

Memprediksi Penyakit Gagal Jantung Menggunakan Algoritma Regresi pada Machine Learning

Given Jeremia^{1*}, Salsabilla Aryaning Putri², Safira Arinta Azzahra³

Program Studi Teknik Informatika, Universitas Surabaya, Surabaya, Jawa Timur

Email: ^{1*}s160419118@student.ubaya.ac.id, ²s160419135@student.ubaya.ac.id, ³s160419158@student.ubaya.ac.id

(Naskah masuk: 25 November 2021, diterima: 25 November 2021)

Abstrak

Penyakit jantung adalah kondisi ketika jantung mengalami sebuah gangguan. Salah satu bentuk dari gangguan tersebut adalah gagal jantung. Gagal jantung adalah kondisi ketika pompa jantung melemah, sehingga tidak mampu mengalirkan darah yang cukup ke seluruh tubuh. Gejala gagal jantung dapat berkembang secara bertahap maupun terjadi secara tiba-tiba. Hal ini sebenarnya bisa dicegah menggunakan teknologi yang ada pada saat zaman ini. Perkembangan teknologi meningkat secara pesat, khususnya pada *machine learning* yang dapat digunakan untuk membantu memprediksi penyakit gagal jantung dengan menggunakan data pasien yang sudah ada. Salah satu algoritma pada *machine learning* yang dapat digunakan untuk memprediksi penyakit gagal jantung adalah Algoritma Regresi. Algoritma Regresi adalah metode yang berfungsi mengembalikan nilai numerik pada setiap sampel yang bertujuan untuk menentukan hubungan sebab dan akibat antar variabel.

Kata Kunci: Penyakit Jantung, Gagal Jantung, Perkembangan Teknologi, Algoritma Regresi, Machine Learning.

Predicting Heart Failure Using Machine Learning Regression Algorithm

Abstract

Heart Disease is any disorder that affects the heart. One form of the disorder is heart failure. Heart failure is a condition when the heart's pump weakens, making it unable to circulate enough blood throughout the body. Symptoms of heart failure can develop gradually or occur suddenly. This can actually be prevented using existing technology at this time. Technological developments are increasing rapidly, especially in machine learning which can be used to help predict heart failure using existing patient data. One algorithm in Machine Learning that can be used to predict heart failure is the Regression Algorithm. Regression Algorithm is a method that functions to return a numerical value in each sample which aims to determine cause and effect relationships between variables.

Keywords: Heart Disease, Heart Failure, Technological Development, Regression Algorithm, Machine Learning.

I. PENDAHULUAN

Jantung menjadi salah satu organ vital di dalam tubuh makhluk hidup dengan fungsi utama memompa darah untuk diedarkan ke seluruh tubuh. Karena kegunaannya yang cukup penting, maka jika terjadi gangguan pada jantung

dapat menyebabkan penurunan kesehatan secara drastis. Penyakit jantung merupakan salah satu penyebab kematian tertinggi di Indonesia (Kemenkes RI, 2017 dalam Pangaribuan et al, 2021). Penyakit jantung adalah kondisi ketika jantung mengalami gangguan fungsi. Terjadinya penyakit jantung disebabkan oleh infeksi atau bisa juga

disebabkan oleh risiko bawaan sejak lahir. Penyakit jantung mempunyai macam yang berbeda-beda, seperti penyakit jantung koroner, serangan jantung, aritmia, gagal jantung, dan masih banyak jenisnya. Dalam banyak kasus, penyakit jantung terjadi akibat penumpukan plak di arteri koroner, sehingga aliran darah ke jantung terhambat dan meningkatkan risiko serangan jantung serta komplikasi lainnya. Ketidakmampuan jantung dalam memompa cukup darah ke tubuh juga mengakibatkan gagal jantung yang dapat dipicu juga oleh faktor risiko penyakit lain seperti diabetes, darah tinggi, atau kondisi gangguan fungsi jantung yang lain. Penulisan artikel ini akan berfokus pada penyakit gagal jantung.

Gagal jantung adalah salah satu sindrom klinis dari berbagai jenis penyakit jantung pada stadium yang parah. Gagal jantung merupakan salah satu jenis penyakit kardiovaskuler yang paling umum. Hal ini telah menjadi masalah kesehatan utama yang terus berkembang. Gagal jantung terjadi ketika ada penyumbatan pada paru-paru atau jaringan tubuh lainnya karena jantung tidak dapat memompa darah dan gagal mengedarkan oksigen ke seluruh tubuh. Gagal jantung akan menyebabkan penurunan bertahap pada curah jantung dan gangguan sirkulasi yang akan memunculkan gejala dan tanda gagal jantung, dan ini disebut sindrom gagal jantung. Kinerja pompa jantung yang buruk dapat mengakibatkan suplai oksigen dari jantung tidak mencukupi kebutuhan metabolisme tubuh, dan mungkin ada obstruksi sirkulasi pulmonal. Gejala gagal jantung bisa bermacam-macam, seperti kelelahan, penurunan kemampuan aktivitas fisik, dan kesulitan bernapas (Yuwandini, 2020).

Dengan kemajuan teknologi, penyakit tersebut bisa dideteksi lebih awal sehingga angka pasien penyakit jantung diharapkan dapat diminimalisir. Mengingat organ jantung merupakan organ yang vital, maka memprediksi risiko gagal jantung haruslah menjadi prioritas di dunia medis. Namun, hingga saat ini, pendeteksian penyakit pada jantung, khususnya gagal jantung, belum mencapai tingkat akurasi yang tinggi (Chicco & Jurman, 2020). Dalam konteks ini, catatan medis elektronik dari pasien menjadi sangat berguna sebagai sumber informasi yang dapat dianalisis untuk mengungkapkan risiko gagal jantung pada pasien.

Machine learning yang diterapkan pada rekam medis dapat menjadi alat yang ampuh untuk memprediksi kelangsungan hidup pasien gagal jantung dan untuk mengidentifikasi tanda-tanda klinis yang paling penting (atau faktor risiko) yang dapat menyebabkan gagal jantung. Ilmuwan dapat menggunakan *machine learning* lebih dari sekadar untuk prediksi klinis. Kecerdasan komputer, khususnya, menunjukkan kekuatan prediktif ketika diterapkan pada catatan medis atau dalam kombinasi dengan pemeriksaan medis. *Machine learning* sendiri bukanlah suatu teknologi yang dapat memberikan prediksi dengan tingkat akurasi 100% dalam hal analisa data. Namun, keakuratan yang dihasilkan oleh *machine learning* sudah cukup untuk

dapat menjadi dasar untuk menarik kesimpulan terkait kondisi pasien di dunia medis (Pangaribuan et al, 2021)

Pada *machine learning* terdapat berbagai macam metode yang dapat digunakan untuk menyelesaikan suatu masalah. Pemilihan metode yang akan digunakan berdasarkan jenis analisis yang dibutuhkan. Algoritma Regresi adalah salah satu metode pada *machine learning* yang bertujuan untuk mengembalikan target numerik pada setiap sampel. Berbeda dengan klasifikasi yang merupakan metode *machine learning* yang bekerja dengan cara memberikan label untuk setiap sampel menggunakan dua atau lebih kelompok yang berbeda.

Dari uraian di atas, penelitian ini bertujuan untuk mengetahui secara pasti bagaimana algoritma regresi dalam *machine learning* dapat digunakan untuk mendeteksi risiko gagal jantung pada manusia.

II. STUDI LITERATUR

Machine Learning

Machine learning atau pembelajaran mesin adalah salah satu bidang kecerdasan buatan (*Artificial Intelligence*). Secara khusus, mempelajari bagaimana komputer dapat belajar dari data untuk meningkatkan kecerdasan buatan. Komputer memiliki kemampuan yang dapat belajar dari pengalaman tugas-tugasnya dan sangat memungkinkan untuk dapat memiliki peningkatan kinerja (Mitchell, 1997 dalam Puteri & Silvanie, 2020). *Machine learning* juga menjadi wujud perkembangan teknologi yang bertujuan untuk mengganti atau meniru perilaku manusia dalam menyelesaikan permasalahan terkomputerisasi sehingga akan tercipta suatu otomatisasi sistem.

Machine learning merupakan kemampuan komputer untuk belajar tanpa penjelasan secara eksplisit atau pemrograman terlebih dahulu pada komputer (Arthur, 1959 dalam Puteri & Silvanie, 2020). Dalam pemrograman tradisional data dan program dijalankan di komputer untuk menghasilkan *output*, sementara dalam pemrograman *machine learning* data dan *output* dijalankan oleh komputer untuk membuat suatu program. Proses pada *machine learning* dapat dibagi menjadi beberapa hal, termasuk pengumpulan data (*data collect*), persiapan input data (*data input preparation*), analisis data input (*analyse input data*), keterlibatan manusia (*human involvement*), algoritma pembelajaran (*train algorithm*), uji algoritma (*test algorithm*), dan mengimplementasikan model (*use it*) (Puteri & Silvanie, 2020).

Seperti namanya, *machine learning* mencoba meniru cara manusia atau makhluk hidup mempelajari dan menyimpulkan proses. Pembelajaran mesin memiliki setidaknya dua aplikasi utama yaitu klasifikasi dan prediksi (Ahmad, 2017). Seperti yang dijelaskan pada paragraf sebelumnya, ciri khas *machine learning* adalah adanya proses pembelajaran, pengajaran, dan pelatihan. Oleh karena itu, *machine learning* membutuhkan informasi pembelajaran atau data untuk dipelajari. Klasifikasi adalah teknik *machine*

learning yang digunakan oleh komputer untuk mengurutkan atau mengklasifikasikan objek menurut karakteristik tertentu ketika orang mencoba membedakan objek satu sama lain. Sedangkan prediksi atau regresi digunakan oleh mesin untuk menebak hasil input berdasarkan informasi yang dipelajari melalui pelatihan.

Machine learning biasanya dibagi menjadi *unsupervised learning* dan *supervised learning* (Pangaribuan et al, 2021). *Supervised learning* adalah metode *machine learning* dalam pemrosesan kumpulan data yang sudah berisi label data dapat dianalisis untuk memungkinkan mesin mengidentifikasi label data dalam proses klasifikasi dan prediksi. Sementara *unsupervised learning* adalah sebuah teknik dalam *machine learning* yang memungkinkan mesin untuk menarik kesimpulan dari data.

Algoritma Regresi

Algoritma memiliki definisi langkah perhitungan penyelesaian masalah. Sedangkan regresi merupakan persamaan matematik yang dapat digunakan untuk memprediksi secara sistematis tentang kemungkinan yang akan terjadi akibat ketergantungan dari variabel tidak bebas terhadap variabel bebas. Algoritma Regresi adalah metode pada *supervised learning* yang mengembalikan target numerik untuk setiap sampel yang bertujuan untuk mencari pola dan menentukan sebuah nilai numerik. Dalam *machine learning*, algoritma regresi memiliki beberapa bentuk diantaranya adalah regresi linear, polinomial, quantile, logistik, ridge dan lasso.

Regresi linier merupakan teknik regresi yang paling sederhana. Dalam metode ini, variabel bersifat berkelanjutan dan hubungan variabel terikat dan variabel bebas diasumsikan linier. Dalam regresi linier asumsi berikut harus dipenuhi: Harus ada hubungan linier antara variabel bebas dan variabel terikat, tidak ada *outlier* pada data, pengamatan data harus independen, serta tidak ada *polyline* dan residual harus berdistribusi normal dengan rata-rata bernilai 0 dan varians konstan. Regresi polinomial adalah teknik regresi untuk menyelesaikan persamaan *nonlinier* dengan menggunakan fungsi polinomial variabel bebas. Jika pada regresi linier garis-garis pada *scatterplot* adalah garis lurus, pada regresi polinomial garis-garis *scatterplot* sedikit melengkung. Regresi logistik adalah regresi dimana variabel terikatnya biner atau ada dua jenis. Namun, jika kita memiliki lebih dari dua jenis variabel terikat, juga dapat digunakan regresi logistik polinomial.

Regresi kuantil merupakan hasil pengembangan lebih jauh dari regresi linier dan sering digunakan ketika ada informasi yang tidak biasa. Ada terlalu banyak *skewness* dan divergensi dalam data. Kelebihan regresi kuantil dibandingkan regresi linier adalah dapat digunakan jika datanya berbeda dan terdapat *outlier*. Distribusi variabel terikat dapat digambarkan dengan menggunakan besaran yang berbeda. Dan itu dapat digunakan jika kemiringan data terlalu tinggi. Jika koefisien kuantil berbeda jauh dengan

koefisien regresi linier, maka dapat menggunakan regresi kuantil.

Sedangkan regresi ridge dapat memberikan estimasi koefisien regresi yang bias dengan melakukan modifikasi pada metode kuadrat terkecil guna mendapatkan pengurangan varian melalui penambahan suatu tetapan ke dalam menstabilkan koefisien yang mampu membantu mengatasi multikolinieritas pada data (Tazliqoh et al, 2015). Sementara itu, regresi lasso adalah teknik regresi berganda yang digunakan untuk mengurangi koefisien aproksimasi menjadi nol, dan *selection* operator adalah pemilihan variabel bebas untuk membuat model dengan variabel terbaik (Robbani et al, 2017).

III. KETENTUAN PENULISAN

A. Penulisan Judul

Memprediksi Penyakit Gagal Jantung Menggunakan Algoritma Regresi pada *Machine Learning*.

B. Penulisan Identitas Penulis

Penulis pertama bernama Given Jeremia seorang mahasiswa semester 5 Program Studi Teknik Informatika di Universitas Surabaya (UBAYA) Jawa Timur. Lahir di Surabaya, 21 Desember 2000.

Penulis kedua bernama Salsabilla Aryaning Putri seorang mahasiswa semester 5 Program Studi Teknik Informatika di Universitas Surabaya (UBAYA) Jawa Timur. Lahir di Surabaya, 6 April 2001.

Penulis pertama bernama Safira Arinta Azzahra seorang mahasiswa semester 5 Program Studi Teknik Informatika di Universitas Surabaya (UBAYA) Jawa Timur. Lahir di Surabaya, 26 April 2001.

C. Penulisan Isi Artikel

Artikel utama ditulis dengan menggunakan huruf *Times New Roman*, ukuran 10 poin, spasi tunggal, dan paragraf rata kiri-kanan (justify) [2]. Artikel secara garis besar terdiri dari subbab Pendahuluan, Metodologi Penelitian, Hasil dan Pembahasan, dan Kesimpulan. Penulisan judul subbab ditulis dengan huruf cetak tebal kapital dengan didahului penomoran menggunakan angka romawi.

Subbab Pendahuluan berisi tentang latar belakang dan analisis permasalahan yang akan diangkat dalam penelitian. Selain itu dalam subbab ini penulis bisa menambahkan studi literatur dari beberapa penelitian terdahulu dengan menuliskan kutipan-kutipan sesuai dengan format.

Subbab Metodologi Penelitian berisi tentang pembahasan metode-metode yang dipakai. Dalam subbab ini harap dijelaskan lebih detail mengenai teori-teori atau algoritma atau metode yang dipakai dalam penelitian. Penulis disarankan pula untuk menuliskan kutipan pustaka dari sumber-sumber jurnal ilmiah dan buku dari terbitan 5 tahun terakhir.

Subbab Hasil dan Pembahasan berisi tentang proses implementasi dan pengujian. Dalam subbab ini harus terlihat

kontribusi penulis terhadap kemajuan ilmu pengetahuan. Selain itu dalam bab ini harus terlihat pengujian-pengujian yang dilakukan. Uraikan proses dan metodenya secara jelas dengan menggunakan blok diagram, *flowchart*, dan lain sebagainya. Data-data berupa gambar, tabel, diagram, dan bentuk data lain yang diambil dari sumber referensi wajib mencantumkan sumbernya secara jelas.

Subbab Kesimpulan berisi tentang penarikan kesimpulan dari penelitian tersebut. Upayakan di bagian kesimpulan ini terdapat data-data yang memperkuat kesimpulan yang diambil. Dengan demikian akan tampak kekuatan dan kekurangan dari penelitian tersebut. Bisa juga dalam subbab ini ditambahkan dengan saran-saran pengembangan penelitian di masa mendatang. Kesimpulan dituliskan dalam bentuk narasi paragraf dan bukan dituliskan dalam bentuk poin.

D. Penulisan Kutipan

Setiap referensi yang tertulis di daftar pustaka wajib dikutip dalam paragraf. Kutipan ditulis dengan menggunakan penomoran dalam tanda kurung siku, misal [1], [2], dan seterusnya sesuai dengan format kutipan IEEE. Penulisan kutipan diurut berdasarkan kemunculan kutipan dalam paragraf dan bukan diurut berdasarkan abjad nama pengarang. Penulis dilarang menambahkan pustaka yang tidak disitasi/dikutip dalam paragraf.

E. Penulisan Daftar Referensi

Daftar referensi hendaknya dipilih sumber referensi yang termutakhir (kurang lebih 5 tahun terakhir) [3]. Daftar referensi tuliskan di bagian akhir dari artikel dengan menggunakan standar penulisan format American Physiological Association (APA). Mohon penulis memperhatikan format penulisan. Penulis disarankan menggunakan aplikasi Mendeley untuk penulisan referensi ini.

IV. METODOLOGI

Tahapan awal dalam melakukan persiapan penelitian yaitu mengidentifikasi masalah. Pada penelitian ini dilakukan identifikasi masalah terhadap seseorang yang memiliki penyakit jantung agar dapat memprediksi penyakit gagal jantung.

Tahapan kedua dalam langkah penelitian ini adalah proses pengumpulan data. Data ini dikumpulkan melalui situs Kaggle (<https://www.kaggle.com>). Data yang digunakan meliputi 11 fitur untuk memprediksi kejadian penyakit jantung. Data asli kemudian dikumpulkan dalam satu file berbentuk csv.

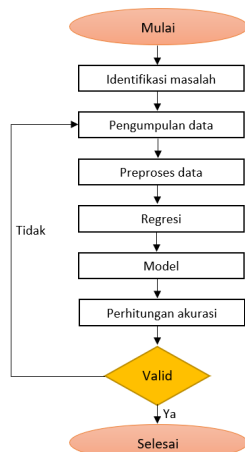
Tahapan ketiga yaitu dimulai dengan proses cleaning data. Proses cleaning adalah proses analisis kualitas dari suatu data dengan cara mengubah, menghapus, dan mengoreksi data yang salah dan bertujuan untuk membersihkan *noise* dan *missing value*. Hal ini dikarenakan proses pengumpulan datanya tidak sempurna sehingga ada

bagian yang tidak saling berkaitan dan hilang. Untuk menangani *missing value* pada kumpulan *big data*, kita dapat mengabaikan bagian data yang hilang. Tetapi, jika data yang kita miliki cenderung kecil, pendekatan yang digunakan adalah dengan mengisi *missing value* dengan memasukan suatu nilai secara manual maupun dengan proses komputasi. Untuk mengatasi *missing value* dapat menggunakan teknik *Detection Outlier (Standard Deviation)*. *Data cleaning* selanjutnya digunakan untuk *data noise*. Penggunaan *data noise* ini bertujuan untuk menyederhanakan proses interpretasi. Salah satu teknik dari *data cleaning* yang digunakan untuk mengatasi *noise* adalah *Standard Scaler* dan *Normalizer*.

Pada tahap Regresi, data dibagi menjadi dua bagian, yaitu *data training* dan *data testing*. *Data training* adalah data yang digunakan sebagai pengujian pada regresi sebagai model data prediksi. Data yang digunakan pada training disebut dengan training set. Sedangkan, data testing digunakan untuk menguji model prediksi dari data training. Metode regresi yang digunakan pada laporan ini adalah Linear Regresi, Polinomial Regresi, dan Logistik Regresi. Data ini diaplikasikan menggunakan data yang sudah diolah pada proses *Training*. Setelah itu, hasil dari Regresi ini diolah kembali di *Model Evaluation*.

Pada tahap kelima, metode yang digunakan adalah *Model Evaluation*. *Model Evaluation* merupakan bagian dari proses pengembangan model data yang membantu untuk menemukan model terbaik dan model yang dipilih akan bekerja untuk memprediksi data di masa depan. Model ini menggunakan *train test split* dengan data yang sudah diolah di *preprocessing*. Dari data yang diambil, 70% dari datasetnya dijadikan *data training* dan sisa 30% digunakan untuk *data testing*.

Pada tahap keenam, Setelah proses pemodelan data, model data diuji keakuratan dan tingkat eror prediksi terhadap data sebenarnya. Perhitungan akurasi ini bertujuan untuk melihat seberapa berpengaruh seluruh atribut dengan prediksi yang dilakukan. Kemudian dilakukan proses evaluasi untuk melihat berapa tingkat akurasi penelitian yang dilakukan dengan *Mean Square Error (MSE)* dan *Accuracy (R²)*. Kedua metode tersebut digunakan untuk mengukur tingkat keakuratan suatu model prediksi. Nilai MSE merepresentasikan rata-rata kesalahan (*error*) absolut antara hasil prediksi dengan nilai sebenarnya.



Gambar 1. Flowchart tahapan

V. HASIL DAN PEMBAHASAN

Hasil Pengumpulan Data

Berdasarkan pencarian data, diperoleh sebuah dataset mengenai *Heart Failure Prediction Dataset* yang didapatkan dari *Kaggle*. Dataset tersebut sering digunakan oleh ahli *machine learning*. Dataset tersebut memiliki 11 atribut yang digunakan untuk memprediksi apakah seseorang mengidap penyakit jantung atau tidak. Pada penelitian kali ini semua atribut digunakan dalam menganalisis resiko penyakit jantung seseorang. Gambar 2 menunjukkan informasi mengenai nilai pada dataset.

(918, 12)												
	Age	Sex	ChestPainType	RestingBP	Cholesterol	FastingBS	RestingECG	MaxHR	ExerciseAngina	Oldpeak	ST_Slope	HeartDisease
0	40	M	ATA	140	289	0	Normal	172	N	0.0	Up	
1	49	F	NAP	160	180	0	Normal	156	N	1.0	Flat	
2	37	M	ATA	130	263	0	ST	98	N	0.0	Up	
3	48	F	ASY	138	214	0	Normal	108	Y	1.5	Flat	
4	54	M	NAP	150	195	0	Normal	122	N	0.0	Up	

Gambar 2. Dataset.

Implementasi Metode Regresi

Setelah dilakukannya ketiga Metode Regresi dengan inputan semua data, dapat dikatakan Logistik Regresi dapat dikatakan sangat baik karena memiliki perbandingan tingkat akurasi yang lebih jauh.

Prediksi Dengan Regresi

Hal yang ditentukan dalam melakukan ketiga Metode Regresi ini adalah jumlah fitur. Dalam keadaan saat ini menggunakan 11 buah feature pada kolom dataset terakhir dengan target 0 dan 1. Artinya 0 berarti tidak memiliki penyakit jantung, sedangkan 1 adalah yang memiliki jantung. Sebelum melakukan training, dilakukanlah sebuah *Detection Outlier* untuk menghilangkan *missing value* yang ada dalam dataset. Setelah melakukan *Detection Outlier* dipisahkan data *feature* dan data target, serta dilakukanlah *standard scaler* dan *normalizer* untuk menghilangkan *noise* pada data *feature*.

Selanjutnya membagi data antara data training dan testing dengan 70% data training dan 30% data testing dari dataset

Testing Output Regresi

Berdasarkan output yang diperoleh dari tahap training, maka tahapan selanjutnya yaitu melakukan analisis dengan 3 Metode Regresi meliputi Linier Regresi, Polynomial Regresi, dan Logistik Regresi. Serta testing menggunakan sebesar 30 persen dari jumlah data.

Hasil Prediksi Menggunakan Dengan Metode Regresi

Hasil prediksi menggunakan ketiga Metode Regresi dilakukan pada tahapan proses training dan testing. Tingkat akurasi dapat ditentukan dari hasil analisa prediksi label target seseorang memiliki penyakit jantung. Berdasarkan hasil yang diperoleh tersebut dapat dihitung tingkat akurasi dan MSE dengan hasil seperti pada tabel berikut.

Metode Regresi	MSE	R ² / Accuracy
Linier Regresi	0.11752794160467865	53.9292 %
Polynomial Regresi	0.11158588020889848	54.8077 %
Logistik Regresi	0.13333333333333333	86.67 %

Berdasarkan hasil dari MSE dan R² atau *accuracy* pada masing-masing Metode Regresi dapat dikatakan Metode Logistik Regresi memiliki *accuracy* yang paling besar yaitu sebesar 86.67%.

Implementasi Pada Aplikasi

Berdasarkan hasil dari prediksi, dinyatakan bahwa Logistik Regresi memiliki *accuracy* paling tinggi sehingga pada dataset ini, Logistik Regresi sudah tepat sehingga bisa di buat aplikasi sederhana dari data pasien 2 pasien yang kita dapatkan. Pasien pertama memiliki penyakit jantung, sedangkan pasien kedua tidak memiliki penyakit jantung. Hal ini sudah disesuaikan dengan *feature dataset* seperti tabel berikut.

Feature	Pasien Pertama	Pasien Kedua
Umur	43	36
Jenis Kelamin	Perempuan	Laki-laki
Jenis Nyeri Dada	TA	ATA

Tekanan Darah	100	120
Nilai Kolesterol	223	267
Gula Puasa	Tidak	Tidak
Hasil Elektrokardiografi Pada Saat Istirahat	Normal	Normal
Detak Jantung Maksimal	142	160
Exercise Angina	Tidak	Tidak
Old Peak	0	3
ST Slope	Up	Flat

Berdasarkan data dari pasien pertama, data pasien tersebut diinputkan ke aplikasi sederhana di Python dan hasil menunjukkan bahwa pasien tidak memiliki penyakit jantung. Hal ini sesuai dengan data yang diberikan seperti pada gambar di bawah ini.

```

Input Age : 43
Gender (M/L) : M
Chest Pain Type (ASY/NAP/ATA/TA) : TA
Resting BP : 100
Cholesterol : 223
Fasting Blood Sugar (Yes/No) : No
Resting ECG (Normal,LVH,ST) : Normal
Max Heart Rate : 142
Exercise Angina (N/Y) : N
Old Peak : 0
ST Slope Flat/Up/Down : Up

--Hasil Prediksi--
[43, 0, 3, 100, 223, 0, 0, 142, 0, 0.0, 1]
Tidak Memiliki Penyakit Jantung

```

Gambar 3. Dataset pasien pertama.

Berdasarkan data dari pasien kedua, data pasien diinputkan ke aplikasi sederhana di Python dan hasil menunjukkan bahwa pasien memiliki penyakit jantung sesuai dengan data yang diberikan seperti pada gambar di bawah ini.

```

Input Age : 36
Gender (M/L) : M
Chest Pain Type (ASY/NAP/ATA/TA) : ATA
Resting BP : 120
Cholesterol : 267
Fasting Blood Sugar (Yes/No) : No
Resting ECG (Normal,LVH,ST) : Normal
Max Heart Rate : 160
Exercise Angina (N/Y) : No
Old Peak : 3
ST Slope Flat/Up/Down : Flat

--Hasil Prediksi--
[36, 0, 2, 120, 267, 0, 0, 160, 1, 3.0, 0]
Memiliki Penyakit Jantung

```

Gambar 4. Dataset pasien kedua.

VI. KESIMPULAN

1. Setelah dilakukan *Detection Outlier* pada dataset *Heart*, ditemukan *missing value* yang dapat dibuktikan dengan total data awal 918 dan setelah dilakukan *Detection Outlier* data berkurang menjadi 899.
2. Model dari *Logistic Regression* lebih sesuai untuk melakukan prediksi pada dataset *heart*. Hal ini dibuktikan dengan hasil akurasi yang sudah dijalankan dari Model *Evaluation* dengan total akurasi 86,67%. Total akurasi ini tertinggi dibandingkan dengan Linear Regresi dan Polinomial Regresi.
3. Aplikasi sederhana yang dibuat bisa memecahkan masalah dengan Metode Logistik Regresi yang memiliki akurasi yang besar.

REFERENSI

- [1] Yuwandini, M., & Nurmaini, S. (2020). *Klasifikasi Gagal Jantung Kongestif Menggunakan Metode Recurrent Neural Network (Rnn)* (Doctoral Dissertation, Sriwijaya University).
- [2] Pangaribuan, J. J., Tanjaya, H., & Kenichi, K. (2021). Mendeteksi Penyakit Jantung Menggunakan Machine Learning Dengan Algoritma Logistic Regression. *Journal Information System Development (ISD)*, 6(2), 1-10.
- [3] Puteri, K., & Silvanie, A. (2020). Machine Learning Untuk Model Prediksi Harga Sembako Dengan Metode Regresi Linear Berganda. *JUNIF: Jurnal Nasional Informatika*, 1(2), 82-94.
- [4] Chicco, D., & Jurman, G. (2020). Machine Learning Can Predict Survival of Patients With Heart Failure From Serum Creatinine And Ejection Fraction Alone.

BMC Medical Informatics And Decision Making, 20(1), 1-16.

- [5] Ahmad, A. (2017). Mengenal artificial intelligence, machine learning, neural network, dan deep learning. *J. Teknol. Indones.*, no. October, 3.
 - [6] Tazliqoh, A. Z., Rahmawati, R., & Safitri, D. (2015). Perbandingan Regresi Komponen Utama dengan Regresi Ridge pada Analisis Faktor-Faktor Pendapatan Asli Daerah (PAD) Provinsi Jawa Tengah. *Jurnal Gaussian*, 4(1), 1-10.
 - [7] Robbani, M., Agustiani, F., & Herrhyanto, N. (2019). Regresi Least Absolute Shrinkage And Selection Operator (Lasso) Pada Kasus Inflasi Di Indonesia Tahun 2014-2017. *Jurnal EurekaMatika*, 7(2), 1-16.
-