

BAB I PENDAHULUAN

1.1 Latar Belakang

Dalam dunia Kesehatan, penyakit jantung menjadi masalah utama. Pada tahun 2015 kurang lebih 17 juta orang meninggal dunia yang diakibatkan oleh penyakit jantung dan rusaknya pembuluh darah dan menurut data *World Health Organization* (WHO) 8.7 juta orang meninggal yang diakibatkan oleh penyakit jantung [1]. Di seluruh dunia, penyakit jantung menjadi penyebab utama kematian termasuk Indonesia, dari 100 penduduk Indonesia dalam semua umur terdapat 1.5 menderita penyakit jantung karena risiko penduduk di Indonesia memiliki penyakit jantung adalah 1.5%. [2][3]. Penyakit jantung ada beberapa macam antara lain penyakit jantung iskemik atau penyakit jantung *aterosklerotik*, penyakit jantung koroner atau bisa disebut sebagai penyakit arteri koroner, dan penyakit jantung yang diakibatkan oleh gangguan otot jantung karena hasil akhir dari akumulasi plak *ateromatosa* dalam dinding-dinding arteri yang memasok darah ke *miokardium*[4]. Jadi sangat penting agar penyakit jantung dideteksi secara dini, Sangat disarankan agar semua orang termasuk yang merasa sangat sehat harus menemui spesialis jantung dua kali setahun untuk menyelidiki kemungkinan adanya faktor yang dapat menyebabkan penyakit jantung. akan tetapi diagnosa sering kali dibuat dengan berfokus pada naluri dan pengalaman dokter, bukan bersumber pada data yang dicakup dalam database. Hal ini menimbulkan dampak yang tidak diinginkan seperti kesalahan dalam mendiagnosis penyakit dan biaya pengobatan yang tidak perlu pada pasien, apalagi biaya pengobatan penyakit jantung sangat mahal[5][6].

Perkembangan Teknologi informasi telah menyebabkan terciptanya aplikasi informatika dalam bidang Kesehatan yang menghasilkan data yang besar yang meliputi rincian rumah sakit, rincian pasien, rincian penyakit dan biaya pengobatan yang dapat digunakan untuk memprediksi penyakit menggunakan Teknik *data mining*[7]. Teknik *data mining* adalah menjadi area penelitian yang penting untuk analisis data besar yang efektif karena kompleksitas dan ukuran data meningkat. *Data mining* digunakan di banyak domain yaitu penambangan gambar, opini penambangan, penambangan web, penambangan teks, penambangan grafik dan sistem data medis. *Data mining* telah menjadi bidang

penelitian medis yang penting untuk menemukan pola yang tidak diketahui dalam data medis[8]. Beberapa Teknik *data mining* adalah pengelompokan, klasifikasi dan prediksi. Teknik *data mining* berguna dalam memprediksi banyak penyakit seperti penyakit jantung[9].

Beberapa penelitian sebelumnya telah dilakukan untuk klasifikasi penyakit jantung, penelitian yang dilakukan oleh Ali Bou Nassif dkk pada tahun 2018 yang membahas tentang penerapan algoritma klasifikasi yang berbeda yaitu SVM, *Naïve Bayes* dan KNN untuk klasifikasi penyakit jantung, penelitian ini menggunakan 10 kfold untuk pengujian model dengan dataset yang digunakan adalah dataset *Cleveland* yang memiliki jumlah data adalah 297 dan 13 fitur, 7 fitur dipilih setelah melalui seleksi fitur *Information gain* dan *Correlation*, menghasilkan bahwa metode *naïve bayes* memiliki akurasi terbaik yaitu 84%[10]. Penelitian juga dilakukan oleh M Ikhsan Perdana Putra dkk pada tahun 2019 yang membahas tentang penerapan metode MKNN untuk klasifikasi penyakit kanker payudara. Dataset dalam penelitian ini diambil dari UCI – *Machine Learning Repository Breast Cancer Winconsin* yang terdiri dari 31 gejala penyakit dan jumlah data 568. Terdiri dari dua kelas yaitu kanker jinak dan kanker ganas. Hasil pengujian pada penelitian ini menunjukkan bahwa nilai K sangat mempengaruhi akurasi. Rata-rata akurasi cenderung menurun jika nilai K dinaikkan dan akurasi akan meningkat jika data latihnya dinaikkan. Hasil akurasi tertinggi pada pengujian ini sebesar 97.61 % dengan K=1 dan data latih 90%[11]. Penelitian lain dilakukan oleh K. Srivastava dan D. K. Choubey pada tahun 2020, Penelitian ini membahas tentang penerapan algoritma *machine learning* dan *data mining* yaitu *Decision Tree*, *Naïve Bayes Classifier*, *K-Nearest Neighbour*, *Random Forest*, *Artificial Neural Network*, dan *Support Vector Machines* untuk prediksi penyakit jantung. Dataset yang digunakan dalam penelitian ini adalah *Cleveland Heart Disease* dari UCI *Machine Learning repository* dengan jumlah data 303 dan 13 fitur dan kelas biner. Hasil dari penelitian ini menyimpulkan dari semua algoritma, *K-Nearest Neighbor* memberikan akurasi tertinggi sebesar 87%[12].

Dalam penelitian ini, peneliti mengusulkan metode MKNN (*Modifed K-Nearest Neighbor*) dengan seleksi fitur *Pearson Correlation* untuk identifikasi penyakit yang diharapkan mampu mengetahui bagaimana implementasi dan

akurasi dari metode MKNN (*Modified K-Nearest Neighbor*) dengan seleksi fitur *Pearson Correlation* untuk Identifikasi penyakit jantung.

1.2 Perumusan Masalah

1.2.1 Permasalahan

Diagnosa sering kali dibuat dengan berfokus pada naluri dan pengalaman dokter, bukan bersumber pada data yang dicakup dalam *database*. Hal ini menimbulkan dampak yang tidak diinginkan seperti kesalahan dalam mendiagnosis penyakit dan biaya pengobatan yang tidak perlu pada pasien. Oleh karena itu diperlukan identifikasi dini penyakit jantung menggunakan metode MKNN (*Modified K-Nearest Neighbor*) seleksi fitur *Pearson Correlation*.

Metode ini dipilih karena penulis ingin mengetahui apakah metode MKNN dan seleksi fitur *Pearson Correlation* mampu jika diterapkan pada dataset penyakit jantung karena berdasarkan penelitian-penelitian sebelumnya metode tersebut mampu memberikan akurasi yang tinggi jika diterapkan pada dataset yang berbeda.

1.2.2 Metode Usulan

Penelitian ini mengusulkan metode MKNN (*Modified K-Nearest Neighbor*) dan seleksi fitur *Pearson Correlation* untuk identifikasi penyakit jantung.

1.2.3 Pertanyaan Peneliti

1. Bagaimana cara melakukan identifikasi penyakit jantung menggunakan metode MKNN (*Modified K-Nearest Neighbor*) dengan seleksi fitur *Pearson Correlation*?
2. Bagaimana akurasi dari metode MKNN (*Modified K-Nearest Neighbor*) dengan seleksi fitur *Pearson Correlation* untuk identifikasi penyakit jantung?

1.3 Tujuan dan Manfaat

1.3.1 Tujuan Peneliti

1. Melakukan Identifikasi penyakit jantung menggunakan metode MKNN (*Modified K-Nearest Neighbor*) dengan seleksi fitur *Pearson Correlation*.
2. Mengetahui akurasi dari metode MKNN (*Modified K-Nearest Neighbor*) dengan seleksi fitur *Pearson Correlation* untuk identifikasi penyakit jantung.

1.3.2 Manfaat Penelitian

Membantu dan mengurangi risiko kesalahan para ahli dalam mengidentifikasi penyakit jantung.

1.4 Batasan – Batasan

Batasan masalah pada penelitian ini adalah sebagai berikut :

1. Data yang akan digunakan pada penelitian ini adalah data *Heart Disease* dari situs *Kaggle Repository*
(<https://www.kaggle.com/datasets/johnsmith88/heart-disease-dataset>)
2. Banyak data adalah 1025 baris dengan 13 fitur

1.5 Sistematika Laporan Skripsi

Sistematika laporan penelitian ini terdiri dari beberapa bab, yaitu :

BAB 1 PENDAHULUAN

Pada bab ini membahas tentang latar belakang, rumusan masalah, tujuan, manfaat, batasan masalah, serta sistematika penulisan laporan.

BAB II KAJIAN PUSTAKA

Bab ini membahas landasan teori utama yaitu teori tentang klasifikasi, penyakit jantung, metode *MinMax*, *Pearson Correlation* dan MKNN yang diperlukan dalam menyelesaikan penelitian dan penelitian sebelumnya yang telah digunakan untuk menyelesaikan dengan metode atau masalah yang sama.

BAB III METODE USULAN

Membahas tentang metode yang diusulkan (algoritma) yang dalam penelitian ini metode yang diusulkan adalah MKNN (*Modified K-Nearest Neighbor*) dan seleksi fitur *Pearson Correlation*, arsitektur (*input*, *proses*, *output*), data penelitian, tahapan penelitian, serta rencana pengujian.

BAB IV HASIL DAN PEMBAHASAN

Pada bab ini membahas tentang hasil implementasi sistem dan hasil yang telah didapatkan dari proses uji coba klasifikasi penyakit jantung berdasarkan dengan skenario pengujian, dalam penelitian ini ada 3 skenario pengujian yaitu perbandingan jumlah fitur, perbandingan jumlah k pada MKNN dan perbandingan jumlah *k-fold* dan tanpa *k-fold*.

BAB V PENUTUP

Pada bab ini membahas tentang kesimpulan dan saran berdasarkan hasil uji coba yang telah dilakukan sebagai pengembangan penelitian terkait untuk berikutnya.

DAFTAR PUSTAKA

Berisi tentang daftar jurnal atau referensi yang dijadikan acuan dalam penelitian ini.