

## Optimasi Hyperparameter CatBoost dengan Particle Swarm Optimization untuk Klasifikasi Hipertensi

Muhammad Iqbal Al Afgany <sup>a,1,\*</sup>, Ani Dijah Rahajoe <sup>b,2</sup>, Henni Endah Wahanani <sup>b,3</sup>

<sup>a</sup>Universitas Pembangunan Nasional "Veteran" Jawa Timur, Jl. Rungkut Madya No. 1, Gn. Anyar, Surabaya, Indonesia

<sup>1</sup> 21081010330@student.upnjatim.ac.id\*; <sup>2</sup> anidijah.if@upnjatim.ac.id; <sup>3</sup> henniendah.if@upnjatim.ac.id

\* Penulis Koresponden

### INFO ARTIKEL

#### Histori Artikel

Pengajuan 2025-10-13  
Diperbaiki 2025-11-26  
Diterima 2025-12-04

#### Kata Kunci

CatBoost,  
Hipertensi,  
Machine Learning,  
Optimasi,  
Particle Swarm  
Optimization (PSO).

### ABSTRAK

Hipertensi merupakan penyakit kardiovaskular yang menyerang 11.952.694 penduduk berusia  $\geq 15$  tahun di Jawa Timur pada tahun 2019, namun hanya 40,1% yang memperoleh layanan kesehatan. Penelitian ini bertujuan menganalisis pengaruh optimasi Particle Swarm Optimization (PSO) terhadap performa algoritma CatBoost dalam klasifikasi tingkat hipertensi. Dataset penelitian menggunakan data gabungan dari Puskesmas Kepatihan Gresik (191 data) dan Kaggle (12.500 data) yang dibagi dengan rasio 80:10:10. PSO digunakan untuk optimasi hyperparameter CatBoost meliputi iterations, depth, learning\_rate, dan l2\_leaf\_reg. Evaluasi model menggunakan metrik akurasi, precision, recall, dan F1-score. Hasil penelitian menunjukkan model CatBoost dengan optimasi PSO mencapai akurasi 96% dengan konfigurasi optimal  $iterations=100$ ,  $depth=3$ ,  $learning\_rate=0.055$ , dan  $l2\_leaf\_reg=3$ , lebih tinggi 2% dibandingkan tanpa optimasi (94%). Penelitian ini membuktikan efektivitas PSO dalam mengoptimalkan hyperparameter CatBoost untuk deteksi dini hipertensi yang lebih akurat.

### ABSTRACT

#### Keyword

CatBoost,  
Hypertension, Machine  
Learning, Optimization,  
Particle Swarm  
Optimization (PSO).

*Hypertension is a cardiovascular disease affecting 11,952,694 residents aged  $\geq 15$  years in East Java in 2019, yet only 40.1% received healthcare services. This study aims to analyze the effect of Particle Swarm Optimization (PSO) on CatBoost algorithm performance in hypertension level classification. The research dataset combined data from Puskesmas Kepatihan Gresik (191 data) and Kaggle (12,500 data) divided with an 80:10:10 ratio. PSO was used for CatBoost hyperparameter optimization including iterations, depth, learning\_rate, and l2\_leaf\_reg. Model evaluation utilized accuracy, precision, recall, and F1-score metrics. Results show that CatBoost with PSO optimization achieved 96% accuracy with optimal configuration of  $iterations=100$ ,  $depth=3$ ,  $learning\_rate=0.055$ , and  $l2\_leaf\_reg=3$ , 2% higher than without optimization (94%). This study proves the effectiveness of PSO in optimizing CatBoost hyperparameters for more accurate early hypertension detection.*

Ini adalah artikel akses terbuka di bawah lisensi [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/).



## 1. Pendahuluan

Hipertensi merupakan gangguan kardiovaskular yang dikarakterisasi oleh kenaikan tekanan darah sistolik dan diastolik di atas batas normal, yang berpotensi menimbulkan kerusakan pada organ-organ vital seperti mata, jantung, otak, dan ginjal. Tekanan darah dikatakan normal pada nilai 120/80 mmHg, sedangkan diagnosis hipertensi ditegakkan ketika tekanan darah mencapai lebih dari 140/90 mmHg[1]. Penyakit ini dikenal sebagai "*silent killer*" karena umumnya tidak memperlihatkan gejala klinis yang jelas, namun dapat memicu komplikasi berbahaya seperti stroke, aterosklerosis, dan kegagalan fungsi ginjal [2]. Etiologi hipertensi dapat dikategorikan menjadi dua kelompok utama, yaitu faktor genetik yang meliputi riwayat hipertensi dalam keluarga, serta faktor non-genetik yang mencakup usia, jenis kelamin, kebiasaan merokok, asupan garam yang tinggi, pola konsumsi makanan berkolesterol tinggi, konsumsi minuman beralkohol, obesitas, tekanan psikologis, dan minimnya aktivitas fisik [3].

Hipertensi menempati posisi ketiga sebagai penyebab kematian tertinggi secara global, dengan jumlah penderita mencapai 1,13 miliar orang di seluruh dunia, di mana 31% di antaranya dialami oleh populasi dewasa [4]. Kawasan Asia mencatat 38,4 juta kasus hipertensi pada tahun 2000, yang diproyeksikan akan meningkat menjadi 67,4 juta pada tahun 2025, dengan wilayah Asia Tenggara memiliki tingkat prevalensi sebesar 25% dari total populasi[5]. Data Jawa Timur tahun 2019 menunjukkan estimasi 11.952.694 individu berusia  $\geq 15$  tahun menderita hipertensi, dengan distribusi 48% laki-laki dan 52% perempuan, namun hanya 40,1% yang memperoleh akses layanan kesehatan[6].

Rendahnya cakupan layanan kesehatan tersebut mengindikasikan pentingnya upaya deteksi dini hipertensi. Deteksi dini atau skrining bertujuan mendeteksi onset penyakit sebelum memasuki fase klinis dengan mengidentifikasi faktor-faktor risiko seperti usia, jenis kelamin, dan riwayat keturunan melalui pemeriksaan tekanan darah [7], [8]. Dalam konteks ini, peran tenaga kesehatan dalam melakukan screening dan pengkajian sistematis menjadi krusial untuk meningkatkan cakupan deteksi hipertensi di masyarakat [7], [8]. Skrining dimanfaatkan sebagai pemeriksaan sederhana pada orang-orang yang secara asimtomatik belum menunjukkan gejala klinis [9].

Hasil Riset Kesehatan Dasar Indonesia, mengindikasikan bahwa hipertensi memerlukan perhatian khusus mengingat tren peningkatan jumlah penderita dan mortalitas yang berkaitan dengan penyakit ini. Terapi hipertensi membutuhkan waktu yang panjang, sementara metode deteksi manual yang masih dominan digunakan menunjukkan urgensi deteksi dini untuk meningkatkan kualitas pelayanan kesehatan[10]. Keterbatasan akses dan minimnya pemahaman masyarakat mengenai hipertensi menjadi hambatan utama dalam upaya penanganannya. Dengan demikian, diperlukan sistem cerdas yang mampu mengklasifikasikan hipertensi secara dini berdasarkan kondisi pasien, mendukung tenaga medis dalam upaya pencegahan, serta meningkatkan efektivitas terapi dan kualitas pelayanan kesehatan.

Beberapa penelitian terdahulu telah mengeksplorasi penggunaan algoritma machine learning untuk klasifikasi hipertensi. Penelitian oleh Purwono [11] menggunakan algoritma Neural Network dengan parameter solver='lbfgs', alpha=1e-5, hidden\_layer\_sizes=(5,2) dan mencapai akurasi 85% untuk prediksi jenis penyakit hipertensi dengan pembagian data 70:30. Sementara itu, penelitian oleh Pramesti Dewi [12] menerapkan algoritma Random Forest dengan optimasi GridSearch CV, yang berhasil meningkatkan akurasi dari 72,3% menjadi 86,1% dengan parameter optimal max\_depth=80, max\_features=3, min\_samples\_leaf=3, min\_samples\_split=8, dan n\_estimators=1000. Meskipun kedua penelitian menunjukkan hasil yang baik, masih terdapat peluang untuk meningkatkan akurasi melalui penggunaan algoritma yang lebih robust dalam menangani data kategorikal dan teknik optimasi hyperparameter yang lebih efisien.

*CatBoost* (Categorical Boosting) merupakan algoritma klasifikasi terkini yang dikembangkan oleh *Yandex* sebagai pengembangan dari algoritma *MatrixNet*. Algoritma *machine learning* ini bersifat *open source* dan dapat diimplementasikan pada berbagai domain untuk menyelesaikan beragam permasalahan. Salah satu keunggulan utama *CatBoost* terletak pada mekanisme *gradient boosting* yang efektif dalam menangani data heterogen serta mampu meningkatkan stabilitas dan performa prediksi. Metodologi ini mengadopsi teknik pengkodean yang efisien dan berpotensi meminimalkan *overfitting*. Keistimewaan lain dari *CatBoost* adalah kemampuannya memproses data kategorikal secara langsung tanpa memerlukan konversi terlebih dahulu, berbeda dengan algoritma lain yang mengharuskan transformasi data non-numerik menjadi format numerik, misalnya melalui penggunaan variabel *dummy*. Proses konversi tersebut dapat berdampak pada akurasi model, sehingga *CatBoost* dirancang untuk mengatasi keterbatasan ini dan menghasilkan tingkat akurasi yang lebih optimal [13].

Hyperparameter merupakan parameter yang ditetapkan oleh pengguna dan berpengaruh terhadap performa model. Proses hyperparameter tuning dilakukan untuk mengidentifikasi kombinasi optimal guna meningkatkan akurasi model melalui penggunaan algoritma optimasi yang bertujuan meminimalkan kesalahan generalisasi, serta membandingkan performa model pada nilai default dengan hasil tuning[14].

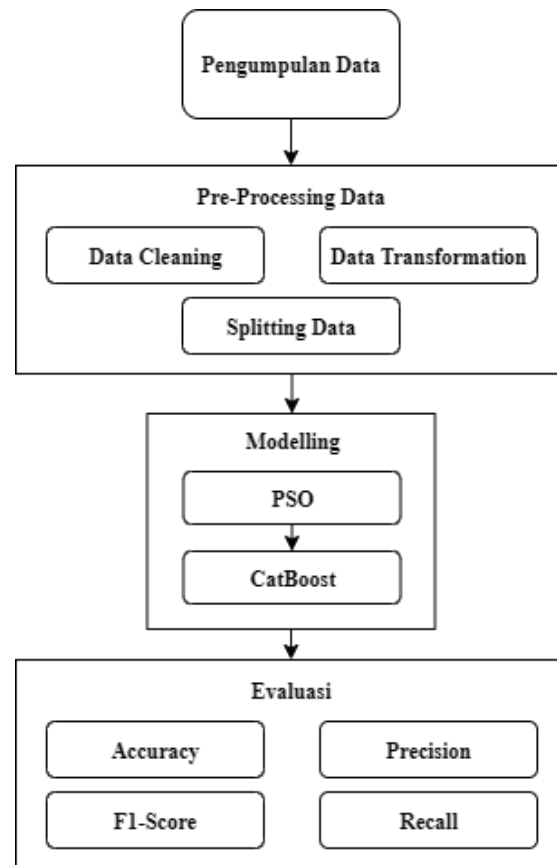
Efektivitas PSO dalam optimasi hyperparameter di bidang medis telah dibuktikan dalam berbagai penelitian. Sari et al[15], menerapkan PSO untuk mengoptimalkan parameter Naive Bayes pada klasifikasi penyakit jantung dengan dataset 303 record, menghasilkan akurasi 96,67%, precision 93,75%, dan recall 100%, menunjukkan peningkatan signifikan dibandingkan Naive Bayes tanpa optimasi PSO. Demikian pula, Alexita et al.[16] menggunakan PSO untuk mengidentifikasi hyperparameter optimal Random Forest pada klasifikasi kanker payudara berbasis citra mammogram, meningkatkan akurasi dari 86,05% menjadi 88,37%. Studi yang dilakukan oleh Demir & Sahin[17] lebih lanjut membuktikan bahwa PSO efektif digunakan untuk mengoptimalkan hyperparameter pada algoritma *CatBoost*. Studi tersebut mengidentifikasi bahwa hyperparameter yang paling krusial dan berpengaruh signifikan, yaitu *depth*, *learning\_rate*, dan *model\_shrink\_rate*, dapat ditentukan menggunakan algoritma PSO. Hasil optimasi menghasilkan model PSO-*CatBoost* dengan akurasi 95% dan Kappa statistic 90%, jauh melampaui model *CatBoost* default yang hanya mencapai akurasi 78% dan Kappa statistic 57%. Lebih lanjut, PSO-*CatBoost* juga menunjukkan peningkatan signifikan pada metrik Precision, Recall, dan F1-Score dengan nilai masing-masing 0.9426, 0.9602, dan 0.9513, yang secara substansial lebih tinggi dibandingkan *CatBoost* default. Temuan ini menegaskan pentingnya tuning hyperparameter dalam meningkatkan kinerja algoritma *machine learning* untuk klasifikasi medis.

Berdasarkan latar belakang di atas, penelitian ini memiliki tujuan untuk mengkaji dampak dari optimasi menggunakan *Particle Swarm Optimization* (PSO) terhadap kinerja algoritma *CatBoost* dalam mengklasifikasikan hipertensi. Proses optimasi dilaksanakan melalui pencarian *hyperparameter* optimal guna mengukur akurasi beserta metrik evaluasi lainnya. Dengan demikian, hasil penelitian ini diharapkan mampu memberikan sumbangsih dalam implementasi teknologi *machine learning* untuk menunjang deteksi dini serta manajemen hipertensi dengan lebih akurat dan efisien.

## 2. Metode

Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksperimen, bertujuan untuk membangun model klasifikasi tingkat hipertensi menggunakan algoritma *CatBoost* yang dioptimasi dengan metode PSO. Metode ini dipilih karena *CatBoost* mampu menangani data kategorikal secara langsung dan memberikan akurasi klasifikasi yang baik ketika dikombinasikan

dengan optimasi hyperparameter menggunakan PSO. Tahapan penelitian dapat dilihat pada Gambar 1.



Gambar 1. Metode Penelitian

Pada Gambar 1 di atas menggambarkan alur proses penelitian yang dilakukan dalam klasifikasi menggunakan kombinasi metode *PSO* dan *CatBoost*. Diagram tersebut menunjukkan tahapan-tahapan yang dilakukan secara sistematis, mulai dari pengumpulan data hingga evaluasi model. Berikut penjelasan masing-masing tahapan dari alur tersebut.

## 2.1 Pengumpulan Data

Penelitian ini menggunakan dataset gabungan dari dua sumber, yaitu data primer dari rekam medis pasien Puskesmas Kepatihan Gresik dan data sekunder dari *Kaggle*. Dataset primer terdiri dari 191 data yang mencakup informasi numerik dan kategorikal, sedangkan dataset sekunder berjumlah 12.500 data yang juga mencakup informasi serupa. Kedua dataset ini digabungkan dan dibagi menjadi data training, validating, dan testing untuk klasifikasi tingkat hipertensi.

## 2.2 Pre-Processing Data

Pre-processing data merupakan tahapan krusial dalam pengembangan model machine learning yang bertujuan untuk mempersiapkan data mentah menjadi format yang sesuai untuk proses pemodelan. Tahapan ini mencakup serangkaian proses transformasi data yang meliputi *data cleaning*, *data transformation*, dan *Splitting data*. Kualitas hasil *pre-processing* akan berpengaruh langsung terhadap performa model yang dibangun. Dalam penelitian ini, *pre-*

*processing* data dilakukan melalui tiga tahapan utama yaitu *data cleaning*, *data transformation*, dan *splitting data*.

### 2.2.1 Data Cleaning

Tahap awal dalam pre-processing data adalah data cleaning. Pada tahap ini, dilakukan penghapusan kolom yang tidak relevan untuk proses klasifikasi. Pada dataset primer, kolom Nama Pasien dihapus karena hanya berisi identitas pasien dan tidak mempengaruhi hasil prediksi. Begitu juga pada dataset sekunder, kolom id dihapus karena hanya berfungsi sebagai identifikasi pasien. Penghapusan kolom-kolom tersebut bertujuan untuk menyederhanakan data, mengurangi kompleksitas model, dan mencegah *overfitting*.

### 2.2.2 Data Transformation

Pada tahap ini, kolom Tingkat Hipertensi yang merupakan target klasifikasi perlu diubah dari bentuk kategorikal menjadi bentuk numerik agar dapat diproses oleh algoritma *machine learning*. Proses konversi ini dilakukan menggunakan metode *Label Encoding*, yaitu teknik yang mengubah setiap kategori menjadi representasi angka. Kategori Normal diubah menjadi nilai 0, Pre-Hipertensi menjadi nilai 1, Hipertensi Tingkat 1 menjadi nilai 2, dan Hipertensi Tingkat 2 menjadi nilai 3. Pengkodean ini dilakukan secara berurutan sesuai dengan tingkat keparahan kondisi hipertensi, sehingga memudahkan model dalam memahami hierarki atau urutan dari setiap kelas target.

### 2.2.3 Splitting Data

Tahap *splitting data* adalah proses membagi dataset menjadi tiga bagian, yaitu data *training* untuk melatih model, data *validation* untuk memantau kinerja model selama pelatihan, dan data *testing* untuk menguji performa model setelah dilatih. Dalam penelitian ini, pembagian data dilakukan dengan rasio 80% untuk *training*, 10% untuk *validation*, dan 10% untuk *testing*. Pembagian data ini bertujuan agar model dapat belajar dengan baik dari data training, dipantau kinerjanya melalui data validation, dan dievaluasi secara objektif menggunakan data *testing* yang belum pernah dilihat sebelumnya.

## 2.3 Modelling

Tahap modelling merupakan inti dari penelitian ini, di mana dilakukan pembangunan model klasifikasi menggunakan algoritma CatBoost yang dikombinasikan dengan metode optimasi Particle Swarm Optimization (PSO). Proses modelling ini terdiri dari dua tahapan utama, yaitu optimasi hyperparameter menggunakan PSO untuk menemukan kombinasi parameter terbaik, dan pelatihan model CatBoost menggunakan parameter optimal yang telah diperoleh. Kedua tahapan ini dirancang untuk menghasilkan model klasifikasi yang akurat dan robust dalam memprediksi tingkat hipertensi.

### 2.3.1 Optimasi Hyperparameter PSO

Optimasi hyperparameter dengan PSO dilakukan untuk meningkatkan performa model CatBoost. PSO mencari kombinasi parameter terbaik seperti *depth*, *learning rate*, dan *l2\_leaf\_reg*.

1. *Depth* mengatur kedalaman pohon keputusan yang mempengaruhi kemampuan model dalam menangkap pola data.
2. *Learning rate* mengontrol seberapa cepat model belajar.
3. *l2\_leaf\_reg* berfungsi mencegah *overfitting* dengan memberikan penalti pada ukuran pohon.

Dalam penelitian ini, proses optimasi PSO dilakukan dengan mengeksplorasi ruang pencarian hyperparameter yang telah ditentukan. Rentang nilai untuk setiap hyperparameter dipilih berdasarkan studi literatur dan karakteristik dataset yang digunakan. Tabel 1 menunjukkan konfigurasi hyperparameter yang digunakan dalam proses optimasi PSO.

**Table 1.** Konfigurasi Hyperparameter untuk Optimasi PSO

Hyperparameter	Nilai yang Diuji
Iterations	100
depth	1, 3, 6
learning_rate	0.0126, 0.055, 0.064
l2_leaf_reg	1, 3, 5

PSO menggunakan pendekatan berbasis populasi untuk mengeksplorasi berbagai kombinasi hyperparameter secara efisien dalam menemukan konfigurasi optimal. Algoritma ini akan mengevaluasi setiap kombinasi dari parameter *depth*, *learning\_rate*, dan *l2\_leaf\_reg* dalam rentang nilai yang telah ditentukan, sementara parameter *iterations* ditetapkan konstan sebesar 100 untuk menjaga konsistensi pelatihan. Proses optimasi menggunakan metrik akurasi pada data validasi sebagai fungsi *fitness* untuk memilih kombinasi terbaik. Dengan membandingkan hasil model menggunakan parameter *default* dan parameter hasil optimasi PSO, dapat dipastikan bahwa model memiliki akurasi yang lebih tinggi dan performa yang lebih stabil dalam menghadapi data baru.

### 2.3.2 Training Model CatBoost

Pada pelatihan model, digunakan model *CatBoost* untuk klasifikasi tingkat hipertensi. Model ini dilatih menggunakan optimasi hyperparameter dengan PSO. Proses pelatihan bertujuan untuk membangun model yang dapat mengenali pola-pola penting dalam data, sehingga mampu memberikan prediksi yang akurat. Setelah pelatihan, model diuji untuk memastikan kinerjanya dalam memprediksi tingkat hipertensi dengan akurasi yang tinggi.

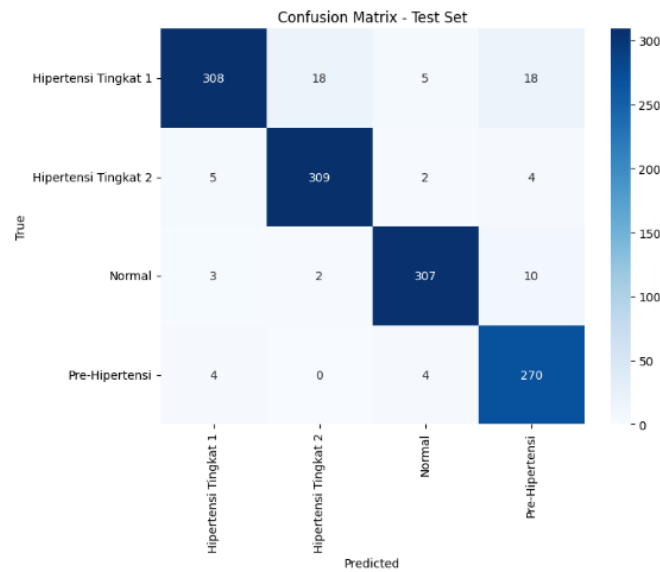
## 2.4 Evaluasi Model

Evaluasi model dilakukan untuk mengukur seberapa akurat model dalam melakukan prediksi. Pada penelitian ini, digunakan empat metrik evaluasi yaitu *Accuracy*, *Precision*, *Recall*, dan *F1-Score*. *Accuracy* mengukur jumlah prediksi yang benar. *Precision* mengukur ketepatan prediksi kelas positif. *Recall* mengukur kemampuan model mendeteksi semua data positif. *F1-Score* adalah gabungan dari *Precision* dan *Recall*. Dengan metrik-metrik ini, performa model dapat dinilai dan ditingkatkan jika diperlukan..

## 3. Hasil dan Analisis

Bab ini menyajikan hasil pengujian dan analisis performa model klasifikasi tingkat hipertensi yang telah dikembangkan. Evaluasi dilakukan terhadap dua model, yaitu model *CatBoost* tanpa optimasi dan model *CatBoost* dengan optimasi PSO. Hasil pengujian ditampilkan dalam bentuk *confusion matrix* dan *classification report* yang mencakup metrik evaluasi berupa *accuracy*, *precision*, *recall*, dan *F1-Score*. Analisis komparatif dilakukan untuk mengetahui efektivitas metode optimasi PSO dalam meningkatkan performa model *CatBoost* untuk klasifikasi tingkat hipertensi.

### 3.1 Pengujian Model CatBoost



**Gambar 2.** Confusion Matrix Model Catboost

Pada Gambar 2 menjelaskan *Confusion matrix* klasifikasi dari model *CatBoost* tanpa optimasi *PSO*. Berdasarkan hasil yang ditampilkan, untuk kelas "Hipertensi Tingkat 1" terdapat 308 data yang terklasifikasi dengan benar, kelas "Hipertensi Tingkat 2" memiliki 309 data yang benar, kelas "Normal" dengan 307 data yang benar, dan kelas "Pre-Hipertensi" dengan 270 data yang benar terklasifikasi.

```

=== Classification Report ===
              precision    recall  f1-score   support

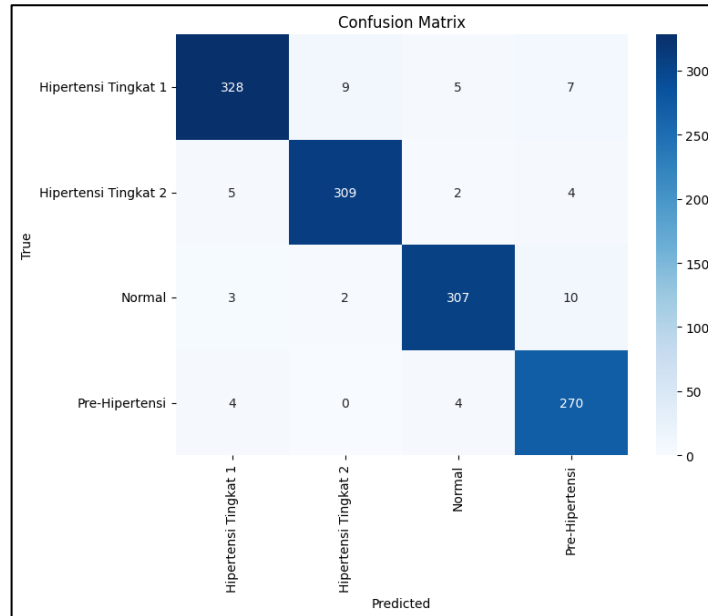
 Hipertensi Tingkat 1      0.96      0.88      0.92      349
 Hipertensi Tingkat 2      0.94      0.97      0.95      320
           Normal          0.97      0.95      0.96      322
 Pre-Hipertensi           0.89      0.97      0.93      278

 accuracy                   0.94      1269
 macro avg                  0.94      0.94      0.94      1269
 weighted avg               0.94      0.94      0.94      1269
    
```

**Gambar 3.** Classification Report Model CatBoost

Pada Gambar 3 menunjukkan *Classification report* bahwa model *CatBoost* tanpa optimasi *PSO* memiliki performa yang sangat baik di semua kelas. Kelas "Normal" mencapai *precision* 0.97, *recall* 0.95, dan *F1-Score* 0.96 dengan jumlah sampel 322. Kelas "Pre-Hipertensi" memperoleh *precision* 0.89, *recall* 0.97, dan *F1-Score* 0.93 dengan jumlah sampel 278. Kelas "Hipertensi Tingkat 1" mencatat *precision* 0.96, *recall* 0.88, dan *F1-Score* 0.92 dengan jumlah sampel 349. Sedangkan kelas "Hipertensi Tingkat 2" mencapai *precision* 0.94, *recall* 0.97, dan *F1-Score* 0.95 dengan jumlah sampel 320. Secara keseluruhan, model memperoleh akurasi sebesar 0.94 dengan nilai *macro average* dan *weighted average* yang juga mencapai 0.94.

### 3.2 Pengujian Model CatBoost + PSO



**Gambar 4.** Confusion Matrix Model CatBoost + PSO

Berdasarkan *confusion matrix* pada Gambar 4, model menghasilkan 1.225 data yang terklasifikasi dengan benar dan 26 data yang salah klasifikasi. Untuk kelas "Normal", terdapat 307 data yang terklasifikasi dengan benar sebagai *True Positive* (TP). Kelas "Pre-Hipertensi" memiliki 270 data yang terklasifikasi dengan benar sebagai TP. Kelas "Hipertensi Tingkat 1" mencatat 328 data yang terklasifikasi dengan benar sebagai TP. Sedangkan kelas "Hipertensi Tingkat 2" memiliki 309 data yang terklasifikasi dengan benar sebagai TP.

```

=== Classification Report (Testing Data) ===
              precision    recall  f1-score   support

Hipertensi Tingkat 1      0.96      0.94      0.95        349
Hipertensi Tingkat 2      0.97      0.97      0.97        320
      Normal              0.97      0.95      0.96        322
      Pre-Hipertensi       0.93      0.97      0.95        278

   accuracy                   0.96       1269
  macro avg              0.96      0.96      0.96       1269
 weighted avg            0.96      0.96      0.96       1269
    
```

**Gambar 5.** Classification Report Model CatBoost + PSO

Gambar 5 menunjukkan bahwa model memiliki performa yang sangat baik di semua kelas. Kelas "Normal" mencapai *precision* 0.97, *recall* 0.95, dan *F1-Score* 0.96 dengan jumlah sampel 322. Kelas "Pre-Hipertensi" memperoleh *precision* 0.93, *recall* 0.97, dan *F1-Score* 0.95 dengan jumlah sampel 278. Kelas "Hipertensi Tingkat 1" mencatat *precision* 0.96, *recall* 0.94, dan *F1-Score* 0.95 dengan jumlah sampel 349. Sedangkan kelas "Hipertensi Tingkat 2" mencapai *precision* 0.97, *recall* 0.97, dan *F1-Score* 0.97 dengan jumlah sampel 320. Secara keseluruhan, model memperoleh akurasi sebesar 0.96 dengan nilai *macro average* dan *weighted average* yang juga mencapai 0.96. Hasil ini menunjukkan bahwa penggunaan optimasi *PSO* berhasil meningkatkan akurasi model dari 94% menjadi 96%, mencerminkan kinerja yang sangat baik dalam klasifikasi.

### 3.3 Analisis Pengujian

Berdasarkan hasil pengujian yang telah dilakukan, maka hasil pengujian tersebut akan disatukan dan diambil nilai akurasi yang paling baik. Kumpulan nilai akurasi dari pengujian sebelumnya ditunjukkan pada Tabel 2. Dari kumpulan nilai akurasi pengujian tersebut, nilai akurasi terbaik diperoleh pada model *CatBoost* dengan optimasi *PSO* menggunakan pembagian data 80:10:10 untuk *training*, *validation*, dan *testing*. Model *CatBoost* tanpa optimasi *PSO* menghasilkan akurasi sebesar 94%, sedangkan setelah dilakukan optimasi hyperparameter menggunakan *PSO*, akurasi model meningkat menjadi 96%. Peningkatan sebesar 2% ini menunjukkan bahwa optimasi *PSO* efektif dalam meningkatkan performa model *CatBoost* untuk klasifikasi tingkat hipertensi

**Table 2.** Perbandingan Metrik Evaluasi Model *CatBoost* Sebelum dan Sesudah Optimasi *PSO*

Model	Accuracy	Precision	Recall	F1-Score	Hyperparameter	Keterangan
CatBoost tanpa <i>PSO</i>	94%	94%	94%	94%	Default	Parameter default
CatBoost dengan <i>PSO</i>	96%	96%	96%	96%	<i>iterations</i> =100 <i>depth</i> =3 <i>learning_rate</i> =0.055 <i>l2_leaf_reg</i> =3	Hyperparameter Optimal

Pada Tabel 2 di atas dapat disimpulkan bahwa model yang menghasilkan nilai akurasi paling baik adalah model *CatBoost* dengan optimasi *PSO* menggunakan pembagian data 80:10:10, dengan akurasi sebesar 96%. Model *CatBoost* tanpa optimasi *PSO* dengan parameter yang sama menghasilkan nilai akurasi sebesar 94% yang ditunjukkan pada Gambar 3. Dimana terdapat 26 data yang salah klasifikasi dari total 1.269 data yang diuji, yang ditunjukkan pada confusion matrix di Gambar 2. Setelah dilakukan optimasi hyperparameter menggunakan *PSO*, akurasi model meningkat menjadi 96% yang ditunjukkan pada Gambar 5, dengan jumlah kesalahan klasifikasi yang berkurang.

Proses optimasi *PSO* menghasilkan konfigurasi hyperparameter terbaik dengan nilai *iterations* = 100, *depth* = 3, *learning\_rate* = 0.055, dan *l2\_leaf\_reg* = 3. Kombinasi parameter optimal ini diperoleh setelah *PSO* melakukan eksplorasi terhadap berbagai kombinasi nilai dalam rentang yang telah ditentukan pada Tabel 1. Parameter *depth* = 3 menunjukkan kedalaman pohon yang optimal untuk menangkap pola data tanpa mengalami *overfitting*, sementara *learning\_rate* = 0.055 memberikan kecepatan pembelajaran yang seimbang. Nilai *l2\_leaf\_reg* = 3 berfungsi sebagai regularisasi yang efektif untuk mencegah model terlalu kompleks.

Berdasarkan analisis tersebut, perbandingan antara model *CatBoost* tanpa *PSO* dan model *CatBoost* dengan optimasi *PSO* menggunakan parameter yang sama menunjukkan bahwa nilai akurasi yang didapatkan lebih baik pada model dengan optimasi *PSO* daripada tanpa optimasi. Maka dari itu, metode optimasi *PSO* berhasil meningkatkan nilai akurasi model *CatBoost* dari 94% menjadi 96%, atau meningkat sebesar 2%.

#### 4. Simpulan

Pengujian optimasi *hyperparameter* menggunakan *Particle Swarm Optimization* pada model *CatBoost* untuk klasifikasi tingkat hipertensi dilakukan melalui beberapa percobaan dengan kombinasi parameter yang berbeda. Hasil optimasi *PSO* mengidentifikasi konfigurasi hyperparameter terbaik dengan nilai *iterations* = 100, *depth* = 3, *learning\_rate* = 0.055, dan *l2\_leaf\_reg* = 3. Tujuan dari percobaan ini adalah untuk menemukan konfigurasi *hyperparameter* terbaik yang menghasilkan akurasi klasifikasi tertinggi. Berdasarkan hasil pengujian, diperoleh bahwa model *CatBoost* dengan optimasi *PSO* memberikan performa terbaik dengan nilai akurasi

sebesar 96%, di mana dari total 1.269 data uji, model berhasil mengklasifikasikan dengan tingkat akurasi yang sangat baik pada semua kelas tingkat hipertensi.

Adapun hasil perbandingan antara model CatBoost dengan optimasi PSO dan CatBoost tanpa optimasi pada parameter pembagian data yang sama, yaitu 80:10:10, menunjukkan bahwa *CatBoost* dengan *PSO* menghasilkan akurasi lebih tinggi sebesar 96%, dibandingkan dengan *CatBoost* tanpa optimasi yang hanya memperoleh akurasi sebesar 94%. Hal ini membuktikan bahwa optimasi hyperparameter menggunakan *PSO*, yang mencakup pencarian nilai optimal untuk parameter *depth*, *learning rate*, dan *l2\_leaf\_reg*, mampu meningkatkan akurasi klasifikasi sebesar 2% dibandingkan penggunaan parameter *default*. Dengan demikian, kombinasi *CatBoost* dan *PSO* dapat dianggap sebagai metode yang efektif dan lebih baik dalam mengklasifikasikan tingkat hipertensi pada dataset yang digunakan dalam penelitian ini.

## Referensi

- [1] Herwati and W. Sartika, "Terkontrolnya Tekanan Darah Penderita Hipertensi Berdasarkan Polidiet Dankebiasaanolahraga Dipadangtahun 2011," *Jurnal Kesehatan Masyarakat Andalas*, vol. 8, no. 1, pp. 8–14, Sep. 2014.
- [2] E. Triyanto, *Pelayanan keperawatan bagi penderita hipertensi secara terpadu*. Yogyakarta: Graha Ilmu, 2014.
- [3] B. Nuraini, "Risk Factors Of Hypertension," *J MAJORITY*, vol. 4, no. 10, pp. 10–19, 2015.
- [4] Y. Tri Gesela Arum, "Hipertensi pada Penduduk Usia Produktif (15-64 Tahun)," *HIGEIA Journal of Public Health Research and Development*, vol. 3, pp. 346–358, 2019, doi: 10.15294/higeia/v3i3/30235.
- [5] M. Yogiantoro, *Hipertensi Esensial Ilmu Penyakit Dalam*, 5th ed. Jakarta: Interna Publishing, 2009.
- [6] Kementerian Kesehatan Republik Indonesia, "Profil Kesehatan Indonesia 2019," Kementerian Kesehatan Republik Indonesia, Jakarta, 2020.
- [7] World Health Organization, "A Global Brief on Hypertension: Silent Killer, Global Public Health Crisis," Geneva, 2013.
- [8] Kementerian Kesehatan Republik Indonesia, "Pedoman Teknis Penemuan dan Tatalaksana Penyakit Hipertensi," Jakarta, 2014.
- [9] R. H. Fletcher, S. W. Fletcher, and E. H. Wagner, *Clinical Epidemiology: The Essentials*, 4th ed. Philadelphia: Lippincott Williams & Wilkins, 2005.
- [10] Y. Susilo, A. Wulandari, and Westriningsih, "Cara Jitu Mengatasi Darah Tinggi (Hipertensi)," pp. 187–194, 2011.
- [11] P. Purwono, P. Dewi, S. K. Wibisono, and B. Putra Dewa, "Model Prediksi Otomatis Jenis Penyakit Hipertensi dengan Pemanfaatan Algoritma Machine Learning Artificial Neural Network," *Jurnal Teknik Informatika*, vol. 7, no. 2, pp. 82–90, Mar. 2022, doi: 10.33506/insect.v7i2.1828.
- [12] P. Dewi, P. Purwono, and S. D. Kurniawan, "Pemanfaatan Teknologi Machine Learning pada Klasifikasi Jenis Hipertensi Berdasarkan Fitur Pribadi," *Smart Comp: Jurnalnya Orang Pintar Komputer*, vol. 11, pp. 172–176, Jul. 2022.

- [13] N. K. Dewi, "Deteksi Fake Follower Instagram Menggunakan Catboost Classifier," UIN Sunan Kalijaga Yogyakarta, Yogyakarta, 2021.
- [14] E. Elgeldawi, A. Sayed, A. R. Galal, and A. M. Zaki, "Hyperparameter tuning for machine learning algorithms used for arabic sentiment analysis," *Informatics*, vol. 8, no. 4, p. 79, Dec. 2021, doi: 10.3390/informatics8040079.
- [15] A. Novita Sari, Z. Abidin, and Bijanto, "Optimasi Algoritma Naive Bayes Dengan Particle Swarm Optimization (Pso) Untuk Klasifikasi Penyakit Jantung," *Jurnal Edukasi Elektromatika*, vol. 6, no. 1, Jun. 2025.
- [16] A. C. Salwa Alexita, P. Kusumaningtyas, and M. Rofi'i, "Optimasi Algoritma Random Forest Menggunakan Pso Untuk Klasifikasi Kanker Payudara Dengan Citra Mammograms," *Teknika STTKD: Jurnal Teknik, Elektronik, Engine*, vol. 11, no. 1, pp. 47–54, Feb. 2025, doi: 10.56521/teknika.v11i1.1346.
- [17] S. Demir and E. K. Sahin, "Predicting occurrence of liquefaction-induced lateral spreading using gradient boosting algorithms integrated with particle swarm optimization: PSO-XGBoost, PSO-LightGBM, and PSO-CatBoost," *Acta Geotech*, vol. 18, no. 6, pp. 3403–3419, Jun. 2023, doi: 10.1007/s11440-022-01777-1.