

BAB I

PENDAHULUAN

1.1 Latar Belakang

Hipertensi adalah penyakit kardiovaskular yang ditandai dengan peningkatan abnormal pada tekanan darah sistolik maupun diastolik, yang dapat menyebabkan kerusakan serius pada berbagai organ tubuh, termasuk mata, jantung, otak, dan ginjal. Tekanan darah normal adalah 120/80 mmHg, dan seseorang dikatakan menderita hipertensi jika tekanan darahnya lebih dari 140/90 mmHg [1]. Kondisi ini sering disebut "silent killer" karena seringkali tidak menunjukkan gejala, meskipun dapat menyebabkan komplikasi serius seperti stroke, pengerasan pembuluh darah (atherosclerosis), dan gagal ginjal [2]. Faktor penyebab hipertensi terbagi menjadi dua kategori, yaitu faktor genetik, seperti riwayat keluarga dengan hipertensi, dan faktor non-genetik, seperti usia, jenis kelamin, kebiasaan merokok, konsumsi garam berlebih, pola makan tinggi kolesterol, konsumsi alkohol, obesitas, stres, serta kurangnya aktivitas fisik [3].

Hipertensi menjadi penyebab kematian tertinggi ke-3 di dunia, dengan prevalensi global mencapai 1,13 miliar orang, dan 31% di antaranya berisiko pada orang dewasa [4]. Di Asia, tercatat 38,4 juta penderita hipertensi pada tahun 2000, diprediksi meningkat menjadi 67,4 juta pada tahun 2025, dengan Asia Tenggara memiliki prevalensi 25% dari total penduduk [5]. Di Jawa Timur pada 2019, diperkirakan ada 11.952.694 orang berusia ≥ 15 tahun yang mengidap hipertensi, dengan proporsi 48% pria dan 52% wanita, namun hanya 40,1% yang mendapatkan layanan kesehatan [6]. Riset Kesehatan Dasar Indonesia, menunjukkan hipertensi perlu mendapat perhatian serius karena meningkatnya jumlah penderita dan angka kematian terkait. Pengobatan hipertensi bersifat jangka panjang, sementara deteksi manual yang masih digunakan membuat deteksi dini sangat diperlukan untuk meningkatkan perawatan kesehatan. [7]. Akses terbatas dan rendahnya pengetahuan masyarakat tentang hipertensi menjadi tantangan utama dalam penanganannya. Oleh karena itu, dibutuhkan sistem cerdas yang dapat mengklasifikasikan hipertensi secara dini berdasarkan kondisi pasien, membantu

petugas medis dalam pencegahan dini, dan meningkatkan efektivitas pengobatan serta perawatan kesehatan.

CatBoost (Categorical Boosting) adalah metode klasifikasi terbaru yang dikembangkan oleh *Yandex*, sebagai penerus algoritma *MatrixNet*. Metode machine learning ini tersedia secara *open source* dan dapat diterapkan di berbagai bidang untuk berbagai masalah. Salah satu keunikan *Catboost* adalah mekanisme *gradient boosting* yang cocok untuk bekerja dengan data heterogen dan dapat meningkatkan stabilitas serta kemampuan prediksi yang baik. Metodelogi ini menggunakan pengkodean yang efisien dan dapat mengurangi *overfitting*. Selain itu, *CatBoost* dapat secara langsung memperhitungkan data non-numerik, berbeda dengan metode lain dimana data-data non-numerik harus terlebih dahulu diterjemahkan ke dalam bahasa angka yaitu salah satunya dengan menggunakan variabel *dummy*. Hal tersebut dapat mempengaruhi keakuratan model, sehingga *CatBoost* diciptakan untuk dapat memperbaiki kekurangan tersebut dan meningkatkan akurasi menjadi lebih baik [8].

Seleksi fitur penting sebelum klasifikasi untuk menghilangkan fitur yang tidak relevan atau redundant, karena jumlah fitur yang besar dapat meningkatkan kerumitan data dan menjadi masalah dalam klasifikasi [9]. *Particle Swarm Optimization* (PSO) adalah algoritma metaheuristik yang terinspirasi oleh perilaku kolektif kawanan burung dan ikan, diperkenalkan oleh Eberhart dan Kennedy pada 1995. PSO digunakan untuk menemukan solusi optimal, dengan partikel yang bergerak melalui ruang pencarian berdasarkan pengalaman terbaik pribadi dan kawanannya. Algoritma ini diterapkan di berbagai bidang, termasuk teknik, ekonomi, dan ilmu komputer.[10]. PSO dapat digunakan untuk mengoptimalkan parameter, meningkatkan bobot atribut, serta melakukan seleksi atribut dan fitur, guna meningkatkan kinerja model, terutama dalam aplikasi klasifikasi [11]. Penelitian yang dilakukan oleh Mayadi et al. [12], menunjukkan bahwa PSO dapat digunakan untuk seleksi fitur dalam meningkatkan kinerja algoritma *ensemble classifier*. Penelitian tersebut menggunakan empat algoritma *ensemble classifier*, yaitu *Random Forest (RF)*, *XGBOOST*, *CatBoost*, dan *LightGBM*, yang awalnya memiliki akurasi 92.68%. Setelah dilakukan optimasi menggunakan PSO dengan 50 partikel dan 35 dimensi, fitur yang digunakan berkurang dari 35 menjadi 8.

Hasilnya, akurasi masing-masing algoritma meningkat sebesar 7.32%, sehingga mencapai akurasi 100%.

Hyperparameter adalah parameter yang ditentukan oleh pengguna dan mempengaruhi kinerja model. Proses *hyperparameter* tuning bertujuan menemukan kombinasi optimal untuk meningkatkan akurasi model, dengan menggunakan algoritma optimasi untuk meminimalkan kesalahan generalisasi dan membandingkan kinerja model pada nilai default dan hasil tuning [13]. Penelitian yang telah dilakukan oleh Demir & Sahin [14], membuktikan bahwa *PSO* dapat digunakan untuk mengoptimalkan *hyperparameter* pada algoritma *CatBoost*. Penelitian tersebut menunjukkan bahwa *hyperparameter* paling penting dan berpengaruh, yaitu *depth*, *learning_rate*, dan *model_shrink_rate*, ditentukan menggunakan algoritma *PSO*. Hasil optimasi ini menghasilkan model *PSO-CatBoost* yang mencapai akurasi 95% dan *Kappa statistic* 90%, dibandingkan dengan model *CatBoost default* yang hanya mencapai akurasi 78% dan *Kappa statistic* 57%. Selain itu, *PSO-CatBoost* juga menunjukkan peningkatan pada nilai *Precision*, *Recall*, dan *F1-Score* masing-masing sebesar 0.9426, 0.9602, dan 0.9513, yang jauh lebih tinggi dibandingkan *CatBoost default*. Hasil ini mengungkapkan pentingnya *tuning hyperparameter* untuk meningkatkan kinerja algoritma *CatBoost*.

Oleh karena beberapa latar belakang di atas, maka pada penelitian ini, penulis berencana untuk melakukan penelitian guna mengetahui pengaruh optimasi metode *Particle Swarm Optimization (PSO)* terhadap nilai akurasi algoritma *CatBoost* untuk klasifikasi hipertensi. Penelitian ini akan menggunakan dataset hipertensi yang berisi data riwayat medis. Data akan dibagi menjadi beberapa kategori untuk mendukung proses klasifikasi. *Particle Swarm Optimization (PSO)* nantinya akan digunakan untuk seleksi fitur dan juga *hyperparameter tuning* pada algoritma *CatBoost*, serta terdapat beberapa skenario pengujian untuk pembagian data latih dan data uji.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, maka perumusan masalah yang dapat diambil adalah sebagai berikut :

1. Apakah *Particle Swarm Optimization* dapat diimplementasikan sebagai seleksi fitur dan *hyperparameter tuning* untuk meningkatkan nilai akurasi pada *CatBoost* pada klasifikasi tingkat hipertensi?
2. Bagaimana pengaruh penggunaan *Particle Swarm Optimization* terhadap kinerja akurasi model *CatBoost* dalam klasifikasi tingkat hipertensi?

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah disampaikan, maka tujuan dari penelitian ini diantaranya :

1. Mengimplementasikan algoritma *Particle Swarm Optimization* sebagai *hyperparameter tuning* dan seleksi fitur untuk meningkatkan nilai akurasi pada *CatBoost* dalam klasifikasi tingkat hipertensi.
2. Mengetahui pengaruh penggunaan *Particle Swarm Optimization* terhadap peningkatan akurasi model CatBoost dalam klasifikasi tingkat hipertensi.

1.4 Manfaat Penelitian

Manfaat yang didapatkan dari penelitian yang dilakukan diantaranya :

1. Menghasilkan model klasifikasi yang akurat untuk mendeteksi risiko hipertensi, sehingga dapat mempercepat proses analisis data pasien. Sehingga tenaga medis dapat lebih fokus pada penanganan dan pengambilan keputusan klinis dibandingkan melakukan analisis secara manual.
2. Menjadi landasan untuk pengembangan sistem diagnosa otomatis atau aplikasi kesehatan yang terintegrasi dengan rekam medis digital, khususnya dalam mendeteksi dan memantau kondisi hipertensi pada pasien.

1.5 Batasan Masalah

Batasan-batasan masalah yang ditemukan dari penelitian ini, antara lain:

1. Dataset yang digunakan dalam penelitian ini terdiri dari dua jenis data, yaitu data primer dan data sekunder. Data primer diperoleh dari Puskesmas Kepatihan Gresik, dengan jumlah 191 data dan 12 fitur. Sementara itu, data sekunder diperoleh dari Kaggle, yang berjumlah 12.500 data dan juga memiliki 12 fitur.

2. Model yang digunakan dalam penelitian ini adalah *CatBoost*, yang akan dikombinasikan dengan *PSO* untuk optimasi *hyperparameter* dan seleksi fitur.
3. Penelitian ini membagi hasil klasifikasi menjadi empat kategori, yaitu Normal, Pre-Hipertensi, Hipertensi 1, dan Hipertensi 2.
4. Hasil dari penelitian ini merupakan akurasi model *CatBoost* setelah diterapkan seleksi fitur dan optimasi *hyperparameter* menggunakan *Particle Swarm Optimization (PSO)* dalam klasifikasi tingkat hipertensi.