

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang Masalah

*Sepsis* merupakan kondisi medis serius yang ditandai oleh respons *imun* ekstrem tubuh terhadap infeksi, yang dapat menyebabkan kegagalan organ dan kematian jika tidak ditangani secara cepat dan tepat [1]. Deteksi dini terhadap *sepsis* menjadi sangat krusial karena semakin awal *intervensi* diberikan, semakin besar peluang pasien untuk selamat [2]. Sayangnya, gejala *sepsis* awal sering kali tidak spesifik dan menyerupai kondisi medis lain, sehingga menyulitkan proses diagnosis secara konvensional.

Perkembangan *machine learning* dalam bidang medis membuka peluang baru untuk melakukan prediksi *sepsis* dini berdasarkan data klinis pasien. Salah satu dataset yang digunakan secara luas adalah dataset *Prediction of Sepsis* dari PhysioNet Computing in Cardiology Challenge 2019 yang tersedia di Kaggle. Dataset ini memuat rekaman waktu nyata dari tanda-tanda vital pasien seperti *heart rate*, *oxygen saturation (O2Sat)*, tekanan darah sistolik dan diastolik, suhu tubuh, serta data tambahan lainnya. Penggunaan metode *ensemble learning* seperti *Random Forest*, *Gradient Boosting*, *XGBoost*, dan *AdaBoost* terbukti dapat meningkatkan *akurasi* prediksi dibandingkan model tunggal [3][4][5]. Oleh karena itu, penelitian ini bertujuan untuk menerapkan dan mengevaluasi kinerja berbagai algoritma *ensemble* dalam mendeteksi *sepsis* sejak dini berdasarkan data tanda-tanda vital pasien.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, rumusan masalah dalam penelitian ini adalah:

1. Bagaimana penerapan metode *ensemble learning* seperti *Random Forest*, *XGBoost*, *Gradient Boosting*, dan *AdaBoost* dalam memprediksi sepsis berdasarkan data tanda vital pasien?
2. Algoritma *ensemble learning* mana yang memberikan kinerja terbaik dalam hal akurasi, sensitivitas, dan spesifisitas dalam mendeteksi sepsis dini?
3. Bagaimana proses *preprocessing* seperti penanganan data hilang dan normalisasi dilakukan untuk mempersiapkan dataset sepsis sebelum pelatihan model dilakukan?
4. Bagaimana hasil prediksi dapat diinterpretasikan secara transparan menggunakan metode interpretabilitas model seperti *SHAP* atau *LIME* agar dapat dimanfaatkan oleh tenaga medis dalam pengambilan keputusan klinis?

## 1.3 Batasan Masalah

1. Dataset yang digunakan adalah *Prediction of Sepsis* dari PhysioNet Challenge 2019 yang diperoleh melalui Kaggle.
2. Fitur yang dianalisis terbatas pada tanda-tanda vital dan parameter klinis yang tersedia dalam dataset.

3. Metode pembelajaran yang diuji terbatas pada algoritma *ensemble learning* seperti *Random Forest*, *XGBoost*, *Gradient Boosting*, dan *AdaBoost*.
4. Evaluasi performa model difokuskan pada metrik akurasi, sensitivitas, dan spesifisitas.

#### 1.4 Tujuan Penelitian

1. Mengimplementasikan metode *ensemble learning* untuk prediksi dini *sepsis* berdasarkan data tanda *vital* pasien.
2. Mengevaluasi dan membandingkan kinerja beberapa *algoritma ensemble* dalam prediksi *sepsis*.
3. Melakukan *preprocessing* dan penanganan data hilang pada *dataset sepsis* secara *sistematis*.
4. Menganalisis *interpretabilitas* model untuk memahami kontribusi masing-masing *fitur* dalam proses *prediksi*.

#### 1.5 Manfaat Penelitian

1. Memberikan kontribusi pada pengembangan sistem deteksi dini *sepsis* berbasis data dengan pendekatan *ensemble learning* yang dapat membantu *klinisi* dalam pengambilan keputusan cepat.
2. Menyediakan evaluasi perbandingan performa berbagai model *ensemble* dalam prediksi *sepsis*, yang dapat menjadi *referensi* untuk penelitian lanjutan.

3. Menawarkan pendekatan *interpretabilitas* yang transparan agar hasil prediksi lebih dapat dipercaya dan dimengerti oleh tenaga medis.
4. Menambah literatur dan pengetahuan terkait penerapan *machine learning* dalam bidang prediksi penyakit kritis di bidang kedokteran.

## 1.6 Sistematika Penulisan

Laporan penelitian ini disusun dengan mengikuti kaidah penulisan ilmiah yang baik dan benar, yang dibagi dalam bab-bab sebagai berikut:

### **BAB I : PENDAHULUAN**

Bab ini berisi tentang latar belakang permasalahan, perumusan masalah, pembatasan dan ruang lingkup masalah, tujuan dan manfaat penelitian, serta sistematika penulisan. Bab ini memberikan gambaran umum mengenai penelitian yang dilakukan serta alasan mengapa penelitian ini penting dilakukan.

### **BAB II : LANDASAN TEORI**

Bab ini menguraikan teori-teori yang mendasari penelitian, termasuk penjelasan tentang sepsis sebagai kondisi medis, serta pembahasan algoritma *ensemble learning* seperti *Random Forest*, *AdaBoost*, *XGBoost*, dan *Gradient Boosting* dalam konteks prediksi medis. Selain itu, bab ini juga membahas studi terdahulu yang relevan untuk memberikan gambaran tentang penelitian sejenis dalam bidang prediksi penyakit menggunakan *machine learning*.

### **BAB III : METODOLOGI PENELITIAN**

Bab ini menjelaskan secara rinci metode yang digunakan dalam penelitian ini, termasuk metode pengumpulan data, deskripsi dataset yang digunakan, serta proses pengembangan model prediksi sepsis menggunakan algoritma *ensemble*. Bab ini juga menjelaskan perangkat lunak yang digunakan, teknik pemrograman, serta tahapan *preprocessing* dan implementasi model.

### **BAB IV : ANALISIS DAN HASIL**

Bab ini membahas proses implementasi algoritma *ensemble learning* untuk prediksi sepsis, mencakup tahap evaluasi performa model menggunakan metrik seperti akurasi, sensitivitas, spesifisitas, serta analisis *confusion matrix*. Hasil yang diperoleh dianalisis secara mendalam, termasuk pembahasan kelebihan dan keterbatasan tiap algoritma yang digunakan.

### **BAB V : PENUTUP**

Bab ini menyajikan kesimpulan dari penelitian yang telah dilakukan, temuan-temuan utama, serta saran-saran untuk pengembangan penelitian selanjutnya. Bab ini juga memberikan rekomendasi terhadap pemanfaatan model prediksi sepsis berbasis *ensemble learning* di dunia medis.