

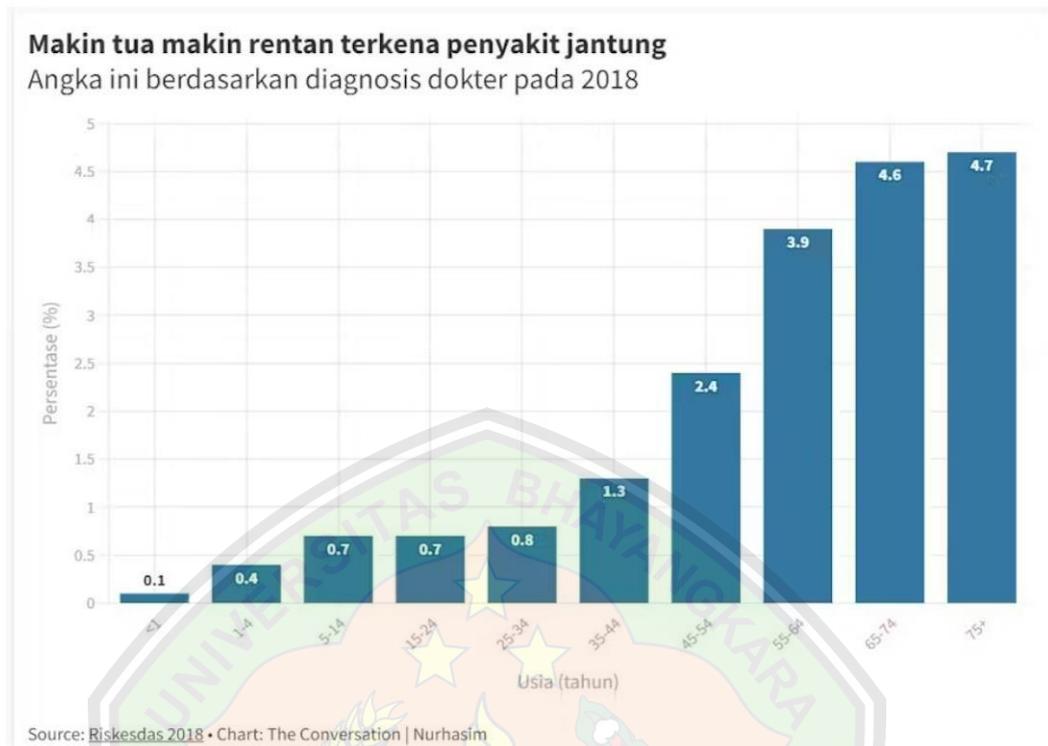
BAB I

PENDAHULUAN

1.1 Latar Belakang

Klasifikasi dalam data mining adalah teknik untuk memisahkan data ke dalam kelas berbeda. Salah satu algoritma klasifikasi adalah random forest. Klasifikasi dengan random forest dilakukan dengan membangun beberapa pohon keputusan yang menghasilkan satu suara prediksi dari suara mayoritas (1). Pohon keputusan dibangun dengan menggunakan *Classification and Regression Trees (CART)*. *Random forest* dapat menghasilkan akurasi yang tinggi dibandingkan dengan metode klasifikasi lainnya. Hanya saja saat *random forest* dijalankan pada *imbalanced* data akan mengakibatkan kemampuan klasifikasi menurun. Dalam algoritma pengklasifikasian termasuk *random forest* menganggap bahwa kelas minoritas ini sebagai *noise* atau *outlier* yang akan diabaikan, sehingga hasil prediksi klasifikasi cenderung pada kelas mayoritas. Akibatnya akurasi dari prediksi untuk kelas minoritas akan jauh lebih kecil dibandingkan kelas mayoritas.

Jantung adalah salah satu organ tubuh yang berperan sangat penting bagi manusia karena berfungsi memompa darah ke seluruh tubuh. Menurut data organisasi Kesehatan dunia (WHO) tahun 2018 bahwa lebih dari 17 juta orang di dunia meninggal akibat penyakit jantung dan pembuluh darah yang merupakan penyebab kematian tertinggi di dunia. Terjadinya penyakit jantung karena penyempitan dan penghambatan pembuluh darah jantung sehingga penyaluran oksigen dan nutrisi ke jantung berkurang. Jika gejala penyakit jantung diketahui maka dapat dilakukan antisipasi (2). Data yang terkumpul dapat dianalisis hingga dapat mengungkapkan informasi berguna yang tersembunyi. Dengan melakukan penambangan data, yang merupakan ilmu baru, kami dapat mengekstrak pengetahuan tersembunyi dari data tersebut. Fungsi inti dari data mining menerapkan berbagai metode dan algoritma untuk menemukan data yang disimpan.



Gambar 1.1 Data Presentasi Penyakit Jantung

Sumber : Penelitian [3]

Dengan memprediksi resiko penyakit jantung dapat mengetahui seseorang terkena penyakit jantung atau tidak dari gejala yang ditimbulkan. Penelitian di bidang ilmu data dan kecerdasan buatan telah memberikan kontribusi yang signifikan dalam pemahaman dan prediksi penyakit, termasuk penyakit jantung [3].

Untuk dilakukan pencegahan lain dengan menggunakan bantuan machine learning untuk membantu klasifikasi terhadap resiko terkena penyakit jantung, yakni akan terkena atau sudah terkena penyakit jantung atau belum terkena penyakit jantung. Untuk memudahkan proses identifikasi orang yang memiliki jantung maka digunakan machine learning. Machine learning merupakan mesin pembelajaran dengan proses komputasi dan menggunakan data input. Machine learning merupakan metode untuk mempermudah proses identifikasi klasifikasi pada penelitian ini. Hal ini terjadi karena penyakit jantung merupakan penyebab kematian yang paling utama.

Salah satu pendekatan yang telah terbukti efektif adalah penggunaan metode *Random forest* dalam analisis prediktif. Metode ini merupakan salah satu teknik machine learning yang mampu menghasilkan prediksi akurat dengan memanfaatkan ensambel pohon keputusan. Penelitian sebelumnya dalam domain ini telah menunjukkan bahwa analisis data medis dengan menggunakan *Random forest* mampu mengidentifikasi pola kompleks dari berbagai variabel klinis yang terkait dengan risiko penyakit jantung. Hal ini memberikan potensi besar untuk membangun model prediksi yang dapat membantu dalam deteksi dini, pengelolaan, dan pencegahan penyakit jantung. Skripsi ini bertujuan untuk mengembangkan dan mengevaluasi model prediksi penyakit jantung berbasis metode *Random forest*. Data medis yang relevan, seperti riwayat kesehatan, hasil tes diagnostik, dan faktor risiko yang terkait, akan digunakan untuk melatih dan menguji model prediksi ini. Penelitian ini diharapkan dapat memberikan kontribusi yang berarti dalam bidang kedokteran predictive *analytics*, dengan menyediakan alat yang dapat membantu praktisi medis dalam melakukan identifikasi dini dan pengelolaan penyakit jantung secara lebih efektif. Maka dari itu pada penelitian skripsi dengan teknik data mining, penulis ingin membuat sistem prediksi tahap awal penyakit Jantung menggunakan algoritma *Random forest*. diharapkan pada penelitian ini bisa mendapatkan nilai akurasi yang lebih tinggi daripada penelitian sebelumnya.

1.2 Identifikasi Masalah

Berdasarkan dari uraian latar belakang yang sudah dijabarkan diatas, dapat diidentifikasi masalah yang ada, diantaranya adalah :

1. Terlambatnya mendiagnosa penyakit Jantung menjadi faktor utama bertambahnya pasien.
2. Kurang Informasi terkait bahayanya penyakit Jantung yang dapat mengakibatkan komplikasi.

1.3 Rumusan Masalah

Berdasarkan latar belakang yang sudah dijabarkan diatas, dapat di Rumuskan masalah yang ada, diantaranya adalah :

1. Bagaimana caranya membuat sistem Prediksi tahap awal penyakit Jantung dengan menggunakan Algoritma *Random forest*?
2. Bagaimana caranya membuat informasi tentang penyakit Jantung?

1.4 Batasan Masalah

Batasan masalah ini dilakukan agar penelitian ini lebih fokus dengan yang tertara dijudul, adapun batasan masalahnya di antaranya :

1. Algoritma yang dilakukan penelitian ini menggunakan algoritma *Random forest*.
2. Penyakit Jantung pada penelitian ini bersifat global tidak spesifik.

1.5 Tujuan dan Manfaat

1.5.1 Tujuan Penelitian

Adapun tujuan dari penelitian ini, diantaranya :

1. Melakukan prediksi resiko penyakit jantung menggunakan metode *Random forest*.
2. Menghitung ketepatan data atau tingkat akuraasi dalam memprediksi resiko penyakit jantung menggunakan metode *Random forest*.

15.2 Manfaat Penelitian

Adapun manfaat dari penelitian ini, di antaranya :

1. Menambah pengetahuan penulis tentang memprediksi resiko penyakit jantung menggunakan metode *Random forest*.
2. Penelitian ini diharapkan dapat membantu dan mempermudah tenaga medis dalam memprediksi resiko penyakit jantung.
3. Sebagai referensi dan sumber Pustaka untuk penulisan penelitian yang sejenis.

1.6 Sistematika Penulisan

Sistematika penulisan adalah penggambaran susunan tulisan itu sendiri, yang dilakukan secara teratur dan rinci untuk memberikan gambaran yang utuh Sistem penulisan Skripsi ini mencakup lima bab berikut :

BAB I PENDAHULUAN

Bab pendahuluan ini mendeskripsikan mengenai latar belakang masalah, identifikasi masalah, rumusan masalah, tujuan penelitian, manfaat penelitian dan sistematika penulisan.

BAB II LANDASAN TEORI

Pada bab ini akan menjelaskan teori-teori seperti Algoritma *Random forest*, Python, *website*, dan penyakit Jantung yang berhubungan dengan judul skripsi ini.

BAB III METODOLOGI PENELITIAN

Pada bab ini menjelaskan mengenai objek penelitian, kerangka penelitian, analisis sistem berjalan, permasalahan, analisis usulan sistem dan analisis kebutuhan sistem.

BAB IV HASIL DAN PEMBAHASAN

Bab ini menjelaskan tentang proses perancangan, implementasi dan pengujian. Mulai dari perancangan dan analisis program, gambaran rancangan secara umum dan analisis kebutuhan perangkat lunak yang digunakan dalam pembuatan sistem ini.

BAB V PENUTUP

Bab ini merupakan bab terakhir yang berisi kesimpulan, dan saran yang di usulkan pengembangan lebih lanjut agar tercapai hasil yang lebih baik.

DAFTAR PUSTAKA

Bab ini berisi tentang sumber-sumber yang diambil oleh penulis dalam mengerjakan penelitian proposal Skripsi.