

BAB I

PENDAHULUAN

Menurut World Health Organization (WHO), penyakit jantung menelan korban sebanyak 17,9 juta jiwa setiap tahunnya (*Cardiovascular Diseases*, n.d.). Di Indonesia, terdapat 15,5 juta kasus penyakit jantung pada tahun 2022. Hal ini menempatkan penyakit jantung sebagai salah satu penyakit yang paling mematikan di Indonesia (*Kematian akibat Penyakit Jantung di Indonesia Terus Meningkat - Dataindonesia.id*, 2023). Penyakit jantung terjadi karena adanya gangguan dari jantung dan pembuluh darah (Ghani et al., 2016).

Sayangnya gejala maupun penyakit jantung jarang disadari oleh penderitanya. Faktor yang melandasi hal tersebut karena penderita umumnya lalai atau tidak memiliki cukup biaya untuk mengecek kesehatan. Tapi di samping itu, sebenarnya penyakit jantung dapat dideteksi dini dengan memperhatikan beberapa gejalanya, seperti nyeri dada di sebelah kiri, irama jantung tak teratur, detak jantung yang terkadang sangat cepat atau terlalu lambat, serta sesak napas (Kevin, 2022). Banyak faktor yang menyebabkan penyakit jantung dapat terjadi, di antaranya keturunan, usia, sosioekonomi, makanan yang dikonsumsi, serta pola hidup. Selain itu, terdapat faktor risiko dominan yang menyebabkan penyakit jantung, yaitu hipertensi, gangguan mental emosional, dan diabetes melitus (Ghani et al., 2016). Oleh karena itu, tindakan pencegahan sangat perlu dilakukan dalam mengidentifikasi faktor risiko penyakit jantung.

Di zaman serba digital ini, identifikasi faktor risiko penyakit jantung dapat dilakukan dengan memanfaatkan teknologi *machine learning*. Dengan *machine learning* maka dapat diidentifikasi apakah seseorang sudah menderita penyakit jantung, akan menderita penyakit jantung, atau belum

menderita penyakit jantung. *Machine learning* merupakan bidang ilmu komputer untuk membuat mesin menjadi cerdas melalui pelatihan berulang kali (Pratama et al., 2017). Penerapan *machine learning* memanfaatkan algoritma dalam membuat klasifikasi untuk mencapai kesimpulan. Dua jenis algoritma yang akan digunakan dalam penelitian ini adalah *K-Nearest Neighbor* (KNN) dan *Random Forest*.

KNN merupakan algoritma yang memanfaatkan sekumpulan data untuk klasifikasi berdasarkan ukuran kemiripannya (Diansyah, 2022). Secara umum, algoritma ini sangat sederhana karena hanya mengukur jarak terdekat dari *query instance* ke *training sample* (Lestari, 2015). Diansyah (2022) pernah menggunakan KNN pada klasifikasi tingkat kepuasan pengguna transportasi publik di Pekanbaru dan menghasilkan tingkat akurasi sebesar 94,12%. Selain KNN, terdapat algoritma *Random Forest* yang menggunakan metode biner rekursif untuk mencapai ambang batas terbaik di dalam memisahkan data (Religia et al., 2021). Dalam penelitian yang dilakukan oleh Purbolaksono et al. (2021), diketahui bahwa metode *Random Forest* memiliki akurasi sebesar 97,8% dalam memprediksi pasien penyakit diabetes dan lebih baik daripada metode *Support Vector Machine* (SVM).

Penyakit jantung menjadi salah satu penyakit mematikan di Indonesia bahkan dunia. Penderita penyakit jantung umumnya tidak menyadari keberadaan penyakit tersebut. Banyak sekali faktor yang dapat menyebabkan penyakit jantung, sehingga perlu sebuah metode yang akurat dalam mengklasifikasi penyakit tersebut. Salah satu cara yang dapat digunakan adalah dengan memanfaatkan *machine learning*. Dalam penelitian ini akan dikembangkan *machine learning* dengan dua buah algoritma sebagai perbandingan akurasinya, yaitu algoritma KNN dan *Random Forest*. Kedua algoritma tersebut sudah pernah digunakan dalam penelitian klasifikasi penyakit diabetes dan memiliki tingkat akurasi sangat

baik. Berdasarkan hal tersebut, maka penelitian ini bertujuan untuk mengklasifikasi faktor risiko penyakit jantung menggunakan algoritma KNN dan *Random Forest*.

1.1 Perumusan Masalah

1. Tingkat kasus penyakit jantung Indonesia sangat tinggi.
2. Para penderita penyakit jantung tidak sadar terhadap penyakit yang dialaminya.
3. Banyaknya faktor yang dapat menyebabkan penyakit jantung.
4. Perlu adanya metode prediksi yang akurat dalam mengklasifikasi faktor risiko penyakit jantung.

1.2 Tujuan

1. Menentukan faktor risiko yang paling memengaruhi penyakit jantung.
2. Menentukan algoritma terbaik antara KNN dan *Random Forest* untuk mengklasifikasi faktor risiko penyakit jantung.

1.3 Batasan Masalah

Batasan masalah dalam penelitian ini terletak pada:

1. Penggunaan dua algoritma saja untuk memprediksi risiko penyakit jantung, yaitu KNN dan *Random Forest*.

1.4 Manfaat Penelitian

1. Penelitian ini dapat memberikan wawasan lebih jauh tentang dua algoritma, yaitu *K-Nearest Neighbor* dan *Random Forest*, terkhusus dalam memprediksi risiko penyakit jantung.
2. Penelitian ini mengembangkan model perkiraan risiko penyakit jantung yang lebih akurat, sehingga dapat membawa manfaat praktis bagi penderita gejala penyakit jantung.