

**KOMBINASI ALGORITMA RELIEFF DAN
MULTIPLE IMPUTATION BY CHAINED EQUATION (MICE)
UNTUK PENANGANAN DATA HILANG
PADA *DATASET* PENYAKIT GINJAL KRONIS**

SKRIPSI

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

Oleh :

Rizki Eka Putri

NIM 08011281924051



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

2023

HALAMAN PENGESAHAN

**KOMBINASI ALGORITMA RELIEFF DAN
MULTIPLE IMPUTATION BY CHAINED EQUATION (MICE)
DALAM PENANGANAN DATA HILANG
PADA DATASET PENYAKIT GINJAL KRONIS**

SKRIPSI

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

Oleh :

**Rizki Eka Putri
NIM. 08011281924051**

Indralaya, 04 April 2023

Pembimbing Pendamping

Pembimbing Utama



Dr. Bambang Suprihatin, S.Si., M.Si

NIP. 197101261994121001



Dr. Anita Desiani, S.Si., M.Kom

NIP. 197712112003122002

**Mengetahui,
Ketua Jurusan Matematika**



Drs. Sugandi Fahdin, M.M

NIP. 195807271986031003

PERNYATAAN KEASLIAN KARYA ILMIAH

Yang bertanda tangan dibawah ini:

Nama Mahasiswa : Rizki Eka Putri
NIM : 08011281924051
Fakultas/Jurusan : Matematika dan Ilmu Pengetahuan Alam/Matematika

Menyatakan bahwa skripsi ini adalah hasil karya saya sendiri dan karya ilmiah ini belum pernah diajukan sebagai pemenuhan persyaratan untuk memperoleh gelar kesarjanaan strata satu (S1) dari Universitas Sriwijaya maupun perguruan tinggi lain.

Semua informasi yang dimuat dalam skripsi ini yang berasal dari penulis lain baik yang dipublikasikan atau tidak telah diberikan penghargaan dengan mengutip nama sumber penulis secara benar. Semua isi dari skripsi ini sepenuhnya menjadi tanggung jawab saya sebagai penulis.

Demikianlah surat pernyataan ini saya buat dengan sebenarnya.

Indralaya, 10 April 2023

Penulis



Rizki Eka Putri

NIM. 08011281924051

LEMBAR PERSEMBAHAN

Kupersembahkan skripsi ini untuk :

Yang Maha Kuasa Allah Subhanahu Wa Ta'ala,

Kedua orang tuaku tersayang,

Adik Laki-lakiku,

Keluarga besarku,

Semua guru dan dosenku,

Sahabat-sahabatku,

Almamaterku

Moto

“Do The Best, Let allah Do The Rest”

KATA PENGANTAR

Puji syukur kehadirat Allah Subhanahu wa Ta'ala yang telah memberikan rahmat dan hidayah-Nya, sehingga penulis dapat menyelesaikan skripsi yang berjudul “Kombinasi Algoritma ReliefF dan *Multiple Imputation by Chained Equation* (MICE) dalam Penanganan Data Hilang pada *Dataset* Penyakit Ginjal Kronis” sebagai salah satu syarat memperoleh gelar sarjana sains bidang studi Matematika di Fakultas MIPA Universitas Sriwijaya.

Penulis menyadari bahwa proses pembuatan skripsi ini merupakan proses pembelajaran yang sangat berharga serta tak lepas dari kekurangan dan keterbatasan. Dengan segala hormat dan kerendahan hati, penulis mengucapkan terima kasih dan penghargaan kepada:

1. Kedua orang tuaku tercinta, Bapak **Sofyan Rifai** dan Ibu **Parida**, yang tak pernah lelah merawat, mendidik, menuntun, memberi nasehat, semangat serta doa untuk penulis. Terima kasih atas segala perjuangan, pengorbanan, serta kasih sayang hingga detik ini dan sampai kapanpun.
2. Bapak **Drs. Sugandi Yahdin, M.M** selaku Ketua Jurusan Matematika FMIPA Universitas Sriwijaya yang telah memberikan arahan dan motivasi kepada penulis selama proses perkuliahan.
3. Ibu **Dr. Dian Cahyawati Sukanda, M.Si** selaku Sekretaris Jurusan Matematika FMIPA Universitas Sriwijaya yang telah mengarahkan urusan akademik kepada penulis.

4. Ibu **Anita Desiani, S.Si., M.Kom** selaku dosen pembimbing utama yang telah bersedia memberikan waktu, tenaga, pikiran, nasehat, dan motivasi untuk memberikan bimbingan dan pengarahan selama proses pembuatan skripsi, kompetisi, dan perjalanan perkuliahan ini.
5. Bapak **Dr. Bambang Suprihatin, S.Si., M.Si** selaku dosen pembimbing pendamping yang telah bersedia memberikan waktu, tenaga, pikiran, nasehat, dan motivasi untuk memberikan bimbingan dan pengarahan selama proses pembuatan skripsi serta proses perkuliahan dengan penuh pengertian dan kesabaran.
6. Ibu **Dr. Yuli Andriani, M.Si** dan Ibu **Eka Susanti, M.Sc** selaku dosen pembahas dan penguji yang telah memberikan tanggapan, kritik, dan saran yang sangat bermanfaat untuk perbaikan dan penyelesaian skripsi ini.
7. Ibu **Novi Rustiana Dewi, S.Si., M.Si** selaku dosen pembimbing akademik yang telah membimbing dan mengarahkan urusan akademik penulis.
8. **Seluruh Dosen di Jurusan Matematika FMIPA** yang telah memberikan ilmu, nasihat, motivasi, serta bimbingan selama proses perkuliahan.
9. Pak **Irwansyah** selaku admin dan Ibu **Hamidah** selaku pegawai tata usaha Jurusan Matematika FMIPA yang telah membantu penulis selama perkuliahan.
10. **Seluruh guru** yang telah memberikan ilmu yang bermanfaat hingga mengantarkan penulis pada pendidikan ini.

11. Adik laki-lakiku tersayang, **Rahmat Dwi Putra**, yang selalu mendoakan dan memberikan perhatian kepada penulis, beserta keluarga besar yang selalu mendukung penulis.
12. **Semua sahabat seperjuangan** dalam masa perkuliahan dan proses skripsi. Terima kasih sudah menjadi orang-orang baik di sekeliling penulis yang selalu mendukung, membantu dengan tulus, dan memberi energi positif.
13. **Keluarga Matematika 2019, BPH Himastik Akselerasi, Keluarga IMMETA SUMSEL 2021/2022** dan **rekan-rekan perlombaan** selama perkuliahan.
14. Kakak-kakak tingkat **angkatan 2017** dan **2018** serta adik-adik tingkat **angkatan 2020** dan **2021**, terima kasih atas segala kebaikan dan bantuan.
15. Semua pihak yang tidak dapat penulis sebutkan satu per satu. Semoga segala kebaikan yang diberikan mendapatkan balasan terbaik dari Allah SWT.

Semoga skripsi ini dapat menambah pengetahuan dan bermanfaat bagi mahasiswa/mahasiswi Jurusan Matematika Fakultas dan Ilmu Pengetahuan Alam Universitas Sriwijaya dan semua pihak yang memerlukan.

Indralaya, Maret 2023

Penulis

**COMBINATION OF RELIEFF ALGORITHM AND
MULTIPLE IMPUTATION BY CHAINED EQUATIONS (MICE)
IN HANDLING MISSING DATA
FOR CHRONICH KIDNEY DISEASE DATASET**

By

**RIZKI EKA PUTRI
NIM. 08011281924051**

ABSTRACT

Missing data is a problem where there are empty values or incomplete values in the data. The University of California Irvine (UCI) Machine Learning Repository dataset that records chronic kidney disease patient data is one of the datasets that has a data loss problem. Lost data handling can be done by two methods, namely deletion and imputation. One of the deletion algorithms is ReliefF, while one of the imputation algorithms is Multiple Imputation by Chained Equation (MICE). This study aims to overcome missing data in the UCI chronic kidney disease dataset using a combination of the ReliefF and MICE algorithms. The results of the combination of algorithms were tested using the K-Nearest Neighbor (KNN) method to determine the increase in classification performance. The results of the ReliefF feature selection show that there are 15 features that influence classification with an increase in accuracy of 23,5%. The results shown from the combination of the two methods are an increase in accuracy, precision, and recall of 25,5%, 19,7%, 23,08%. The chronic kidney disease dataset experienced significant increases in accuracy, precision, and recall. It can be concluded that the combination of the ReliefF and MICE algorithms can improve classification performance in the UCI chronic kidney disease dataset.

Keywords : Missing Data, Feature Selection, Imputation, ReliefF, MICE.

**KOMBINASI ALGORITMA RELIEFF DAN
MULTIPLE IMPUTATION BY CHAINED EQUATIONS (MICE)
DALAM PENANGANAN DATA HILANG
PADA DATASET PENYAKIT GINJAL KRONIS**

Oleh

**RIZKI EKA PUTRI
NIM. 08011281924051**

ABSTRAK

Data hilang adalah suatu keadaan dimana terdapat nilai yang kosong atau nilai yang tidak lengkap dalam data. *Dataset University of California Irvine (UCI) Machine Learning Repository* yang merekam data pasien penyakit ginjal kronis merupakan salah satu *dataset* yang memiliki permasalahan data hilang. Penanganan data hilang dapat dilakukan dengan dua metode, yaitu *deletion* dan imputasi. Salah satu algoritma *deletion* adalah ReliefF, sedangkan salah satu algoritma imputasi adalah *Multiple Imputation by Chained Equation (MICE)*. Penelitian ini bertujuan untuk mengatasi data hilang pada *dataset* UCI penyakit ginjal kronis menggunakan kombinasi algoritma ReliefF dan MICE. Hasil dari kombinasi algoritma diuji coba menggunakan metode *K-Nearest Neighbor (KNN)* untuk mengetahui peningkatan kinerja klasifikasi. Hasil seleksi fitur ReliefF menunjukkan ada 15 fitur yang berpengaruh terhadap klasifikasi dengan peningkatan akurasi sebesar 23,5%. Hasil yang ditunjukkan dari kombinasi kedua metode tersebut adalah peningkatan akurasi, presisi, dan *recall* sebesar 25,5%, 19,7%, 23,08%. *Dataset* penyakit ginjal kronis mengalami peningkatan akurasi, presisi, dan *recall* yang signifikan, dapat disimpulkan bahwa kombinasi algoritma ReliefF dan MICE dapat meningkatkan kinerja klasifikasi pada *dataset* UCI penyakit ginjal kronis.

Kata Kunci : Data Hilang, Seleksi Fitur, Imputasi, ReliefF, MICE.

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PENGESAHAN	ii
LEMBAR PERSEMBAHAN	iii
KATA PENGANTAR	iv
ABSTRACT	vii
ABSTRAK	viii
DAFTAR ISI	ix
DAFTAR TABEL	x
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Perumusan Masalah	4
1.3 Pembatasan Masalah	4
1.4 Tujuan	4
1.5 Manfaat	5
BAB II TINJAUAN PUSTAKA	6
2.1 Data Hilang	6
2.2 Seleksi Fitur	6
2.3 Algoritma ReliefF	7
2.4 <i>Multiple Imputation by Chained Equation (MICE)</i>	9
2.5 Pengujian Kinerja Kombinasi Algoritma	12
BAB III METODELOGI PENELITIAN	15
3.1 Tempat	15
3.2 Waktu	15
3.3 Alat	15
3.4 Metode Penelitian	15
BAB IV HASIL DAN PEMBAHASAN	20
4.1 Deskripsi Data	20
4.2 Seleksi Fitur Algoritma ReliefF	23
4.2.1 Contoh Perhitungan Manual Algoritma ReliefF	23
4.2.2 Hasil Seleksi Fitur dengan Algoritma ReliefF	34
4.3 Imputasi Data Hilang	36
4.3.1 Imputasi Data Hilang pada Fitur yang Memiliki Data Hilang Kurang Dari atau Sama Dengan 10%	37
4.3.2 Imputasi Data Hilang pada Fitur yang Memiliki Data Hilang Lebih Dari 10%	41
4.4 Pengujian Kinerja Kombinasi Algoritma	50
4.5 Analisis dan Interpretasi Hasil	55
BAB V KESIMPULAN DAN SARAN	58
5.1 Kesimpulan	58
5.2 Saran	58
DAFTAR PUSTAKA	59

DAFTAR TABEL

Tabel 2.1 Model <i>confussion matrix</i>	13
Tabel 2.2 Kategori evaluasi algoritma (Han <i>et al.</i> , 2022).....	14
Tabel 4.1 <i>Dataset</i> pasien penyakit ginjal kronis	21
Tabel 4.2 Keterangan fitur	22
Tabel 4.3 Data contoh perhitungan manual menggunakan algoritma Relief-F	23
Tabel 4.4 Nilai bobot <i>al</i> pada tiap iterasi	26
Tabel 4.5 Nilai bobot <i>hemo</i> pada tiap iterasi.....	28
Tabel 4.6 Nilai bobot <i>wc</i> pada tiap iterasi.....	30
Tabel 4.7 Nilai bobot <i>htn</i> pada tiap iterasi	33
Tabel 4.8 Hasil bobot setiap fitur pada data contoh.....	33
Tabel 4.9 Nilai bobot setiap fitur pada dataset penelitian.....	34
Tabel 4.10 Hasil perbandingan nilai akurasi pada setiap seleksi fitur	35
Tabel 4.11 Fitur yang dipertahankan setelah seleksi fitur ReliefF.....	36
Tabel 4.12 Keterangan fitur <i>pcc</i>	37
Tabel 4.13 Keterangan fitur <i>htn</i>	38
Tabel 4.14 Keterangan fitur <i>dm</i>	39
Tabel 4.15 Keterangan fitur <i>cad</i>	39
Tabel 4.16 Keterangan fitur <i>appet</i>	40
Tabel 4.17 Keterangan fitur <i>pe</i>	40
Tabel 4.18 Keterangan fitur <i>ane</i>	40
Tabel 4.19 Imputasi data hilang kurang dari atau sama dengan 10%	41
Tabel 4.20 Data contoh perhitungan manual menggunakan algoritma MICE.....	42

Tabel 4.21 Data imputasi sementara	43
Tabel 4.22 Tabel pembantu untuk mencari persamaan regresi fitur <i>hemo</i>	44
Tabel 4.23 Nilai baru fitur <i>hemo</i> iterasi 1	46
Tabel 4.24 Tabel pembantu untuk mencari persamaan regresi fitur <i>wc</i>	46
Tabel 4.25 Nilai baru fitur <i>wc</i> iterasi 1	48
Tabel 4.26 Perbandingan nilai sementara dengan imputasi MICE Iterasi 1	48
Tabel 4.27 Hasil imputasi MICE pada data contoh	49
Tabel 4.28 Hasil imputasi MICE pada <i>dataset</i> penelitian.....	50
Tabel 4.29 <i>Confussion matrix</i> sebelum penanganan data hilang	51
Tabel 4.30 <i>Confussion matrix</i> setelah penanganan data hilang	53
Tabel 4.31 Perbandingan hasil pengujian kinerja kombinasi algoritma	55
Tabel 4.32 Hasil perhitungan rata-rata presisi dan <i>recall</i>	57
Tabel 4.33 Perbandingan hasil penelitian dengan penelitian lainnya	57

BAB I

PENDAHULUAN

1.1 Latar Belakang

Salah satu permasalahan yang sering muncul pada data adalah adanya *missing data* atau data hilang (Darwanto *et al.*, 2021). Data hilang dapat terjadi karena beberapa hal, diantaranya adalah kesalahan prosedur entri data manual, kesalahan perangkat dan kesalahan pengukuran (Desiani *et al.*, 2021). Permasalahan data hilang memiliki dampak negatif, diantaranya mengurangi keakuratan model prediksi dan kesalahan dalam interpretasi data (Oktavianti & Yanti, 2022).

Ada dua metode dalam mengatasi data hilang, yang pertama *deletion* dan yang kedua imputasi. Metode *deletion* adalah penanganan data hilang dengan cara menghapus instan (*record*) atau fitur yang memiliki data hilang. *Deletion* pada *record* dapat dilakukan secara langsung jika persentase data hilang sebesar 1-5%, sedangkan *deletion* pada fitur tidak dapat dilakukan secara langsung karena dapat menurunkan kinerja klasifikasi pada data (Desiani *et al.*, 2021). *Deletion* pada fitur dilakukan dengan menghilangkan fitur yang tidak penting atau disebut dengan seleksi fitur (Jupriyadi, 2018). Salah satu algoritma seleksi fitur adalah algoritma ReliefF. Algoritma ReliefF adalah metode seleksi fitur berbasis filter yang memberikan peringkat pada fitur yang paling signifikan berdasarkan tingkat relevannya (Zainudin *et al.*, 2017). Kelebihan dari algoritma ReliefF adalah tidak membatasi tipe data yang digunakan dan dapat bekerja pada data yang tidak lengkap (Osanaiye *et al.*, 2019).

Beberapa penelitian menggunakan algoritma ReliefF menunjukkan hasil kinerja yang baik. Baliarsingh *et al.*, (2019) menerapkan algoritma ReliefF pada klasifikasi data penyakit colon tumor, menunjukkan peningkatan akurasi hingga sebesar 84,26%. Kilicarlan *et al.*, (2020) menggabungkan Algoritma ReliefF dan *Convolutional Neural Network* (CNN) untuk mendiagnosa dan mengklasifikasi data *Central Nervous System* (CNS), menunjukkan peningkatan akurasi sebesar 18,75%. Yahdin *et al.*, (2021) menerapkan algoritma ReliefF pada data *tracer* alumni Universitas Sriwijaya, menunjukkan peningkatan akurasi sebesar 0,98% dengan metode klasifikasi *Naïve Bayes* dan peningkatan akurasi sebesar 5,96% dengan metode klasifikasi KNN.

Metode kedua untuk menangani data hilang adalah imputasi. Imputasi adalah metode penanganan data hilang yang bekerja dengan mengestimasi nilai yang hilang dengan suatu nilai tertentu yang dianggap sesuai (Darwanto *et al.*, 2021). Imputasi dilakukan dengan dua cara, yang pertama *single imputation* (imputasi tunggal) dan yang kedua *multiple imputation* (imputasi ganda) (Oktavianti & Yanti, 2022). Imputasi tunggal dilakukan dengan mengisi setiap nilai hilang secara langsung, dengan pendekatan nilai tunggal berupa *mean*, modus, ataupun median (Khan & Hoque, 2020). Pada imputasi ganda setiap data hilang diisi dengan nilai yang berbeda-beda (Austin *et al.*, 2021). Salah satu algoritma imputasi ganda yang sering digunakan adalah algoritma *Multiple Imputation by Chained Equation* (MICE). MICE adalah algoritma imputasi ganda yang mengestimasi setiap nilai yang hilang menggunakan persamaan regresi secara iteratif pada setiap fitur. (Wulff & Ejlskov, 2017).

Beberapa penelitian menggunakan algoritma MICE menunjukkan hasil kinerja yang baik. Wu *et al.*, (2019) membandingkan enam cara imputasi termasuk algoritma MICE pada data pengobatan kanker payudara dan tingkat kesuburan yang meningkatkan akurasi hingga 74,7%. Hegde *et al.* (2019) menerapkan imputasi MICE pada data penyakit disglukemia, penerapannya meningkatkan hasil akurasi hingga lebih dari 80%. Rafsunjani & Safa (2019) membandingkan lima cara imputasi termasuk algoritma MICE pada data sistem tekanan udara yang meningkatkan akurasi hingga 94,79%.

Salah satu *dataset* bidang kesehatan yang memiliki masalah data hilang yaitu *dataset* penyakit ginjal kronis yang berasal dari *University of California Irvine (UCI) Machine Learning Repository*, dengan persentase data hilang sebesar 10,12%. *Dataset* tersebut dapat diakses melalui laman (https://archive.ics.uci.edu/ml/datasets/Chronic_Kidney_Disease) yang terdiri atas 400 *record*, sebanyak 250 *record* adalah penderita penyakit ginjal kronis dan 150 *record* tidak menderita penyakit ginjal kronis. *Dataset* tersebut memiliki 25 fitur, terdiri dari 24 fitur yang digunakan untuk mendiagnosa penyakit ginjal kronis dan 1 fitur *class* yang dijadikan sebagai penentu apakah pasien menderita penyakit ginjal kronis.

Penelitian ini mengkombinasikan algoritma ReliefF dan MICE untuk menangani data hilang pada *dataset* UCI penyakit ginjal kronis. Pada penelitian ini ReliefF digunakan untuk menyeleksi fitur pada *dataset*, sedangkan MICE digunakan untuk mengimputasi data hilang. Hasil penanganan data hilang dengan kombinasi kedua algoritma kemudian dilakukan pengujian dengan metode klasifikasi, untuk melihat bagaimana kombinasi algoritma ReliefF dan MICE

mempengaruhi kinerja klasifikasi. Pengukuran pada penelitian ini dilakukan dengan mengukur akurasi, presisi dan *recall* menggunakan algoritma *K-Nearest Neighbors* (KNN).

1.2 Perumusan Masalah

Rumusan masalah dari penelitian ini adalah bagaimana penerapan kombinasi algoritma ReliefF dan MICE dalam penanganan data hilang untuk meningkatkan kinerja klasifikasi pada *dataset* UCI penyakit ginjal kronis.

1.3 Pembatasan Masalah

Beberapa pembatasan masalah dalam penelitian ini adalah :

1. Penelitian ini hanya membahas mengenai penanganan data hilang pada *dataset* penyakit ginjal kronis yang berasal dari *UCI Machine Learning Repository* yang memiliki dua label kelas yaitu C_1 (pasien yang menderita penyakit ginjal kronis) dan C_2 (pasien yang tidak menderita penyakit ginjal kronis) tanpa membahas mengenai pasien yang tidak menderita penyakit ginjal.
2. Pengujian kinerja kombinasi algoritma dilakukan hanya menggunakan metode KNN tanpa membahas bagaimana proses klasifikasi.
3. Ukuran pengujian kinerja kombinasi algoritma pada penelitian ini menggunakan akurasi, presisi dan *recall*.

1.4 Tujuan

Tujuan penelitian ini adalah mengatasi data hilang pada *dataset* UCI penyakit ginjal kronis menggunakan kombinasi algoritma ReliefF dan MICE dengan melihat nilai akurasi, presisi, dan *recall* yang diperoleh.

1.5 Manfaat

Manfaat dari hasil penelitian ini adalah :

1. Memiliki *dataset* penyakit ginjal kronis UCI *Machine Learning Repository* yang lengkap.
2. Memiliki referensi data untuk mempelajari penyakit ginjal kronis yang lengkap.
3. Memberikan referensi alternatif penelitian baru dalam menangani data hilang.
4. Sebagai rekomendasi di bidang kesehatan untuk memprediksi penyakit ginjal kronis secara dini.

DAFTAR PUSTAKA

- Abbasi, Z., & Rahmani, M. (2019). An instance selection algorithm based on ReliefF. *International Journal on Artificial Intelligence Tools*, 28(1).
- Abedalkhader, W., & Abdulrahman, N. (2017). Missing data classification of chronic kidney disease. *International Journal of Data Mining & Knowledge Management Process*, 7(5/6), 55–61.
- Aleryani, A., Wang, W., & de la Iglesia, B. (2020). Multiple Imputation ensembles (mie) for dealing with missing data. *SN Computer Science*, 1(3).
- Austin, P. C., White, I. R., Lee, D. S., & van Buuren, S. (2021). Missing data in clinical research: a tutorial on multiple imputation. *Canadian Journal of Cardiology*, 37(9), 1322–1331.
- Baliarsingh, S. K., Ding, W., Vipsita, S., & Bakshi, S. (2019). A memetic algorithm using emperor penguin and social engineering optimization for medical data classification. *Applied Soft Computing Journal*, 85, 105773.
- Cherrington, M., Thabtah, F., Lu, J., & Xu, Q. (2019). Feature selection: Filter methods performance challenges. *2019 International Conference on Computer and Information Sciences, ICCIS 2019*, 1–4.
- Darwanto, A. R. S., Taza Luzia Viarindita, & Yekti Widyaningsih. (2021). Analisis regresi logistik binomial dan algoritma random forest pada proses pengklasifikasian penyakit ginjal kronis. *Jurnal Statistika Dan Aplikasinya*, 5(1), 1–14.
- Desiani, A., Dewi, N. R., Fauza, A. N., Rachmatullah, N., Arhami, M., & Nawawi, M. (2021). Handling missing data using combination of deletion technique, mean, mode and artificial neural network imputation for heart disease dataset. *Science and Technology Indonesia*, 6(4), 303–312.
- Ebiaredoh-Mienye, S. A., Swart, T. G., Esenogho, E., & Mienye, I. D. (2022). A machine learning method with filter-based feature selection for improved prediction of chronic kidney disease. *Bioengineering*, 9(8).
- Eiras-Franco, C., Guijarro-Berdiñas, B., Alonso-Betanzos, A., & Bahamonde, A. (2021). Scalable feature selection using ReliefF aided by locality-sensitive hashing. *International Journal of Intelligent Systems*, 36(11), 6161–6179.
- Ginkel, J. R. Van, Linting, M., Rippe, R. C. A., Voort, A. Van Der, Ginkel, J. R. Van, Linting, M., Rippe, R. C. A., & Voort, A. Van Der. (2019). Rebutting existing misconceptions about multiple imputation as a method for handling missing data rebutting existing misconceptions about multiple imputation as a method for. *Journal of Personality Assessment*, 0(0), 1–12.
- Han, J., Pei, J., & Tong, H. (2022). *Data mining: concepts and techniques*. Morgan kaufmann.

- Hasnain, M., Pasha, M. F., Ghani, I., Imran, M., Alzahrani, M. Y., & Budiarto, R. (2020). Evaluating trust prediction and confusion matrix measures for web services ranking. *IEEE Access*, 8, 90847–90861.
- Hegde, H., Shimpi, N., Panny, A., Glurich, I., Christie, P., & Acharya, A. (2019). MICE vs PPCA: Missing data imputation in healthcare. *Informatics in Medicine Unlocked*, 17(November), 100275.
- Jupriyadi. (2018). Implementasi Seleksi Fitur Menggunakan Algoritma FVBRM Untuk Klasifikasi Serangan Pada Intrusion Detection System (Ids). *Seminar Nasional Teknologi Informasi (SEMNASTEK)*, 17(January 2018), 1–6. <https://jurnal.umj.ac.id/index.php/semnastek/article/view/3452/2601>
- Khan, S. I., & Hoque, A. S. M. L. (2020). SICE: an improved missing data imputation technique. *Journal of Big Data*, 7(1).
- Kilicarlan, S., Adem, K., & Celik, M. (2020). Diagnosis and classification of cancer using hybrid model based on ReliefF and convolutional neural network. *Medical Hypotheses*, 137(December 2019), 109577.
- Mera-Gaona, M., Neumann, U., Vargas-Canas, R., & López, D. M. (2021). Evaluating the impact of multivariate imputation by MICE in feature selection. *PLoS ONE*, 16(7 July), 1–28.
- Oktavianti, S. A., & Yanti, T. S. (2022). Penanganan data hilang menggunakan metode markovichain monte carlo (MCMC). *Bandung Conference Series: Statistics*, 2(2), 409–415.
- Osanaiye, O., Ogundile, O., Aina, F., & Periola, A. (2019). Feature selection for intrusion detection system in a cluster-based heterogeneous wireless sensor network. *Facta Universitatis - Series: Electronics and Energetics*, 32(2), 315–330.
- Prasetio, R. T. (2020). Seleksi fitur dan optimasi parameter KNN berbasis algoritma genetika pada dataset medis. *Jurnal Responsif: Riset Sains Dan Informatika*, 2(2), 213–221.
- Rafsunjani, S., & Safa, R. S. (2019). *An empirical comparison of missing value imputation techniques on APS failure prediction*. February, 21–29.
- Stief, A., Ottewill, J. R., & Baranowski, J. (2020). ReliefF-based feature ranking and feature selection for monitoring induction motors. *2018 23rd International Conference on Methods & Models in Automation & Robotics (MMAR)*, 675215, 171–176.
- Sun, L., Kong, X., Xu, J., Xue, Z., Zhai, R., & Zhang, S. (2019). A hybrid gene selection method based on relieff and ant colony optimization algorithm for tumor classification. *Scientific Reports*, 9(1), 1–14.
- Susanti, Shantika, M., & Evy, S. (2018). K nearest neighbor dalam imputasi missing data. *Buletin Ilmiah Math. Stat. Dan Terapannya (Bimaster)*, 07(1), 9–14.

- Urbanowicz, R. J., Meeker, M., La Cava, W., Olson, R. S., & Moore, J. H. (2018). Relief-based feature selection: Introduction and review. *Journal of Biomedical Informatics*, 85(July), 189–203.
- Wahyuningsih, S., & Utari, D. R. (2018). Perbandingan metode k-nearest neighbor , naive bayes dan decision tree untuk prediksi kelayakan pemberian kredit. *Konferensi Nasional Sistem Informasi 2018 STMIK Atma Luhur Pangkalpinang*, 8 – 9 Maret 2018, 619–623.
- Wibawa, M. S., Maysanjaya, I. M. D., & Putra, I. M. A. W. (2017). Boosted classifier and features selection for enhancing chronic kidney disease diagnose. *2017 5th International Conference on Cyber and IT Service Management, CITSM 2017*.
- Wu, X., Akbarzadeh Khorshidi, H., Aickelin, U., Edib, Z., & Peate, M. (2019). Imputation techniques on missing values in breast cancer treatment and fertility data. *Health Information Science and Systems*, 7(1), 1–8.
- Wulff, J. N., & Ejlskov, L. (2017). Multiple imputation by chained equations in praxis: Guidelines and review. *Electronic Journal of Business Research Methods*, 15(1), 41–56.
- Yahdin, S., Desiani, A., Gofar, N., & Agustin, K. (2021). Application of the relief-f algorithm for feature selection in the prediction of the relevance education background with the graduate employment of the universitas sriwijaya. *Computer Engineering and Applications Journal*, 10(2), 71–80.
- Zainudin, M. N. Shah, Sulaiman, M. N., Mustapha, N., Perumal, T., Nazri, A. S. A., Mohamed, R., & Manaf, S. A. (2017). Feature selection optimization using hybrid relief-f with self-adaptive differential evolution. *International Journal of Intelligent Engineering and Systems*, 10(2), 21–29.
- Zainudin, Muhammad Noorazlan Shah, Sulaiman, N., Mustapha, N., Perumal, T., & Mohamed, R. (2018). Two-stage feature selection using ranking self-adaptive differential evolution algorithm for recognition of acceleration activity. *Turkish Journal of Electrical Engineering and Computer Sciences*, 26(3), 1378–1389.