

IMPLEMENTASI ALGORITMA NUTCRACKER OPTIMIZER PADA LIGHTGBM DALAM MEMPREDIKSI PENYAKIT PARU-PARU

Lianto, Abdul Rahman

Informatika, Universitas Multi Data Palembang
Jalan Rajawali 14, 30113 Palembang, Indonesia
lianto_2226250063@mhs.mdp.ac.id

ABSTRAK

Penyakit paru-paru merupakan salah satu ancaman kesehatan global yang memerlukan penanganan cepat dan akurat. Dalam mengatasi keterbatasan diagnosis konvensional, penerapan *Machine Learning* telah menjadi komponen penting untuk mendukung sistem deteksi dini secara lebih objektif. Salah satu algoritma klasifikasi yang andal dalam mengolah data medis adalah *Light Gradient Boosting Machine* (LightGBM). Namun, kinerja LightGBM sangat bergantung pada pengaturan *hyperparameter*, di mana konfigurasi yang tidak tepat dapat memicu kesalahan diagnosis klinis yang fatal. Oleh karena itu, penelitian ini bertujuan mengoptimalkan *hyperparameter* LightGBM menggunakan algoritma metaheuristik *Nutcracker Optimizer* (NO) guna mencapai presisi klasifikasi yang maksimal. Metode penelitian yang diterapkan meliputi tahap prapemrosesan data, pembagian himpunan data menggunakan skema *Hold-Out* (80:20), serta penerapan *K-Fold Cross Validation* ($K=5$) pada proses optimasi. Hasil pengujian menunjukkan bahwa integrasi algoritma NO berhasil meningkatkan akurasi model dari 93,25% menjadi 94,18%, dengan nilai presisi 94,90% dan *F1-Score* 93,94%. Signifikansi utama terlihat pada penurunan kasus *False Positive* secara drastis, yakni dari 86 menjadi 11 kasus. Kesimpulannya, model LightGBM yang dioptimasi dengan NO terbukti efektif dalam menekan risiko diagnosis berlebih (*overdiagnosis*), sehingga layak digunakan sebagai pendukung keputusan klinis.

Kata kunci: Penyakit Paru-Paru, LightGBM, Nutcracker Optimizer, Klasifikasi, Optimasi.

1. PENDAHULUAN

Kesehatan paru-paru merupakan aspek vital yang kini menghadapi ancaman serius akibat pergeseran gaya hidup modern. Urgensi ini tercermin dari data *Global Cancer Statistics* (GLOBOCAN) 2022, yang menempatkan kanker paru-paru sebagai penyebab utama kematian akibat kanker di dunia, dengan estimasi 1,8 juta kasus atau setara 18,7% dari total kematian global [1]. Sayangnya, tingginya mortalitas ini tidak diimbangi dengan kesadaran deteksi dini. Data dari RSUP Dr. Hasan Sadikin Bandung menunjukkan kondisi kritis, di mana 97% pasien baru terdiagnosis pada stadium IV [2]. Rendahnya partisipasi skrining ini umumnya dipicu oleh kurangnya pemahaman terhadap faktor risiko serta persepsi bahwa prosedur medis itu rumit dan membebani [3]. Keterlambatan diagnosis ini pada akhirnya sangat membatasi efektivitas dan opsi penanganan medis [4].

Padahal, identifikasi dini terhadap faktor risiko merupakan langkah pencegahan yang paling efisien. Informasi risiko penyakit paru-paru tidak hanya berada pada hasil laboratorium, melainkan juga tersembunyi dalam pola hidup sehari-hari, seperti riwayat merokok, konsumsi alkohol, tingkat stres, dan penyakit bawaan. Namun, menganalisis korelasi antarvariabel gaya hidup tersebut secara manual sangat rentan terhadap subjektivitas. Oleh karena itu, penerapan teknologi *Data Mining* dan *Machine Learning* sangat esensial. Teknologi ini mampu mengekstraksi data pola hidup yang kompleks menjadi informasi prediksi yang objektif, akurat, dan mudah diakses sebagai sarana skrining awal [5].

Saat ini, diagnosis awal sering kali masih bergantung pada metode konvensional atau sistem pakar berbasis memori (*Case Based Reasoning/ CBR*). Meskipun cukup membantu, metode ini memiliki keterbatasan dalam hal efisiensi pemrosesan data berskala besar dan pencapaian akurasi tingkat tinggi [4]. Sebagai alternatif, algoritma *Light Gradient Boosting Machine* (LightGBM) menawarkan kecepatan komputasi dan keandalan tinggi dalam menangani data tabular medis, bahkan telah terbukti mencapai akurasi 97% pada studi klasifikasi status kesehatan sebelumnya [6]. Akan tetapi, permasalahan utama dari implementasi LightGBM adalah efektivitas modelnya yang sangat bergantung pada konfigurasi *hyperparameter* [7]. Mengandalkan pengaturan bawaan (*default*) atau pencarian manual sering kali gagal menemukan parameter yang optimal, sehingga berisiko memicu kesalahan klasifikasi yang berakibat fatal dalam pengambilan keputusan klinis [6].

Sebagai rencana penyelesaian masalah atas kendala parameter tersebut, pendekatan optimasi berbasis metaheuristik sangat dibutuhkan untuk mengeksplorasi ruang solusi secara otomatis [8]. Salah satu algoritma terbaru yang terbukti tangguh adalah *Nutcracker Optimizer* (NO), yang diadaptasi dari kecerdasan spasial burung pemecah biji dalam mencari dan menyimpan makanan [8]. Algoritma NO memiliki kemampuan yang sangat baik dalam menyeimbangkan fase eksplorasi dan eksploitasi, sehingga berpotensi besar menutupi kelemahan LightGBM [9]. Lebih lanjut, untuk memastikan keandalan sistem yang dibangun, skema evaluasi akan divalidasi secara ketat menggunakan metode *K-Fold Cross Validation* sehingga diharapkan dapat menghasilkan solusi

diagnosis pendukung yang akurat, stabil, dan efisien. Berdasarkan latar belakang dan permasalahan tersebut, penelitian ini bertujuan untuk mengembangkan model prediksi risiko penyakit paru-paru yang presisi dengan mengintegrasikan algoritma LightGBM dan optimasi *Nutcracker Optimizer* (NO).

2. TINJAUAN PUSTAKA

Prediksi penyakit paru-paru dengan *Machine Learning* banyak mengeksplorasi algoritma tradisional seperti *Random Forest* (RF), *C4.5*, *K-Nearest Neighbor* (KNN), dan *Support Vector Machine* (SVM). Performa algoritma-algoritma tersebut cukup kompetitif, namun hasilnya sangat bervariasi (91%–99%) karena sangat bergantung pada keseimbangan data [8] [14]. Di sisi lain, penggunaan model probabilitas tunggal seperti *Naïve Bayes* dan Regresi Linier cenderung tertahan pada ambang batas akurasi menengah di kisaran 86%–90% [15] [16] [10].

Untuk mengatasi keterbatasan *single classifier* tersebut, tren penelitian bergeser pada teknik *ensemble* dan *boosting*. Penggabungan model dengan *Adaboost* atau *Extreme Gradient Boosting* (XGBoost) terbukti jauh lebih superior dan mampu mendongkrak performa diagnosis secara signifikan dibandingkan algoritma konvensional [11] [12]. Selain pemanfaatan arsitektur *boosting*, penelitian berbasis Jaringan Syaraf Tiruan (JST) mengonfirmasi bahwa penyematan algoritma optimasi (*optimizer*) memegang peranan krusial untuk mencapai performa puncak model melalui penyesuaian parameter secara otomatis [13] [14].

Berpijak pada pentingnya komputasi *boosting* dan algoritma optimasi, perkembangan terkini (*state of the art*) menunjukkan bahwa *Light Gradient Boosting Machine* (LightGBM) sangat unggul pada pemrosesan data medis tabular dengan rekor akurasi mencapai 97%–98% [6] [7]. Dalam proses pencarian hiperparameter, *Nutcracker Optimizer* (NO) juga mulai diakui keandalannya dalam mengeksplorasi ruang parameter kompleks [9]. Akan tetapi, sebagian besar sistem deteksi penyakit paru-paru saat ini belum ada yang memfokuskan kajian pada integrasi antara kecepatan pemrosesan LightGBM dengan kapabilitas pencarian optimal milik NO. Celah penelitian (*research gap*) inilah yang menjadi landasan pelaksanaan penelitian ini guna menghasilkan model prediktif yang efisien dan berpresisi tinggi. Untuk merealisasikan arsitektur model berdasarkan celah penelitian tersebut, diperlukan pemahaman teknis yang komprehensif mengenai cara kerja masing-masing algoritma. Oleh karena itu, tinjauan teoretis berikut akan menguraikan konsep dasar dari *Light Gradient Boosting Machine* (LightGBM) sebagai algoritma klasifikasi utama, serta *Nutcracker Optimizer* (NO) sebagai metode optimasi pencarian hiperparameternya.

2.1. Light Gradient Boosting Machine (LightGBM)

Light Gradient Boosting Machine (LightGBM) merupakan algoritma pembelajaran mesin yang menggunakan kerangka kerja *gradient boosting* berbasis pohon keputusan. Berbeda dengan algoritma *boosting* konvensional yang menumbuhkan pohon secara mendatar (*level-wise*), LightGBM menerapkan strategi pertumbuhan *leaf-wise* yang tumbuh secara vertikal dengan memilih daun yang memiliki penurunan nilai *loss* paling maksimal untuk dikembangkan [6]. Karakteristik utama yang membuat LightGBM unggul adalah efisiensi penggunaan memori yang lebih rendah serta kecepatan pelatihan yang jauh lebih singkat dibandingkan metode *boosting* lainnya, sehingga sangat andal untuk klasifikasi data medis berskala besar [6] [7].

Secara matematis, tujuan LightGBM adalah meminimalkan fungsi objektif. Jika diberikan dataset $\{(x_i, y_i)\}$, fungsi objektif pada iterasi ke- t didefinisikan seperti persamaan (1):

$$Obj^{(t)} = \sum_{i=0}^n [-y_i \ln(p_i) - (1 - y_i) \ln(1 - p_i)] + \Omega(f_t) \quad (1)$$

Berbeda dengan algoritma lain, LightGBM menggunakan pendekatan berbasis gradien dan hessian. Untuk menentukan percabangan (*split*) terbaik pada pohon, LightGBM menghitung nilai Gain dengan persamaan (2):

$$Gain = \frac{1}{2} \left[\frac{G_L^2}{H_L + \lambda} + \frac{G_R^2}{H_R + \lambda} - \frac{(G_L + G_R)^2}{H_L + H_R + \lambda} \right] - \gamma \quad (2)$$

2.2. Nutcracker Optimizer (NO)

Algoritma *Nutcracker Optimization* (NO) adalah algoritma metaheuristik yang diinspirasi oleh perilaku cerdas burung pemecah biji (*Nutcracker*) dalam mempertahankan kelangsungan hidup. Algoritma ini memodelkan dua strategi utama, yaitu mekanisme pencarian makanan (*foraging*) dan mekanisme penyimpanan serta pemulihan cadangan makanan (*caching*) [8]. Dalam proses optimasi, algoritma ini secara dinamis menyeimbangkan tahap eksplorasi untuk memetakan seluruh ruang solusi dan tahap eksploitasi untuk memperhalus pencarian pada area potensial. Penerapan NO terbukti efektif dalam memecahkan masalah optimasi parameter yang kompleks, sehingga mampu meningkatkan akurasi klasifikasi pada berbagai model *machine learning* secara signifikan [8] [9].

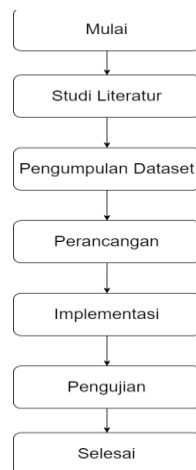
Pada fase pencarian makanan, individu memperbarui posisinya berdasarkan posisi terbaik saat ini (*current best*) atau posisi acak menggunakan persamaan (3):

$$X_i^{t+1} = \begin{cases} X_i^t + r_1 \times (X_{best}^t - X_i^t) & \text{Jika } r_2 \geq P_{pa} \\ X_i^t + r_1 \times (X_A^t - X_B^t) & \text{Jika } r_2 \leq P_{pa} \end{cases} \quad (3)$$

Pada fase penyimpanan, algoritma melakukan pencarian lokal di sekitar solusi yang ditemukan (seolah burung mengingat lokasi penyimpanan) menggunakan persamaan (4):

$$X_i^{t+1} = X_i^t + \mu \times (X_{best}^t - X_i^t) + \sigma \times (X_{RP}^t - X_i^t) \quad (4)$$

3. METODE PENELITIAN



Gambar 1. Tahapan Metode Penelitian

Dari flowchart dari Gambar 1, metodologi penelitian disusun sebagai kerangka kerja sistematis untuk menyelesaikan rumusan masalah secara terstruktur. Penerapan metodologi ini bertujuan untuk menjamin validitas hasil penelitian serta meminimalisir potensi kesalahan selama proses eksperimen berlangsung. Tahapan penelitian dirancang secara berurutan, mulai dari studi literatur, pengumpulan data, perancangan model, implementasi,

hingga evaluasi akhir. Rincian dari masing-masing tahapan metodologi tersebut diuraikan sebagai berikut:

3.1. Studi Literatur

Tahapan studi literatur dilakukan untuk membangun dasar pemahaman yang kuat sebelum memulai penelitian. Penulis mengumpulkan dan mempelajari informasi dari berbagai sumber terpercaya, seperti buku dan jurnal ilmiah. Fokus utamanya adalah memahami cara kerja algoritma *Light Gradient Boosting Machine* (LightGBM) serta algoritma *Nutcracker Optimizer* (NO) yang akan digunakan untuk mencari pengaturan terbaik model secara otomatis. Pengetahuan teknis ini dilengkapi dengan pendalaman mengenai hubungan antara pola hidup (*lifestyle*) dan risiko penyakit paru-paru. Hal ini mencakup identifikasi fitur perilaku dominan—seperti kebiasaan merokok, tingkat aktivitas fisik, dan usia—sebagai prediktor utama dalam mendeteksi penyakit. Terakhir, dilakukan penentuan metode pengukuran kinerja model yang tepat untuk memastikan hasil penelitian bersifat objektif dan dapat dipertanggungjawabkan secara ilmiah.

3.2. Pengumpulan Dataset

Pada tahap ini, dilakukan pengumpulan data primer yang akan digunakan untuk melatih dan menguji model. Dataset yang digunakan adalah data publik dari situs repositori Kaggle yang berjudul "*Dataset predik terkena penyakit paru-paru*". Dataset ini bersifat kategorikal yang merepresentasikan kebiasaan dan pola hidup seseorang, dengan total himpunan data sebanyak 30.000 *record*. Berdasarkan distribusi kelasnya, dataset ini cukup seimbang, yang terdiri dari 15.648 data dengan label kelas "Tidak" (berisiko) dan 14.352 data berlabel kelas "Ya" (berisiko penyakit paru-paru).

Tabel 1. Dataset terkena penyakit paru-paru

No	Usia	Jenis_Kelamin	Merokok	Bekerja	Rumah_Tangga	Aktivitas_Bergadang	Aktivitas_Olahraga	Asuransi	Penyakit_Bawaan	Hasil
1	Tua	Pria	Pasif	Tidak	Ya	Ya	Sering	Ada	Tidak	Ya
2	Tua	Pria	Aktif	Tidak	Ya	Ya	Jarang	Ada	Ada	Tidak
3	Muda	Pria	Aktif	Tidak	Ya	Ya	Jarang	Ada	Tidak	Tidak
4	Tua	Pria	Aktif	Ya	Tidak	Tidak	Jarang	Ada	Ada	Tidak
5	Muda	Wanita	Pasif	Ya	Tidak	Tidak	Sering	Tidak	Ada	Ya
6	Muda	Wanita	Pasif	Ya	Tidak	Tidak	Sering	Tidak	Ada	Tidak
7	Tua	Wanita	Pasif	Tidak	Ya	Tidak	Sering	Tidak	Tidak	Ya
8	Muda	Pria	Aktif	Tidak	Ya	Ya	Sering	Tidak	Tidak	Tidak
9	Tua	Wanita	Aktif	Ya	Ya	Ya	Jarang	Ada	Ada	Ya
10	Muda	Wanita	Pasif	Ya	Tidak	Ya	Jarang	Ada	Ada	Ya
11	Tua	Wanita	Pasif	Ya	Ya	Tidak	Sering	Ada	Ada	Ya
12	Tua	Wanita	Aktif	Tidak	Ya	Tidak	Jarang	Ada	Tidak	Tidak
13	Muda	Pria	Aktif	Tidak	Ya	Ya	Jarang	Ada	Tidak	Tidak
14	Tua	Wanita	Aktif	Ya	Tidak	Ya	Jarang	Ada	Ada	Tidak
15	Muda	Wanita	Pasif	Ya	Tidak	Ya	Sering	Tidak	Ada	Ya
16	Muda	Wanita	Pasif	Ya	Tidak	Ya	Jarang	Ada	Ada	Ya
17	Tua	Wanita	Pasif	Ya	Ya	Tidak	Sering	Ada	Ada	Ya
18	Tua	Wanita	Aktif	Tidak	Ya	Tidak	Jarang	Ada	Tidak	Tidak
19	Muda	Pria	Aktif	Tidak	Ya	Ya	Jarang	Ada	Tidak	Tidak
20	Tua	Wanita	Aktif	Ya	Tidak	Ya	Jarang	Ada	Ada	Tidak

Berdasarkan cuplikan data yang disajikan pada Tabel 1, terlihat struktur dataset yang digunakan dalam penelitian ini terdiri dari beberapa atribut klinis dan demografis pasien. Data tersebut mencakup variabel numerik dan kategorikal, di antaranya adalah usia, jenis kelamin, merokok, bekerja, rumah tangga, aktivitas begadang, aktivitas olah raga, asuransi, dan penyakit bawaan yang terekam dalam format tabular. Setiap baris dalam gambar tersebut merepresentasikan satu sampel pasien, di mana kolom terakhir bertindak sebagai label target untuk menentukan klasifikasi penyakit paru-paru.

3.3. Perancangan Model

Perancangan model diawali dengan input dataset penyakit paru-paru. Data mentah tersebut kemudian memasuki tahap *Preprocessing* yang terdiri dari dua langkah utama. Pertama, pembersihan data (*data cleaning*) untuk mengeliminasi duplikasi dan menangani nilai yang hilang (*missing values*). Kedua, transformasi fitur kategorikal menjadi format numerik biner (nilai 1 atau 0) menggunakan teknik *Label Encoding* agar data dapat diproses secara optimal oleh algoritma LightGBM.

Setelah tahap *preprocessing*, dilakukan skema pembagian data bertingkat. Pertama, dataset dibagi menggunakan metode *Hold-Out Split* (rasio 80:20), di mana 80% dialokasikan sebagai Data Latih dan 20% disimpan secara terisolasi sebagai Data Uji. Selanjutnya, skema *K-Fold Cross Validation* ($K=5$) diterapkan khusus pada Data Latih. Data latih ini kemudian menjadi dasar untuk tahap optimasi *hyperparameter* menggunakan *Nutcracker Optimizer* (NO). Pada tahap ini, algoritma NO menginisialisasi populasi parameter dan secara iteratif mencari kombinasi konfigurasi terbaik (seperti *learning rate*, *num leaves*, dan *max depth*) untuk model LightGBM. Setiap kombinasi dievaluasi berdasarkan rata-rata akurasi *K-Fold* yang berfungsi sebagai nilai *fitness* untuk memandu pencarian solusi pada iterasi berikutnya.

3.4. Implementasi

Tahapan implementasi merupakan proses teknis penerjemahan rancangan model yang telah disusun ke dalam baris kode program (*coding*). Seluruh proses komputasi dibangun menggunakan bahasa pemrograman Python yang dijalankan pada lingkungan pengembangan Google Colab. Pada tahap ini, logika algoritma *Nutcracker Optimizer* (NO)—mulai dari inisialisasi populasi hingga mekanisme pencarian solusi—diprogram dan diintegrasikan secara langsung dengan arsitektur algoritma LightGBM. Variabel-variabel *hyperparameter* model didefinisikan sebagai target optimasi yang akan dicari nilai terbaiknya. Proses ini diakhiri dengan pelatihan model final menggunakan seluruh Data Latih (80%) dipadukan dengan kombinasi parameter optimal yang telah ditemukan oleh algoritma NO.

3.5. Pengujian dan Evaluasi

Tahap pengujian merupakan langkah akhir untuk mengukur keandalan sistem yang telah dibangun. Kinerja model NO-LightGBM final akan diuji secara buta (*blind test*) menggunakan Data Uji (20%) yang telah dipisahkan sejak awal penelitian. Evaluasi kinerja diukur menggunakan standar metrik berbasis *Confusion Matrix*, yang meliputi *Accuracy*, *Precision*, *Recall*, dan *F1-Score*. Hasil pengujian ini kemudian dikomparasikan secara langsung dengan hasil pengujian model LightGBM standar (menggunakan *hyperparameter default*) untuk membuktikan signifikansi peningkatan performa yang dihasilkan dari optimasi berbasis burung *Nutcracker* tersebut.

4. HASIL DAN PEMBAHASAN

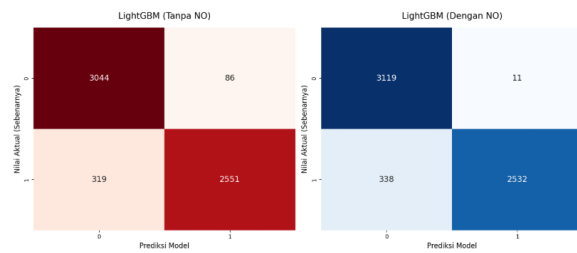
Pengujian sistem dilakukan untuk mengukur dan membandingkan kinerja algoritma LightGBM standar (*baseline*) dengan arsitektur LightGBM yang telah dioptimasi menggunakan algoritma metaheuristik *Nutcracker Optimizer* (NO). Pengujian ini menggunakan skema *blind test* pada 20% Data Uji yang sebelumnya telah diisolasi dari proses pelatihan model.

Pada tahap optimasi *hyperparameter*, algoritma NO bertugas mengeksplorasi ruang pencarian untuk menemukan kombinasi parameter komputasi yang paling ideal bagi model. Berdasarkan hasil iterasi pencarian, algoritma NO berhasil menetapkan konfigurasi optimal pada LightGBM dengan nilai *learning rate* sebesar 0,05, *num leaves* sebesar 31, dan batas kedalaman pohon keputusan (*max depth*) pada angka 7. Parameter optimal inilah yang kemudian digunakan pada model final untuk memprediksi probabilitas penyakit paru-paru.

Tabel 2. Tabel hasil

Metrik Evaluasi	LightGBM	LightGBM & NO
<i>Accuracy</i>	93.25%	94.18%
<i>Precision</i>	93.63%	94.90%
<i>Recall</i>	93.07%	93.94%
<i>F1-Score</i>	93.20%	94.13%

Dari Tabel 2 dapat dilihat bahwa algoritma NO mampu meningkatkan performa LightGBM berdasarkan metrik evaluasi *Accuracy*, *Precision*, *Recall*, dan *F1-Score* dibandingkan dengan model *baseline* (tanpa optimasi). Peningkatan yang konsisten pada seluruh metrik ini khususnya pada *F1-Score* menunjukkan bahwa model memiliki keseimbangan prediksi yang sangat baik dan terhindar dari bias kelas. Hal ini membuktikan bahwa algoritma *Nutcracker Optimizer* terbukti efektif dalam memaksimalkan keandalan dan akurasi LightGBM untuk sistem deteksi dini penyakit paru-paru.



Gambar 2. Grafik Confusion Matrix

Berdasarkan Gambar 2, terlihat jelas adanya pergeseran distribusi prediksi antara model *baseline* dan model optimasi. Pada model LightGBM standar, tingkat kesalahan klasifikasi tercatat sebanyak 86 kasus *False Positive* (FP) dan 319 kasus *False Negative* (FN). Setelah dioptimasi menggunakan algoritma NO, model menunjukkan perubahan karakteristik prediksi yang tajam; jumlah kasus FP berhasil ditekan secara drastis menjadi hanya 11 kasus, meskipun hal ini memberikan efek pertukaran (*trade-off*) berupa sedikit peningkatan pada kasus FN menjadi 338 kasus. Dinamika ini membuktikan bahwa penyematan *Nutcracker Optimizer* (NO) mengarahkan model untuk menjadi jauh lebih presisi, ketat, dan selektif sebelum mengklasifikasikan seseorang terkena penyakit. Penurunan drastis pada rasio *False Positive* ini sangat menguntungkan untuk meminimalisir terjadinya diagnosis palsu (*overdiagnosis*), sehingga dapat menghindarkan pasien sehat dari beban psikologis maupun tindakan medis lanjutan yang tidak perlu.

5. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian dan pembahasan yang telah dilakukan, dapat disimpulkan bahwa implementasi algoritma *Nutcracker Optimizer* (NO) terbukti efektif dalam meningkatkan performa model *Light Gradient Boosting Machine* (LightGBM) untuk memprediksi penyakit paru-paru. Peningkatan ini ditunjukkan oleh kenaikan nilai akurasi dari 93,25% pada model *baseline* menjadi 94,18% pada model yang telah dioptimasi, dengan nilai *F1-Score* yang mencapai 94,13%. Secara spesifik, integrasi algoritma NO mampu menekan angka kesalahan klasifikasi pada kategori *False Positive* secara signifikan, yakni dari 86 kasus menjadi hanya 11 kasus. Hal ini membuktikan bahwa mekanisme pencarian parameter berbasis perilaku burung *nutcracker* berhasil menemukan titik optimum global pada hiperparameter LightGBM, sehingga menghasilkan sistem prediksi yang lebih presisi, andal, dan mampu meminimalisir risiko kesalahan diagnosis medis (*overdiagnosis*).

Untuk pengembangan penelitian selanjutnya, disarankan agar pengujian dilakukan pada dataset medis dengan tingkat ketidakseimbangan data yang lebih ekstrem guna menguji ketangguhan model dalam kondisi yang lebih kompleks. Selain itu, perlu dilakukan studi komparasi antara *Nutcracker Optimizer* dengan algoritma metaheuristik terbaru lainnya untuk mengevaluasi efisiensi waktu komputasi dalam proses optimasi. Terakhir, pengembangan sistem dapat diarahkan pada pembuatan antarmuka

aplikasi berbasis web atau *mobile* agar hasil prediksi ini dapat diimplementasikan secara praktis sebagai alat bantu bagi tenaga medis maupun masyarakat umum dalam melakukan deteksi dini penyakit paru-paru.

DAFTAR PUSTAKA

- [1] J. Ferlay, M. E. Rebecca, and L. S. Mph, "Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *Am. Cancer Soc.*, vol. 74, no. 3, pp. 205–313, 2024, doi: 10.3322/caac.21834.
- [2] T. Kosawara, J. Irawan, and F. Surya, "Karakteristik karsinoma paru bukan sel kecil dan hasil pemeriksaan EFGR di RSUP dr. Hasan sadikin periode tahun 2018-2020," *Med. Kartika J. Kedokt. dan Kesehat.*, vol. 5, no. 2, pp. 150–161, 2022, doi: 10.35990/mk.v5n2.p150-161.
- [3] S. A. Amalia, I. Windarti, N. Ayu, V. Irawati, and A. Infianto, "Faktor-faktor risiko pada pasien kanker paru dan skrining di indonesia," *Heal. Med. Sci.*, vol. 3, no. 2, pp. 1–8, 2026, doi: 10.47134/phms.v3i2.560.
- [4] S. R. Maulana, L. Affandi, and M. Haniah, "Sistem pakar diagnosa penyakit paru-paru menggunakan metode case based reasoning," *JIP (Jurnal Inform. Polinema)*, vol. 9, no. 2, pp. 193–200, 2023, doi: 10.33795/jip.v9i2.1225.
- [5] A. Y. Permana, H. N. Fazri, M. F. N. Athoilah, M. Robi, and R. Firmansyah, "Penerapan data mining dalam analisis prediksi kanker paru menggunakan algoritma random forest," *J. Ilm. Tek. Inform. dan Komun.*, vol. 3, no. 2, pp. 27–41, 2023, doi: 10.55606/juitik.v3i2.472.
- [6] M. Thoriqulhaq, M. Idhom, and K. M. Hindrayani, "Implementasi algoritma LightGBM untuk prediksi status gizi bayi dan balita di desa Doko kabupaten Kediri," *J-TETA J. Tek. Terap.*, vol. 4, no. 1, pp. 65–73, 2025.
- [7] M. Soni, M. W. Bhatt, and P. Ofori-amanfo, "Light-GBM based minority oversampling model using biomedical data analysis for breast cancer classification," *Discov. Appl. Sci.*, vol. 7, no. 763, 2025, doi: 10.1007/s42452-025-07390-7.
- [8] A. F. Mohamed *et al.*, "Boosted nutcracker optimizer and chaos game optimization with cross vision transformer for medical image classification," *Egypt. Informatics J.*, vol. 26, p. 100457, 2024, doi: 10.1016/j.eij.2024.100457.
- [9] Q. Tang *et al.*, "Application of the NOA-optimized random forest algorithm to fluid identification — low-porosity and low-permeability reservoirs," *Processes*, vol. 13, no. 7, pp. 1–14, 2025, doi: 10.3390/pr13072132.
- [10] M. A. R. Wahid, A. Nugroho, and A. H. Anshor, "Prediksi penyakit kanker paru-paru dengan algoritma regresi linier," *Bull. Inf. Technol.*, vol. 4, no. 1, pp. 63–74, 2023, doi: 10.47065/bit.v3i1.501.
- [11] M. B. Sadr and M. S. Hasibuan, "Prediksi

- penyakit paru-paru menggunakan algoritma naïve bayes dan adaboost,” *J. Teknol. Inf. Magister Darmajaya*, vol. 1, no. 1, 2024.
- [12] R. Harahap, M. Irpan, M. A. Dinata, L. Efrizoni, and Rahmaddeni, “Perbandingan algoritma random forest dan xgboost untuk klasifikasi penyakit paru-paru berdasarkan data,” *J. Ilm. betrik*, vol. 15, no. 02, pp. 130–141, 2025, doi: 10.36050/3v3xwn06.
- [13] V. R. Danestiara and A. N. Fadhillah, “Prediksi penyakit paru menggunakan algoritma deep neural network,” *In search*, vol. 21, no. 1, pp. 174–181, 2022, doi: 10.37278/insearch.v22i1.955.
- [14] F. Khadafi and P. P. P. A. N. W. F. I. R. H. Zer, “Optimasi akurasi backpropagation dengan adaptive moment estimation terhadap kasus prediksi deteksi penyakit paru-paru,” *J. JISIILKOM (Jurnal Inov. Sist. Inf. Ilmu Komputer)*, vol. 3, no. 1, 2025.