

OPTIMASI RANDOM FOREST TERHADAP DATA PENYAKIT LIVER MENGUNAKAN FIREFLY ALGORITHM

Nemesius Sunarjo¹⁾, Danang Aditya Nugraha²⁾, Heri Santoso³⁾

Universitas PGRI Kanjuruhan Malang^{1,2,3)}

email : nemesiussunarjo2002@gmail.com

Abstrak

Penyakit liver merupakan satu diantara penyakit yang amat berbahaya bagi kelangsungan hidup manusia. Dalam upaya untuk mengetahui penyakit liver sejak dini maka diperlukan metode klasifikasi. Peneliti melakukan pengujian dan klasifikasi penyakit liver dengan algoritma Random Forest yang kemudian di Optimasi dengan algoritma Firefly. Tujuan daripada studi ini ialah guna mengetahui bagaimana implementasi firefly algorithm dalam mengoptimalkan akurasi algoritma random forest pada penyakit liver. Data yang digunakan sebanyak 1700 data dengan 11 atribut. Temuan studi ini dengan algoritma Random Forest menghasilkan akurasi sebesar 87.24% sedangkan ketika dioptimasi menggunakan Firefly Algoritma menghasilkan akurasi sebesar 93.24%. Hasil penelitian menunjukan adanya peningkatan akurasi algoritma Random Forest dan di optimasi menggunakan Firefly Algorithm.

Kata Kunci : penyakit liver; klasifikasi; optimasi; random forest; firefly algorithm.

Abstract

Liver disease is one of the most dangerous diseases for human survival. In an effort to find out liver disease early on, a classification method is needed. Researchers conducted testing and classification of liver disease with the Random Forest algorithm which was then optimized with the Firefly algorithm. The purpose of this study is to learn how the application of the firefly algorithm in optimizing the accuracy of the random forest algorithm in liver disease. The data used is 1700 data with 11 attributes. The findings of this study with the Random Forest algorithm produced an accuracy of 87.24% while when optimized using the Firefly Algorithm produced an accuracy of 93.24%. The findings demonstrated a rise in the precision of the Random Forest algorithm and optimized using Firefly Algorithm.

Keywords : liver disease; classification; optimization; random forest; firefly algorithm.

1. PENDAHULUAN

Berdasarkan WHO (2013), “Penyakit hati sering disebut sebagai pembunuh diam-diam karena pada umumnya berkembang tanpa menimbulkan gejala yang jelas sehingga banyak penderita tidak menyadari kondisi serius yang dialaminya hingga mencapai tahap yang berbahaya”. Berdasarkan Kesuma, (203) “Jumlah penderita penyakit liver di Indonesia diperkirakan mencapai 28 juta orang, dan kondisi tersebut menjadikan penyakit ini sebagai salah satu dari sepuluh besar penyakit dengan tingkat risiko kematian yang sangat tinggi, sehingga tidak mengherankan apabila angka kematian akibat liver terus menunjukkan peningkatan dari tahun ke tahun”. Permasalahan utama yang kerap muncul pada penderita gangguan hati adalah kesulitan dalam mendeteksinya sejak tahap awal, sehingga penelitian ini diarahkan untuk mengeksplorasi serta menentukan model klasifikasi yang paling sesuai dalam mengidentifikasi penyakit liver secara lebih dini guna mendukung proses diagnosis awal yang lebih akurat.

Sebagai upaya untuk memperoleh hasil temuan yang lebih tepat dan dapat diandalkan, metode Random Forest dirancang dengan cara mengintegrasikan beragam masukan yang

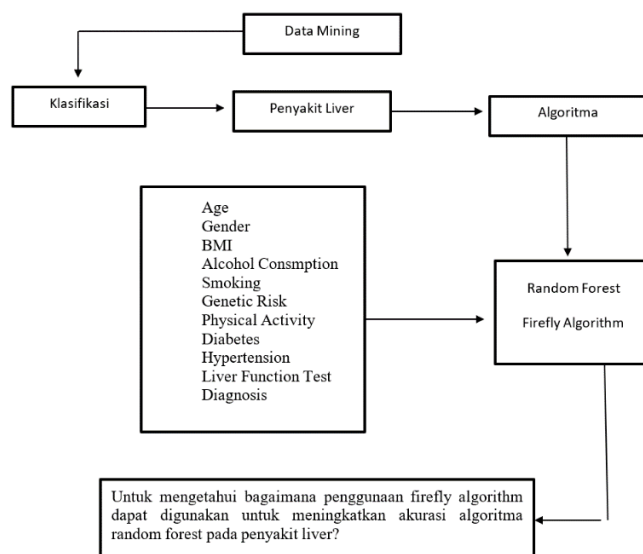
dihasilkan dari sejumlah pohon keputusan, di mana algoritma ini kemudian membangun sekumpulan pohon secara simultan dengan memanfaatkan data maupun atribut yang dipilih secara acak, sehingga menghasilkan kumpulan model prediktif yang sering dianalogikan sebagai sebuah “hutan” karena terdiri atas banyak pohon keputusan yang saling mendukung dalam meningkatkan akurasi prediksi. Berdasarkan Handayani & Charis Fauzan, (2024) “Proses pembentukan pohon keputusan pada algoritma Random Forest dilakukan melalui tahapan pemilihan atribut akar yang ditentukan berdasarkan nilai gain dari perhitungan entropi, kemudian prosedur ini diterapkan secara berulang pada setiap atribut yang tersedia hingga akhirnya seluruh instance yang ada dalam data terklasifikasi ke dalam kelas yang sama secara konsisten”.

Namun demikian, merujuk pada sejumlah penelitian terdahulu yang menjadi acuan, dapat dipahami bahwa temuan yang dihasilkan masih belum sepenuhnya sejalan dengan ekspektasi peneliti, sebab tingkat akurasi yang diperoleh relatif rendah dan menimbulkan interpretasi yang cenderung ambigu. Berdasarkan Styawati et al., (2021) “Optimasi pada dasarnya adalah suatu proses sistematis yang bertujuan untuk menemukan solusi paling optimal dari suatu permasalahan kompleks yang dianggap menarik untuk dikaji, di mana pencarian solusi tersebut dapat dilaksanakan dengan memanfaatkan sejumlah agen yang saling berinteraksi dan bekerja sama, sehingga pada akhirnya membentuk suatu sistem agen yang bersifat adaptif serta mampu berkembang sesuai dengan dinamika permasalahan yang dihadapi”. Temuan studi mengindikasikan bahwa, dibanding particle swarm optimization serta genetika, metode firefly secara signifikan lebih efektif dalam mengidentifikasi nilai optimal global dengan nilai evaluasi yang lebih tinggi. Firefly Algorithm (FA) memiliki manfaat utama dalam mempercepat proses pencarian solusi karena setiap variabel yang terlibat dalam persamaan diarahkan untuk bergerak menuju jawaban terbaik, sehingga dalam penelitian ini algoritme tersebut diterapkan sebagai pendekatan optimasi guna memperoleh solusi paling optimal pada permasalahan klasifikasi penyakit liver.

2. METODE / ALGORITMA

1. Representase Pengetahuan

Penelitian ini berfokus pada upaya optimasi kinerja algoritma Random Forest dalam mengolah data terkait penyakit liver dengan memanfaatkan pendekatan Firefly Algorithm, di mana terdapat sejumlah tahapan prosedural yang secara umum diterapkan dan diuraikan pada bagian berikutnya.



Gambar 1. Tahapan Penelitian

2. Random Forest

Kelompok pembelajaran ensemble mencakup metode klasifikasi serta regresi Random Forest. Berdasarkan Widya Kayohana, (2024) “Metode *Random Forest* pada dasarnya adalah hasil pengembangan lanjutan dari algoritma Decision Tree, di mana setiap pohon keputusan yang terbentuk tidak hanya dilatih menggunakan sampel individu yang diambil secara acak melalui teknik *bootstrap aggregating (bagging)*, tetapi juga dikonstruksi dengan melakukan pemilihan fitur secara acak (*random feature selection*) sehingga menghasilkan kombinasi pohon yang lebih beragam, stabil, serta mampu meningkatkan performa prediksi dibandingkan dengan penggunaan satu pohon keputusan tunggal”.

Rumus menentukan pohon keputusan dari RF

Rumus entropy tertera pada persamaan 2.3.

$$Entropi (S) = \sum_{i=1}^n -p_i * \log_2 p_i \quad (2.3)$$

Keterangan:

A: Fitur/variabel yang dicatat dalam setiap contoh di S.

S : sekelompok fakta atau kejadian yang dicatat selama penelitian.

Pi: Proporsi setiap partisi Si pada keseluruhan himpunan S dinyatakan sebagai persentase ataupun bagian.

n : total partisi dalam kasus mengatur segmentasi ataupun pembagian S.

Menentukan Gain menerapkan pers. berikut:

$$Gain (S, A) = Entropy (S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy (S_i) \quad (2.4)$$

Keterangan:

A : Fitur/variabel yang dicatat dalam setiap contoh di S.

S : Keseluruhan entitas atau data yang dilihat dalam konteks penelitian disebut sebagai himpunan kasus.

|S| : mengindikasikan berapa banyak contoh yang ada di seluruh himpunan S.

|Si| : mengindikasikan berapa banyak contoh yang berada di bawah partisi nomor satu atribut A.

3. Firefly Algorithm

Algoritma Firefly merupakan salah satu metode metaheuristik yang dikembangkan dengan meniru perilaku alami kunang-kunang, di mana mekanisme utama dari algoritma ini didasarkan pada kecenderungan serangga tersebut untuk saling tertarik terhadap intensitas cahaya yang dipancarkan, sehingga prinsip ketertarikan berdasarkan kecerahan tersebut dijadikan dasar dalam membangun prosedur pencarian solusi optimal pada algoritma ini. Berdasarkan Technology, (2023) “Dalam algoritma kunang-kunang, terdapat dua faktor utama yang menjadi komponen krusial dalam menentukan kinerjanya, yaitu intensitas cahaya yang dihasilkan oleh masing-masing kunang-kunang sebagai indikator kualitas solusi, serta tingkat daya tarik antar kunang-kunang yang berperan dalam mengatur pergerakan mereka menuju solusi yang lebih optimal”.

3. HASIL DAN PEMBAHASAN

a. Pengumpulan Data

Pendekatan pengumpulan data pada studi ini ialah tinjauan literatur. Dalam investigasi ini, 1.700 titik data akan diperiksa. Karakteristik fisik termasuk usia, jenis kelamin, BMI, penggunaan alkohol, merokok, risiko genetik, aktivitas fisik, diabetes, hipertensi, tes fungsi hati, dan diagnosis, semuanya dapat dimasukkan ke dalam data. Penelitian ini memiliki tujuan utama untuk melakukan proses klasifikasi data yang berkaitan dengan penyakit liver dengan memanfaatkan indikator-indikator yang telah ditentukan sebelumnya, di mana metode Random

Forest dipilih sebagai algoritma dasar dalam klasifikasi, kemudian kinerjanya dioptimalkan melalui penerapan Firefly Algorithm sehingga diharapkan mampu menghasilkan peningkatan signifikan pada tingkat akurasi klasifikasi yang diperoleh.

Tabel 1. Atribut Penelitian

No	Atribut
1	Age
2	Gender
3	BMI
4	Alcohol Consmption
5	Smoking
6	Genetic Risk
7	Physical Activity
8	Diabetes
9	Hypertension
10	Liver Function Test
11	Diagnosis

b. Preprocessing Data

Preprocessing merupakan tahapan terpenting dalam memproses data mining agar menghindari permasalahan yang akan mengganggu hasil dari proses data mining. Tujuan dari *preprocessing* data adalah untuk mempersiapkan data agar sesuai dengan kebutuhan analisis serta menjamin kualitas data yang baik. Secara umum *preprocessing* data terdiri dari beberapa proses yaitu *Cleaning* digunakan untuk mendeteksi, memperbaiki atau menghapus serta memeriksa data yang tidak konsisten atau tidak akurat, *integration* digunakan untuk menggabungkan data dari beberapa sumber ke dalam penyimpanan data yang sesuai, *transformation* digunakan untuk menormalisasi serta pengumpulan data menjadi sama, *reduction* digunakan menguraikan data kedalam bentuk ukuran yang lebih kecil tetapi tetap menghasilkan hasil analisis yang sama.

c. Rancangan Pengolahan Data untuk Optimasi Parameter SVM dengan Algoritma Firefly pada Penyakit Liver**1 Pengumpulan Data :**

- Pengumpulan data dilakukan berdasarkan indikator yang mencakup atribut fisik seperti : *Gender, Age, Alcohol Consmption, BMI, Genetic Risk, Smoking, Diabetes, Physical Activity, Liver Function Test, Hypertension, serta Diagnosis.*
- Pastikan dataset bersih dari missing Values dan outlier untuk menghindari distorsi hasil analisis.

2 Normalisasi Data ; Normalisasi atau standarisasi atribut-atribut tertentu dari dataset. Normalisasi adalah proses mengubah nilai-nilai dalam suatu dataset kedalam ke dalam rentang tertentu, umumnya antara 0 dan 1 dalam kasus Min-Max Scaling. Normalisasi dapat membantu algoritma pembelajaran mesin bekerja lebih baik, terutama ketika atribut memiliki skala yang berbeda.**3 Pembagian dataset:** Dataset dipisahkan menjadi dua kategori: pengujian dan pelatihan. Model data mining dibuat dengan menggunakan data pelatihan, dan kinerja model selama pengujian dinilai dengan menggunakan data pengujian. Penentuan data training serta data testing pada data ekspor dijalankan secara random sebanyak tujuh kali percobaan dengan presentasi berbeda yaitu : (60% dan 40% , 65% dan 35%, 70% dan 30%, 75% dan 25% 80% dan 20% , 85% dan 15% serta 90% dan 10%).**4 Model Random Forest;** Tentukan Parameter Random Forest yang Akan Dioptimasi, Identifikasi parameter utama Random Forest yang berpengaruh besar terhadap performa model.

- 5 Optimasi dengan Firefly Algorithm
 - Implementasikan algoritma firefly untuk optimasi Random Forest menggunakan library python.
 - Proses optimasi bertujuan untuk menemukan kombinasi parameter yang menghasilkan kinerja terbaik pada model.
- 6 Evaluasi Model dan seleksi solusi terbaik
 - Pilih kombinasi parameter RF dengan intensitas cahaya tertinggi (fungsi objektif maksimum).
 - Latih model RF menggunakan parameter terbaik pada *training set*, lalu evaluasi performa pada *test set* untuk mendapatkan akurasi akhir.

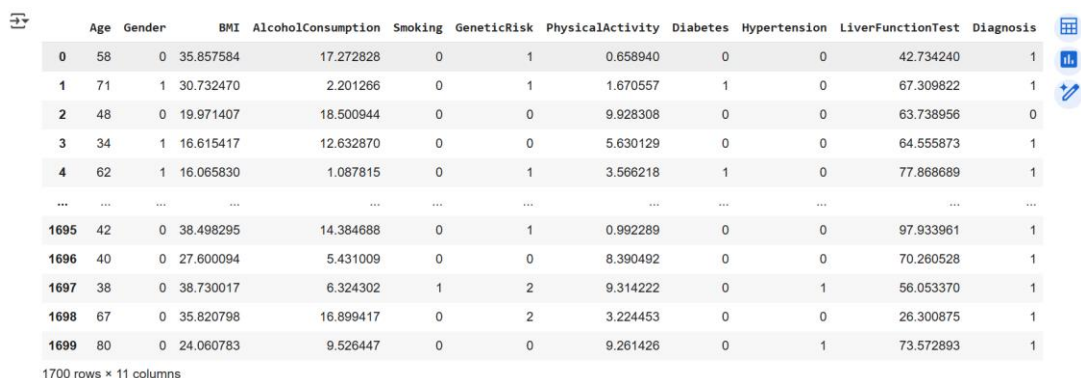
d. Analisis Dan Hasil

1) Pengumpulan Data

Data pada studi ini diambil dengan studi literatur, dimana data yang digunakan merupakan data siap pakai terkait penyakit liver. Data ini diperoleh dari database yang tersedia di situs <https://www.kaggle.com/datasets/rabieelkharoua/predict-liver-disease-1700-records-dataset>. Total terdapat 1700 data yang hendak dianalisis pada studi ini. Data tersebut dapat meliputi bermacam atribut fisik, *semisal Gender, Age, Alcohol Consumption, BMI, Genetic Risk, Smoking, Diabetes, Physical Activity, Liver Function Test, Hypertension, serta Diagnosis*.

2) Preprocessing Data dan Normalisasi Data

Pada proses normalisasi untuk mengubah nilai-nilai dalam suatu dataset kedalam ke dalam rentang tertentu, umumnya antara 0 dan 1 dalam kasus Min-Max Scaling.



	Age	Gender	BMI	AlcoholConsumption	Smoking	GeneticRisk	PhysicalActivity	Diabetes	Hypertension	LiverFunctionTest	Diagnosis
0	58	0	35.857584	17.272828	0	1	0.658940	0	0	42.734240	1
1	71	1	30.732470	2.201266	0	1	1.670557	1	0	67.309822	1
2	48	0	19.971407	18.500944	0	0	9.928308	0	0	63.738956	0
3	34	1	16.615417	12.632870	0	0	5.630129	0	0	64.555873	1
4	62	1	16.065830	1.087815	0	1	3.566218	1	0	77.868689	1
...	...	...	...	...	...	...	...	...	...	...	...
1695	42	0	38.498295	14.384688	0	1	0.992289	0	0	97.933961	1
1696	40	0	27.600094	5.431009	0	0	8.390492	0	0	70.260528	1
1697	38	0	38.730017	6.324302	1	2	9.314222	0	1	56.053370	1
1698	67	0	35.820798	16.899417	0	2	3.224453	0	0	26.300875	1
1699	80	0	24.060783	9.526447	0	0	9.261426	0	1	73.572893	1

1700 rows x 11 columns

Gambar 2. Hasil Normalisasi Data

3) Pengujian Random Forest

Pada tahap ini adalah menganalisis data dalam klasifikasi algoritma random forest pada penyakit Liver. Dalam tahap ini dilakukan perhitungan untuk mendapatkan hasil akurasi terbaik.

Tabel 2. Tabel Hasil Pengujian Random Forest

Data Training	Data Testing	Accuracy
60%	40%	86.62%
65%	35%	87.06%
70%	30%	88.82%
75%	25%	88.71%
80%	20%	87.94%
85%	15%	90.20%
90%	10%	93.53%

4) Pengujian Optimasi Parameter Random Forest menggunakan firefly Algorithm

Penerapan Algoritma Firefly dalam mengoptimalkan parameter Random Forest. Implementasi algoritma firefly dalam mengoptimalkan parameter random forest. Parameter random forest yang umumnya di optimalkan untuk mencari parameter terbaik seperti :

- jumlah pohon($n_estimator$): mengindikasikan jumlah pohon keputusan yang dibangun dalam hutan acak.
- kedalaman maksimum pohon(max_depth), membatasi kedalaman setiap pohon keputusan.
- Jumlah fitur yang dipertimbangkan saat split ($max_features$), menentukan banyak fitur yang dipertimbangkan untuk memecah node.
- Minimum sampel split ($min_samples_split$), jumlah minimum sampel yang di perlukan untuk memecah node internal
- Minimum sampel leaf ($min_samples_leaf$), jumlah minimum sampel yang harus ada di daun($leaf_node$).

Gunakan algoritma firefly untuk mencari kombinasi parameter yang menghasilkan model random forest untuk menentukan akurasi terbaik.

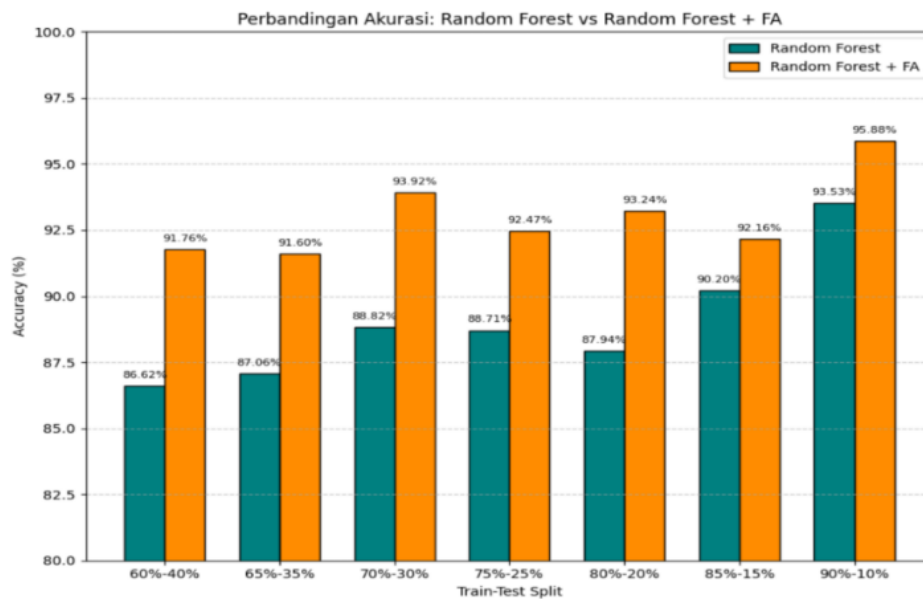
Tabel 3. Hasil Pengujian RF + FA

Data training	Testing	Accuracy
60%	40%	0.9176
65%	35%	0.9160
70%	30%	0.9392
75%	25%	0.9247
80%	20%	0.9324
85%	15%	0.9216
90%	10%	0.9588

5) Pembahasan

Perbandingan akurasi pada grafik 5.1, Pengujian Random Forest melalui data training 60% serta data testing 40% mendapatkan temuan akurasi 86.62% , Ketika menggunakan Random Forest + Firefly Algorithm dengan data training yang sama mendapatkan akurasi sebesar 91.76%, dengan selisih 5.14%. Pengujian melalui data training 65% serta data testing 35% mendapat akurasi Random Forest sebesar 87.06% sedangkan menggunakan Random Forest + Firefly Algorithm sebesar 91.60%, dengan selisih 4.54%. Pengujian dengan data training 70% serta data testing 30% Random Forest mendapat akurasi sejumlah 88.82% sedangkan Random Forest + Firefly sebesar 93.92%, dengan selisih 5.1%.

Pengujian dengan data training 75% serta data testing 25% mendapat hasil akurasi Random Forest sejumlah 88.71% sedangkan Random Forest + Firefly Algorithm 92.47%, dengan selisih 3.76%. Pengujian melalui data training 80% serta data testing 20% mendapatkan temuan akurasi Random Forest 87.94% sedangkan Random Forest + Firefly Algorithm 93.24%, dengan selisih 5.3%. Pengujian dengan data training 85% serta data testing 15% mendapatkan hasil akurasi Random Forest 90.20% sedangkan ketika menggunakan Random Forest + Firefly Algorithm 92.16%, dengan selisih 1.96%. Pengujian melalui data training 90% serta data testing 10% mendapat hasil akurasi Random Forest 93.55% dan ketika menggunakan Random Forest + Firefly sebesar 95.88%, dengan selisih 2.35%.. Dari penelitian tersebut dapat disimpulkan bahwa algoritma optimasi yang digunakan bisa meningkatkan akurasi dari algoritma Random Forest.



Grafik 3. Perbandingan Pengujian RF + FA

4. KESIMPULAN

Kesimpulan

Temuan akurasi dari studi lebih signifikan pada pembagian data training serta data testing 60% serta 40%, 70% serta 30% juga 80% dan 20% dengan selisih 5%-an. Hyperparameter terbaik dari Random Forest setelah di Optimasi adalah $n_estimator=90$, $max_depth=18$, $min_samples_split=22$, $min_samples_leaf=3$, $max_features=1$. Ini berada di data training 80% dan data testing 20% peningkatannya sangat signifikan di angka 5,31%. Secara keseluruhan, penggunaan algoritma Firefly sebagai Algoritma Optimasi terbukti mampu meningkatkan performa model klasifikasi secara menyeluruh, baik dari aspek akurasi keseluruhan maupun dari aspek evaluasi klasifikasi report untuk masing-masing kelas.

Saran

Untuk saran bagi peneliti selanjutnya guna mengembangkan Algoritma Random Forest dalam memprediksi: a). Disarankan untuk mengeksplorasi dan membandingkan algoritma optimasi lain seperti PSO, GA, ataupun Grid Search untuk melihat potensi peningkatan performa lebih lanjut. B). Untuk meningkatkan akurasi dan stabilitas model, sebaiknya dilakukan penambahan jumlah data supaya mendapat output yang optimal, sebab makin banyak data maka output akurasi yang didapat makin tinggi.

REFERENSI

- [1] Ali, M. (2022). Optimasi Pemograman Sistem Pengendalian Mesin CNC Pengebor PCB Berdasar Metode Firefly Algorithm. *ALINIER: Journal of Artificial Intelligence & Applications*, 3(2), 28–37. <https://doi.org/10.36040/alinier.v3i2.5840>
- [2] Azmi, B. N., Hermawan, A., & Avianto, D. (2023). JTIM : Jurnal Teknologi Informasi dan Multimedia Analisis Pengaruh Komposisi Data Training dan Data Testing pada Penggunaan PCA dan Algoritma Decision Tree untuk Klasifikasi Penderita Penyakit Liver. 4(4), 281–290.
- [3] Cahyanti, F. L. D., Sarasati, F., Astuti, W., & Firasari, E. (2023). Klasifikasi Data Mining Dengan Algoritma Machine Larning Untuk Prediksi Penyakit Liver. *Technologia : Jurnal Ilmiah*, 14(2), 134. <https://doi.org/10.31602/tji.v14i2.10093>

- [4] Falatehan, A. I., Hidayat, N. dan Brata, K. C.(2018) “Sistem Pakar Diagnosis Penyakit Hati Menggunakan Metode Fuzzy Tsukamoto Berbasis Android,” Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer (J-PTIIK) Universitas Brawijaya, 2(8), hal. 2373–2381. Tersedia pada: <http://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/1773>.
- [5] Handayani, P., & Charis Fauzan, A. (2024). KLIK: Kajian Ilmiah Informatika dan Komputer Machine Learning Klasifikasi Status Gizi Balita Menggunakan Algoritma Random Forest. Media Online), 4(6), 3064–3072. <https://doi.org/10.30865/klik.v4i6.1909>
- [6] Hasbi, H., & Sasongko, T. B. (2024). Optimasi Performa Random Forest dengan Random Oversampling dan SMOTE pada Dataset Diabetes. Jurnal Media Informatika Budidarma, 8(3), 1756. <https://doi.org/10.30865/mib.v8i3.7855>
- [7] Hay, S.I., Abajobir, A.A., Abate, K. H., Abbafati, C., Abbas, K.M., Abbd-Allah, F., ...& Ciobanu, L. G. (2017). Global, regional, and national disability-adjusted life-years (DALYs) for 333 disease and injuries and healthy life expectancy (HALE) for 195 countries and territories, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. The Lancet, 390(10100), 1260-1344.
- [8] Herliawan, I., Iqbal, M., Gata, W., Rifai, A., & Purnama, J. J. (2020). Classification of Liver Disease By Applying Random Forest. JTIK (Jurnal Ilmu Pengetahuan Dan Teknologi Komputer), 6(1), 89–94. <https://doi.org/10.33480/jitik.v6i1.1424>
- [9] Jamiyanti, E., Jannah, S. W., & Hasan, F. (2024). Perbandingan Performa Optimasi MPPT Menggunakan Algoritma Particle Swarm Optimization dan Firefly Algorithm pada Photovoltaic dalam Kondisi Bayangan Partial. 1(2), 91–101.
- [10] Kesuma, M. (2023). Prediksi Penyakit Liver Menggunakan Algoritma Random Forest. Jurnal Informasi Dan Komputer, 11(2), 2023.
- [11] Lubis, A. I., Erdiansyah, U. dan Siregar, R. (2022) “Komparasi Akurasi pada Naive Bayes dan Random Forest dalam Klasifikasi Penyakit Liver,” Journal of Computing Engineering, System and Science (CESS), 7(1), hal. 81–89.
- [12] Pahlevi, O.-, Amrin, A.-, & Handrianto, Y.-. (2023). Implementasi Algoritma Klasifikasi Random Forest Untuk Penilaian Kelayakan Kredit. Jurnal Infotech, 5(1), 71–76. <https://doi.org/10.31294/infotech.v5i1.15829>
- [13] Pujianto, R., Yusup, D., & Padilah, T. N. (2022). Analisis Sentimen Opini Publik Tentang Vaksin Booster Menggunakan Metode Support Vector Machine dan firefly Algorithm. Jurnal Ilmiah Wahana Pendidikan, Desember, 2022(23), 363–373. <https://doi.org/10.5281/zenodo.7397891>
- [14] Purwaningsih, E., & Nurelasari, E. (2022). Penerapan Algoritma C4.5 Dalam Pengukuran Tingkat Kepuasan Pelanggan Terhadap Layanan Resto Cepat Saji. Reputasi: Jurnal Rekayasa Perangkat Lunak, 3(1), 40–44. <https://doi.org/10.31294/reputasi.v3i1.1199>
- [15] Simanjuntak, L. P. (2024). Penerapan Data Mining Dalam Proses Hukum Pidana Bagi Pelaku Kekerasan Pada Wanita Menggunakan Algoritma C4. 5. ADA Journal of Information System ..., 1(2), 79–92. <https://journals.adaresearch.or.id/adajisr/article/view/41%0Ahttps://journals.adaresearch.or.id/adajisr/article/download/41/31>
- [16] Styawati, Andi Nurkholis, Zaenal Abidin, & Heni Sulistiani. (2021). Optimasi Parameter Support Vector Machine Berbasis Algoritma Firefly Pada Data Opini Film. Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi), 5(5), 904–910. <https://doi.org/10.29207/resti.v5i5.3380>
- [17] Technology, C. (2023). Fa-Knn : Hybrid Algoritma Untuk. 5(2), 71–80.
- [18] Via, Y. V., & Arianti, R. W. (2020). Optimasi Algoritma Firefly Pada Penentuan

- Kandidat Solusi Awal. Prosiding Seminar Nasional Informatika Bela Negara, 1, 205–208. <https://doi.org/10.33005/santika.v1i0.65>
- [19] Wibawa, A. P., Guntur, M., Purnama, A., Fathony Akbar, M., & Dwiyanto, F. A. (2018). Metode-metode Klasifikasi. Prosiding Seminar Ilmu Komputer Dan Teknologi Informasi, 3(1), 134–138.
- [20] Widya Kayohana, K. (2024). Klasifikasi Penyakit Hati Menggunakan Random Forest Dan Knn. JATI (Jurnal Mahasiswa Teknik Informatika), 8(4), 7924–7929. <https://doi.org/10.36040/jati.v8i4.10457>
- [21] Wijaya, H. D., & Dwiasnati, S. (2020). Implementasi Data Mining dengan Algoritma Naïve Bayes pada Penjualan Obat. Jurnal Informatika, 7(1), 1–7. <https://doi.org/10.31311/ji.v7i1.6203>
- [22] <https://www.revou.co/id/kosakata/random-forest>