

# BAB I

## PENDAHULUAN

### 1.1. Latar Belakang

Penyakit kardiovaskular adalah sekelompok penyakit jantung dan pembuluh darah yang mencakup berbagai kondisi seperti penyakit jantung koroner, penyakit serebrovaskular, penyakit arteri perifer, penyakit jantung rematik, dan penyakit jantung bawaan. Menurut data dari WHO (*World Health Organization*), penyakit ini merupakan penyebab utama kematian di seluruh dunia yang merenggut sekitar 17,9 juta jiwa tiap tahunnya. Serangan jantung dan *stroke* mendominasi penyebab dari kematian akibat penyakit kardiovaskular [1].

Penyakit kardiovaskular masih menjadi salah satu ancaman global dimana penyakit ini berperan utama sebagai penyebab kematian nomor satu di seluruh dunia. Kelompok usia produktif sering kali menjadi target utama, sehingga tingginya angka kematiannya memberikan dampak ekonomi dan sosial yang signifikan bagi masyarakat. Berdasarkan data dari RISKESDAS (Riset Kesehatan Dasar) tahun 2018, jumlah kasus penyakit kardiovaskular di Indonesia meningkat setiap tahunnya. Tercatat bahwa sebanyak 2.784.064 penduduk Indonesia sudah menderita penyakit kardiovaskular. Selain itu, data dari RISKESDAS juga menunjukkan peningkatan prevalensi penyakit ini seperti hipertensi meningkat dari 25,8 % pada tahun 2013 menjadi 34,1 % pada tahun 2018 [2].

Terdapat dua kategori faktor risiko yang menyebabkan terjadinya penyakit kardiovaskular: faktor risiko yang tidak dapat diubah dan faktor risiko yang dapat diubah. Faktor risiko yang tidak dapat diubah seperti riwayat penyakit keluarga, usia, jenis kelamin, dan obesitas. Sementara itu, faktor risiko yang bisa diubah mencakup *dislipidemia* (kondisi kolesterol yang tidak normal), diabetes melitus (peningkatan kadar glukosa), hipertensi (peningkatan tekanan darah), stress, pola makan yang tidak sehat dan kurangnya aktivitas fisik [3]. Dalam beberapa kasus, faktor sosial juga dapat menjadi penentu terjadinya penyakit kardiovaskular. Penduduk perkotaan lebih banyak menderita penyakit kardiovaskular dengan prevalensi 1,6 % dibandingkan dengan penduduk pedesaan yang hanya 1,3 % [4].

Saat ini, organisasi kesehatan dunia (WHO) sedang berupaya untuk mengurangi sekitar 25 % kematian dini akibat penyakit kardiovaskular. WHO mendukung pemerintah dalam mencegah, menangani dan mengendalikan penyakit kardiovaskular dengan cara mengembangkan strategi global untuk mengurangi kejadian, dan mortalitas dari penyakit ini. Strategi yang dikembangkan tersebut meliputi pengurangan faktor risiko, mengembangkan standar perawatan, meningkatkan kapasitas sistem perawatan pasien, serta memantau pola serta tren penyakit untuk dijadikan masukan dalam tindakan nasional dan global [1]. Diperlukan pencegahan dan pengendalian yang bertujuan untuk mencegah dan mengurangi kasus baru. Pencegahan kasus baru berfokus pada perubahan perilaku dan perubahan lingkungan agar lebih sehat serta mengidentifikasi faktor risiko penyakit sehingga dapat dilakukan pencegahan sejak dini.

Dalam perkembangan teknologi yang ada pada zaman sekarang, beberapa rumah sakit sudah bertahap dalam mengadopsi sistem machine learning. Machine Learning adalah cabang dari kecerdasan buatan (AI) yang semakin banyak digunakan dalam bidang kedokteran [32]. Secara khusus, implementasi machine learning berhasil meningkatkan kinerja para ahli medis dan meningkatkan efisiensi sistem perawatan kesehatan. Peningkatan signifikan dalam akurasi diagnosis sudah ditunjukkan melalui kinerja model machine learning bersama dengan para tenaga medis [31].

Sudah terdapat beberapa penelitian yang berkaitan tentang penyakit kardiovaskular, salah satunya telah dilakukan oleh Elias Dritsas dan rekan-rekannya (2022) dengan judul “*Cardiovascular Disease Risk Prediction with Supervised Machine Learning Techniques*”, dengan jumlah data sebanyak 70.000. Peneliti menggunakan sistem untuk memprediksi dan mengklasifikasi penyakit kardiovaskular menggunakan beberapa algoritma dan mendapatkan akurasi diantaranya Naïve Bayes (59.59 %), Support Vector Machine (70.61 %), Random Forest (70.86 %), Logistic Regression (72.06 %) [5]. Hasil tertinggi dalam penelitian ini didapatkan oleh algoritma *Logistic Regression*. Namun, di sini masih terdapat peluang untuk meningkatkan akurasi dari algoritma ini dengan menggunakan optimasi *Grid Search CV*.

Penelitian lain tentang penggunaan *Grid Search CV* sudah dilakukan oleh I Ketut Adian Jayaditya dan rekan-rekannya (2023) dengan judul “Implementasi Random Forest pada Klasifikasi Penyakit Kardiovaskular dengan Hyperparameter Tuning Grid Search”. Pada penelitian tersebut, penulis menggunakan 70.000 data dan algoritma yang digunakan adalah Random Forest. Akurasi yang didapatkan sekitar 69.65 % tanpa *Grid Search* sedangkan hasil 73.06 % menggunakan *Grid Search*. Hal ini menunjukkan bahwa *Grid Search* merupakan salah satu metode yang bisa digunakan untuk memberikan estimasi yang lebih akurat terkait tentang kinerja model.

Berdasarkan hasil yang didapatkan dari penelitian terdahulu, maka penelitian ini bertujuan untuk mengembangkan algoritma *Logistic Regression* yang akan digunakan dalam klasifikasi penyakit kardiovaskular. Algoritma yang telah dikembangkan, selanjutnya akan dilakukan optimasi yang diharapkan dapat meningkatkan akurasi. Algoritma yang dilatih dan dianalisis diharapkan dapat memberikan hasil yang lebih akurat dibandingkan dengan penelitian sebelumnya agar dapat dilakukan proses klasifikasi dan pencegahan penyakit kardiovaskular dengan lebih tepat.

## 1.2. Rumusan Masalah

Berdasarkan latar belakang masalah yang ada, dapat dirumuskan masalah sebagai berikut :

- a. Bagaimana performa dari algoritma *Logistic Regression* dalam mengklasifikasi penyakit kardiovaskular?
- b. Bagaimana optimasi *grid search cv* mampu meningkatkan akurasi dan kinerja model dalam klasifikasi penyakit kardiovaskular?
- c. Bagaimana perbandingan hasil performa model sebelum dan sesudah dilakukan optimasi dengan *grid search CV*?
- d. Bagaimana perbandingan dari hasil penelitian yang dibangun dan penelitian terdahulu?

### 1.3. Tujuan Penelitian

Penelitian ini bertujuan untuk mengembangkan algoritma *Logistic Regression* dengan optimasi *grid search CV* yang dapat mengklasifikasi penyakit kardiovaskular secara akurat. Algoritma dan optimasi yang dikembangkan diharapkan dapat membantu dalam klasifikasi, prediksi serta pencegahan penyakit kardiovaskular sehingga dapat memberikan hasil yang baik dalam perawatan kesehatan.

### 1.4. Batasan Masalah

Adapun batasan masalah dalam penelitian ini adalah :

- a. Penelitian ini berfokus pada target klasifikasi penyakit kardiovaskular yang hanya terdiri dari 2 kelas (menderita dan tidak menderita).
- b. Menggunakan dataset yang diambil dari situs *Kaggle* dengan judul "*Cardiovascular Disease Dataset*" [6].
- c. Total data yang digunakan sebanyak 70.000 yang terdiri dari 11 fitur dan 1 target.
- d. Menggunakan "Google Colaboratory" dengan bahasa pemrograman "Python"