

BAB I

PENDAHULUAN

1.1. Latar Belakang Masalah

Aritmia merupakan gangguan irama jantung yang dapat berdampak serius, terutama jika tidak terdeteksi secara dini [1]. Pada usia remaja, aritmia sering kali tidak menunjukkan gejala yang jelas, sehingga diperlukan metode diagnostik berbasis data yang mampu mendeteksi pola-pola dalam sinyal elektrokardiogram (ECG). Dalam beberapa tahun terakhir, pemanfaatan pendekatan berbasis machine learning semakin luas digunakan untuk meningkatkan ketepatan dalam klasifikasi aritmia. Berbagai penelitian sebelumnya menunjukkan bahwa model seperti Support Vector Machine (SVM) dan Random Forest mampu memberikan performa yang cukup baik [2] [3].

Namun, salah satu tantangan utama dalam klasifikasi aritmia adalah kompleksitas pola data serta ketidakseimbangan kelas. Sinyal ECG biasanya bersifat nonlinier dan sangat bervariasi antar individu, sehingga model sederhana sering kali tidak mampu menangkap seluruh informasi penting dalam data [4]. Selain itu, jumlah data pada kelas aritmia umumnya lebih sedikit dibandingkan kelas normal, sehingga model sering kali menjadi cenderung bias terhadap kelas mayoritas [5].

Untuk mengatasi tantangan tersebut, metode ensemble seperti stacked generalization (stacking) dapat digunakan sebagai solusi untuk meningkatkan akurasi serta kemampuan model dalam mengenali data baru. Pendekatan stacking memungkinkan penggabungan beberapa model pembelajar dasar (base learners) sehingga menghasilkan prediksi yang lebih optimal [6]. Dalam konteks ini, penggabungan CatBoost dengan model seperti SVM atau Random Forest, baik melalui pendekatan *stacking* maupun *voting*, berpotensi menghasilkan kinerja klasifikasi yang lebih optimal. SVM unggul dalam membedakan kelas dengan margin maksimal, sedangkan Random Forest mampu menangani data dengan noise dan fitur yang kompleks [7].

Model CatBoost yang merupakan algoritma *gradient boosting* berbasis pohon keputusan mulai banyak digunakan karena kemampuannya dalam menangani fitur kategorikal serta performa tinggi pada dataset yang kompleks [8]. CatBoost juga mengimplementasikan teknik regularisasi yang membuatnya lebih stabil terhadap overfitting dibandingkan algoritma boosting lainnya [9]. Meskipun CatBoost telah terbukti unggul dalam beberapa studi klasifikasi, integrasinya dalam kerangka stacking bersama SVM dan Random Forest untuk klasifikasi aritmia, khususnya pada remaja, masih belum banyak dieksplorasi [10][11].

Di Indonesia, penyakit jantung termasuk salah satu penyebab utama kematian. Data dari Riset Kesehatan Dasar. Penyakit jantung mencapai lebih dari 1,5% dari total populasi, dan angkanya terus mengalami peningkatan setiap tahunnya [12]. Meskipun aritmia pada remaja belum banyak dilaporkan secara spesifik, faktor risiko seperti stres, kelelahan fisik, konsumsi makanan tinggi gula dan lemak, serta penggunaan gadget yang berlebihan dapat mempengaruhi kesehatan jantung remaja secara tidak langsung [13][14].

Selain itu, banyak studi yang menggunakan data klinis dari rumah sakit, yang

mengakibatkan keterbatasan penelitian berbasis populasi non-klinis seperti pelajar atau remaja [15]. Oleh sebab itu, penelitian ini dilakukan untuk mengevaluasi efektivitas pendekatan stacked generalization dengan mengombinasikan model CatBoost, Support Vector Machine (SVM), dan Random Forest dalam klasifikasi aritmia pada remaja.

1.2. Rumusan Masalah

Berdasarkan permasalahan diatas, maka rumusan masalah penelitian ini adalah:

1. Bagaimana performa algoritma CatBoost secara individual dalam melakukan klasifikasi aritmia pada data ECG remaja?
2. Bagaimana pengaruh penerapan metode stacked generalization dengan integrasi CatBoost–SVM dan CatBoost–Random Forest terhadap hasil klasifikasi aritmia?
3. Bagaimana hasil klasifikasi aritmia menggunakan pendekatan voting ensemble antara model CatBoost–SVM dan CatBoost–Random Forest?

Pendekatan mana yang menghasilkan kinerja paling optimal dalam klasifikasi aritmia jika ditinjau dari metrik evaluasi seperti akurasi, F1-score, dan log loss?

1.3. Batasan Masalah

Batasan masalah dari penelitian ini adalah:

1. Penelitian ini dilakukan di Sekolah Menengah Kejuruan (SMK) Swasta Teladan 2 yang berlokasi di Kota Medan
2. Subjek dalam penelitian ini berjumlah 101 siswa dan siswi yang berada pada rentang usia 15 hingga 19 tahun
3. Pengumpulan data dilakukan dalam kondisi istirahat sambil duduk tenang tanpa aktivitas
4. Pengumpulan data dilakukan menggunakan perangkat Smart Holter ECG 5-lead sebanyak 3 unit yang terintegrasi dengan Raspberry Pi 4 dan kabel elektroda standar. Menggunakan bahasa pemrograman Python pada Visual Studio Code, dengan penerapan algoritma CatBoost, serta integrasinya bersama SVM dan Random Forest dalam skema stacked generalization dan voting ensemble.

1.4. Tujuan Penelitian

Tujuan dari penelitian ini adalah sebagai berikut:

1. Penelitian ini bertujuan untuk mengembangkan model klasifikasi aritmia dengan menggunakan algoritma CatBoost, baik secara mandiri maupun dikombinasikan dengan SVM dan Random Forest
2. Mengevaluasi dan membandingkan kinerja metode ensemble stacking dan voting dalam klasifikasi aritmia berdasarkan metrik akurasi, F1-score, dan log loss. Menentukan pendekatan machine learning yang paling sesuai untuk klasifikasi aritmia berdasarkan data ECG remaja.

1.5. Manfaat Penelitian

Manfaat yang diperoleh dari penelitian ini adalah:

1. Penelitian ini diharapkan dapat memberikan kontribusi terhadap pengembangan sistem pendukung diagnosis aritmia secara otomatis dengan memanfaatkan algoritma machine learning.
2. Perbandingan antara metode ensemble seperti stacking dan voting berbasis kombinasi CatBoost dengan SVM dan Random Forest.
3. Mengidentifikasi pendekatan algoritmik yang paling efektif dan efisien dalam mengklasifikasikan sinyal ECG pada populasi remaja.