

# **BAB I**

## **PENDAHULUAN**

Pada bab I memaparkan latar belakang penelitian, rumusan permasalahan, batasan masalah dalam penelitian, tujuan dan manfaat, serta sistematika penulisan skripsi.

### **1.1 Latar Belakang**

Jantung merupakan bagian tubuh manusia yang terletak di pusat sistem peredaran darah. Jantung bertugas untuk mengangkut dan mengalirkan darah ke semua bagian tubuh manusia [1]. Sistem pembuluh darah besar yang tidak berfungsi dengan baik dapat menimbulkan gangguan pada jantung dan peredaran darah [2]. Pada tahun 2008, sebanyak 17,3 juta orang meninggal karena penyakit jantung. Studi terbaru yang dilakukan pada tahun 2018 oleh *World Health Organization* (WHO) menunjukkan bahwa 56,9 juta kematian yang terjadi di dunia selama tahun 2016 disebabkan oleh penyakit jantung [3]. Meskipun jenis penyakit ini tidak dapat menular, namun termasuk penyakit yang mematikan pertama di dunia [4].

Kasus penyakit jantung semakin tinggi seiring meningkatnya pola hidup kurang sehat yang memberikan dampak buruk pada kesehatan masyarakat [5]. Penyebab utama penyakit jantung antara lain mengonsumsi alkohol, memiliki kolesterol tinggi, merokok, menjalani diet yang tidak sehat, dan obesitas. Selain itu, peningkatan risiko penyakit jantung juga disebabkan oleh bertambahnya usia dan tingginya tekanan darah [4]. Angka kematian akibat penyakit jantung semakin tinggi dikarenakan penderita terlambat melakukan pengobatan [5]. Berdasarkan kondisi tersebut, maka diperlukan diagnosis awal penyakit jantung untuk mengetahui kemungkinan bahwa seseorang menderita penyakit jantung atau tidak [2]. Dengan melakukan diagnosis awal, para penderita penyakit jantung dapat segera mengambil tindakan pengobatan. Hal tersebut dapat mempercepat proses penyembuhan dan meminimalisir risiko kematian akibat penyakit ini [6].

Dalam penelitian terdahulu yang pernah dilakukan oleh Rezaul Karim dkk membandingkan algoritma *Decision Tree*, *Naive Bayes*, *SVM*, *Random Forest*,

*Logistic Regression*, dan KNN guna mengetahui algoritma terbaik dalam mengklasifikasi penyakit jantung. Data yang digunakan adalah *Framingham heart study dataset* dari website Kaggle. Hasil penelitian menunjukkan bahwa tiga algoritma terbaik dengan nilai akurasi yang sama sebesar 87% adalah *Logistic Regression*, KNN, dan SVM [7]. Penelitian selanjutnya yang dilakukan oleh Devansh Shah dkk menggunakan algoritma *Decision Tree*, KNN, *Naive Bayes*, dan *Random Forest* untuk mengklasifikasi penyakit jantung. *Dataset* yang digunakan adalah *Cleveland* dari website *UCI Machine Learning*. Perhitungan jarak pada algoritma KNN menggunakan *Euclidean Distance*. Hasil penelitian menunjukkan bahwa algoritma KNN ( $k = 7$ ) menghasilkan akurasi tertinggi sebesar 90.789% [8]. Penelitian selanjutnya yang dilakukan oleh Sikha Suhani Bhuyan dan Ashis Kumar Mishra membandingkan algoritma *Logistic Regression*, *LGBM Classifier*, KNN, *Random Forest*, *Decision Tree*, SVM, dan *Naive Bayes* untuk memprediksi penyakit jantung. *Dataset* yang digunakan adalah *Cleveland* dari website *UCI Machine Learning*. Perhitungan jarak pada algoritma KNN menggunakan *Euclidean Distance*. Hasil penelitian menunjukkan bahwa algoritma KNN memiliki akurasi tertinggi sebesar 85.71% [9].

Berdasarkan uraian penelitian terdahulu, dimana diperoleh algoritma terbaik untuk melakukan klasifikasi terhadap penyakit jantung yaitu *K-Nearest Neighbor* menggunakan perhitungan jarak berbasis *Euclidean Distance* dengan tipe data numerik. Pada penelitian sebelumnya tidak menjelaskan tentang penerapan *K-Nearest Neighbor* pada data bertipe nominal (kategorikal). Penelitian ini membahas tentang diagnosis terhadap penyakit jantung menggunakan algoritma *K-Nearest Neighbor* dengan *dataset* yang digunakan bertipe nominal (kategorikal). Prinsip kerja algoritma KNN adalah menentukan jarak paling dekat antara data latih dengan data uji menggunakan parameter  $k$  (tetangga terdekat) [1]. Penelitian yang dilakukan oleh Chaoqun Li dan Hongwei Li menunjukkan bahwa perhitungan jarak pada data nominal dapat menggunakan *Overlap Metric*, dimana jarak antar dua data bernilai 0 jika memiliki nilai yang sama. Sedangkan jika nilai kedua data berbeda, maka jarak *Overlap* bernilai 1 [10]. Penelitian ini dilakukan dengan judul “Diagnosis Penyakit Jantung menggunakan Algoritma *K-Nearest Neighbor* berbasis *Overlap Metric*” yang dapat menghasilkan nilai akurasi yang baik.

## **1.2 Perumusan Masalah**

### **1.2.1 Permasalahan**

Angka kematian akibat penyakit jantung semakin tinggi dikarenakan penderita terlambat melakukan pengobatan. Berdasarkan kondisi tersebut, maka diperlukan diagnosis awal penyakit jantung untuk mengetahui kemungkinan bahwa seseorang menderita penyakit jantung atau tidak. Dengan melakukan diagnosis awal, para penderita penyakit jantung dapat segera mengambil tindakan pengobatan. Hal tersebut dapat mempercepat proses penyembuhan dan meminimalisir risiko kematian akibat penyakit ini.

### **1.2.2 Solusi**

Penerapan algoritma *K-Nearest Neighbor* dengan perhitungan jarak berbasis *Overlap Metric* untuk mendiagnosis penyakit jantung.

### **1.2.3 Pertanyaan Penelitian**

1. Berapa proporsi pembagian data latih dan data uji (70%:30%, 80%:20%, dan 90%:10%) yang menghasilkan nilai performa (*accuracy*, *precision*, *recall*, dan *F-score*) tertinggi untuk algoritma KNN dengan perhitungan jarak berbasis *Overlap Metric* dalam mendiagnosis penyakit jantung ?
2. Berapa nilai *k* terbaik yang digunakan untuk algoritma KNN ?

## **1.3 Tujuan dan Manfaat**

### **1.3.1 Tujuan**

Tujuan dari pelaksanaan penelitian ini adalah mengetahui proporsi yang tepat untuk pembagian data latih dan data uji, nilai performa (*accuracy*, *precision*, *recall*, dan *F-score*), dan nilai *k* terbaik untuk algoritma KNN berbasis *Overlap Metric* guna melakukan diagnosis penyakit jantung.

### **1.3.2 Manfaat**

Hasil dari penelitian ini dapat menjadi suatu pilihan atau solusi untuk membuat sistem diagnosis penyakit jantung. Sistem tersebut dapat mempermudah kinerja dokter spesialis jantung dalam mendiagnosis awal penyakit jantung yang diderita oleh pasien.

## **1.4 Batasan Masalah**

Batasan masalah pada penelitian ini sebagai berikut:

1. *Dataset* yang digunakan bernama *Heart Failure Prediction* diperoleh dari *website* Kaggle (<https://www.kaggle.com/datasets/fedesoriano/heart-failure-prediction>). *Dataset* tersebut terdiri dari 918 data, 11 atribut (*Sex, ChestPainType, RestingBP, Cholesterol, FastingBS, RestingECG, MaxHR, ExerciseAngina, Oldpeak, ST\_Slope*), dan 2 kelas sebagai hasil klasifikasi yaitu nilai 1 berarti positif penyakit jantung dan 0 berarti negatif.
2. Pada penelitian ini tidak membahas tentang data *outlier*.

### 1.5 Sistematika Penulisan Skripsi

Secara umum penulisan skripsi dibagi menjadi beberapa bab yang membahas hal-hal sebagai berikut:

#### BAB I PENDAHULUAN

Pada bab I memaparkan latar belakang penelitian, rumusan permasalahan, batasan masalah, tujuan dan manfaat penelitian, serta sistematika penulisan skripsi.

#### BAB II TEORI PENUNJANG

Pada bab II menerangkan teori penunjang penelitian, yaitu penyakit jantung, klasifikasi, algoritma *K-Nearest Neighbor*, diskritisasi data, perhitungan jarak berbasis *Overlap Metric*, dan tabel *Confusion Matrix*.

#### BAB III METODE PENELITIAN

Pada bab III menjelaskan alur dalam pelaksanaan penelitian ini. Tahapan tersebut adalah pengumpulan data, *data preprocessing*, pembagian data latih dan data uji, klasifikasi dengan algoritma *K-Nearest Neighbor*, serta evaluasi kinerja model klasifikasi.

#### BAB IV HASIL DAN PEMBAHASAN

Pada bab IV membahas secara rinci tentang proses klasifikasi penyakit jantung menggunakan algoritma *K-Nearest Neighbor* dengan perhitungan jarak berbasis *Overlap Metric*, penentuan parameter *k* terbaik, dan evaluasi kinerja dari model klasifikasi.

#### BAB V PENUTUP

Pada bab V menguraikan kesimpulan yang didapatkan dari hasil analisa uji coba penelitian dan saran untuk pengembangan penelitian selanjutnya.