

BAB 1

PENDAHULUAN

1.1. Latar Belakang

Penyakit jantung termasuk kedalam kelompok penyakit *Cardiovascular (CVD)*. Di negara-negara maju bahkan di dunia, penyakit Cardiovascular ini menjadi penyakit dengan paling mematikan dan berbahaya. Jantung sendiri mempunyai 4 bagian ruang, yaitu ruang atas dan ruang bawah. Beberapa gejala yang terjadi pada orang terkena penyakit jantung yaitu kelelahan, keringan yang berlebihan, dan sebagainya. Gejala-gejala ini bisa berbeda-beda tergantung pada setiap orang[1]. Sebanyak 17.9 juta orang meninggal di tahun 2017 menurut data yang diperoleh Organisasi Kesehatan Dunia (WHO)[2]. Di tahun 2030 jumlah penderita penyakit jantung akan meningkat menjadi 23.3 juta orang[3].

Di Amerika Serikat, kasus kematian akibat jantung menjadi peringkat teratas. Sebagian orang tidak menyadari penyebab terjadinya penyakit jantung. Di tahun 2018 menurut data Riset Kesehatan Dasar (Riskesdas) penderita penyakit jantung di Indonesia sebanyak 4.2 juta orang yang terkena penyakit jantung. Tetapi sorotan utama terletak pada negara-negara yang berkembang di mana penyakit jantung mengalami kenaikan kasus[4]. Faktor-faktor yang membuat terjadinya penyakit jantung adalah kurang olahraga, kurang tidur, terlalu banyak minum alkohol, merokok, keturunan, diabetes dan sebagainya[5]. Dokter sangat bisa terbantu dalam memprediksi dini penyakit jantung melalui pembelajaran mesin yang bisa mengenali pola yang tidak terlihat dan juga akan memberikan pemahaman secara klinis[6].

Data mining mempunyai peranan yang sangat besar dalam dunia kedokteran, pendidikan, teknik, bisnis, dan lainnya[7]. Peranan penting dari data mining karena data mining dapat menentukan pola dalam kumpulan data besar menggunakan teknik yang menggabungkan pembelajaran mesin, statistik, dan basis data. Data mining untuk klasifikasi, prediksi, dan memperkirakan untuk memperoleh informasi yang berguna[8]. *Logistic regression* adalah algoritma yang dibuat untuk nebak hasil. Hasil yang dikeluarkan dari *logistic regression* adalah nilai biner yaitu angka 0 atau 1. Di dalam dunia kesehatan, algoritma ini sangat cocok dalam memprediksi penyakit, terutama penyakit jantung[9][10].

Pada penelitian sebelumnya yang dilakukan oleh Ahmad Yogiarto, dkk[11] untuk mengimplementasikan metode *K-Nearest Neighbors* untuk klasifikasi penyakit

jantung yang mendapatkan nilai akurasi sekitar 64.03% tetapi nilai akurasi tersebut masih belum optimal untuk mengklasifikasikan penyakit jantung.

Sedangkan penelitian yang dilakukan oleh Arni Sepharni, dkk[12] untuk klasifikasi penyakit jantung menggunakan C4.5” Penelitian ini melakukan perhitungan dengan menggunakan metode C4.5 yang menghasilkan tingkat akurasi sekitar 79% Sehingga masih diperlukan nilai yang tinggi untuk memprediksi penyakit jantung dengan algoritma yang lain. Perlu dilakukan pengembangan algoritma lain untuk memprediksi penyakit jantung. Dengan demikian penulis tertarik untuk melakukan penelitian dengan judul **Implementasi Algoritma Logistic Regression Dalam Memprediksi Penyakit Jantung.**

1.2. Rumusan Masalah

Rumusan masalah penelitian antara lain:

1. Bagaimana tingkat akurasi algoritma *Logistic Regression* dalam memprediksi penyakit jantung?
2. Bagaimana memprediksi penyakit jantung dengan menggunakan algoritma *Logistic Regression*?
3. Faktor-faktor apa saja yang dapat menyebabkan orang terkena penyakit jantung?

1.3. Tujuan dan Manfaat

1.3.1. Tujuan

Tujuan dari penelitian adalah sebagai berikut:

1. Memprediksi penyakit penyakit jantung dengan menggunakan metode *Logistic Regression*.
2. Implementasi algoritma *Logistic Regression* untuk prediksi penyakit jantung.
3. Mengukur akurasi prediksi penyakit jantung dengan menggunakan metode tersebut.

1.3.2. Manfaat

Manfaat yang diharapkan dari dilakukannya antara lain:

1. Membantu para tenaga medis atau dokter untuk memprediksi dini penyakit jantung.
2. Membantu masyarakat untuk untuk mengetahui akan bahayanya penyakit jantung.

1.4. Batasan Masalah

Batasan Masalah penelitian adalah sebagai berikut:

1. Penelitian menggunakan algoritma *Logistic Regression* untuk memprediksi penyakit jantung.
2. Menggunakan bahasa pemrograman python di jupyter notebook.
3. Berkonsentrasi pada logistic regression dalam memprediksi penyakit jantung.
4. Parameter yang digunakan normal dan terkena penyakit jantung.
5. Dataset dipakai berbentuk csv yang diperoleh dari <https://www.kaggle.com/datasets/eishkaran/heart-disease>

1.5. Keterbaruan

Tabel 1.1 Keterbaruan

No	Penelitian Sebelumnya	Keterbaruan
1	Aditya Pratama et al [13] melakukan penelitian untuk memprediksi penyakit diabetes dengan algoritma logistic regression. Hasil penelitian menunjukan logistic regression dapat memprediksi penyakit diabetes dengan akurasi sebesar 82%.	Meskipun algoritma yang digunakan sama dengan penelitian sebelumnya yaitu logistic regression. Tetapi akurasi yang dihasilkan dari penelitian tersebut masih rendah. Kami ingin meningkatkan akurasi nya lebih tinggi lagi.
2	Fitran Dwi Pramakrisna et al [14] mengimplementasi algoritma logistic regression untuk mengklasifikasikan sms. Hasil akurasi yang didapatkan adalah sekitar 97%.	Meskipun algoritma yang dipakai sama dengan penelitian sebelumnya yaitu logistic regression. Akan tetapi dataset yang dipakai pada penelitian sebelumnya terkait dengan spam sms. Kami ingin memakai dataset medis jantung yang mengandung fitur-fitur klinis pasien jantung.

No	Penelitian Sebelumnya	Keterbaruan
3	Yahya Anugerah Dwi et al [15] menerapkan algoritma logistic regression untuk mengklasifikasikan diagnose kanker payudara. Hasil yang didapatkan dari penelitian adalah dengan akurasi 98%.	Meskipun algoritma yang digunakan sama dengan penelitian sebelumnya yaitu logistic regression. Tetapi dataset yang dipakai pada penelitian sebelumnya yaitu dataset pasien yang terkena penyakit diabetes. Kami ingin memakai dataset pasien terkena penyakit jantung
4	Suhyyah et al [16] melakukan penelitian untuk mengklasifikasi penyakit stroke dengan menggunakan algoritma logistic regression. Hasil yang didapatkan dari penelitian dengan logistic regression adalah dengan akurasi 94%.	Meskipun algoritma yang digunakan sama seperti dengan penelitian sebelumnya yaitu logistic regression, akan tetapi dataset yang dipakai sebelumnya adalah dataset pasien penyakit stroke. Kami ingin memakai dataset pasien yang terkena penyakit jantung.
5	Hendra et el [17] melakukan penelitian untuk mendeteksi spam sms dengan algoritma logistic regression. Hasil yang diperoleh dari penelitian ini mendapatkan nilai akurasi sebesar 96%	Meskipun algoritma yang digunakan sama seperti dengan penelitian sebelumnya yaitu logistic regression, akan tetapi dataset yang dipakai penelitian sebelumnya adalah dataset spam sms. Kami ingin memakai daset jantung pasien.