

# BAB I PENDAHULUAN

## 1.1 Latar Belakang

Penyakit jantung merupakan salah satu permasalahan kesehatan yang signifikan secara global. Berdasarkan data dari *World Health Organization* (WHO), penyakit jantung menjadi penyebab kematian nomor satu di dunia. Identifikasi dan klasifikasi penyakit jantung dengan akurasi tinggi merupakan langkah penting dalam diagnosis dini, pengobatan yang tepat, serta pencegahan yang efektif[1].

Dalam beberapa tahun terakhir, metode *machine learning* telah menjadi fokus penelitian yang kuat dalam bidang klasifikasi penyakit jantung[1]. Salah satu metode yang telah terbukti efektif dalam berbagai aplikasi adalah eXtreme Gradient Boosting (XGBoost). XGBoost merupakan algoritma yang berbasis pada pohon keputusan dan digunakan untuk tugas klasifikasi dan regresi. Keunggulan XGBoost terletak pada kemampuannya menghasilkan model yang akurat dengan waktu pelatihan yang relatif cepat[2][3].

Namun, meskipun XGBoost telah berhasil digunakan dalam berbagai domain, implementasinya untuk klasifikasi penyakit jantung masih membutuhkan penelitian lebih lanjut. Penelitian terbaru menunjukkan bahwa dengan memanfaatkan dataset yang lebih besar dan lebih kompleks, serta menggunakan teknik penyetelan parameter yang cermat, XGBoost dapat memberikan performa yang lebih baik dalam klasifikasi penyakit jantung[4].

Oleh sebab itu, penelitian ini memiliki tujuan untuk mengimplementasikan algoritma eXtreme Gradient Boosting (XGBoost) untuk klasifikasi penyakit jantung. Penelitian ini akan memanfaatkan dataset yang tersedia di website Kaggle dan teknik penyetelan parameter yang optimal untuk meningkatkan akurasi klasifikasi penyakit jantung. Untuk itu, penelitian ini juga akan membandingkan performa XGBoost dengan metode klasifikasi lainnya yang umum digunakan dalam domain ini, seperti Random Forest dan Support Vector Machine.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dikemukakan di atas, terdapat empat pokok masalah pada penelitian ini yaitu:

1. Sejauh mana akurasi klasifikasi penyakit jantung yang dapat dicapai dengan menggunakan metode XGBoost?
2. Variabel-variabel apa saja yang sangat mempengaruhi seseorang terkena penyakit jantung?
3. Bagaimana membandingkan performa metode XGBoost dengan metode klasifikasi lainnya, seperti Random Forest dan Support Vector Machine, dalam klasifikasi penyakit jantung?
4. Bagaimana hasil dari tuning parameter untuk meningkatkan kinerja metode XGBoost dalam klasifikasi penyakit jantung?

## 1.3 Batasan Masalah

Dalam penelitian ini memiliki beberapa batasan untuk di teliti, berikut batasan masalah penelitian yang dilakukan yaitu :

1. Penelitian ini akan difokuskan pada penggunaan algoritma Extreme Gradient Boosting (XGBoost) untuk klasifikasi penyakit jantung. Penelitian tidak akan membahas diagnosis atau pengobatan penyakit jantung secara keseluruhan.
2. Penelitian ini akan menggunakan dataset yang didapat dari website Kaggle (<https://www.kaggle.com/datasets/rashikrahmanpritom/heart-attack-analysis-prediction-dataset>) untuk klasifikasi penyakit jantung. Namun, penelitian ini tidak akan mencakup pengumpulan data baru.
3. Penelitian ini akan membandingkan performa metode XGBoost dengan metode klasifikasi lainnya, seperti Random Forest dan Support Vector Machine (SVM). Namun, penelitian ini tidak akan mempertimbangkan metode klasifikasi lainnya yang mungkin juga digunakan dalam domain ini.
4. Penelitian ini akan mencoba menyetel parameter untuk model XGBoost saja

dengan bantuan *Randomized Search Cross Validation* untuk mencapai performa yang optimal dalam klasifikasi penyakit jantung. Namun, untuk metode Random Forest dan SVM tidak dilakukan penyetelan parameter yang baru.

#### 1.4 Tujuan Penelitian

Berdasarkan rumusan dan batasan masalah di atas, maka tujuan dalam penelitian ini adalah untuk:

1. Tujuan utama penelitian ini adalah mengimplementasikan metode XGBoost dalam klasifikasi penyakit jantung. Penelitian akan mencakup langkah-langkah implementasi yang tepat, termasuk preproses data, pelatihan model, dan evaluasi performa.
2. Meningkatkan akurasi klasifikasi penyakit jantung dengan memanfaatkan metode XGBoost. Dengan penyetelan parameter yang cermat dan pemanfaatan dataset yang ada, diharapkan metode XGBoost dapat memberikan hasil yang lebih akurat dalam mengklasifikasikan penyakit jantung.
3. Membandingkan performa metode XGBoost dengan metode klasifikasi lainnya, seperti Random Forest dan Support Vector Machine, dalam klasifikasi penyakit jantung.
4. Mendapatkan kombinasi parameter yang tepat untuk mendapatkan hasil klasifikasi yang akurat.
5. Hasil penelitian ini diharapkan dapat memberikan rekomendasi yang berharga dalam pengembangan sistem klasifikasi penyakit jantung yang lebih akurat dan efisien.

#### 1.5 Manfaat Penelitian

Berdasarkan tujuan di atas, maka manfaat yang diharapkan dalam penelitian ini adalah sebagai berikut:

1. Manfaat Teoritis

Penelitian ini akan memberikan pemahaman yang lebih baik tentang keefektifan metode Extreme Gradient Boosting (XGBoost) dalam klasifikasi penyakit jantung. Hal ini akan membantu dalam pengembangan dan pemilihan metode klasifikasi yang lebih tepat dan efisien di masa depan.

## 2. Manfaat Praktis bagi Objek Penelitian

Penelitian ini dapat memberikan wawasan tentang faktor-faktor yang berhubungan dengan penyakit jantung. Hal ini dapat digunakan untuk mengidentifikasi populasi yang rentan dan menerapkan upaya pencegahan yang lebih efektif, seperti pengaturan gaya hidup dan pengawasan kesehatan.

## 3. Manfaat bagi Peneliti Selanjutnya

Untuk ikut serta mengembangkan metode eXtreme Gradient Boosting serta menjadi pembanding dengan metode klasifikasi yang lainnya.

### 1.6 Sistematika Penulisan

Bab satu dari skripsi ini adalah pendahuluan, yang memberikan gambaran umum tentang isi skripsi sesuai dengan judulnya. Penulis menyajikan ringkasan dari setiap isi bab per bab yang dibagi menjadi enam bagian, yaitu latar belakang masalah, rumusan masalah, tujuan penelitian, Batasan masalah, manfaat penelitian, dan sistematika penulisan.

Bab dua merupakan tinjauan pustaka yang mencakup tinjauan terhadap penelitian sebelumnya yang menjadi dasar penulisan skripsi ini, seperti penelitian tentang analisis XGBoost. Dalam bab ini juga akan diuraikan perbedaan dengan penelitian sebelumnya. Selain itu, bab ini juga memuat dasar teori skripsi ini, meliputi teori tentang Penyakit Jantung, Preprocessing, Feature Selection, Pearson Correlation, machine learning, klasifikasi, ensemble learning XGBoost, Random Forest, SVM, dan Confusion Matrix.

Bab tiga merupakan metode penelitian yang memuat alur penelitian, metode pengumpulan data, jenis dan sumber data. Dalam bab ini juga akan diuraikan teknik pengambilan data, variabel penelitian, definisi operasional, metode analisis data



dan langkah penelitian.

Bab empat merupakan bagian dari skripsi yang berfokus pada analisis data dan pembahasan. Pada bab ini, penulis menyajikan data yang telah dikumpulkan dan melakukan analisis mendalam terhadapnya. Selain itu, bab ini juga berisi penjelasan mengenai hasil dari analisis yang telah dilakukan dan pembahasannya secara mendetail.

Bab lima merupakan bagian penutup yang menyajikan rangkuman dari hasil penelitian yang didapatkan dari bab analisis data dan pembahasan. Di dalam bab ini, terdapat kesimpulan yang diambil dari analisis data yang telah dilakukan dalam penelitian. Selain itu, bab ini juga memberikan saran-saran yang bermanfaat bagi pihak-pihak lain di masa yang akan datang.

