

KLASIFIKASI ANEMIA MENGGUNAKAN METODE *RANDOM FOREST*

SKRIPSI



**UIN SUNAN AMPEL
S U R A B A Y A**

Disusun Oleh

LAILI NUR FARIDA

09010220007

**PROGRAM STUDI MATEMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI SUNAN AMPEL
SURABAYA**

2025

PERNYATAAN KEASLIAN

Saya yang bertanda tangan di bawah ini,

Nama : LAILI NUR FARIDA

NIM : 09010220007

Program Studi : Matematika

Angkatan : 2020

Menyatakan bahwa saya tidak melakukan plagiat dalam penulisan skripsi saya yang berjudul "KLASIFIKASI ANEMIA MENGGUNAKAN METODE *RANDOM FOREST*". Apabila suatu saat nanti terbukti saya melakukan tindakan plagiat, maka saya bersedia menerima sanksi yang telah ditetapkan.

Demikian pernyataan keaslian ini saya buat dengan sebenar-benarnya.

Surabaya, 20 Desember 2024

Yang menyatakan,



LAILI NUR FARIDA

NIM. 09010220007

LEMBAR PERSETUJUAN PEMBIMBING

Skripsi oleh

Nama : LAILI NUR FARIDA
NIM : 09010220007
Judul skripsi : KLASIFIKASI ANEMIA MENGGUNAKAN METODE
RANDOM FOREST

telah diperiksa dan disetujui untuk diujikan.

Pembimbing I




Dr. Dian Candra Rini Novitasari, M.Kom
NIP. 198511242014032001

Pembimbing II



Wika Dianita Utami, M.Sc
NIP. 199206102018012003

Mengetahui,
Ketua Program Studi Matematika
UIN Sunan Ampel Surabaya



Yuniar Farida, MT
NIP. 197905272014032002

PENGESAHAN TIM PENGUJI SKRIPSI

Skripsi oleh

Nama : LAILI NUR FARIDA
NIM : 09010220007
Judul Skripsi : *KLASIFIKASI ANEMIA MENGGUNAKAN METODE
RANDOM FOREST*

Telah dipertahankan di depan Tim Penguji
pada tanggal 20 Desember 2024

Mengesahkan,
Tim Penguji

Penguji I




Aris Fanani, M.Kom
NIP. 198701272014031002

Penguji II



Dian Yuliati, M.Si
NIP. 198707142020122015

Penguji III



Dr. Dian Candra Kini Novitasari, M.Kom
NIP. 198511242014032001

Penguji IV



Wika Dianita Utami, M.Sc
NIP. 199206102018012003

Mengetahui,

Dekan Fakultas Sains dan Teknologi
JIN Surab Ampel Surabaya



Hamdani, M.Pd
NIP. 198607312000031002



UIN SUNAN AMPEL
SURABAYA

KEMENTERIAN AGAMA
UNIVERSITAS ISLAM NEGERI SUNAN AMPEL SURABAYA
PERPUSTAKAAN

Jl. Jend. A. Yani 117 Surabaya 60237 Telp. 031-8431972 Fax.031-8413300
E-Mail: perpus@uinsby.ac.id

LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI
KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

Sebagai sivitas akademika UIN Sunan Ampel Surabaya, yang bertanda tangan di bawah ini, saya:

Nama : LAILI NUR FARIDA
NIM : 09010220007
Fakultas/Jurusan : Sains dan Teknologi / Matematika
E-mail address : lailifarida24@gmail.com

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Perpustakaan UIN Sunan Ampel Surabaya, Hak Bebas Royalti Non-Eksklusif atas karya ilmiah :

Sekripsi Tesis Desertasi Lain-lain (.....)

yang berjudul :

KLASIFIKASI ANEMIA MENGGUNAKAN METODEE RANDOM
FOREST

beserta perangkat yang diperlukan (bila ada). Dengan Hak Bebas Royalti Non-Eksklusif ini Perpustakaan UIN Sunan Ampel Surabaya berhak menyimpan, mengalih-media/format-kan, mengelolanya dalam bentuk pangkalan data (database), mendistribusikannya, dan menampilkan/mempublikasikannya di Internet atau media lain secara *fulltext* untuk kepentingan akademis tanpa perlu meminta ijin dari saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan atau penerbit yang bersangkutan.

Saya bersedia untuk menanggung secara pribadi, tanpa melibatkan pihak Perpustakaan UIN Sunan Ampel Surabaya, segala bentuk tuntutan hukum yang timbul atas pelanggaran Hak Cipta dalam karya ilmiah saya ini.

Demikian pernyataan ini yang saya buat dengan sebenarnya.

Surabaya, 8 Januari 2025

Penulis

(LAILI NUR FARIDA)

nama terang dan tanda tangan

ABSTRAK

KLASIFIKASI ANEMIA MENGGUNAKAN METODE *RANDOM FOREST*

Anemia adalah kelainan darah yang disebabkan karena kurangnya sel darah merah dan kadar hemoglobin. Berdasarkan survei yang dilakukan oleh *World Health Organization* (WHO), anemia memengaruhi 1.62 miliar orang di seluruh dunia. Sebanyak 93% pasien penyakit kronis mengalami anemia. Anemia dapat didiagnosis menggunakan *Complete Blood Count* (CBC) yang bertujuan untuk mengevaluasi jumlah total dan karakteristik komponen sel dalam darah. Tujuan dari penelitian adalah untuk melakukan klasifikasi *Random Forest* (RF) terhadap anemia. Penelitian ini menggunakan data numerik berupa karakteristik sel darah, yang terdiri dari 269 pasien *Normocytic Normochromic Anemia* (NNA), 189 pasien *Iron Deficiency Anemia* (IDA), dan 336 pasien *Healthy* (HE). Pengolahan data yang dilakukan pada penelitian ini terdiri atas beberapa tahap, yaitu input data, dilanjutkan dengan analisis deskriptif dan proses *data encoding*. Setelah itu, pembagian data menggunakan *K-Fold Cross Validation* dengan k sejumlah 5. Tahap selanjutnya adalah proses klasifikasi menggunakan RF dengan uji coba parameter pada data latih, yaitu jumlah pohon keputusan yang akan dibangun, kedalaman maksimal dari setiap pohon, dan jumlah minimal sampel pada setiap daun. Berdasarkan parameter tersebut diperoleh hasil klasifikasi berupa diagnosis HE, IDA, dan NNA menggunakan model RF, yang kemudian dilakukan evaluasi hasil menggunakan *Confusion Matrix*, sehingga diperoleh hasil evaluasi klasifikasi berupa akurasi, sensitivitas, dan spesifisitas. Penelitian ini memperoleh hasil terbaik pada k sejumlah 5 dengan parameter n estimators, max depth, dan min samples leaf berturut-turut sebanyak 10, 90, dan 4. Nilai akurasi, sensitivitas, dan spesifisitas yang dihasilkan masing-masing sebesar 100%.

Kata kunci: Anemia, *Iron Deficiency Anemia*, Klasifikasi, *Normocytic Normochromic Anemia*, *Random Forest*

ABSTRACT

CLASSIFICATION OF ANEMIA USING RANDOM FOREST METHOD

Anemia is a blood disorder caused by a lack of red blood cells and hemoglobin levels. Based on a survey conducted by the World Health Organization (WHO), anemia affects 1.62 billion people worldwide. As many as 93% of chronic disease patients experience anemia. Anemia can be diagnosed using Complete Blood Count (CBC) which aims to evaluate the total number and characteristics of cell components in the blood. The purpose of this study was to conduct a Random Forest (RF) classification of anemia. This study used numerical data in the form of blood cell characteristics, consisting of 269 Normocytic Normochromic Anemia patients, 189 Iron Deficiency Anemia patients and 336 healthy patients. Data processing carried out in this study consists of several stages, namely data input, continued with descriptive analysis and data encoding process. After that, data division using K-Fold Cross Validation with K of 5. The next stage is the classification process using RF with parameter testing on training data, namely the number of decision trees to be built, the maximum depth of each tree and the minimum number of samples on each leaf. Based on these parameters, the classification results were obtained in the form of HE, IDA, and NNA diagnoses using the RF model, which were then evaluated using the *Confusion Matrix*, so that the classification evaluation results were obtained in the form of accuracy, sensitivity, and specificity. This study obtained the best results at k of 5 with parameters n estimators, max depth, and min samples leaf of 10, 90, and 4, respectively. The accuracy, sensitivity, and specificity values were each 100%.

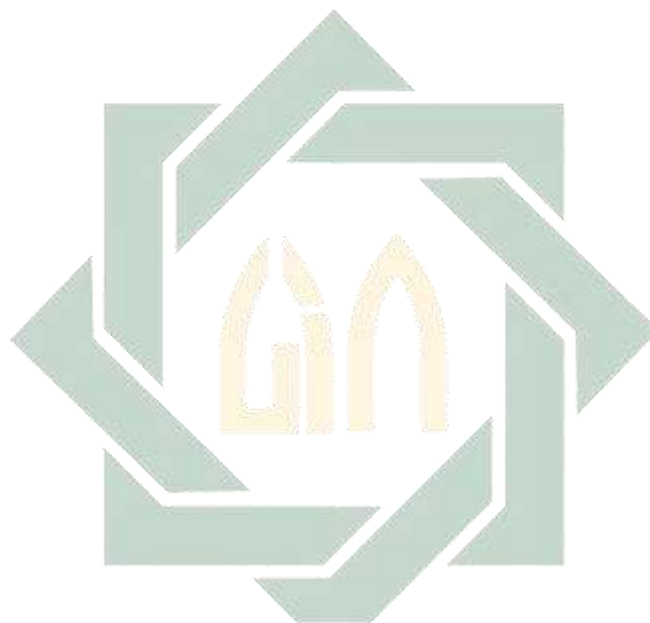
Keywords: Anemia, Iron Deficiency Anemia, Classification, Normocytic Normochromic Anemia, Random Forest

DAFTAR ISI

HALAMAN JUDUL	i
LEMBAR PERSETUJUAN PEMBIMBING	ii
PENGESAHAN TIM PENGUJI SKRIPSI	iii
HALAMAN PERNYATAAN KEASLIAN	iv
MOTTO	v
HALAMAN PERSEMBAHAN	vi
KATA PENGANTAR	vii
DAFTAR ISI	ix
DAFTAR TABEL	xii
DAFTAR GAMBAR	xiv
ABSTRAK	xv
ABSTRACT	xvi
I PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	7
1.3. Tujuan Penelitian	8
1.4. Manfaat Penelitian	8
1.5. Batasan Masalah	9
1.6. Sistematika Penulisan	9
II TINJAUAN PUSTAKA	11
2.1. Anemia	11
2.2. <i>Data Encoding</i>	14
2.3. <i>K-Fold Cross Validation</i>	15
2.4. <i>Random Forest</i>	16
2.4.1. Pengambilan sampel	17
2.4.2. Pembentukan DT	17
2.4.3. Pembentukan hutan	19

2.4.4. Proses klasifikasi	19
2.5. <i>Confusion Matrix</i>	20
2.6. Integrasi Keilmuan	23
III METODE PENELITIAN	26
3.1. Jenis Penelitian dan Sumber Data	26
3.2. Tahapan Penelitian	26
IV HASIL DAN PEMBAHASAN	31
4.1. Input Data	31
4.2. Analisis Deskriptif Data	33
4.3. <i>Data Encoding</i>	35
4.4. Pembagian Data	36
4.5. Proses Klasifikasi	37
4.5.1. Pengambilan sampel	38
4.5.2. Pembentukan DT	40
4.5.3. Pembentukan hutan	56
4.5.4. <i>Majority Voting</i>	63
4.6. Hasil Klasifikasi	64
4.7. Evaluasi Hasil	76
4.8. Analisis Hasil	77
4.9. Aplikasi Klasifikasi Anemia	80
4.10. Integrasi Keilmuan	82
V PENUTUP	84
5.1. Kesimpulan	84
5.2. Saran	85
DAFTAR PUSTAKA	85
A Lampiran A: DT ke-1	95
B Lampiran B: DT ke-2	96
C Lampiran C: DT ke-3	97
D Lampiran D: DT ke-4	98
E Lampiran E: DT ke-5	99

F	Lampiran F: DT ke-6	100
G	Lampiran G: DT ke-7	101
H	Lampiran H: DT ke-8	102
I	Lampiran I: DT ke-9	103
J	Lampiran J: DT ke-10	104



UIN SUNAN AMPEL
S U R A B A Y A

DAFTAR TABEL

2.1	Nilai batas normal karakteristik anemia	14
2.2	Tabel <i>Confusion Matrix</i>	21
3.1	Variasi Jumlah Parameter Uji Coba	29
4.1	Data Karakteristik Anemia	31
4.2	Keterangan variabel pada data anemia	32
4.3	Hasil Analisis Deskriptif Data	34
4.4	Hasil <i>Encoding</i> Anemia	35
4.5	Jumlah Data Latih dan Uji pada Setiap <i>Fold</i> dengan $k=5$	36
4.6	Data pelatihan $k=5$	37
4.7	<i>Bootstrap</i> Pohon Pertama	38
4.8	<i>Bootstrap</i> Pohon Kedua	39
4.9	<i>Bootstrap</i> Pohon Kesepuluh	39
4.10	<i>Threshold</i> Variabel HCT	41
4.11	Nilai Entropi dan IG pada Setiap <i>Threshold</i> Variabel HCT	43
4.12	Nilai Entropi dan IG pada Setiap Variabel di Simpul Akar Pohon Pertama	44
4.13	<i>Threshold</i> Variabel MCH dengan $HCT \leq 40.45$	46
4.14	Nilai entropi dan IG pada setiap <i>threshold</i> variabel MCH dengan $HCT \leq 40.45$	48
4.15	Nilai Entropi dan IG pada Setiap Variabel di Simpul Cabang Pertama Pohon Pertama	49
4.16	<i>Threshold</i> Variabel NAP dengan $HCT \leq 40.45$ dan $MCH > 25.9$	51
4.17	Nilai entropi dan IG pada setiap <i>threshold</i> variabel NAP dengan $HCT \leq 40.45$ dan $MCH > 25.9$	53
4.18	Nilai Entropi dan IG pada Setiap Variabel di Simpul Cabang Kedua Pohon Pertama	54
4.19	Data Pengujian $k=5$	63

4.20 Hasil <i>Majority Voting</i>	64
4.21 Hasil Uji Coba dengan $k=5$	65
4.22 Rata-rata hasil uji coba tiap <i>fold</i> dengan n estimators=10 dan <i>max depth</i> =30	67
4.23 Rata-rata hasil uji coba tiap <i>fold</i> dengan n estimators=10 dan <i>max depth</i> =60	68
4.24 Rata-rata hasil uji coba tiap <i>fold</i> dengan n estimators=10 dan <i>max depth</i> =90	69
4.25 Rata-rata hasil uji coba tiap <i>fold</i> dengan n estimators=60 dan <i>max depth</i> =30	70
4.26 Rata-rata hasil uji coba tiap <i>fold</i> dengan n estimators=60 dan <i>max depth</i> =60	71
4.27 Rata-rata hasil uji coba tiap <i>fold</i> dengan n estimators=60 dan <i>max depth</i> =90	72
4.28 Rata-rata hasil uji coba tiap <i>fold</i> dengan n estimators=110 dan <i>max depth</i> =30	73
4.29 Rata-rata hasil uji coba tiap <i>fold</i> dengan n estimators=110 dan <i>max depth</i> =60	74
4.30 Rata-rata hasil uji coba tiap <i>fold</i> dengan n estimators=110 dan <i>max depth</i> =90	75
4.31 Tabel Hasil <i>Confusion Matrix</i>	76
4.32 Perbandingan Hasil Penelitian terkait Klasifikasi Anemia	79

UIN SUNAN AMPEL
S U R A B A Y A

DAFTAR GAMBAR

1	Visualisasi <i>K-Fold Cross Validation</i>	16
2	Arsitektur metode <i>Decision Tree</i>	17
3	Arsitektur metode <i>Random Forest</i>	20
4	Diagram Alir Penelitian	27
5	Simpul Akar DT ke-1	45
6	Simpul Cabang Pertama DT ke-1	50
7	Simpul Cabang Kedua DT ke-1	55
8	DT ke-1	56
9	DT ke-2	57
10	DT ke-3	58
11	DT ke-4	58
12	DT ke-5	59
13	DT ke-6	60
14	DT ke-7	60
15	DT ke-8	61
16	DT ke-9	62
17	DT ke-10	62
18	Hasil Perbandingan Akurasi, Sensitivitas, dan Spesifisitas Berdasarkan <i>N Estimators</i> 10 dengan <i>Max Depth</i> 30, 60, dan 90	78
19	Tampilan Awal pada Aplikasi	80
20	Tampilan Kedua pada Aplikasi Sebelum Input Data	81
21	Tampilan Kedua pada Aplikasi Saat Input Data	81
22	Tampilan Kedua pada Aplikasi dengan Hasil Klasifikasi	82

DAFTAR PUSTAKA

- Abdi, J., Hadavimoghaddam, F., Hadipoor, M., and Hemmati-Sarapardeh, A. (2021). Modeling Of CO₂ Adsorption Capacity by Porous Metal Organic Frameworks using Advanced Decision Tree-Based Models. *Scientific Reports*, 11(1):1–14.
- Aboelnaga, E. (2024). Anemia Types Classification.
- Agnello, L., Giglio, R. V., Bivona, G., Scazzone, C., Gambino, C. M., Iacona, A., Ciaccio, A. M., Sasso, B. L., and Ciaccio, M. (2021). The Value of a Complete Blood Count (CBC) for Sepsis Diagnosis and Prognosis. *Diagnostics*, 11(10):1–19.
- Al-akhras, M., Hindi, K. E., Habib, M., and Abu, B. (2021). Instance Reduction for Avoiding Overfitting in Decision Trees. *De Gruyter*, 30(1):438–459.
- Ali, K., Shaikh, Z. A., Khan, A. A., and Laghari, A. A. (2022). Multiclass Skin Cancer Classification using Efficientnets – A First Step Towards Preventing Skin Cancer. *Neuroscience Informatics*, 2(4):1–10.
- Alshwaiyat, N. M., Ahmad, A., and Al-Jamal, H. A. N. (2023). Effect of Diet-Induced Weight Loss on Iron Status and Its Markers among Young Women with Overweight/Obesity and Iron Deficiency Anemia: A Randomized Controlled Trial. *Frontiers in Nutrition*, 10(May):1–14.
- Al-shehari, T. and Alsowail, R. A. (2021). An Insider Data Leakage Detection

- Using One-Hot Encoding, Synthetic Minority Oversampling and Machine Learning Techniques. *Entropy*, 23(10):1–24.
- Asian, J., Dholah Rosita, M., and Mantoro, T. (2022). Sentiment Analysis for the Brazilian Anesthesiologist Using Multi-Layer Perceptron Classifier and Random Forest Methods. *Jurnal Online Informatika*, 7(1):132.
- Azad, C., Bhushan, B., Sharma, R., Shankar, A., Singh, K. K., and Khamparia, A. (2022). Prediction Model using SMOTE, Genetic Algorithm and Decision Tree (PMSGD) for Classification of Diabetes Mellitus. *Multimedia Systems*, 28(4):1289–1307.
- Azhari, M., Situmorang, Z., and Rosnelly, R. (2021). Perbandingan Akurasi, Recall, dan Presisi Klasifikasi pada Algoritma C4.5, Random Forest, SVM dan Naive Bayes. *Jurnal Media Informatika Budidarma*, 5(2):640.
- Berg, M., Hartley, B., Ratio, E., Repayment, P., Liu, Y., and Steffen, B. (2021). Loan Default Prediction using Decision Trees and Random Forest : A Comparative Study.
- Chicco, D. and Jurman, G. (2020). The Advantages of the Matthews Correlation Coefficient (MCC) Over F1 Score and Accuracy in Binary Classification Evaluation. *BMC Genomics*, 21(1):1–13.
- Chicco, D. and Jurman, G. (2023). The Matthews Correlation Coefficient (MCC) Should Replace The ROC AUC as The Standard Metric for Assessing Binary Classification. *BioData Mining*, 16(1):1–23.
- Dahouda, M. K. and Joe, I. (2021). A Deep-Learned Embedding Technique for Categorical Features Encoding. *IEEE Access*, 9:114381–114391.

- Dewi, E. S. and Imah, E. M. (2020). Comparison of Machine Learning Algorithms for Autism Spectrum Disorder Classification. *International Joint Conference on Science and Engineering (IJCSE 2020) Comparison*, 196(Ijcse):152–159.
- El-Askary, N. S., Salem, M. A., and Roushdy, M. I. (2019). Feature Extraction and Analysis for Lung Nodule Classification using Random Forest. *ACM International Conference Proceeding Series*, pages 248–252.
- Ghosh, R., Ghosh, C., Hazra, A., Das, D. D., and Sarkar, D. T. (2023). Anaemia and Its Homoeopathic Management: Review Article. *International Journal of Homoeopathic Sciences*, 7(3):279–286.
- Görtler, J., Hohman, F., Moritz, D., Wongsuphasawat, K., Ren, D., Nair, R., Kirchner, M., and Patel, K. (2022). Neo: Generalizing Confusion Matrix Visualization to Hierarchical and Multi-Output Labels. *Conference on Human Factors in Computing Systems - Proceedings*.
- Greener, J. G., Kandathil, S. M., Moffat, L., and Jones, D. T. (2022). A Guide to Machine Learning for Biologists. *Nature Reviews Molecular Cell Biology*, 23:40–55.
- Hoenemann, C., Ostendorf, N., Zarbock, A., Doll, D., Hagemann, O., Zimmermann, M., and Luedi, M. (2021). Reticulocyte and Erythrocyte Hemoglobin Parameters for Iron Deficiency and Anemia Diagnostics in Patient Blood Management. A Narrative Review. *Journal of Clinical Medicine*, 10(18).
- Holladay, J. T., Wilcox, R. R., Koch, D. D., and Wang, L. (2021). Review and Recommendations for Univariate Statistical Analysis of Spherical Equivalent Prediction Error for IOL Power Calculations. *Journal of Cataract and Refractive Surgery*, 47(1):65–77.

- Jalolov, T. S. (2024). Use of Spss Software in Psychological Data Analysis. *PSIXOLOGIYA VA SOTSIOLOGIYA ILMYIY JURNALI*, 8:1–6.
- Jatmiko, Y. A., Padmadisastra, S., and Chadidjah, A. (2019). Analisis Perbandingan Kinerja Cart Konvensional, Bagging dan Random Forest pada Klasifikasi Objek: Hasil dari Dua Simulasi. *Media Statistika*, 12(1):1.
- Jeong, B., Cho, H., Kim, J., Kwon, S. K., Hong, S. W., Lee, C. S., Kim, T. Y., Park, M. S., Hong, S., and Heo, T. Y. (2020). Comparison Between Statistical Models and Machine Learning Methods on Classification for Highly Imbalanced Multiclass Kidney Data. *Diagnostics*, 10(6):1–17.
- Karagül Yıldız, T., Yurtay, N., and Öne, B. (2021). Classifying Anemia Types using Artificial Learning Methods. *Engineering Science and Technology, an International Journal*, 24(1):50–70.
- K.T, N., Prasad, K., and Singh, B. M. K. (2022). Analysis of Red Blood Cells from Peripheral Blood Smear Images for Anemia Detection: A Methodological Review. *Medical and Biological Engineering and Computing*, 60(9):2445–2462.
- Luque, A., Carrasco, A., Martín, A., and de las Heras, A. (2019). The Impact of Class Imbalance in Classification Performance Metrics Based on the Binary Confusion Matrix. *Pattern Recognition*, 91:216–231.
- Marcot, B. G. and Hanea, A. M. (2021). What is an Optimal Value of K in K-Fold Cross-Validation in Discrete Bayesian Network Analysis? *Computational Statistics*, 36(3):2009–2031.
- Marik, P. E. and Flemmer, M. (2021). Iron Deficiency Anemia in Inflammatory Bowel Diseases—A Narrative Review. *Nutrients*, 13(6):1–14.

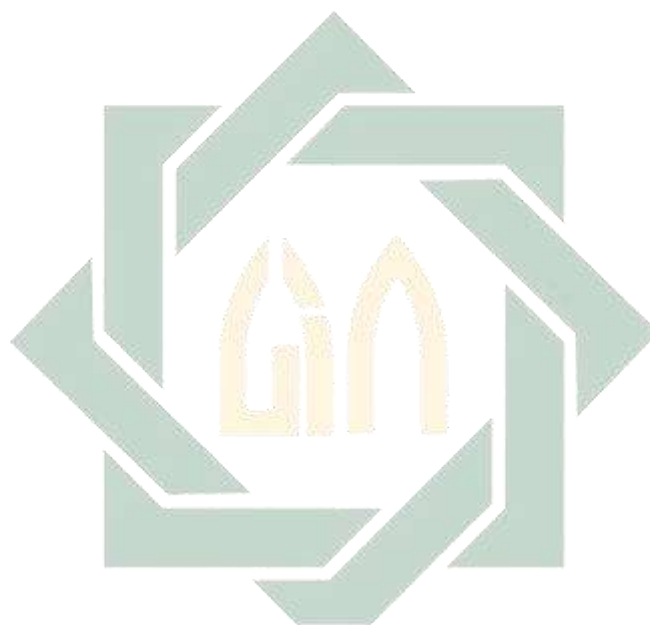
- Mishra, S., Mallick, P. K., Jena, L., and Chae, G. S. (2020). Optimization of Skewed Data using Sampling-Based Preprocessing Approach. *Frontiers in Public Health*, 8(July):1–7.
- Modrow, J. C., Luz, N. A., and Martins, V. R. (2023). Anemia in Patients Admitted to A Hospital in Teixeira Soares-PR. *A Look At Development*, (2004).
- Mohammed, M. R. and Mahmood, B. (2022). Morphological Types of Anemia Associated with Chronic Renal Diseases. *Open Access Macedonian Journal of Medical Sciences*, 10(B):905–908.
- Noor, N. B., Anwar, M. S., and Dey, M. (2019). Comparative Study between Decision Tree, SVM and KNN to Predict Anaemic Condition. *BECITHCON 2019 - 2019 IEEE International Conference on Biomedical Engineering, Computer and Information Technology for Health*, pages 24–28.
- Normawati, D. and Ismi, D. P. (2019). K-Fold Cross Validation for Selection of Cardiovascular Disease Diagnosis Features by Applying Rule-Based Datamining. *Signal and Image Processing Letters*, 1(2):23–35.
- Pal, K. and Patel, B. V. (2020). Data Classification with K-Fold Cross Validation and Holdout Accuracy Estimation Methods with 5 Different Machine Learning Techniques. *Proceedings of the 4th International Conference on Computing Methodologies and Communication, ICCMC 2020*, (Iccmc):83–87.
- Pandey, V. (2021). Study of Spectrum of Anemia in Various Age Groups. *IP Archives of Cytology and Histopathology Research*, 6(3):197–202.
- Parmar, A., Katariya, R., and Patel, V. (2019). A Review on Random Forest: An

- Ensemble Classifier. *Lecture Notes on Data Engineering and Communications Technologies*, 26:758–763.
- Pavliukovych, N. D., Pavliukovych, O., and Shuper, V. O. (2022). Structure of Anemic Syndrome in Patients with Chronic Forms of Coronary Artery Disease. *Jurnal Kedokteran, Biologi dan Olahraga Ukraina*, 7(40):95–99.
- Quist, J., Taylor, L., Staaf, J., and Grigoriadis, A. (2021). Random Forest Modelling of High-Dimensional Mixed-Type Data for Breast Cancer Classification. *Cancers*, 13(5):1–15.
- Rahardja, U. (2022). Application of the C4.5 Algorithm for Identifying Regional Zone Status using a Decision Tree in the Covid-19 Series. *Aptisi Transactions on Technopreneurship (ATT)*, 4(2):164–173.
- Ramadhan, B., Firdaus, D., and Rafi, A. R. (2023). Teknik SMOTE sebagai Solusi Imbalance Class dalam Model Deteksi Intrusi DDoS dengan Metode PCA-Random Forest. *MIND Journal*, 8(1):52–64.
- Sandag, G. A. (2020). Prediksi Rating Aplikasi App Store Menggunakan Algoritma Random Forest. *CogITo Smart Journal*, 6(2):167–178.
- Sasikala, N., Banu, G. R., Babiker, T., and Rajpoot, P. (2021). A Role of Data Mining Techniques to Predict Anemia Disease. *International Journal of Computer Applications*, 174(20):16–20.
- Satniyazovna, A. S. (2022). Anemia is A Chronic Disease. *The American Journal of Medical Sciences and Pharmaceutical Research*, 04(09):12–19.
- Seo, I. H. and Lee, Y. J. (2022). Usefulness of Complete Blood Count (CBC)

- to Assess Cardiovascular and Metabolic Diseases in Clinical Settings: A Comprehensive Literature Review. *Biomedicines*, 10(11).
- Shaik, A. B. and Srinivasan, S. (2019). *A Brief Survey on Random Forest Ensembles in Classification Model*, volume 56. Springer Singapore.
- Song, M., Jung, H., Lee, S., Kim, D., and Ahn, M. (2021). Diagnostic Classification and Biomarker Identification of Alzheimer's Disease with Random Forest Algorithm. *Brain Sciences*, 11(4):1–14.
- Sujatha, R., Aarthy, S. L., Chatterjee, J. M., Alaboudi, A., and Jhanjhi, N. Z. (2021). A Machine Learning Way to Classify Autism Spectrum Disorder. *International Journal of Emerging Technologies in Learning*, 16(6):182–200.
- 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.
- Talarico, V., Giancotti, L., Mazza, G. A., Miniario, R., and Bertini, M. (2021). Iron Deficiency Anemia in Celiac Disease. *Nutrients*, 13(5):1–11.
- Tembusai, Z. R., Mawengkang, H., and Zarlis, M. (2021). K-Nearest Neighbor with K-Fold Cross Validation and Analytic Hierarchy Process on Data Classification. *International Journal of Advances in Data and Information Systems*, 2(1):1–8.
- Thotakura, M., Manikanta, V., and Bharathi, Y. K. (2021). Prevalence of Morphological Patterns of Anemia in Geriatric Population. *International Journal of Clinical and Diagnostic Pathology*, 4(3):178–180.

- Torrez, M., Chabot-Richards, D., Babu, D., Lockhart, E., and Foucar, K. (2022). How I Investigate Acquired Megaloblastic Anemia. *International Journal of Laboratory Hematology*, 44(2):236–247.
- Vohra, R., Hussain, A., Dudyala, A. K., Pahareeya, J., and Khan, W. (2022). Multi-Class Classification Algorithms for the Diagnosis of Anemia in an Outpatient Clinical Setting. *PLoS ONE*, 17(7 July):1–18.
- Wayahdi, M. R., Syahputra, D., Hafiz, S., and Ginting, N. (2020). Evaluation of the K-Nearest Neighbor Model with K-Fold Cross Validation on Image Classification. *Jurnal Infokum*, 9(1):1–6.
- Westphal, E. and Seitz, H. (2021). A Machine Learning Method for Defect Detection and Visualization in Selective Laser Sintering Based on Convolutional Neural Networks. *Additive Manufacturing*, 41(March).
- Yang, H., Luo, Y., Ren, X., Wu, M., He, X., Peng, B., Deng, K., Yan, D., Tang, H., and Lin, H. (2021). Risk Prediction of Diabetes: Big Data Mining with Fusion of Multifarious Physical Examination Indicators. *Information Fusion*, 75(March):140–149.
- Yang, J., Li, Q., Feng, Y., and Zeng, Y. (2023). Iron Deficiency and Iron Deficiency Anemia: Potential Risk Factors in Bone Loss. *International Journal of Molecular Sciences*, 24(8):1–20.
- Yilmaz, A. E. and Demirhan, H. (2023). Weighted Kappa Measures for Ordinal Multi-Class Classification Performance. *Applied Soft Computing*, 134:110020.
- Zhou, X., Lu, P., Zheng, Z., Tolliver, D., and 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*, page 106931.



UIN SUNAN AMPEL
S U R A B A Y A