

KLASIFIKASI DIAGNOSA PENYAKIT TIROID MENGGUNAKAN METODE RANDOM FOREST

Raja Darmawan¹, Yulison Herry Chrisnanto², Gunawan Abdillah³

^{1,2,3}, Informatika, Fakultas Sains dan Informatika, Universitas Jenderal Achmad Yani

Jln. Terusan Jend, Sudirman, Kota Cimahi, Jawa Barat 40525

¹ rajadarmawan20@if.unjani.ac.id, ² yhc@if.unjani.ac.id*, ³ gunawanabdillah03@gmail.com

Abstract

The thyroid is an important gland in the human neck that regulates metabolism through its hormones. Hormonal disorders in the thyroid can have a significant impact on health. The symptoms are often difficult to identify as they resemble lifestyle complaints. To overcome this problem, machine learning is needed to detect thyroid disease diagnosis based on symptoms. This research aims to develop machine learning techniques that can help classify thyroid disease diagnoses using the random forest method. Tests were conducted 6 times by changing parameter values to determine their effect on model accuracy. The results showed that parameters such as the number of decision trees, maximum depth, and minimum number of leaf samples can affect the accuracy of the model. The test results show that the highest accuracy is obtained in the third test with a data division of 80/20 with an accuracy of 99%. This study concludes that the random forest method is effective in improving the accuracy of thyroid disease diagnosis and emphasizes the importance of parameter adjustment for optimal results.

Keywords : *Thyroid Disease, Classification, Random Forest*

Abstrak

Tiroid adalah kelenjar penting di leher manusia yang mengatur metabolisme melalui hormon-hormonnya. Gangguan hormon pada tiroid dapat berdampak signifikan pada kesehatan. Gejalanya sering kali sulit diidentifikasi karena menyerupai keluhan yang diakibatkan gaya hidup. Untuk mengatasi masalah ini, diperlukan pembelajaran mesin untuk mendeteksi diagnosis penyakit tiroid berdasarkan gejala. Penelitian ini bertujuan untuk mengembangkan teknik machine learning yang dapat membantu klasifikasi diagnosa penyakit tiroid menggunakan metode random forest. Pengujian dilakukan sebanyak 6 kali dengan mengubah nilai parameter untuk mengetahui pengaruhnya terhadap akurasi model. Hasil penelitian menunjukkan bahwa parameter seperti jumlah pohon keputusan, kedalaman maksimum, dan jumlah minimum sampel daun dapat mempengaruhi akurasi model. Hasil pengujian menunjukkan akurasi tertinggi diperoleh pada pengujian ketiga dengan pembagian data 80/20 didapatkan hasil akurasi 99%. Penelitian ini menyimpulkan bahwa metode random forest efektif dalam meningkatkan akurasi diagnosa penyakit tiroid serta menekankan pentingnya penyesuaian parameter untuk hasil yang optimal.

Kata kunci : *Penyakit Tiroid, Klasifikasi, Random Forest*

1. PENDAHULUAN

Bagian tubuh manusia yang paling penting adalah tiroid, yang terletak di bawah jakun pada leher. Kelenjar ini, yang memiliki bentuk seperti kupu-kupu, menghasilkan hormon tiroid dan bertanggung jawab untuk menjaga metabolisme [1]. Sangat penting untuk mengamati dan

mengukur berbagai gangguan hormonal pada manusia. Faktor penyebab tiroid adalah fleksibilitas hormon manusia yang terlalu tinggi atau rendah. Perawatan kesehatan penting untuk menyeimbangkan variasi hormon yang terlalu tinggi atau rendah. Faktor risiko untuk gangguan hormon cukup banyak, sehingga dokter membutuhkan perhatian yang lebih besar untuk

memantau dan menemukan diagnosis yang tepat pada waktu yang tepat [2]. Karena gejala gangguan fungsi tiroid sangat mirip dengan keluhan yang disebabkan oleh gaya hidup modern, gangguan fungsi tiroid sering kali sulit diidentifikasi. Akibatnya, pasien sering mengabaikan masalahnya dan tidak memeriksanya ke dokter [3].

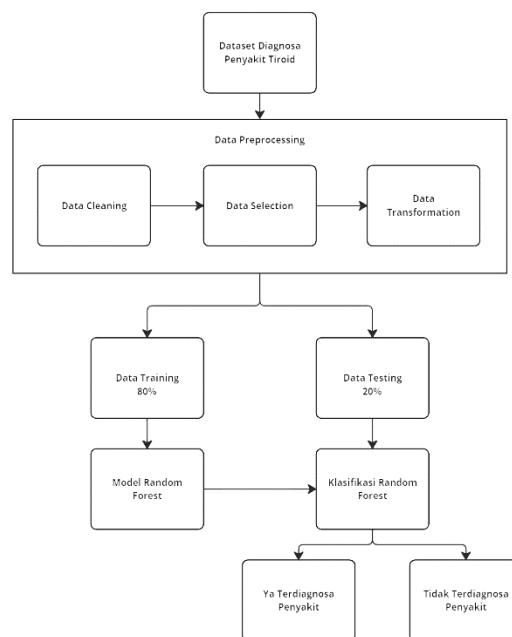
Oleh karena itu, Teknik data mining diperlukan untuk menganalisa data penyakit ini [4]. Dalam teknik data mining, terdapat berbagai algoritma klasifikasi seperti *Naive Bayes*, *Decision Tree*, *Neural Network*, *K-NN*, *Random Forest*, dan lainnya yang digunakan untuk menggambarkan serta membedakan kelas data atau konsep.[5]. Teknik data mining yang digunakan adalah algoritma klasifikasi random forest. Random forest adalah konsep pembelajaran kelompok yang menggunakan hasil dari beberapa pohon keputusan untuk meningkatkan akurasi. Metode ini bekerja dengan membangun pohon keputusan secara acak dari kumpulan data. Teknik ini dapat menghasilkan tingkat akurasi yang tinggi dan menghindari masalah overfitting, tetapi juga memerlukan waktu pelatihan yang lebih lama [6].

Pada penelitian terdahulu membahas banyaknya atribut yang memiliki *missing value* sangat mempengaruhi pembuatan model. Untuk menangani dataset yang mengandung data dengan *missing value*, diterapkan algoritma K-Nearest Neighbors Imputation (KNNImputer). Metode ini menghasilkan akurasi sebesar 93% [7]. Penelitian tambahan tentang pengujian dengan menggunakan dataset penyakit tiroid sebanyak 500 data, algoritma *Naive Bayes* memiliki akurasi sebesar 63% pada data *testing* dan 64% pada data *training*, sedangkan algoritma *K-Nearest Neighbors* (KNN) memiliki akurasi sebesar 63% [8].

Penelitian sebelumnya membahas pengujian parameter-parameter yang berbeda dalam metode Random Forest untuk menemukan konfigurasi parameter yang optimal dalam klasifikasi penyakit jantung. Hasil akurasi terbaik yang dicapai adalah 92% dengan menggunakan data training dan testing yang dibagi dengan rasio 70/30 [9]. Penelitian terdahulu lainnya menunjukkan pengaturan parameter dilakukan dengan fokus pada penerapan algoritma Random Forest yang berhasil meraih tingkat akurasi sebesar 82,6087% [10]. Penelitian selanjutnya tentang pengaturan parameter dilakukan menggunakan algoritma Grid Search untuk mencari hyper-parameter terbaik pada algoritma Random Forest [11].

Berdasarkan penelitian terdahulu yang telah disebutkan, penelitian ini bertujuan untuk mengetahui bagaimana penerapan metode random forest dalam proses klasifikasi diagnosa penyakit tiroid dapat menghasilkan performa terbaik serta perubahan nilai pada parameter dapat berpengaruh pada model random forest.

2. METODOLOGI PENELITIAN



Gambar 1. Alur Penelitian

2.1. Pengumpulan Data

Pada penelitian ini data yang digunakan didapatkan dari Kaggle. Data ini menangkap berbagai atribut terkait kondisi tiroid untuk diagnosa medis. Ini mencakup informasi demografis seperti usia dan jenis kelamin. Fitur riwayat kesehatan mencakup asupan tiroksin, obat antitiroid, dan operasi sebelumnya. Status kesehatan pasien terkini mengenai penyakit, kehamilan, adanya penyakit gondok, tumor, atau kondisi hipofisis dicatat. Selain itu, ia mencatat apakah pasien mencurigai *hipotiroidisme* atau *hipertiroidisme*. Hasil laboratorium seperti kadar TSH, T3, TT4, T4U, FTI disertakan jika diukur. Klasifikasi biner menunjukkan ada tidaknya *hipertiroidisme*. Sumber rujukan juga ditunjukkan Dataset berisi atribut-atribut.

2.2. Preprocessing

Pada tahap ini dilakukan untuk memeriksa dan memperbaiki kesalahan pada dataset. Teknik yang digunakan pada penelitian ini yaitu cleaning, selection, dan transformation. *Cleaning*

diperlukan untuk mengatasi data yang hilang menggunakan hasil rata-rata data. Lalu proses *selection* dilakukan untuk menunjukkan kontribusi relative setiap fitur dalam membuat keputusan. Semakin tinggi nilai penting suatu fitur, semakin besar pengaruh fitur tersebut dalam model. Selain itu, proses *transformation* digunakan untuk merubah data dengan tipe *categorical* menjadi *numerik* yaitu 0 dan 1.

2.3. Split Data

Langkah berikutnya adalah melakukan pemisahan data, di mana data akan dibagi menjadi dua bagian, yaitu data uji dan data latih. Pembagian data pada penelitian ini menggunakan 80/20. 80% data *training* dan 20% data *testing*.

2.4. Klasifikasi Random Forest

Pada tahap ini dilakukan pembuatan model random forest untuk klasifikasi diagnose penyakit tiroid. Random forest ini memiliki beberapa parameter yang diujikan untuk mengoptimalkan model. Beberapa parameter tersebut adalah *n_estimator* (jumlah pohon keputusan), *max_depth* (kedalaman maksimum pohon keputusan), *min_sample_leaf* (jumlah minimum sampel daun), dan *random state* (nilai inisialisasi).

2.5. Evaluasi

Setelah itu dilakukan proses evaluasi terhadap performa model random forest. Evaluasi ini melibatkan penggunaan *matriks* seperti confusion matrix untuk mengukur metrik-metrik penting seperti akurasi, presisi, dan recall. Evaluasi ini membantu dalam memahami seberapa baik model tersebut dalam mengklasifikasikan data.

3. HASIL DAN PEMBAHASAN

3.1. Pengumpulan Data

TABEL I. DATASET

No	Atribut	Keterangan
1	Age	Usia pasien
2	Sex	Jenis kelamin pasien
3	Thyroxine	Apakah pasien dalam pengobatan tiroksin

4	Queryonthyroxine	Apakah pasien dalam pengobatan tiroksin
5	onantithyroidmedication	Apakah pasien dalam pengobatan antitiroid
6	Sick	Apakah pasien sedang sakit
7	Pregnant	Apakah pasien sedang hamil
8	thyroidsugery	Apakah pasien telah menjalani operasi tiroid
9	I131treatment	Apakah pasien sedang menjalani pengobatan I131
10	Queryhypothyroid	Apakah pasien yakin menderita hipotiroid
11	Queryhyperthyroid	Apakah pasien yakin menderita hipertiroid
12	Lithium	Apakah pasien menggunakan lithium
13	Goitre	Apakah pasien memiliki gondok
14	Tumor	Apakah pasien memiliki tumor
15	Hypopituitary	Apakah pasien memiliki kelenjar hipofisis
16	Psych	Apakah pasien memiliki masalah kejiwaan

17	TSHmeasured	Apakah TSH diukur dalam darah
18	TSH	Kadar TSH dalam darah
19	T3measured	Apakah T3 diukur dalam darah
20	T3	Kadar T3 dalam darah
21	TT4measured	Apakah TT4 diukur dalam darah
22	T4	Kadar T4 dalam darah
23	T4Umeasured	Apakah T4U diukur dalam darah
24	T4U	Kadar T4U dalam darah
25	FTImeasured	Apakah FTI diukur dalam darah
26	FTI	Kadar FTI dalam darah
27	TBGmeasured	Apakah TBG diukur dalam darah
28	TBG	Kadar TBG dalam darah
29	Referral source	Sumber rujukan
30	binaryClass	Kelas biner

3.2. Preprocessing

Pada penelitian ini, sebelum data masuk ke tahap pemodelan, dilakukan terlebih dahulu proses pra-pemrosesan data. Pada tahapan *cleaning* dilakukan proses pengisian melengkapi data yang hilang. Berikut hasil data sebelum dan sesudah di *cleaning* pada gambar 2 dan gambar 3.

Age	0
sex	150
thyroxine	0
queryonthyroxine	0
onantithyroidmedication	0
sick	0
pregnant	0
thyroidsurgery	0
I131treatment	0
queryhypothyroid	0
queryhyperthyroid	0
lithium	0
goitre	0
tumor	0
hypopituitary	0
psych	0
TSHmeasured	0
TSH	0
T3measured	0
T3	0
TT4measured	0
T4	0
T4Umeasured	0
T4U	0
FTImeasured	0
...	0
TBGmeasured	0
TBG	0
referral source	0
binaryClass	0

Gambar 2. Sebelum *Cleaning*

Age	0
sex	0
thyroxine	0
queryonthyroxine	0
onantithyroidmedication	0
sick	0
pregnant	0
thyroidsurgery	0
I131treatment	0
queryhypothyroid	0
queryhyperthyroid	0
lithium	0
goitre	0
tumor	0
hypopituitary	0
psych	0
TSHmeasured	0
TSH	0
T3measured	0
T3	0
TT4measured	0
T4	0
T4Umeasured	0
T4U	0
FTImeasured	0
...	0
TBGmeasured	0
TBG	0
referral source	0
binaryClass	0

Gambar 3. Sesudah *Cleaning*

Pada data sebelum dilakukan *cleaning* data, terdapat 150 *missing value* pada atribut *sex*. Kemudian, *cleaning* data dilakukan dengan mengisi nilai yang hilang menggunakan rata-rata pada atribut *sex* tersebut.

Selanjutnya, pada tahapan *selection* data dilakukan untuk menunjukkan kontribusi relatif setiap fitur dalam pembuatan keputusan. Semakin tinggi nilai kepentingan suatu fitur, semakin besar pengaruh fitur tersebut dalam model. Dengan memahami nilai kepentingan masing-masing fitur

3.3. Split Data

Setelah melalui tahapan-tahapan tersebut, dataset kemudian dibagi menjadi data *training* dan data *testing*. Dalam penelitian ini, pembagian dataset dilakukan dengan dua cara. Pembagian pertama adalah 80/20, di mana 80% digunakan untuk data *training* dan 20% untuk data *testing*. Pembagian kedua adalah 70/30, dengan 70% untuk data pelatihan dan 30% untuk data pengujian. Kedua pembagian ini bertujuan untuk menguji kinerja model dalam melakukan klasifikasi. Berikut adalah pembagian data yang dilakukan dalam tabel 2.

TABEL II. SPLIT DATA

Data Training	Data Testing
70%	30%
80%	20%

3.4. Klasifikasi Random Forest

Dalam penelitian ini, metode random forest digunakan untuk melakukan klasifikasi penyakit tiroid. Pada random forest, terdapat beberapa parameter yang diuji untuk meningkatkan kinerja dan mengoptimalkan model. Berikut adalah pengujian dengan pengaturan parameter random forest yang telah dilakukan untuk klasifikasi penyakit tiroid pada tabel 3 dan tabel 4.

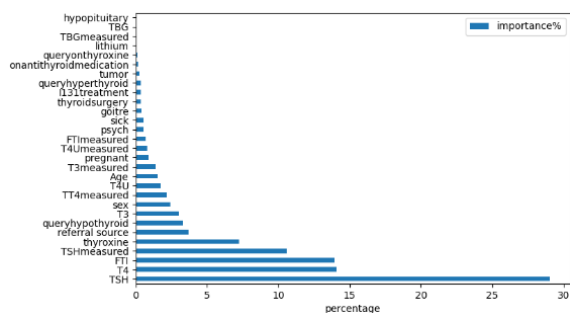
TABEL III. PENGUJIAN PERTAMA

Split Data 80/20			
Pengujian	1	2	3
n_estimator	50	100	150
max_depth	5	10	15
min_samples_leaf	5	10	15
random_state	42	42	42
Accuracy	95%	98%	99%

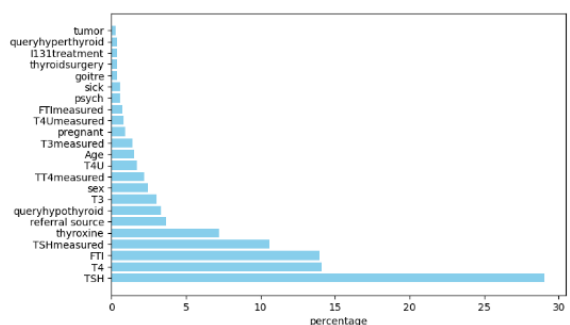
TABEL IV. PENGUJIAN KEDUA

Split Data 70/30			
Pengujian	4	5	6
n_estimator	50	100	150
max_depth	5	10	15
min_samples_leaf	5	10	15
random_state	42	42	42
Accuracy	97%	98%	98%

Pada pengujian pertama dan kedua, diterapkan berbagai parameter yaitu ($n_estimators$) jumlah pohon keputusan diuji dengan nilai (50, 100, 150), (max_depth)



Gambar 4. Sebelum *Selection*



Gambar 5. Sesudah *Selection*

Pada tahapan transformation data, dilakukan perubahan data kategorikal menjadi *numerik* dengan mengonversi nilai "*false*" dan "*true*" menjadi "0" dan "1". Hasil data yang sudah di *transformation* dapat dilihat dalam Gambar 6.

Age	int64
sex	int32
thyroxine	int32
querythyroxine	int32
onantithyroidmedication	int32
sick	int32
pregnant	int32
thyroidsurgery	int32
I131treatment	int32
queryhypothyroid	int32
queryhyperthyroid	int32
lithium	int32
goitre	int32
tumor	int32
hypopituitary	int32
psych	int32
TSHmeasured	int32
TSH	float64
T3measured	int32
T3	float64
TT4measured	int32
T4	float64
T4Umeasured	int32
T4U	float64
FTImeasured	int32
...	
TBG	int32
referral source	int32
binaryClass	int32

Gambar 6. Transformation

kedalaman maksimum pohon keputusan diuji dengan nilai (5, 10, 15), (*min_samples_leaf*) jumlah minimum sampel diuji dengan nilai (5, 10, 15), serta *random_state* dengan nilai 42. Setiap kombinasi parameter ini diuji untuk mengevaluasi pengaruhnya terhadap akurasi model. Hasil pengujian menunjukkan bahwa variasi parameter menghasilkan tingkat akurasi yang beragam. Hal ini menggaris bawahi pentingnya pemilihan parameter yang tepat untuk meningkatkan kinerja model. Dengan mengoptimalkan nilai *n_estimators*, *max_depth*, dan *min_samples_leaf*, model dapat mencapai tingkat akurasi yang lebih tinggi, menunjukkan betapa signifikan dampak tuning parameter terhadap hasil akhir.

3.5. Evaluasi

Pada penelitian ini, dilakukan 6 pengujian menggunakan metode klasifikasi random forest. Pengujian ke-3 dengan pembagian data 80/20 menghasilkan akurasi terbaik. Berikut adalah *classification report* untuk pengujian ke-3 yang ditampilkan dalam gambar 7.

	precision	recall	f1-score	support
0	0.95	0.94	0.95	65
1	0.99	1.00	0.99	689
accuracy			0.99	754
macro avg	0.97	0.97	0.97	754
weighted avg	0.99	0.99	0.99	754

Gambar 7. *Classification Report* Pengujian Ke-3

Pengujian akurasi tersebut didapatkan hasil 99%. Hasil perhitungan manual *confusion matrix* dapat dilihat pada tabel 6.

$$Accuracy = \frac{686+61}{686+61+4+3} = \frac{747}{754} = 0.9907(7)$$

$$Precision = \frac{686}{686+4} = \frac{686}{690} = 0.9942 \quad (8)$$

$$Recall = \frac{686}{686+3} = \frac{686}{690} = 0.9957 \quad (9)$$

$$f1 \text{ score} = 2 \times \frac{0.9942+0.9957}{0.9942+0.9957}$$

$$= 2 \times \frac{0.9899}{1.9899} = 0.9950 \quad (10)$$

TABEL V. HASIL PENGUJIAN

	Nilai
Akurasi	99,07%
Precision	99,42%
Recall	99,57%
f1 score	99,50%

4. Kesimpulan dan Saran

Berdasarkan penelitian yang telah dilakukan, dapat disimpulkan bahwa perubahan nilai pada parameter (*max_depth* dan *min_samples_leaf*) dapat mempengaruhi metode random forest dalam mengklasifikasi diagnosa penyakit tiroid dapat dilihat pada pengujian yang dilakukan. Penelitian ini melibatkan 6 pengujian, dengan hasil terbaik diperoleh pada pengujian ke-3 dalam pembagian data 80/20 yang menghasilkan akurasi 99,07%, *precision* 99,42%, *recall* 99,57%, dan *f1 score* 99,50%. Pembagian data 80/20 ini lebih tinggi dibandingkan dengan pembagian data 70/30 pada pengujian ke-6 yang mencapai akurasi terbaik 98%.

Saran untuk peneliti selanjutnya, agar menggunakan teknik SMOTE (Synthetic Minority Over-sampling Technique) untuk menyeimbangkan data. Teknik SMOTE ini penting karena membantu mengatasi masalah ketidakseimbangan kelas dengan menghasilkan sampel sintetis dari kelas minoritas. Dengan penerapan SMOTE, model pembelajaran mesin dapat mengenali pola dari kedua kelas dengan lebih efektif, sehingga meningkatkan akurasi dalam diagnosa penyakit tiroid.

Daftar Pustaka:

- [1] D. Sartika and Y. Yupianti, "Klasifikasi Penyakit Tiroid Menggunakan Algoritma C4.5 (Studi Kasus: Rumah Sakit Umum Daerah (RSUD) Hasanuddin Damrah Manna)," *Rekayasa*, vol. 13, no. 1, pp. 71–76, 2020, doi: 10.21107/rekayasa.v13i1.5912.
- [2] D. C. Yadav and S. Pal, "To generate an ensemble model for women thyroid prediction using data mining techniques," *Asian Pacific Journal of Cancer Prevention*, vol. 20, no. 4, pp. 1275–1281, 2019, doi: 10.31557/APJCP.2019.20.4.1275.
- [3] S. Agustiani, A. Mustopa, A. Saryoko, W. Gata, and S. K. Wildah, "Penerapan Algoritma J48 Untuk Deteksi Penyakit Tiroid," *Paradigma - Jurnal Komputer dan Informatika*, vol. 22, no. 2, pp. 153–160, 2020, doi: 10.31294/p.v22i2.8174.
- [4] D. Sartika and Y. Yupianti, "Klasifikasi Penyakit Tiroid Menggunakan Algoritma C4.5," *Journal of Science and Technology*, vol. 13, no. 1, pp. 71–76, 2020.
- [5] Ardiyansyah, P. A. Rahayuningsih, and R. Maulana, "Analisis Perbandingan Algoritma Klasifikasi Data Mining Untuk

- Dataset Blogger Dengan Rapid Miner,” *Jurnal Khatulistiwa Informatika*, vol. VI, no. 1, pp. 20–28, 2018.
- [6] E. A. F. Elmuna, *Optimasi metode random forest menggunakan Principal Component Analysis untuk memprediksi harga rumah*. 2023. [Online]. Available: <http://etheses.uin-malang.ac.id/50422/>
- [7] L. Mutawali, W. Murniati, and K. Kunci, “PENERAPAN KNNIMPUTER DALAM MENGOLAH DATA MISSING VALUE UNTUK MEMBANTU MENINGKATKAN AKURASI SUPPORT VECTOR MACHINE KLASIFIKASI PENYAKIT TIROID,” 2022. [Online]. Available: <https://archive.ics.uci.edu/ml/datasets/t-hyroid+diseas>
- [8] C. Untuk and K. Harga, “Perbandingan Kinerja Algoritma Naive Bayes dan,” vol. 7, no. 1, pp. 20–24, 2020.
- [9] “18 klasifikasi penyakit jantung menggunakan random forest classifier”.
- [10] P. Tamba, “PREDIKSI PENYAKIT GAGAL JANTUNG DENGAN MENGGUNAKAN RANDOM FOREST,” *Jurnal Sistem Informasi dan Ilmu Komputer Prima*, vol. 5, no. 2, 2022.
- [11] M. Azhar and H. F. Pardede, “Klasifikasi Dialek Pengujar Bahasa Inggris Menggunakan Random Forest,” *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 5, no. 2, p. 439, Apr. 2021, doi: 10.30865/mib.v5i2.2754.
- [12] Oon Wira Yuda, Darmawan Tuti, Lim Sheih Yee, and Susanti, “Penerapan Penerapan Data Mining Untuk Klasifikasi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Random Forest,” *SATIN - Sains dan Teknologi Informasi*, vol. 8, no. 2, pp. 122–131, 2022, doi: 10.33372/stn.v8i2.885.
- [13] R. R. Andarista and A. Jananto, “Penerapan Data Mining Algoritma C4. 5 Untuk Klasifikasi Hasil Pengujian Kendaraan Bermotor,” *Jurnal Tekno Kompak*, vol. 16, no. 2, pp. 29–43, 2022, [Online]. Available: <https://ejurnal.teknokrat.ac.id/index.php/teknokompak/article/view/1525%0Ahttps://ejurnal.teknokrat.ac.id/index.php/teknokompak/article/download/1525/944>
- [14] F. M. Hana, “Klasifikasi Penderita Penyakit Diabetes Menggunakan Algoritma Decision Tree C4.5,” *Jurnal SISKOM-KB (Sistem Komputer dan Kecerdasan Buatan)*, vol. 4, no. 1, pp. 32–39, 2020, doi: 10.47970/siskom-kb.v4i1.173.
- [15] T. S. Lestari and D. A. N. Sirodj, “Klasifikasi Penipuan Transaksi Kartu Kredit Menggunakan Metode Random Forest,” *Jurnal Riset Statistika*, vol. 1, no. 2, pp. 160–167, 2022, doi: 10.29313/jrs.v1i2.525.
- [16] A. Setiawan, R. Febrio Waleska, M. Adji Purnama, and L. Efrizoni, “Sidomulyo barat, Sidomulyo Bar,” 2024. [Online]. Available: <http://e-journal.stmiklombok.ac.id/index.php/jire> ISSN.2620-6900
- [17] R. Sumanti, P. Studi Sistem Informasi, U. Darwan Ali Jln Batu Berlian No, and S. - Kalimantan Tengah, “KLASIFIKASI DATA TINGKAT KERAWANAN KEBAKARAN MENGGUNAKAN ALGORITMA CART,” 2024. [Online]. Available: <http://e-journal.stmiklombok.ac.id/index.php/jire>
- [18] Suci Amaliah, M. Nusrang, and A. Aswi, “Penerapan Metode Random Forest Untuk Klasifikasi Varian Minuman Kopi di Kedai Kopi Konijiwa Bantaeng,” *VARIANSI: Journal of Statistics and Its application on Teaching and Research*, vol. 4, no. 3, pp. 121–127, 2022, doi: 10.35580/variensiunm31.
- [19] A. Primajaya and B. N. Sari, “Random Forest Algorithm for Prediction of Precipitation,” *Indonesian Journal of Artificial Intelligence and Data Mining*, vol. 1, no. 1, p. 27, 2018, doi: 10.24014/ijaidm.v1i1.4903.
- [20] M. R. Givari, M. R. Sulaeman, and Y. Umaidah, “Perbandingan Algoritma SVM, Random Forest Dan XGBoost Untuk Penentuan Persetujuan Pengajuan Kredit,” *Nuansa Informatika*, vol. 16, no. 1, pp. 141–149, 2022, doi: 10.25134/nuansa.v16i1.5406.