

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang Masalah

Jantung sebuah rongga organ berotot yang memompa darah melalui pembuluh darah dengan kontraksi berirama yang terus berulang merupakan salah satu organ manusia yang berperan dalam sistem peredaran darah. Darah kemudian menyuplai oksigen dan nutrisi pada tubuh, dimana darah juga membantu menghilangkan sisa-sisa dari metabolisme. Letak Jantung berada di rongga dada sekitar sebelah kiri. Jantung sebagai salah satu organ terpenting dalam tubuh memiliki resiko kematian jika ada kelainan yang terjadi pada jantung. Beberapa masalah pada jantung dibagi menjadi dua yaitu penyakit jantung dan serangan jantung. WHO berdasarkan data menyatakan bahwa ada sebanyak 7,3 juta penduduk di dunia yang meninggal dikarenakan penyakit jantung. (World Health Organization, 2020). Tipe penyakit jantung terjadi dikarenakan jantung tidak dapat melaksanakan tugasnya dengan baik, seperti Otot jantung yang lemah atau adanya celah antara serambi kanan dan serambi kiri. (Tegar, 2019)

Mengambil data dari tulisan Center for Disease Control and Prevention, penyakit jantung merupakan penyebab utama meninggalnya wanita, pria dan beberapa kelompok ras dan juga etnis di Amerika Serikat, paling tidak ada satu orang yang meninggal setiap 36 detik dari penyakit jantung. Penyakit jantung yang terjadi perlu didiagnosa oleh dokter dengan menjalankan serangkaian tes dan evaluasi, yaitu pemeriksaan fisik, tes darah, tes dalam kategori non invasif seperti stress test dan elektrokardiogram dan tes invasif seperti katerisasi pada jantung. Semua tes ini akan menghasilkan data pasien seperti kadar kolesterol, tekanan darah dan beberapa data lainnya yang kemudian dapat membantu dalam proses mendiagnosis penyakit pasien. (Donovan et al., 2018)

Industri kesehatan telah berkembang sangat pesat salah satunya dikarenakan teknologi yang berkembang pesat, sehingga perkembangan ini membuka pintu untuk melakukan lebih banyak penelitian. Kesehatan merupakan hal yang penting untuk dijaga, dimana kesalahan pengobatan atau pencegahan dapat menyebabkan kehilangan nyawa, terutama pada penyakit jantung. Peningkatan teknologi pada

industri kesehatan salah satunya adalah melakukan digitalisasi informasi medis terutama data pasien, dengan harapan informasi yang telah digitalisasi dapat digunakan untuk penelitian dan hasil penelitian tersebut dapat meningkatkan layanan industri kesehatan.(NEJM Catalyst, 2018) Tentunya perkembangan teknologi yang pesat ini mempunyai beberapa masalah yang perlu diperhatikan, seperti bagaimana penanganan informasi medis yang didapatkan akan mempunyai ukuran data yang cukup besar dikarenakan perkembangan industri, hal yang perlu kita perhatikan adalah bagaimana bisa kita uraikan, kemudian analisa sehingga kita mendapatkan informasi yang penting dan berguna dimana kemudian dapat diaplikasikan dalam layanan pada industri kesehatan.

Salah satu teknologi yang digunakan untuk memprediksi penyakit jantung adalah machine learning. Machine Learning sendiri bukan merupakan teknologi yang memiliki keakuratan 100% untuk melakukan analisa data dan menarik kesimpulan. Namun keakuratan yang dihasilkan cukup efektif sehingga machine learning telah terbukti untuk membantu di bidang kesehatan.

Pada penelitian sebelumnya yang dilakukan oleh (Ashoka, 2022) implementasi klasifikasi menggunakan algoritma k-NN, Decision Tree, dan Random Forest sudah pernah dilakukan dengan menggunakan objek air dengan menggunakan perbandingan akurasi terhadap data yang dilakukan dengan judul “Otomasi dan Analisis Hasil Prediksi Penelitian Kualitas Air Bersih Antar Classifier Menggunakan Machine Learning.” menyatakan bahwa algoritma k-NN, Decision Tree dan Random Forest didapatkan nilai sebesar 1.0 atau 100%, hal ini berarti machine learning sudah mampu membaca kasus pada penelitian ini dengan baik.

Penelitian kedua (Haris, 2021) mengangkat tema radioaktif dengan judul “Otomasi Penentuan Nilai Radioaktivitas Radioisotop  $^{241}\text{Am}$ ,  $^{60}\text{Co}$ ,  $^{137}\text{Cs}$ , dan  $^{22}\text{Na}$  menggunakan Multi-output Machine Learning.” hasil algoritma machine learning k-NN mendapat perolehan proses uji sebesar 2.47% untuk target data puncak dan 43.57% untuk target data NPA. Hal ini dipengaruhi tingkatan kompleks cara kerja pada algoritma dan memiliki bentuk data berupa kenaikan dan penurunan nilai metrik yang diperoleh.

Selanjutnya Penelitian ketiga (Rajendra et al., 2021) adalah penelitian yang berjudul "Prediction of Diabetes Using Logistic Regression And Ensemble Techniques." Dalam penelitian ini, para peneliti menggunakan *Logistic Regression* untuk memprediksi risiko diabetes pada populasi tertentu. Metode ini melibatkan pengumpulan data pasien termasuk variabel seperti usia, indeks massa tubuh (BMI), tekanan darah, dan level glukosa darah. *Logistic Regression* kemudian diterapkan untuk membuat model prediktif yang dapat menilai probabilitas seseorang dalam memiliki penyakit diabetes.

Ada banyak algoritma klasifikasi yang dapat dipakai di machine learning, tetapi dalam kasus ini penulis menggunakan algoritma *Logistic Regression* dan K-Nearest Neighbor (KNN). Alasan penulis memilih algoritma KNN dan *Logistic Regression* adalah karena berdasarkan literasi penulis terhadap penelitian terdahulu, kedua algoritma tersebut telah terbukti memiliki akurasi yang baik saat digunakan untuk memprediksi suatu objek dan telah terbukti efektif dalam klasifikasi binary, seperti prediksi penyakit jantung. *Logistic Regression* unggul dalam memodelkan hubungan linier antara variabel input dan output, sehingga cocok untuk dataset yang terstruktur dengan baik. KNN, di sisi lain, bekerja berdasarkan kedekatan data, yang memungkinkan analisis berbasis pola non-linear. Dengan membandingkan kedua algoritma, penelitian ini dapat mengevaluasi keunggulan model linier (*Logistic Regression*) dan model berbasis jarak (KNN) dalam konteks prediksi penyakit jantung. Penulis juga menggunakan dataset yang diambil dari UCI Machine Learning Repository.

Penelitian ini diharapkan dapat membantu peneliti di bidang penyakit jantung dalam memilih model-model atau metode metode yang tepat didalam penelitian nya dan dapat membantu peningkatan layanan ahli medis kesehatan di seluruh dunia.

## 1.2 Identifikasi Masalah

Berdasarkan latar belakang yang telah dijabarkan, maka dapat diidentifikasi beberapa masalah sebagai berikut:

1. Penyakit jantung menjadi penyebab utama kematian di berbagai negara. Meskipun upaya pencegahan telah dilakukan, angka kematian akibat penyakit jantung terus meningkat dari tahun ke tahun. Oleh karena itu,

diperlukan penelitian untuk meningkatkan pemahaman dan prediksi terhadap penyakit ini guna mengurangi angka kematian.

2. Penyakit jantung juga merupakan penyakit yang menempati peringkat ke 2 dalam 10 penyakit paling berbahaya di dunia. Oleh karena itu, diperlukan teknologi yang dapat dengan cepat memprediksi penyakit jantung.
3. Proses diagnosa penyakit jantung memerlukan serangkaian tes yang meliputi pemeriksaan fisik, tes darah, elektrokardiogram, dan lain-lain. Namun, beberapa metode tersebut mungkin tidak efisien atau kurang akurat dalam mendeteksi penyakit jantung pada tahap awal. Oleh karena itu, diperlukan penelitian untuk mengevaluasi dan membandingkan metode diagnostik yang berguna untuk meningkatkan keakuratan dan efisiensi dalam mendeteksi penyakit jantung.
4. Teknologi kesehatan akan terus berkembang, termasuk penggunaan teknik machine learning dalam prediksi penyakit. Namun, penerapan teknologi ini dalam mendiagnosis penyakit jantung masih perlu dieksplorasi lebih lanjut. Penelitian ini bertujuan untuk memanfaatkan kemajuan teknologi dalam pengembangan model prediktif yang lebih akurat dan efektif untuk mendiagnosis penyakit jantung.
5. Untuk mengembangkan model prediktif yang baik, diperlukan data yang berkualitas dan representatif tentang pasien penyakit jantung. Namun, seringkali data kesehatan tersedia dalam jumlah terbatas atau tidak lengkap. Oleh karena itu, diperlukan penelitian untuk mengidentifikasi sumber data yang dapat dipercaya dan mengembangkan strategi untuk mengatasi keterbatasan data tersebut.

### **1.3 Pembatasan Masalah**

Berdasarkan latar belakang masalah dan identifikasi masalah, maka pembatasan masalah dibatasi pada:

1. Penelitian masih dilakukan dengan menggunakan dataset yang tersedia pada UCI Machine Learning Repository dan dapat digunakan secara umum di Kaggle.

2. Dataset yang digunakan merupakan data pasien penyakit jantung yang dikumpulkan dari beberapa rumah sakit dan instansi. Diantaranya adalah :
  - a) Hungarian Institute of Cardiology, Budapest.
  - b) University Hospital, Zurich, Switzerland.
  - c) University Hospital, Basel, Switzerland
  - d) V.A. Medical Center, Long Beach and Cleveland Clinic Foundation
3. Dataset yang digunakan terdiri dari 1025 data pasien dengan 76 atribut, tetapi hanya 14 atribut yang digunakan
4. Dataset yang digunakan merupakan rekam medis pasien dari beberapa rumah sakit tetapi identitas dari pasien yang bersangkutan dirahasiakan untuk privasi
5. Dataset yang digunakan berisi atribut data dengan variable berikut: *Age, Sex, Chest Pain Type, Resting Blood Pressure, Serum cholestoral dalam mg/dl, Fasting blood sugar > 120 mg/dl, Resting electrocardiographic result, Maximum Heart Rate Achieved (Thalac), The Slope of the peak exercise ST segment, Exercise Induced Angina, Old Peak, CA (Number of Major Vessels), Thal.*
6. Penelitian ini membuat model machine learning menggunakan algoritma *Logistic Regression* dan Klasifikasi *Nearest Neighbor*.

#### **1.4 Perumusan Masalah**

Berdasarkan latar belakang yang telah diuraikan diatas maka, bisa dirumuskan permasalahan yang dapat diambil yaitu Bagaimana hasil analisis perbandingan prediksi penyakit jantung menggunakan metode *Logistic Regression*, dan *K-Nearest Neighbor* (KNN)?.

#### **1.5 Tujuan Penelitian**

Penelitian ini mempunyai tujuan yaitu untuk mendapatkan analisa data dengan metode *Logistic Regression*, dan *K-Nearest Neighbor* (KNN) pada klasifikasi penyakit jantung, dan hasil akurasi dari metode-metode yang digunakan tersebut, serta bagaimana hasil perbandingan antara *Logistic Regression*, dan *K-Nearest Neighbor* (KNN) yang digunakan dan metode terbaik dari prediksi penyakit jantung.



## 1.6 Manfaat Penelitian

Menciptakan data yang akurat mengenai hasil perbandingan klasifikasi dalam memprediksi penyakit jantung. Sesuai dengan tujuan penelitian, maka penulis berharap dapat memperoleh manfaat sebagai berikut:

1. Penelitian ini diharapkan dapat memberikan pemahaman dan pandangan tentang penyakit jantung dan betapa pentingnya kita menjaga pola hidup dengan sehat.
2. Penelitian ini juga diharapkan dapat memberikan kontribusi terhadap pemahaman kita tentang faktor-faktor risiko yang terkait dengan perkembangan penyakit tersebut, seperti gaya hidup, genetika, dan lingkungan.
3. Mengkaji dan memahami metode *Logistic Regression*, dan *K-Nearest Neighbor* (KNN) dengan baik dan juga menambah wawasan serta ilmu pengetahuan dalam merancang dan membangun sistem informasi Memprediksi penyakit jantung di dunia.
4. Diharapkan dengan dibuatnya perbandingan model data mining klasifikasi dalam memprediksi penyakit jantung di maka kita sebagai manusia dapat lebih menjaga pola hidup sehat dengan baik.

*Intelligentia - Dignitas*