</i> node 1.....	26
Tabel 4.4 Perhitungan <i>entropy</i> dan <i>gain</i> node 1.1.....	28
Tabel 4.5 <i>Confusion matrix</i> metode <i>decision tree</i>	32
Tabel 4.6 <i>Bootstrap sampling</i> pohon pertama.....	33
Tabel 4.7 Perhitungan <i>entropy</i> dan <i>gain</i> untuk variabel X_5, X_6, X_{10}, X_{14} ...	38
Tabel 4.8 Perhitungan <i>entropy</i> dan <i>gain</i> untuk variabel X_2, X_8, X_{11}, X_{16} ...	39
Tabel 4.9 Perhitungan <i>entropy</i> dan <i>gain</i> untuk variabel X_7, X_9, X_{11}, X_{15} ...	40
Tabel 4.10 Perhitungan <i>entropy</i> dan <i>gain</i> untuk variabel X_1, X_6, X_{12}, X_{13}	42
Tabel 4.11 <i>Confusion matrix</i> metode <i>random forest</i>	44
Tabel 4.12 Perbandingan dua metode	45

DAFTAR GAMBAR

Tabel 2.1 Struktur <i>Decision Tree</i>	10
Tabel 2.2 Struktur <i>Random forest</i>	16
Tabel 4.1 Grafik jumlah data <i>train</i> positif dan negatif.....	23
Tabel 4.2 Pohon keputusan <i>root node</i>	28
Tabel 4.3 Pohon keputusan node 1.1.....	29
Tabel 4.4 Pengkondisian pohon keputusan metode <i>decision tree</i> menggunakan software rapid miner	30
Tabel 4.5 Pohon keputusan <i>root node</i>	37
Tabel 4.6 Pohon keputusan node 1.1.....	38
Tabel 4.7 Pohon keputusan node 1.1.1.....	40
Tabel 4.8 Pohon keputusan node 1.1.1.1.....	41
Tabel 4.9 Pohon keputusan pertama	43
Tabel 4.10 Pohon keputusan kedua.....	45
Tabel 4.11 Pohon keputusan ketiga.....	46

DAFTAR LAMPIRAN

Lampiran 1. Pengkondisian pohon keputusan metode <i>decision tree</i> menggunakan software rapid miner	53
Lampiran 2. Pengkondisian pohon keputusan pertama <i>random forest</i> menggunakan software rapid miner	55
Lampiran 3. Pengkondisian pohon keputusan kedua <i>random forest</i> menggunakan software rapid miner	57
Lampiran 4. Pengkondisian pohon keputusan ketiga <i>random forest</i> menggunakan software rapid miner	59

BAB I

PENDAHULUAN

1.1 Latar Belakang

Menurut *American Diabetes Association* dalam penelitian (Zhang & Tan, 2000) Diabetes merupakan suatu penyakit metabolik dengan karakteristik hiperglikemia yang disebabkan oleh kelainan sekresi insulin atau berkurangnya produksi insulin, kerja insulin yang lambat atau bisa karena kedua-duanya. Setiap tahun pasien positif diabetes mengalami peningkatan yang cukup signifikan. Menurut Organisasi Kesehatan Dunia pasien penderita diabetes terus mengalami peningkatan hingga mampu mencatat angka mencapai 422 juta orang di dunia yang mana hal ini menjadi empat kali lipat dari pada 30 tahun lalu (Najib et al., 2019).

Ciri-ciri umum orang yang memiliki penyakit diabetes adalah sering buang air kecil, gampang merasakan haus, merasa cepat lapar, mengalami penurunan berat badan secara drastis, kulit menjadi kering dan gatal, luka yang sulit untuk sembuh, mengalami gangguan penglihatan, mengalami kesemutan atau kebas, mudah terjadi pembengkakan pada kaki dan tangan, badan lemas serta mengalami sakit kepala yang berlebihan, obesitas, dan infeksi jamur atau bakteri. Pengklasifikasian secara tepat orang-orang yang memiliki hasil tes laboratorium apakah positif atau negatif memiliki penyakit diabetes penting dilakukan untuk memperoleh penanganan yang tepat.

Terdapat beberapa penelitian terkait dengan metode yang akan digunakan dalam penelitian ini.

Tabel 1.1 Tinjauan Studi Terdahulu

Sumber/judul	Deskripsi	Algoritma
(Apriyani & Kurniati, 2020) / Perbandingan Metode Naïve Bayes dan Support Vector Machine dalam Klasifikasi Penyakit Diabetes Militus	Data berasal dari rekam medik Rumah Sakit Siti Khadijah-Palembang yang berjumlah 613 record dengan atribut sebanyak 9 atribut dalam kurun waktu hampir 3 tahun. Hasil akurasi, presisi, recall, specificity dan juga F1 Score dari data tersebut dengan menggunakan metode naïve bayes yaitu 92.07%, 93.08%, 97.00%, 75.00%, 95.00%. jika menggunakan metode Support Vector Machine Kernel Polynomial hasil akurasi, presisi, recall, specificity dan juga F1 Score sebesar 96.72%, 99.42%, 96.10%, 97.14%, 97.73%. sedangkan jika menggunakan metode Support Vector Machine Kernel RBF hasil akurasi, presisi, recall, specificity dan juga F1 Score sebesar 80.89%, 100%, 80.89%, -, 89.43%	naïve bayes, Support Vector Machine Kernel Polynomial, dan Support Vector Machine Kernel RBF
(Putry et al., 2022) / Komparasi Algoritma Knn Dan Naïve Bayes Untuk Klasifikasi Diagnosis Penyakit Diabetes Melitus	Penelitian ini akan mengklasifikasikan diagnosis penyakit diabetes melitus. Dimana yang menjadi data masukan atau input dalam penelitian ini adalah data diagnosis penyakit diabetes melitus dengan variable-variabel penentunya yaitu <i>Outcome, Glucose, Blood Pressure, Insulin, BMI, Age, Diabetes Pedigree Function</i> . Hasil akurasi, presisi dan recall pada data ini dengan menggunakan metode <i>naïve bayes</i> sebesar 80%, 86%, dan 86%. Sedangkan jika menggunakan metode <i>K-Nearest Neighbor</i> hasil akurasi, presisi dan recall sebesar 75%, 80%, dan 86%.	naïve bayes, K-Nearest Neighbor (KNN)
(Naik & Patel, 2013) / Tumor Detection and Classification using Decision Tree in Brain MRI	Data ini menggunakan 124 gambar MRI. Dari 124 gambar tersebut dibagi menjadi 73 gambar <i>training</i> dan 51 gambar <i>testing</i> . Hasil akurasi presisi, recall dan specificity dengan menggunakan metode naïve bayes adalah 88.2%, 91%, 91%, dan 83%. Sedangkan jika menggunakan metode	naïve bayes, Decision Tree

	Decision Tree maka hasil akurasi, presisi, recall dan specificity sebesar 96%, 100%, 93% dan 100%.	
(Otok & Nidhomuddin, 2015) / Random forest dan Multivariate adaptive Regression spline (Mars) Binary response Untuk Klasifikasi Penderita hiv/Aids Di Surabaya	Data yang digunakan dalam penelitian ini adalah data sekunder berupa data kasus penderita HIV/AIDS di Kota Surabaya yang didapatkan dari skripsi S1 ITS Surabaya yang disusun oleh Romaiza Millah Hanifa pada tahun 2013. Banyaknya data yang digunakan pada penelitian ini sebanyak 218 sampel dengan 13 variabel yang terdiri dari klien dengan status HIV/AIDS negatif dan klien dengan status HIV/AIDS positif. Hasil akurasi, recall dan specificity dengan menggunakan metode MARS sebesar 80.28%, 94.12%, dan 31.25%. jika menggunakan metode Random Forest maka hasil akurasi, recall dan specificity sebesar 97.8%, 100%, dan 95.55%. sedangkan jika menggunakan metode Random Forest MARS hasil akurasi, recall dan specificity sebesar 91%, 98.82, dan 62.50%.	MARS, Random Forest, Random Forest MARS

Decision tree merupakan salah satu metode klasifikasi yang menggunakan representasi struktur pohon dimana setiap node mempresentasikan atribut, dimana cabangnya mempresentasikan nilai dari variabel prediktor, dan daun mempresentasikan kelasnya. Menurut Eska (2016) pada *Decision tree* terdapat tiga simpul, yaitu simpul akar (*root node*), simpul percabangan/ internal (*branch/ internal node*) dan simpul daun (*leaf node*). Salah satu algoritma yang terdapat pada metode *decision tree* adalah algoritma C4.5. Algoritma C4.5 merupakan suatu algoritma yang dapat digunakan untuk membentuk pohon keputusan sehingga mampu membuat prediksi atau mengklasifikasi. Proses dalam algoritma ini yaitu dengan memilih salah satu atribut yang akan digunakan sebagai akar pohon, kemudian akan dibuat cabang di dalam akar tersebut. Langkah berikutnya

yaitu membagi setiap kasus yang ada dalam cabang, kemudian proses akan terus diulang sampai setiap kasus berada dalam tiap-tiap kelas yang telah ditentukan (Parung, 2018).

Metode klasifikasi lain yang digunakan dalam penelitian ini yaitu *random forest*. *Random forest* adalah suatu metode ensemble yang terdiri dari sekumpulan pohon keputusan, dimana pohon keputusan tersebut digunakan untuk mengklasifikasi data ke suatu kelas. Berdasarkan Paul et al., (2018) kinerja klasifikasi dari *random forest* meningkat dengan bertambahnya jumlah pohon yang dibentuk. Metode *decision tree* dan *random forest* memiliki peranan untuk mengambil keputusan dalam pengklasifikasian sebuah data.

Menurut Azhari et al., (2021) beberapa metode klasifikasi antara lain *Decision Tree*, *rule-based classifiers*, *Bayesian classifier* *Support Vector Machines*, *Artificial Neural Networks*, *Lazy Learners*, dan *ensemble methods*.

1.2 Perumusan Masalah

Rumusan masalah dalam penelitian ini adalah :

1. Bagaimana klasifikasi penyakit diabetes dengan menggunakan metode *decision tree* C4.5 dan *random forest*?
2. Manakah metode yang memiliki tingkat ketepatan yang lebih baik dalam mengklasifikasikan penyakit diabetes?

1.3 Batasan Masalah

Batasan-batasan masalah dalam penelitian ini antara lain :

1. Tingkat ketepatan diukur menggunakan akurasi, presisi, recall, specificity dan F1 score.
2. Data di partisi menjadi 80% data latih (416 pengamatan) , dan 20% data uji (104 pengamatan).

1.4 Tujuan Penelitian

1. Mengklasifikasi penyakit diabetes menggunakan metode *decision tree* C4.5 dan *random forest*.
2. Membandingkan tingkat akurasi metode *decision tree* C4.5 dan *random forest*.

1.5 Manfaat Penelitian

1. Menjadi media belajar bagi penulis dan juga pembaca terkait penggunaan metode *decision tree* C4.5 dan *random forest* dalam melakukan klasifikasi.
2. Sebagai bahan referensi untuk penelitian yang membahas penyakit diabetes.

DAFTAR PUSTAKA

- Alsalman, Y. S., Khamees Abu Halemah, N., Alnagi, E. S., & Salameh, W. (2019). Using Decision Tree and Artificial Neural Network to Predict Students Academic Performance. *2019 10th International Conference on Information and Communication Systems, ICICS 2019*, 104–109. <https://doi.org/10.1109/IACS.2019.8809106>
- Anggraeni, D., & Ramadhani. (2018). *Analisa Decision Tree Untuk Prediksi Diagnosa Diabetes Mellitus*. 9986(September).
- Apriyani, H., & Kurniati, K. (2020). Perbandingan Metode Naïve Bayes Dan Support Vector Machine Dalam Klasifikasi Penyakit Diabetes Melitus. *Journal of Information Technology Ampera*, 1(3), 133–143. <https://doi.org/10.51519/journalita.volume1.issue3.year2020.page133-143>
- Azhari, M., Situmorang, Z., & Rosnelly, R. (2021). Perbandingan Akurasi, Recall, dan Presisi Klasifikasi pada Algoritma C4.5, Random Forest, SVM dan Naive Bayes. *Media Informatika Budidarma*, 5(2), 640–651. <https://doi.org/10.30865/mib.v5i2.2937>
- Chan, A. S. (2018). Prediksi Kedatangan Wisatawan Pada Pariwisata Kota Batam Dengan Menggunakan Teknik Knowledge Data Discovery. *Jurnal Ilmiah Informatika*, 6(01), 11. <https://doi.org/10.33884/jif.v6i01.432>
- Chandrasekar, P., Qian, K., Shahriar, H., & Bhattacharya, P. (2017). Improving the Prediction Accuracy of Decision Tree Mining with Data Preprocessing. *Proceedings - International Computer Software and Applications Conference*, 2, 481–484. <https://doi.org/10.1109/COMPSAC.2017.146>
- Cho, N. H., Shaw, J. E., Karuranga, S., Huang, Y., da Rocha Fernandes, J. D., Ohlrogge, A. W., & Malanda, B. (2018). IDF Diabetes Atlas: Global estimates of diabetes prevalence for 2017 and projections for 2045. *Diabetes Research and Clinical Practice*, 138, 271–281. <https://doi.org/10.1016/j.diabres.2018.02.023>
- Dennedy, M. C., Rizza, R. A., & Dinneen, S. F. (2015). Classification and Diagnosis of Diabetes Mellitus. *Endocrinology: Adult and Pediatric*, 1–2(January), 662-671.e2. <https://doi.org/10.1016/B978-0-323-18907-1.00038-X>
- Eska, J. (2016). Penerapan Data Mining Untuk Prediksi Penjualan Wallpaper Menggunakan Algoritma C4.5. *JURTEKSI (Jurnal Teknologi Dan Sistem Informasi)*, 2, 9 – 13. <https://doi.org/10.31227/osf.io/x6svc>

- Hamoud, A. K., Hashim, A. S., & Awadh, W. A. (2018). Predicting Student Performance in Higher Education Institutions Using Decision Tree Analysis. *International Journal of Interactive Multimedia and Artificial Intelligence*, 5(2), 26. <https://doi.org/10.9781/ijimai.2018.02.004>
- Hana, F. M. (2020). Klasifikasi Penderita Penyakit Diabetes Menggunakan Algoritma Decision Tree C4.5. *Jurnal SISKOM-KB (Sistem Komputer Dan Kecerdasan Buatan)*, 4(1), 32–39. <https://doi.org/10.47970/siskom-kb.v4i1.173>
- Jiang, W., Liu, G., Zhao, X., & Yang, F. (2019). Cross-Subject Emotion Recognition with a Decision Tree Classifier Based on Sequential Backward Selection. *Proceedings - 2019 11th International Conference on Intelligent Human-Machine Systems and Cybernetics, IHMSC 2019, 1*, 309–313. <https://doi.org/10.1109/IHMSC.2019.00078>
- Kurniawan, H. (2020). Deteksi Twitter Bot Menggunakan Klasifikasi Decision Tree. *Jurnal Sustainable: Jurnal Hasil Penelitian Dan Industri Terapan*, 09(2), 31–37. <http://ejournal.bsi.ac.id/ejurnal/index.php/cakrawala/article/view/3680/2624%0Ahttp://j-ptiik.ub.ac.id>
- Liu, J., Ning, B., & Shi, D. (2019). Application of Improved Decision Tree C4.5 Algorithms in the Judgment of Diabetes Diagnostic Effectiveness. *Journal of Physics: Conference Series*, 1237(2). <https://doi.org/10.1088/1742-6596/1237/2/022116>
- Mardi, Y. (2017). Data Mining : Klasifikasi Menggunakan Algoritma C4.5. *Edik Informatika*, 2(2), 213–219. <https://doi.org/10.22202/ei.2016.v2i2.1465>
- Naik, J., & Patel, P. S. (2013). *Tumor Detection and Classification using Decision Tree in Brain MRI*. 49–53.
- Najib, A., Nurcahyono, D., & Setiawan, R. P. P. (2019). Klasifikasi Diagnosa Penyakit Diabetes Mellitus (Dm) Menggunakan Algoritma C4.4. *Just TI (Jurnal Sains Terapan Teknologi Informasi)*, 11(2), 47. <https://doi.org/10.46964/justti.v11i2.153>
- NAsrullah, A. H. (2021). *IMPLEMENTASI ALGORITMA DECISION TREE UNTUK*. 7(2), 45–51.
- Nasution, M. Z. F., Sitompul, O. S., & Ramli, M. (2018). PCA based feature reduction to improve the accuracy of decision tree c4.5 classification. *Journal of Physics: Conference Series*, 978(1). <https://doi.org/10.1088/1742-6596/978/1/012058>
- Novianti, B., Rismawan, T., & Bahri, S. (2016). Implementasi Data Mining

- Dengan Algoritma C4.5 Untuk Penjurusan Siswa (Studi Kasus: Sma Negeri 1 Pontianak). *Jurnal Coding, Sistem Komputer Untan*, 04(3), 75–84.
- Nugroho, K., Noersasongko, E., Purwanto, Muljono, Fanani, A. Z., Affandy, & Basuki, R. S. (2019). Improving random forest method to detect hatespeech and offensive word. *2019 International Conference on Information and Communications Technology, ICOIACT 2019*, 514–518. <https://doi.org/10.1109/ICOIACT46704.2019.8938451>
- Otok, B. W., & Nidhomuddin. (2015). Random Forest Dan Multivariate Adaptive Regression Spline (Mars) Binary Response Untuk Klasifikasi Penderita Hiv / Aids Di Surabaya. *Statistika Fakultas Matematika Dan Ilmu Pengetahuan Alam Institut Teknologi Sepuluh November*, 1(3), 50–57.
- Parung, F. (2018). Penerapan algoritma decision tree c4.5 dalam penerimaan guru pada smk sirajul falah parung. *CKI On SPOT*, 11(2), 192–198.
- Paul, A., Mukherjee, D. P., Das, P., Gangopadhyay, A., Chintia, A. R., & Kundu, S. (2018). Improved Random Forest for Classification. *IEEE Transactions on Image Processing*, 27(8), 4012–4024. <https://doi.org/10.1109/TIP.2018.2834830>
- Punthakee, Z., Goldenberg, R., & Katz, P. (2018). Definition, Classification and Diagnosis of Diabetes, Prediabetes and Metabolic Syndrome. *Canadian Journal of Diabetes*, 42, S10–S15. <https://doi.org/10.1016/j.jcjd.2017.10.003>
- Putry, N. M., Sari, B. N., Kom, M., Informatika, T., & Karawang, U. S. (2022). *KOMPARASI ALGORITMA KNN DAN NAÏVE BAYES UNTUK KLASIFIKASI DIAGNOSIS PENYAKIT DIABETES MELITUS*. 10(1).
- Religia, Y., Nugroho, A., & Hadikristanto, W. (2021). Analisis Perbandingan Algoritma Optimasi pada Random Forest untuk. *JURNAL RESTI*, 1(10), 187–192.
- Saifullah, S., Zarlis, M., Zakaria, Z., & Sembiring, R. W. (2017). Analisa Terhadap Perbandingan Algoritma Decision Tree Dengan Algoritma Random Tree Untuk Pre-Processing Data. *J-SAKTI (Jurnal Sains Komputer Dan Informatika)*, 1(2), 180. <https://doi.org/10.30645/j-sakti.v1i2.41>
- Setio, P. B. N., Saputro, D. R. S., & Bowo Winarno. (2020). Klasifikasi Dengan Pohon Keputusan Berbasis Algoritme C4.5. *PRISMA, Prosiding Seminar Nasional Matematika*, 3, 64–71.
- Siallagan, R. A., & Fitriyani. (2021). Prediksi Penyakit Diabetes Mellitus. *Jurnal Responsif*, 3(1), 45–46.
- Sunjana. (2010). Aplikasi mining data mahasiswa dengan metode klasifikasi decision tree. *Snati*.

- Syamsu, S., Muhajirin, M., & Wijaya, N. S. (2019). Rules Generation Untuk Klasifikasi Data Bakat dan Minat Berdasarkan Rumpun Ilmu Dengan Decision Tree. *Inspiration: Jurnal Teknologi Informasi Dan Komunikasi*, 9(1), 40. <https://doi.org/10.35585/inspir.v9i1.2495>
- Tanujayaa, L. B. C., Susanto, B., & Saragiha, A. (2020). Perbandingan Metode Regresi Logistik dan Random Forest untuk Klasifikasi Fitur Mode Audio Spotify. *Indonesian Journal of Data and Science (IJODAS)*, 1(2715–9930), 68–78.
- Utomo, D. P., & Mesran, M. (2020). Analisis Komparasi Metode Klasifikasi Data Mining dan Reduksi Atribut Pada Data Set Penyakit Jantung. *Jurnal Media Informatika Budidarma*, 4(2), 440. <https://doi.org/10.30865/mib.v4i2.2080>
- Widiyati, D. K., Wati, M., & Pakpahan, H. S. (2018). Penerapan Algoritma ID3 Decision Tree Pada Penentuan Penerima Program Bantuan Pemerintah Daerah di Kabupaten Kutai Kartanegara. *Jurnal Rekayasa Teknologi Informasi (JURTI)*, 2(2), 126. <https://doi.org/10.30872/jurti.v2i2.1864>
- Wulandari, P., Sugiyanto, Z., & Kresnowati, L. (2015). Analisis Faktor Penyebab Kadar Gula darah Pada Penderita Diabetes Melitus (DM) Tipe-2 Di RSUD Tugurejo Semarang. *Jurnal Kesehatan*, 14(2), 353–360.
- Yezli, S., Yassin, Y., Mushi, A., Balkhi, B., & Khan, A. (2021). Insulin Knowledge , Handling , and Storage among Diabetic Pilgrims during the Hajj Mass Gathering. *Journal of Diabetes Research*, 2021.
- Zeng, G. (2020). On the confusion matrix in credit scoring and its analytical properties. *Communications in Statistics - Theory and Methods*, 49(9), 2080–2093. <https://doi.org/10.1080/03610926.2019.1568485>
- Zhang, X. F., & Tan, B. K. H. (2000). Effects of an ethanolic extract of *Gynura procumbens* on serum glucose, cholesterol and triglyceride levels in normal and streptozotocin-induced diabetic rats. *Singapore Medical Journal*, 41(1), 9–13.
- Zhou, X., Lu, P., Zheng, Z., Tolliver, D., & Keramati, A. (2020). Accident Prediction Accuracy Assessment for Highway-Rail Grade Crossings Using Random Forest Algorithm Compared with Decision Tree. *Reliability Engineering and System Safety*, 200, 9. <https://doi.org/10.1016/j.ress.2020.106931>