

**KOMBINASI METODE MICE DAN ADASYN
UNTUK PENANGANAN DATA HILANG
DAN KETIDAKSEIMBANGAN DATA
PADA KLASIFIKASI PENYAKIT JANTUNG**

SKRIPSI

Oleh:

ANANDA PRATIWI

08011282025028



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

2024

SURAT PERNYATAAN KEASLIAN KARYA ILMIAH

Yang bertanda tangan dibawah ini:

Nama Mahasiswa : Ananda Pratiwi
NIM : 08011282025028
Fakultas/Jurusan : Matematika

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

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

Demikianlah surat pernyataan ini saya buat dengan sebenarnya.

Indralaya, 18 Juli 2024



Ananda Pratiwi
NIM. 08011282025028

LEMBAR PENGESAHAN

**KOMBINASI METODE MICE DAN ADASYN
UNTUK PENANGANAN DATA HILANG
DAN KETIDAKSEIMBANGAN DATA
PADA KLASIFIKASI PENYAKIT JANTUNG**

SKRIPSI

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

Oleh :

**ANANDA PRATIWI
08011282025028**

Pembimbing Kedua



**Drs. Ali Amran, M.T.
NIP. 196612131994021001**

**Indralaya, Juli 2024
Pembimbing Utama**



**Dr. Anita Desiani, S.Si., M.Kom.
NIP. 197712112003122002**

**Mengetahui,
Ketua Jurusan Matematika**



**Dr. Dian Cahyawati S., S.Si., M.Si
NIP. 197303212000122001**

HALAMAN PERSEMBAHAN

Kupersembahkan skripsi ini untuk:

Yang Maha Kuasa Allah Subhanahu Wa Ta'ala,

Kedua orang tuaku tersayang,

Satu-satunya saudariku,

Keluarga besarku,

Semua guru dan dosenku,

Sahabat-sahabatku,

Almamaterku

Motto

“If it’s meant to be, it will be”

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 Metode MICE dan ADASYN Untuk Penanganan Data Hilang dan Ketidakseimbangan Data pada Klasifikasi Penyakit Jantung” 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 tua tercinta, Bapak **M. Nasir** dan Ibu **Atikah**, 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. Satu-satunya adik tersayang, **Nazwa Dwy Maharani**, yang selalu mendoakan dan memberikan perhatian kepada penulis, beserta keluarga besar yang selalu mendukung penulis.
2. Bapak **Prof. Hermansyah, S.Si., Ph.D** selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam, Bapak **Dr. Hasanudin, S.Si., M.Si.** selaku Wakil Dekan I Bidang Akademik, Bapak **Dr. M. Hendri, S.T., M.Si.** selaku Wakil Dekan II Bidang Umum, Keuangan, dan Kepegawaian, dan Bapak **Dr. Fiber Monado, M.Si.** selaku Wakil Dekan III Bidang Kemahasiswaan dan Alumni Universitas Sriwijaya.

3. Ibu **Dr. Dian Cahyawati Sukanda, S.Si., M.Si** selaku Ketua Jurusan Matematika FMIPA Universitas Sriwijaya dan Dosen Pembimbing Akademik yang telah memberikan arahan dan motivasi kepada penulis selama proses perkuliahan dan Ibu **Des Alwin Zayanti, M.Si** selaku Sekretaris Jurusan Matematika FMIPA Universitas Sriwijaya yang telah mengarahkan urusan akademik kepada penulis.
4. Ibu **Dr. Anita Desiani, S.Si., M.Kom** dan Bapak **Drs. Ali Amran, M.T** selaku dosen pembimbing 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. Ibu **Dr. Yuli Andriani, S.Si., M.Si** dan Bapak **Drs. Endro Setyo Cahyono, M.Si** selaku dosen pembahas dan penguji yang telah memberikan tanggapan, kritik, dan saran yang sangat bermanfaat untuk perbaikan dan penyelesaian skripsi ini. Ibu **Dra. Ning Eliyati, M.Pd** dan Bapak **Dr. Bambang Suprihatin, S.Si., M.Si** selaku tim pelaksana Seminar penulis.
6. **Seluruh Dosen di Jurusan Matematika FMIPA** yang telah memberikan ilmu, nasihat, motivasi, serta bimbingan selama proses perkuliahan. Pak **Irwansyah** selaku admin dan Ibu **Hamidah** selaku pegawai tata usaha Jurusan Matematika FMIPA yang telah membantu penulis selama perkuliahan.
7. **Semua sahabat seperjuangan, Komputasi 2020, BPH Himastik Akselerasi, BPH Himastik Sinergi Cita, KPU FMIPA UNSRI, Keluarga Matematika 2020** dan rekan-rekan selama perkuliahan dan proses skripsi. Terima kasih

sudah menjadi orang-orang baik di sekeliling penulis yang selalu mendukung, membantu dengan tulus, dan memberi energi positif.

8. Kakak-kakak tingkat angkatan 2017, 2018, dan 2019 serta adik-adik tingkat angkatan 2021, 2022 dan 2023, terima kasih atas segala kebaikan dan bantuan.
9. 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, Juli 2024

Penulis

**COMBINATION OF MICE AND ADASYN METHODS
FOR HANDLING MISSING DATA
AND DATA IMBALANCE
IN HEART DISEASE CLASSIFICATION**

By :

**ANANDA PRATIWI
NIM.08011282025028**

ABSTRACT

The quality of data is determined by several factors, namely data completeness and data balance. The University of California Irvine (UCI) heart disease dataset has issues with missing data and data imbalance, which, if not addressed, can lead to reduced model prediction accuracy and errors in data interpretation. Missing data can be handled using several methods, one of which is data imputation. For missing data less than or equal to 5%, it can be handled using the mode method for nominal attributes. For missing data greater than 5%, the Multiple Imputation by Chained Equations (MICE) method is used. Data imbalance can be addressed using several methods, one of which is oversampling. One method that can be used in the oversampling technique is the Adaptive Synthetic Sampling Approach (ADASYN). To see how handling missing data and data imbalance affects the performance of heart disease classification, performance testing was conducted using the Random Forest, Support Vector Machine (SVM), and Multi-Layer Perceptron (MLP) classification methods. There was an increase in classification performance results after handling missing data and data imbalance. Accuracy increased by 18.48%, precision increased by 18.5%, and recall increased by 18.4%. Based on the obtained results, it can be concluded that the MICE and ADASYN methods can improve classification performance on the UCI heart disease dataset.

Keywords : Missing Data, Data Imbalance, MICE, ADASYN, Classification

**KOMBINASI METODE MICE DAN ADASYN
UNTUK PENANGANAN DATA HILANG
DAN KETIDAKSEIMBANGAN DATA
PADA KLASIFIKASI PENYAKIT JANTUNG**

Oleh :

**ANANDA PRATIWI
NIM.08011282025028**

ABSTRAK

Kualitas suatu data ditentukan oleh beberapa hal yakni kelengkapan data dan keseimbangan data. *Dataset* penyakit jantung *Univercity of California Irvine* (UCI) memiliki permasalahan data hilang dan data tidak seimbang yang mana apabila tidak ditangani dapat menyebabkan kurangnya keakuratan model prediksi dan kesalahan dalam menginterpretasi data. Untuk mengatasi data hilang dapat dilakukan dengan beberapa metode, salah satunya adalah imputasi data. Untuk data hilang kurang dari atau sama dengan 5% dapat ditangani menggunakan metode modus pada atribut nominal. Untuk penanganan data hilang lebih dari 5% menggunakan metode MICE. Ketidakseimbangan data dapat ditangani menggunakan beberapa metode, salah satunya yakni *oversampling*. Salah satu metode yang dapat digunakan dalam teknik *oversampling* adalah metode *Adaptive Synthetic Sampling Approach* (ADASYN). Untuk melihat bagaimana penanganan data hilang dan ketidakseimbangan data mempengaruhi kinerja dari klasifikasi penyakit jantung, dilakukan pengujian kinerja klasifikasi menggunakan metode *Random Forest*, *Support Vector Machine* (SVM) dan *Multi Layer Perceptron* (MLP). Terdapat kenaikan hasil kinerja klasifikasi setelah dilakukan penanganan data hilang dan ketidakseimbangan data. Akurasi meningkat sebesar 18,48%, kenaikan *presisi* sebesar 18,5% *serta* recall sebesar 18,4%. Berdasarkan hasil yang diperoleh dapat disimpulkan bahwa metode MICE dan ADASYN dapat meningkatkan kinerja klasifikasi pada *dataset* penyakit jantung UCI.

Kata Kunci : Data Hilang, Ketidakseimbangan Data, MICE, ADASYN, Klasifikasi

DAFTAR ISI

HALAMAN JUDUL	i
LEMBAR PENGESAHAN	ii
HALAMAN PERSEMBAHAN	iv
KATA PENGANTAR.....	v
ABSTRACT	viii
ABSTRAK	ix
DAFTAR ISI.....	x
DAFTAR TABEL	xii
DAFTAR GAMBAR.....	xiv
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 Ketidakeimbangan Data.....	6
2.3 Metode <i>Multiple Imputation by Chained Equation</i> (MICE).....	7
2.4 Metode <i>Adaptive Synthetic Sampling Approach</i> (ADASYN).....	10
2.5 Pengukuran Kinerja Klasifikasi.....	13
BAB III METODOLOGI PENELITIAN	15
3.1 Tempat	15
3.2 Waktu.....	15
3.3 Alat	15
3.4 Metode Penelitian	15
BAB IV HASIL DAN PEMBAHASAN.....	19
4.1 Deskripsi Data	19
4.2 Imputasi Data Hilang.....	21
4.2.1 Imputasi Data Hilang pada Atribut yang Memiliki Data Hilang Kurang Dari atau Sama Dengan 5%.....	21

4.2.2	Imputasi Data Hilang pada Atribut dengan Data Hilang Lebih Dari 5%.....	22
4.3	Penanganan Ketidakseimbangan Data	33
4.3.1	Perhitungan Manual ADASYN.....	33
4.3.2	Penerapan Metode ADASYN pada Data Penelitian.....	45
4.4	Pengukuran Kinerja Kombinasi Metode	48
4.5	Analisis dan Interpretasi Hasil	53
BAB V	KESIMPULAN DAN SARAN	57
5.1	Kesimpulan.....	57
5.2	Saran.....	57
DAFTAR PUSTAKA	58

DAFTAR TABEL

Tabel 2.1 Confusion Matrix pada Klasifikasi	13
Tabel 2.2 Kategori Nilai Akurasi	14
Tabel 4.1 Dataset Penyakit Jantung	19
Tabel 4.2 Keterangan Atribut	20
Tabel 4.3 Keterangan Atribut restecg.....	21
Tabel 4.4 Imputasi Data Hilang Kurang dari atau sama dengan 5%	22
Tabel 4.5 Data Perhitungan Manual.....	22
Tabel 4.6 Imputasi Data Hilang Sementara dengan Mean.....	24
Tabel 4.7 Persamaan Regresi Atribut trestbps	27
Tabel 4.8 Nilai baru atribut trestbps iterasi 1	29
Tabel 4.9 Persamaan Regresi Atribut chol	29
Tabel 4.10 Nilai Baru Atribut chol Iterasi 1	29
Tabel 4.11 Persamaan Regresi Atribut thalach	29
Tabel 4.12 Nilai Baru Atribut thalach Iterasi 1	31
Tabel 4.13 Perbandingan Nilai Sementara dengan Imputasi MICE Iterasi 1	31
Tabel 4.14 Hasil imputasi MICE.....	32
Tabel 4.15 Dataset Hasil Imputasi Menggunakan Metode MICE	33
Tabel 4.16 Data Perhitungan Manual.....	33
Tabel 4.17 Evaluasi NN pada Dataset Kesatu (D1).....	36
Tabel 4.18 NN untuk Setiap Data Kategori 1	36
Tabel 4.19 Density Distribution untuk Data Kategori 1	37
Tabel 4.20 Jumlah Duplikasi Data Sintesis.....	37

Tabel 4.21 Data Sintetis Baru Setelah ADASYN	45
Tabel 4.22 Data Hasil Penelitian ADASYN	46
Tabel 4.23 Confusion Matrix Metode Random Forest Sebelum Diolah.....	48
Tabel 4.24 Confusion Matrix Metode Random Forest Setelah Diolah.....	50
Tabel 4.25 Confusion Matrix Metode SVM Sebelum Diolah	51
Tabel 4.26 Confusion Matrix Metode SVM Setelah Diolah.....	52
Tabel 4.27 Confusion Matrix metode MLP Sebelum Diolah	52
Tabel 4.28 Confusion Matrix metode MLP Setelah Diolah.....	53
Tabel 4.29 Rata-rata Hasil Akurasi, Presisi dan Recall.....	53
Tabel 4.30 Perbandingan Hasil Penelitian dengan Penelitian Lain	55

DAFTAR GAMBAR

Gambar 4. 1 Perbandingan Jumlah Data Sebelum dan Setelah Proses ADASYN pada Data	47
Gambar 4. 2 Grafik Perbandingan Rata-rata Hasil Klasifikasi.....	55

BAB I

PENDAHULUAN

1.1 Latar Belakang

Permasalahan kualitas data dapat ditentukan oleh beberapa hal, diantaranya adalah data hilang dan ketidakseimbangan data. Data hilang merupakan masalah umum yang hampir muncul di semua kumpulan data, karena kurangnya data atau informasi berupa nilai atribut yang hilang sehingga menyebabkan data tidak lengkap (Liu *et al.*, 2016; Desiani *et al.*, 2021). Permasalahan data hilang dapat menyebabkan kurangnya keakuratan model prediksi dan kesalahan dalam menginterpretasi data (Chen *et al.*, 2017).

Ketidakseimbangan data merupakan masalah yang terjadi apabila jumlah untuk setiap label kelas tidak merata, dengan jumlah data suatu kelas lebih banyak dibandingkan dengan kelas yang lain (Ali, Najib, et al., 2019). Data yang tidak seimbang dapat menyebabkan prediksi menjadi lebih cenderung ke kelompok mayoritas dan tidak sensitif terhadap kelompok minoritas (Ebenuwa et al., 2019). Akibatnya, tingkat keakuratan prediksi seperti akurasi, presisi dan *recall* yang baik hanya dihasilkan untuk kelas mayoritas sedangkan untuk kelas minoritas mendapatkan tingkat keakuratan prediksi yang buruk (Thabtah et al., 2020).

Dataset penyakit jantung merupakan salah satu *dataset* yang mengandung banyak data hilang dan data tidak seimbang yang dapat diperoleh dari website *University of California (UCI) Machine Learning Repository* dan dapat diakses melalui laman <https://archive.ics.uci.edu/dataset/45/-heart+disease>. *Dataset* penyakit jantung memiliki 11 atribut, terdiri dari 10 atribut yang digunakan untuk

mendiagnosa penyakit jantung dan 1 atribut *num* sebagai penentu apakah pasien menderita penyakit jantung. Label dengan kategori sehat atau 0 memiliki persentase data sebesar 54% sedangkan kategori sakit atau 1 memiliki persentase data sebesar 46%. Atribut yang mengalami data hilang dalam *dataset* ini sebanyak 9 atribut dengan persentase sebesar 19%.

Terdapat beberapa metode untuk mengatasi data hilang, salah satunya adalah metode imputasi data. Metode imputasi terbagi menjadi dua jenis yakni metode imputasi tunggal (*single imputation*) dan metode imputasi ganda (*multiple imputation*). Imputasi tunggal bekerja dengan cara mengganti semua data yang hnilai tunggal untuk semua data yang hilang dari variabel tersebut, dengan pendekatan berupa *mean*, modus, atau median (Azifah *et al.*, 2021). Imputasi *mean* dan modus hanya dapat digunakan apabila data yang hilang pada atribut tidak lebih dari 5%. *Mean* digunakan pada atribut yang bersifat numerik, sedangkan modus digunakan pada atribut yang bersifat nominal (Desiani *et al.*, 2021).

Untuk penanganan data hilang lebih dari 5% dapat menggunakan imputasi ganda (Desiani *et al.*, 2021). Imputasi ganda adalah metode yang dapat digunakan untuk mengatasi kekurangan yang terdapat pada metode imputasi tunggal (Lee *et al.*, 2022). Pada imputasi ganda, metode yang paling sering digunakan adalah *Multiple Imputation by Chained Equation* (MICE) (Jäger *et al.*, 2021).

MICE bekerja dengan mencocokkan model prediksi untuk setiap variabel dengan nilai hilang menggunakan model penduga regresi pada variabel lain dalam data (Laqueur *et al.*, 2022). Beberapa penelitian menggunakan MICE menunjukkan hasil kinerja yang baik, diantaranya adalah penelitian Mera-Gaona *et al.*, (2021)

menggunakan MICE pada *dataset* penyakit jantung dengan menghasilkan akurasi sebesar 85,8%. Selain itu, Wu *et al.*, (2019) menggunakan MICE pada *dataset* kanker payudara dengan meningkatkan akurasi sebesar 24%. Namun, penelitian-penelitian tersebut tidak menangani ketidakseimbangan data.

Untuk masalah ketidakseimbangan data dapat ditangani menggunakan beberapa metode, salah satunya yakni *oversampling* (Thabtah *et al.*, 2020). *Oversampling* merupakan metode pembangkitan data secara acak pada kelas minoritas sehingga jumlah data pada kelas minoritas tidak jauh berbeda dengan jumlah data pada kelas mayoritas (Ali, Najib, et al., 2019). Salah satu metode yang dapat digunakan dalam metode *oversampling* adalah metode *Adaptive Synthetic Sampling Approach* (ADASYN).

ADASYN merupakan metode sintesis data dengan membuat berbagai jumlah sampel baru sesuai dengan distribusi sampel minoritas (Guan *et al.*, 2023). Metode ADASYN telah banyak digunakan untuk penelitian seperti penelitian Ramadhan (2021) dengan menggunakan metode ADASYN dan klasifikasi *Support Vector Machine* (SVM) meningkatkan akurasi sebesar 4%. Metode *K-Nearest Neighbor* (KNN) juga cenderung mengalami peningkatan akurasi saat menerapkan ADASYN dengan peningkatan akurasi sebesar 15,64% (Kurniawati *et al.*, 2018).

Penelitian ini berfokus pada metode untuk mengatasi data hilang dan ketidakseimbangan data pada *dataset* penyakit jantung. Untuk penanganan data hilang, penelitian ini memadukan modus dan MICE. Modus diterapkan pada atribut yang mengalami data hilang kurang dari 5%. Untuk atribut yang mengalami data hilang lebih dari 5%, penelitian ini menerapkan metode MICE. Penanganan

ketidakseimbangan data pada penelitian ini menerapkan metode ADASYN. Hasil penanganan data hilang dan ketidakseimbangan data akan dilihat pengaruhnya pada kinerja klasifikasi. Adapun metode klasifikasi yang diterapkan yakni *Random Forest*, *Support Vector Machine* (SVM) dan *Multi Layer Perceptron* (MLP) untuk melihat bagaimana penanganan data hilang dan ketidakseimbangan data mempengaruhi kinerja dari klasifikasi penyakit jantung.

1.2 Perumusan Masalah

Rumusan masalah dalam penelitian ini adalah bagaimana cara meningkatkan kinerja klasifikasi penyakit jantung dengan mengaplikasikan kombinasi penanganan data hilang dan ketidakseimbangan data menggunakan metode MICE dan ADASYN yang dilihat dari keakuratan prediksi penyakit jantung sesuai nilai akurasi, presisi, dan *recall* yang diperoleh.

1.3 Pembatasan Masalah

1. Penanganan ketidakseimbangan data dengan metode ADASYN menggunakan jarak *Euclid*.
2. Pengujian kinerja kombinasi metode dilakukan dengan membandingkan tiga metode yakni *Random Forest*, *Support Vector Machine* (SVM) dan *Multi Layer Perceptron* (MLP) tanpa membahas bagaimana proses klasifikasi.
3. Pengujian kinerja menggunakan nilai akurasi, presisi dan *recall* yang diperoleh.

1.4 Tujuan

Tujuan dari penelitian adalah untuk meningkatkan meningkatkan kinerja klasifikasi penyakit jantung dengan mengaplikasikan kombinasi penanganan data

hilang dan ketidakseimbangan data menggunakan metode MICE dan ADASYN yang dilihat dari keakuratan prediksi penyakit jantung sesuai nilai akurasi, presisi, dan *recall* yang diperoleh.

1.5 Manfaat

1. Mendapatkan *dataset* yang lengkap tanpa mengandung data hilang dan ketidakseimbangan data yang dapat digunakan untuk klasifikasi prediksi penyakit jantung.
2. Menambah referensi di bidang kesehatan dalam memprediksi penyakit jantung.

DAFTAR PUSTAKA

- Aditsania, A., Adiwijaya, & Saonard, A. L. (2017). Handling Imbalanced Data in Churn Prediction Using ADASYN and Backpropagation Algorithm. *Proceeding - 2017 3rd International Conference on Science in Information Technology: Theory and Application of IT for Education, Industry and Society in Big Data Era, ICSITech 2017*.
- Ali, H., Najib, M., Salleh, B. M., Hussain, K., Salleh, M., Ahmad, A., Ullah, A., Muhammad, A., Naseem, R., & Khan, M. (2019). A Review on Data Preprocessing Methods for Class Imbalance Problem. *International Journal of Engineering & Technology*, 8(3), 390–397.
- Ali, H., Salleh, M. N. M., Saedudin, R., Hussain, K., & Mushtaq, M. F. (2019). Imbalance Class Problems in Data Mining: A Review. *Indonesian Journal of Electrical Engineering and Computer Science*, 14(3), 1552–1563.
- Azifah, N., Pauzi, M., Wah, Y. B., Deni, S. M., & Khatijah, S. (2021). Comparison of Single and MICE Imputation Methods for Missing Values. *Science & Technology*, 29(2), 979–998.
- Chen, M., Hao, Y., Hwang, K., Wang, L., & Wang, L. (2017). Disease Prediction by Machine Learning over Big Data from Healthcare Communities. *IEEE Access*, 5(c), 8869–8879.
- Deshmukh, R., Gourkhede, P., & Rangari, S. (2019). Heart Disease Prediction Using Artificial Neural Network. *Ijarccce*, 8(1), 85–89.
- 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.
- Dinh, D. T., Huynh, V. N., & Sriboonchitta, S. (2021). Clustering Mixed Numerical and Categorical Data With Missing Values. *Information Sciences*, 571, 418–442.
- Douzas, G., Bacao, F., & Last, F. (2018). Improving Imbalanced Learning Through a Heuristic Oversampling Method Based on K-Means and SMOTE. *Information Sciences*, 465, 1–20.
- Dütsch, I., & Gediga, G. (2019). Confusion Matrices and Rough Set Data Analysis. *Journal of Physics: Conference Series*, 1229(1).
- Ebenuwa, S. H., Sharif, M. S., Alazab, M., & Al-Nemrat, A. (2019). Variance Ranking Attributes Selection Techniques for Binary Classification Problem in Imbalance Data. *IEEE Access*, 7, 24649–24666.

- Guan, S., Yang, H., & Wu, T. (2023). Transformer Fault Diagnosis Method Based on TLR-ADASYN Balanced Dataset. *Scientific Reports*, 13(1), 1–15.
- Haixiang, G., Yijing, L., Shang, J., Mingyun, G., Yuanyue, H., & Bing, G. (2017). Learning from Class-Imbalanced Data: Review of Methods and Applications. *Expert Systems with Applications : An International Journal*, 73, 220–239.
- Hasan, M. K., Alam, M. A., Roy, S., Dutta, A., Jawad, M. T., & Das, S. (2021). Missing Value Imputation Affects the Performance of Machine Learning: A review and Analysis of the Literature (2010–2021). *Informatics in Medicine Unlocked*, 27, 100799.
- Hayaty, M., Muthmainah, S., & Ghufran, S. M. (2021). Random and Synthetic Over-Sampling Approach to Resolve Data Imbalance in Classification. *International Journal of Artificial Intelligence Research*, 4(2), 86.
- Jäger, S., Allhorn, A., & Bießmann, F. (2021). A Benchmark for Data Imputation Methods. 4(July), 1–16.
- Jakobsen, J. C., Glud, C., Wetterslev, J., & Winkel, P. (2017). When and How Should Multiple Imputation Be Used for Handling Missing Data in Randomised Clinical Trials - A Practical Guide With Flowcharts. *BMC Medical Research Methodology*, 17(1), 1–10.
- Khan, S. I., Sayed, A., & Hoque, L. (2020). SICE : an Improved Missing Data Imputation Technique. *Journal of Big Data*, 7(37), 2–21.
- Kurniawati, Y. E., Permanasari, A. E., & Fauziati, S. (2018). Adaptive Synthetic-Nominal (ADASYN-N) and Adaptive Synthetic-KNN (ADASYN-KNN) for Multiclass Imbalance Learning on Laboratory Test Data. *Proceedings - 2018 4th International Conference on Science and Technology, ICST 2018*, 1, 1–6.
- Laqueur, H. S., Shev, A. B., & Kagawa, R. M. C. (2022). SuperMICE: An Ensemble Machine Learning Approach to Multiple Imputation by Chained Equations. *American Journal of Epidemiology*, 191(3), 516–525.
- Lee, D. H., Woo, S. E., Jung, M. W., & Heo, T. Y. (2022). Evaluation of Odor Prediction Model Performance and Variable Importance According to Various Missing Imputation Methods. *Applied Sciences (Switzerland)*, 12(6).
- Lin, S., Lu, C., Shi, H., & Liu, X. (2020). Telecom Fraud Identification Based on ADASYN and Random Forest. *2020 5th International Conference on Computer and Communication Systems : ICCCS 2020*, 448–452.
- Liu, D., Liang, D., & Wang, C. (2016). A Novel Three-Way Decision Model Based on Incomplete Information System. *Knowledge-Based Systems*, 91, 32–45.
- Majid, A. M., & Utomo, W. H. (2021). Application of Discretization and Adaboost

- Method to Improve Accuracy Of Classification Algorithms in Predicting Diabetes Mellitus. *ICIC Express Letters, Part B: Applications*, 12(12), 1177–1184.
- 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, 1–28.
- Nadzurah, Z. A., Amelia Ritahani, I., & Nurul, A. (2018). Performance Analysis of Machine Learning Algorithms for Missing Value Imputation. *International Journal of Advanced Computer Science and Applications*, 9(6).
- Nalluri, S., Vijaya Saraswathi, R., Ramasubbareddy, S., Govinda, K., & Swetha, E. (2018). Heart Disease Prediction Using Data Mining Techniques. *Advances in Intelligent Systems and Computing*, 09, 903–912.
- Pedersen, A. B., Mikkelsen, E. M., Cronin-Fenton, D., Kristensen, N. R., Pham, T. M., Pedersen, L., & Petersen, I. (2017). Missing Data and Multiple Imputation in Clinical Epidemiological Research. *Clinical Epidemiology*, 9, 157–166.
- Ramadhan, N. G. (2021). Comparative Analysis of ADASYN-SVM and SMOTE-SVM Methods on the Detection of Type 2 Diabetes Mellitus. *Scientific Journal of Informatics*, 8(2), 276–282.
- Ramprakash, P., Sarumathi, R., Mowriya, R., & Nithyavishnupriya, S. (2020). Heart Disease Prediction Using Deep Neural Network. *Proceedings of the 5th International Conference on Inventive Computation Technologies, ICICT 2020*, 666–670.
- Seliem, M. M. (2022). Handling Outlier Data as Missing Values by Imputation Methods: Application of Machine Learning Algorithms. *Turkish Journal of Computer and Mathematics Education*, 13(01), 273–286.
- Siddappa, N. G., & Kampalappa, T. (2019). Adaptive Condensed Nearest Neighbor for Imbalance Data Classification. *International Journal of Intelligent Engineering and Systems*, 12(2), 104–113.
- Thabtah, F., Hammoud, S., Kamalov, F., & Gonsalves, A. (2020). Data Imbalance in Classification: Experimental Evaluation. *Information Sciences*, 513, 429–441.
- 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., & Rodiah, D. (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 and Applications Journal*, 10(2), 71–80.