

BAB I

PENDAHULUAN

1.1. Latar Belakang

Organ jantung sangat penting bagi kelangsungan hidup semua makhluk hidup karena berperan dalam memompa darah, nutrisi, dan oksigen ke seluruh organisme. Pada manusia, jantung dianggap sebagai organ vital karena fungsinya yang sangat penting untuk mengedarkan darah ke seluruh bagian tubuh agar organ lain dapat berfungsi dengan baik. Apabila pasokan darah yang disalurkan oleh jantung terganggu akibat adanya sumbatan pada pembuluh darah utama atau kegagalan pada jantung dalam memompa darah ke seluruh tubuh, kondisi ini disebut sebagai penyakit jantung koroner yang bisa menimbulkan masalah kesehatan yang serius bahkan bisa berujung pada kematian. (Diyan, Yayi, Made 2018).

Di era modern seperti saat ini, perkembangan teknologi semakin pesat. Gaya hidup masyarakat berkembang dari yang semula tradisional dan sederhana menjadi modern yang canggih dan efisien. Perkembangan teknologi ini juga memiliki dampak yang besar pada kehidupan, khususnya pada bidang kesehatan. Penyakit jantung koroner tetap menjadi salah satu penyebab utama kematian di Indonesia. Gejala-gejala yang menunjukkan faktor risiko sangat membantu dalam mendiagnosis apakah seseorang mengalami penyakit jantung atau tidak. Selain itu, keterbatasan informasi dan media serta keterlambatan pemeriksaan awal untuk penyakit jantung juga menjadi masalah. Keterbatasan akses untuk memperoleh informasi mengenai serangan jantung berdampak pada peningkatan angka kematian tiap tahunnya. Oleh karena itu, diperlukan sebuah sistem klasifikasi yang dapat memberikan informasi mengenai serangan jantung dan memeriksa klasifikasi secara dini terkait kondisi serangan jantung yang dialami oleh seseorang. Diperlukannya metode yang sesuai untuk mengelola pengetahuan dari para pakar agar hasil klasifikasi sistem

menjadi akurat. (Mufti, Kusrini, Sudarmawan 2019).

Penelitian ini dilakukan berdasarkan beberapa penelitian sebelumnya seperti penelitian yang dilakukan oleh Mufti, dkk. mengenai Perancangan Sistem Klasifikasi Penyakit Jantung Menggunakan *Naïve Bayes* pada tahun 2019 menyimpulkan bahwa hasil dari kinerja metode *naïve bayes* terhadap diagnosis penyakit jantung memiliki tingkat akurasi senilai 90,61%, rata-rata hasil nilai presisi senilai 87,44% dan rata-rata nilai *recall* senilai 87,95% (Mufti, Kusrini 2019). Lalu terdapat penelitian yang dilakukan oleh Silmi, dkk. Pada tahun 2022 Mengenai *Hybrid Machine Learning Model* Untuk Memprediksi Penyakit Jantung Dengan Menggunakan Metode *Logistic Regression* Dan *Random Forest* menyimpulkan bahwa penggunaan algoritma *logistic regression* memiliki tingkat akurasi sebesar 84,48% dan presisi sebesar 82,41% sedangkan algoritma *Random Forest* memiliki akurasi sebesar 83,16%, dan presisi sebesar 81,31% (Silmi 2022).

Kemudian terdapat penelitian yang dilakukan oleh Ichsan Gunawan, dkk. pada tahun 2020 mengenai “Peningkatan Kinerja Akurasi Prediksi Penyakit *Diabetes Mellitus* Menggunakan Metode *Grid Search* pada Algoritma *Logistic Regression*” menjelaskan bahwa hasil kinerja algoritma *logistic regression* untuk mendeteksi penyakit *diabetes melitus* tanpa menggunakan *Grid Search* dan menggunakan *Grid Search* memiliki perbedaan yang signifikan (Ichsan, Dedy, Mardianto 2020). Terdapat penelitian yang dilakukan oleh Lia Andiani, Sukemi, Dian Palupi Rini pada tahun 2019 yang berjudul “Analisis Penyakit Jantung Menggunakan Metode KNN Dan *Random Forest*”. Berdasarkan penelitian yang dilakukan, hasil tingkat akurasi data yang dihasilkan dari metode KNN lebih tinggi dari hasil tingkat akurasi data dari algoritma *random forest* dengan persentase akurasi metode KNN sebesar 93% dan *Random Forest* sebesar 72% (Lia, Sukemi, Dian 2019). Dan terdapat penelitian yang dilakukan oleh Delima, dkk. Mengenai Implementasi *Data Mining* Untuk Memprediksi Penyakit Jantung Menggunakan Metode *K-Nearest Neighbor* dan *Logistic Regression* menunjukkan bahwa perbandingan

algoritma *K-Nearest Neighbor* dan *Logistic Regression* mendapatkan hasil akhir yaitu akurasi Algoritma *Logistic Regression* sebesar 88% dan algoritma *K-Nearest Neighbor* sebesar 83%. Dengan demikian dapat disimpulkan bahwa algoritma *Logistic Regression* yang terbaik dalam memprediksi penyakit serangan jantung dibanding algoritma *K-Nearest Neighbor* (Delima, dkk. 2022).

Berdasarkan data yang didapat pada penelitian sebelumnya, pada penelitian ini penulis akan melakukan sebuah penelitian yang berjudul “Analisis Tingkat Akurasi Data Menggunakan Metode *Naïve Bayes*, *K-Nearest neighbor*, Dan *Logistic Regression* Untuk Mendiagnosis Penyakit Jantung”. Penelitian ini menggunakan konsep *data mining* dimana data akan dilakukan pra proses terlebih dahulu sebelum dilakukan implementasi dari metode yang digunakan. Selanjutnya adalah menghitung tingkat akurasi data dari beberapa metode yang digunakan pada penelitian sebelumnya diantaranya *naïve bayes*, *K-Nearest neighbor*, dan *Logistic Regression*. Ketiga metode ini dipilih dikarenakan metode ini memiliki kecocokan untuk mengkaji permasalahan yang ada. Hal ini diperkuat dengan adanya penelitian terdahulu yang memiliki similaritas terkait permasalahan yang terjadi dengan metode yang digunakan, sehingga metode ini sudah ter-uji dan sesuai dengan penelitian ini. Selain itu penelitian terdahulu juga memiliki beberapa perbedaan dengan penelitian yang sedang dilakukan oleh penulis seperti sumber *dataset* yang digunakan berasal dari studi kasus yang berbeda. Hal ini membuat hasil tingkat akurasi data dari setiap metode akan berbeda juga mengingat jumlah dan atribut dari *dataset* yang digunakan juga berbeda.

1.2. Rumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan pada bagian sebelumnya pada penelitian “Analisis Implementasi Metode *Naïve Bayes*, *K-Nearest neighbor*, Dan *Logistic Regression* Untuk Mengklasifikasi Penyakit Jantung”, maka dapat di buat beberapa rumusan masalah sebagai berikut:

1. Bagaimanakah hasil dari kinerja algoritma *naïve bayes*, *k-nearest neighbor*, dan *logistic regression* untuk melakukan klasifikasi penderita penyakit jantung?
2. Diantara ketiga algoritma *naïve bayes*, *k-nearest neighbor*, dan *logistic regression*, algoritma manakah yang memiliki performa terbaik yang sehingga dapat digunakan untuk mendiagnosis pasien penyakit jantung?

1.3. Tujuan

Berdasarkan permasalahan yang telah dipaparkan sebelumnya, penelitian ini memiliki beberapa tujuan yang digunakan untuk memecahkan permasalahan yang dihadapi diantaranya:

1. Untuk mengetahui kinerja dari masing-masing algoritma *naïve bayes*, *k-nearest neighbor*, dan *logistic regression*.
2. Untuk mengetahui algoritma apakah yang memiliki kinerja terbaik dalam mendiagnosis penyakit jantung.

1.4. Manfaat

Di samping tujuan penelitian yang telah disampaikan sebelumnya, penelitian ini juga memiliki beberapa manfaat bagi beberapa pihak diantaranya:

1.4.1. Bagi Penulis

1. Penulis dapat mengimplementasikan beberapa metode yang digunakan untuk mendiagnosis penyakit jantung.
2. Penulis dapat mengetahui diantara beberapa algoritma yang digunakan pada penelitian ini, algoritma apakah yang memiliki kinerja terbaik untuk mendiagnosis penyakit jantung?

1.4.2. Bagi Pengguna

Pengguna dapat menggunakan sistem yang telah dibuat untuk mendiagnosis penyakit jantung menggunakan metode dengan tingkat akurasi data tertinggi sehingga proses klasifikasi menjadi lebih mudah dan akurat.

1.5. Batasan Masalah

Berdasarkan rumusan masalah yang telah dijelaskan pada bagian sebelumnya, terdapat beberapa batasan masalah yang digunakan untuk membatasi cakupan pada penelitian “Analisis Implementasi Metode *Naïve Bayes*, *K-Nearest Neighbour*, dan *Logistic Regression* Untuk Mengklasifikasi Penyakit Jantung” diantaranya:

1. *Dataset* yang digunakan adalah data yang berasal dari *website* penyedia *dataset* secara *open source* yaitu data pasien dari *Kaggle* berbentuk file *comma separated value* atau “.csv” dengan jumlah data sebesar 1025 *records*.
2. Terdapat 3 algoritma yang akan digunakan untuk menghitung kinerja dari algoritma tersebut terhadap data dari pasien diantaranya *naïve bayes*, *k-nearest neighbour*, dan *logistic regression*.
3. Penelitian ini dilakukan dengan menggunakan bahasa *python 3* dengan kode editor *Jupyter Notebook*.
4. Terdapat 14 atribut dari *dataset* yang akan digunakan untuk menghitung tingkat kinerja algoritma beserta data uji klasifikasi penyakit jantung diantaranya: *Age*, *Sex*, *Cp*, *Trestbps*, *Chol*, *Fbs*, *Restecg*, *Thalach*, *Exang*, *Oldpeak*, *Slope*, *Ca*, *Thal*, dan *Target*.
5. Terdapat beberapa aspek yang diuji untuk menguji kinerja dari beberapa algoritma yang telah disebutkan sebelumnya yaitu akurasi, presisi, *recall*, dan *F1-Score*.
6. Hasil yang didapat pada penelitian ini hanya digunakan pada saat pengujian pertama saja pada setiap skenario pengujian.