

ANALISIS FAKTOR PREDIKSI DIAGNOSA TINGKAT SERANGAN JANTUNG MENGGUNAKAN METODE REGRESSION

^{1*}Muhammad Ardiansyah Sembiring

Universitas Sumatera Utara

Email: adinmantap88@gmail.com

Abstract: Heart disease, which is also known as cardiovascular, is a variety of conditions where there is narrowing or blockage of the blood vessels which can cause heart attacks, chest pain, stroke. Heart disease can occur in anyone of any age, gender, occupation, and lifestyle. In addition, heart disease cannot be cured. This condition requires careful treatment and monitoring throughout life. When this treatment fails, sufferers have to undergo surgical operations which are quite expensive and complicated. A report from WHO in September 2009 stated that this disease was the first cause of death to date. In 2004, an estimated 17.1 million people died from heart disease. This figure represents 29% of global causes of death, with details of 7.2 million people dying from heart disease and 5.7 million people dying from stroke. This can be prevented by reducing risk factors. The role of information technology can be realized by data retrieval techniques. research to shorten the time and selection of factors for early detection of heart attacks.

Keywords: heart; regression; information Technology

Abstrak: Penyakit jantung atau *heart disease*, yang juga dikenal dengan istilah kardiovaskuler adalah berbagai kondisi dimana terjadi penyempitan atau penyumbatan pembuluh darah yang dapat menyebabkan serangan jantung, nyeri dada, stroke. Penyakit jantung dapat terjadi oleh siapapun disegala usia, jenis kelamin, pekerjaan, dan gaya hidup. Selain itu penyakit jantung tidak dapat disembuhkan. Kondisi ini membutuhkan pengobatan dan pemantauan hati-hati sepanjang hidupnya. Ketika pengobatan ini gagal, penderita harus melakukan operasi bedah yang cukup mahal dan rumit. Laporan dari WHO September 2009, menyebutkan bahwa penyakit tersebut merupakan penyebab kematian pertama sampai saat ini. Pada 2004, diperkirakan 17,1 juta orang meninggal karena Penyakit Jantung. Angka ini merupakan 29% dari penyebab kematian global, dengan perincian 7,2 juta meninggal karena penyakit jantung dan 5,7 juta orang meninggal karena stroke. Hal ini dapat dicegah dengan mengurangi faktor-faktor resiko. Peran teknologi informasi dapat diwujudkan dengan teknik mencari data riset untuk mempersingkat waktu dan pemilihan faktor-faktor pendeteksi dini serangan jantung.

Kata kunci: jantung; regresi; teknologi informasi

PENDAHULUAN

Di Indonesia, penyakit jantung merupakan salah satu permasalahan kesehatan yang dapat berakibat fatal bahkan kematian. Kasus penyakit jantung menjadi salah satu penyebab tingkat kematian tertinggi, bersaing dengan kanker dan hipertensi terutama di negara-negara sedang berkembang di mana masyarakat sedang mengalami perubahan pola hidup [1].

Beberapa perubahan pola hidup dan kebiasaan diet yang tidak terkontrol dapat meningkatkan risiko seseorang terkena penyakit jantung, dan hal ini sering disebut sebagai faktor risiko. Penyakit Jantung Koroner (PJK) adalah kondisi di mana pembuluh darah arteri yang mengalirkan darah ke otot jantung mengalami penyempitan akibat pengendapan kalsium dan lemak (aterosklerosis), mengurangi suplai oksigen dan nutrisi ke jantung secara optimal [2].

Penyakit Jantung Koroner (PJK) atau penyakit kardiovaskular yang menjadi masalah kesehatan utama di era modern. Menurut laporan Organisasi Kesehatan Dunia (WHO) pada September 2009, PJK merupakan penyebab kematian utama global. Pada tahun 2004, diperkirakan 17,1 juta orang meninggal karena penyakit jantung, menyumbang 29% dari total kematian global, dengan sebagian besar kasus disebabkan oleh penyakit jantung dan stroke.

Data dari Riset Kesehatan Dasar tahun 2007 yang dilakukan oleh Kementerian Kesehatan RI menunjukkan bahwa stroke merupakan penyebab kematian utama di Indonesia dengan persentase 26,9%, sedangkan penyakit jantung menyumbang 9,3%. Jika kedua penyakit kardiovaskular tersebut tetap menjadi penyebab utama kematian di Indonesia, maka jumlah kematian akibat keduanya mencapai 36,2% dari total kematian.

Beberapa kasus penyakit jantung seringkali tidak menunjukkan gejala apa pun dan dapat menyebabkan kematian secara mendadak, sementara beberapa kasus lainnya mungkin terjadi ketika pasien sedang melakukan aktivitas fisik yang berat.

Munculnya penyakit jantung secara tiba-tiba menjadi antisipasi bagi masyarakat mengingat pola konsumsi makanan yang tidak teratur dan mayoritas makanan berlemak dapat berpotensi menimbulkan penyakit jantung. Penelitian ini dilakukan untuk mengestimasi penyebab penyakit jantung berasal dari faktor apa saja. Hal tersebut dilakukan dengan analisis dari data yang ada. Proses analisis data dalam penelitian ini menggunakan metode regresi dengan tujuan mendapatkan estimasi yang akurat, dimana metode regresi telah banyak digunakan dalam penelitian berbasis machine learning untuk estimasi dan analisis data. Metode ini dianggap memiliki keuntungan karena mampu memprediksi berdasarkan pola interaktif data yang relevan dari data yang sudah ada sebelumnya. Selain itu, dalam penelitian ini, metode regresi juga digunakan untuk mengestimasi laju pertumbuhan penduduk dengan data yang bersumber dari BPS Simalungun, memberikan hasil yang cukup akurat dan dapat digunakan sebagai dasar untuk pembangunan software estimasi selanjutnya.

METODE

Metode riset memuat hal-hal penting terkait pelaksanaannya penelitian dan tahapan-tahapan yang akan kami tempuh dalam penelitian ini untuk mencapai suatu hasil atau output yang diinginkan. Berikut ini hal-hal yang mencakup kedalam metode riset antara lain:

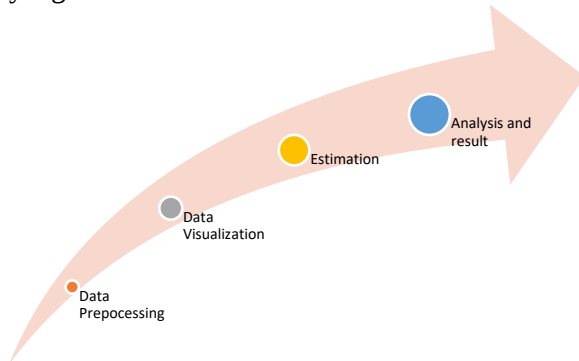
Variabel yang Terdapat dalam Riset

Dalam machine learning, penelitian ini dikategorikan sebagai penelitian supervised learning yaitu sebuah teknik model dimana data yang akan kami proses memiliki label/target/class dengan tujuan untuk mengetahui hubungan kasualitas antara variabel bebas dan variabel yang dijadikan target [3]. maka dari itu variabel x meliputi cp, trtbps, fbs, chol. Serta yang menjadi variabel y adalah Age.

Tahapan-Tahapan Riset

Tahapan yang akan ditempuh dalam penelitian ini digambarkan dengan flowchart

sehingga dapat terlihat dengan jelas tahapan yang akan dilakukan antara lain:



Gambar 1. Tahapan Sistem Estimasi

Pada gambar tersebut, digambarkan dimana tahap yang pertama yang akan dilakukan adalah Data Propocessing dimana dalam proses tersebut terdapat data cleaning untuk melakukan hadling missing pada data ,data selection untuk menyeleksi data berdasarkan kolom yang akan di teliti dan data transformation . Tahapan kedua adalah melakukan data visualisation yaitu menyajikan repretansi dari sebarang data yang digunakan .Tahapan berikutnya adalah melakukan pemisahan data set menjadi 2 bagian yaitu, data training dan data testing. Setelah itu, memasuki tahapan proses estimasi dengan menggunakan regresi yang meliputi model regresi dan menguji data tersebut dengan data testing. Kemudian tahapan terakhir adalah tahapan analisis dan evaluasi data dengan menggunakan 4 nilai eror yaitu MAE, MSE, RMSE dan R2-Score, dengan bertujuan untuk mendapatkan hasil akurasi dari proses estimasi sehingga diperoleh metode terbaik dengan hasil evaluasi yang terbaik pula.

Data Riset

Data riset yang akan digunakan adalah dataset *heart rating* yang berasal dari *kaggle.com* pada tahun 2022. Data tersebut berjumlah sebanyak 303 baris data. Terdapat 6 kolom (*cp, trtbps, fbs, chol, thalachh, output*) yang akan digunakan untuk pengujian dan 1 kolom (*age*) sebagai kolom targer/label. Sampel dataset yang akan digambarkan dalam penelitian ini menggunakan 6 data teratas seperti yang terdapat dalam

	age	cp	trtbps	fbs	chol	thalachh	output
0	63	3	145	1	233	150	1
1	37	2	130	0	250	187	1
2	41	1	130	0	204	172	1
3	56	1	120	0	236	178	1
4	57	0	120	0	354	163	1
...	...	...	...	...	...	...	...
298	57	0	140	0	241	123	0
299	45	3	110	0	264	132	0
300	68	0	144	1	193	141	0
301	57	0	130	0	131	115	0
302	57	1	130	0	236	174	0

Pembagian Data

Pembagian data dalam penelitian ini menggunakan *data training* dan *data testing* dengan rasio sebesar 90:10 Dimana *data training* memiliki kisaran data sebanyak 90 dari dataset serta *data testing* sebanyak 10 dari dataset. Algoritma yang menggunakan rasio 90:10 memiliki performa yang lebih baik dibandingkan dengan algoritma lainnya dimana tingkat akurasi dapat diperoleh sebesar 99,40%..

Estimasi dengan Model Regresi

Regresi merupakan alat yang paling populer yang digunakan untuk memperoleh hasil evaluasi pengatuh sebuah variabel *x* (variabel bebas) dan variabel *y* (variabel terikat).

Linear Regression

Linear regression atau regresi linier terbagi atas 2 yaitu *simple linear regression* dan *multiple linear regression*. Menurut penelitian bahwa regresi linear sederhana adalah model yang menganalisis hubungan antara 1 variabel prediktor dan 1 variabel respon [4]. Dalam penelitian tersebut juga menyatakan bahwa prediksi menggunakan regresi linear sederhana memperoleh hasil prediksi dengan kategori baik karena telah menunjukkan hasil prediksi terbaik selama tahun 2021 pada periode bulan januari.

Adapun persamaan dalam regresi linear sederhana beriku ini :

$$Y = a + bx$$

Dimana :

Y : variabel dependen

a : titik perpotongan garis terhadap koordinat y

b : koefisien variabel independen

x : variabel independen

Kemudian *multiple linear regression* atau regresi linear berganda menurut merupakan suatu model analisis yang menggambarkan hubungan variabel respon (Y) dengan variabel – variabel yang mempengaruhinya (x) dimana variabel (x) tersebut berjumlah lebih dari satu. Persamaan yang digunakan dalam regresi linear berganda adalah:

$$Y = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + \dots + b_nX_n$$

Dimana:

Y: variabel respon

X : variabel prediktor yang mempengaruhi variabel respon

b₀ : intercept

b : koefisien

Model Pengukuran Akurasi Menggunakan Nilai Error

Data yang telah diolah menggunakan model regresi kemudian diukur nilai akurasinya. Pengolahan data yang bersifat estimasi yang masih mengandung unsur ketidakpastian maka diperlukan untuk memasukkan nilai *error* sebagai pengukuran akurasi dari estimasi tersebut. Nilai *Error* adalah selisih antara nilai pengamatan yang sebenarnya dengan nilai prediksi. Untuk estimasi yang diukur menggunakan MAE, MSE, dan RMSE nilai yang paling baik adalah nilai yang paling kecil, semakin kecil nilai yang dihasilkan berarti hasil estimasi semakin bagus. Berbeda dengan *R2-Score* nilai koefisien determinan yang mendekati 1 mengartikan bahwa variabel independennya memberikan semua informasi yang diperlukan dalam mengestimasi variabel dependennya. nilai *error* yang akan digunakan adalah 4 model pengukuran akurasi nilai estimasi diantaranya :

Mean Absolute Error (MAE)

$$MAE = \frac{1}{n} \sum_{i=1}^n |A_i - F_i|$$

Dimana :

n : ukuran sampel

A_i : nilai data aktual ke-i

F_i : nilai data estimasi ke-i

Mean Square Error (MSE)

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$$

Dimana :

n : jumlah data

Y_i : nilai yang diamati

$\hat{Y}_i$: nilai estimasi

Root Mean Square Error (RMSE)

$$RMSE = \left(\frac{\sum (y_i - \hat{y}_i)^2}{n} \right)^{1/2}$$

Dimana :

y_i : nilai hasil pengamatan

$\hat{y}_i$: nilai hasil estimasi

n : jumlah data

R2-Score

$$R^2 = 1 - \frac{RSS}{TSS}$$

Dimana :

R² : koefisien determinan

RSS : jumlah kuadrat

TSS : total kuadrat

HASIL DAN PEMBAHASAN

Pada pembahasan dalam penelitian ini akan dilakukan visualisasi data agar data dapat lebih mudah untuk diamati serta dapat mengetahui hubungan antara tiap variabel x terhadap variabel y. Selanjutnya, data akan diuji dengan 4 bentuk pembagian data dengan rasio yang berbeda – beda dimana didalamnya

akan dilakukan pengujian data menggunakan model regresi dan diukur dengan 4 model nilai evaluasi. Hal tersebut bertujuan untuk menemukan model regresi manakah yang paling baik dengan menghasilkan nilai akurasi yang paling akurat untuk digunakandalam estimasi Age . Tahapan pembahasan yang dilakukan adalah sebagai berikut :

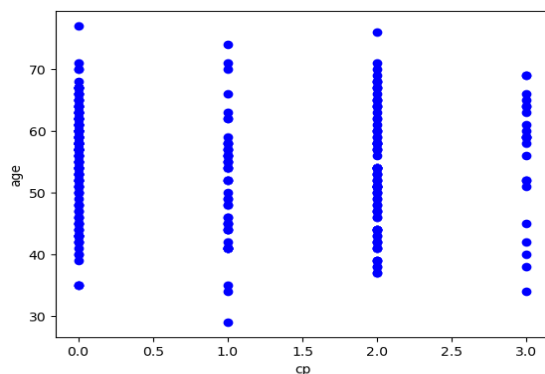
1. Visualisasi Data

Visualisasi data bertujuan untuk memudahkan seseorang dalam mengamati dan memahami sebuah data. Terdapat berbagai model yang dapat digunakan untuk menyajikan sebuah data kedalam bentuk visualnya. *Scatter plot* menggunakan *library* yaitu “matplotlib.pyplot”.

Pada penelitian ini visualisasi data menggunakan *scatter plot* yang bertujuan untuk melihat plot sebaran data dan mengamati seberapa besar hubungan pengaruh antara sumbu x dengan sumbu y [5]. Data *heart* yang digunakan untuk mengestimasi karbon dioksida kendaraan telah divisualisasikan terhadap sumbu x nya. Visualisasi data disajikan dalam bentuk antar variabel x dengan variabel y dan penyajian *scatter* tiap variabel menggunakan warna yang berbeda agar lebih mudah dipahami dan mudah diingat. Visualisasinya data tersebut sebagai berikut ini :

Scatter Plot Cp Terhadap Age

Berikut ini tampilan visualisasi data pada kolesterol sebagai sumbu y terhadap Age sebagai sumbu x.

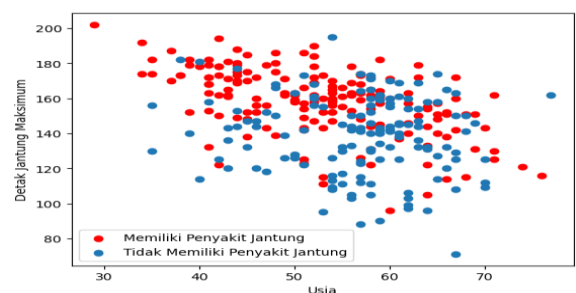


Gambar 2.Scatter cp Terhadap Age

Visualisasi plot titik sembarangan dari sumbu y (Age) terhadap sumbu x(cp) disajikan kedalam scatter berwarna biru hubungan yang ditunjukkan masih bersifat positif karena masih menunjukkan peningkatan baik dari variable age maupun variable cp.

Scatter Plot Detak Jantung Terhadap Age

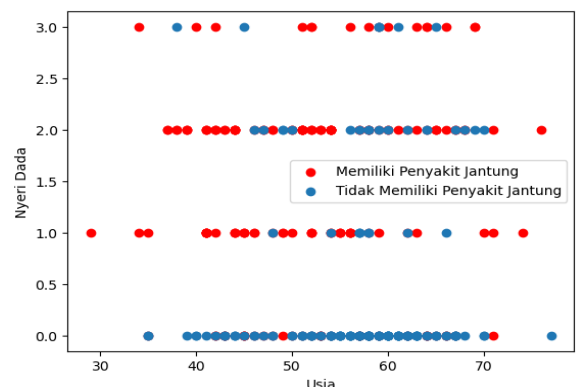
Berikut Tampilan visualisasi data pada Detak jantung sebagai(sumbu y) terhadap Age(sumbuh x)



Gambar 3 .Scatter Detak Jantung Terhadap Age

Visualisasi pada sumbu y (Detak jantung) terhadap sumbu x(age) menggunakan scatter plot berwarna biru merah ini berdasarkan gambar diatas d peroleh persebaran yang tidak menentu semakin tinggi detak jantung maka disini digambarkan semakin tinggi juga memiliki penyakit jantung.

Scatter Plot Nyeri Dada Terhadap Age

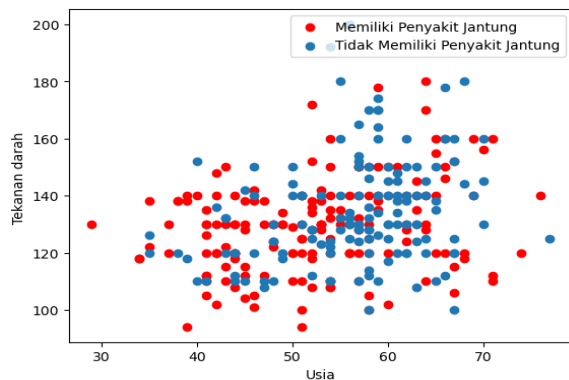


Gambar 4 . Scatter Nyeri Dada Terhadap Age

Visualisasi pada sumbu y (Nyeri pada dada) dan sumbu x(age) menggunakan scatter plot biru merah ini di liat dari scatternya

memiliki garis lurus kesamping dan d mana di lihat bahwa nyeri dada tidak mempengaruhi dan tidak dapat memprediksi kita memiliki sakit jantung tidaknya.

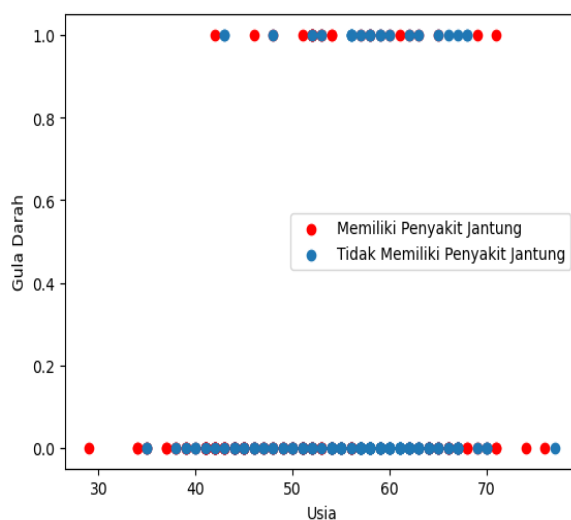
Scatter Plot Tekanan Darah Terhadap Age



Gambar 5. Scatter Tekanan Dara Terhadap Age

Visualisasi pada sumbu y (Tekanan Darah) terhadap sumbu x (age) menggunakan scatter plot berwarna biru merah ini berdasarkan gambar diatas d peroleh persebaran yang tidak menentu semakin tinggi detak jantung maka disini digambarkan semakin tinggi juga memiliki penyakit jantung.

Scatter Plot Tekanan Darah Terhadap Age



Gambar 6. Scatter Tekanan Darah Terhadap Age

Visualisasi pada sumbu y (Gula Drah) dan sumbu x (age) menggunakan scatter plot biru merah ini di liat dari scatternya memiliki garis lurus kesamping dan d mana di lihat bahwa nyeri dada tidak mempengaruhi dan tidak dapat memprediksi kita memiliki sakit jantung tidaknya.

2. Pengujian Akurasi Pada Rasio Data 90:10

Pengujian data pertama menggunakan rasio 90:10, dimana *data training* sebesar 90% dari total data dan *data testing* sebesar 10% menggunakan 7 model regresi diperoleh hasil seperti pada tabel dibawah ini :

Tabel 2. Pengujian Akurasi Rasio 90:10

MODEL	MAE	MSE	RMS E	r2-score
Linear regression	6.423	62.038	7.876	0.159
	92332	33774	44195	15122

Ketentuan yang digunakan dalam melihat hasil estimasi yang baik adalah untuk nilai pengukuran menggunakan MAE, MSE, RMSE nilai yang paling kecil yang dihasilkan merupakan nilai dengan hasil estimasi yang paling bagus. Sedangkan untuk R2-Score nilai yang paling mendekati angka 1 (satu) dan menjauhi 0 (nol) merupakan nilai dengan estimasi yang paling bagus. Berdasarkan tabel diatas, pada model pengukuran MAE nilai estimasi.

KESIMPULAN

Berdasarkan hasil persamaan regresi binary yang dihasilkan , 7 variabel yang paling berpengaruh terhadap akurasi penentu tingkat serangan jantung dalam model persamaan regresi .Diagnosa merupakan variable dependen yang hasilnya tergantung pada variabelnya. Kesimpulan dari penelitian ini akan didapatkan dari berbagi uji coba dan hasil terbaik dari masing – masing literasi .Dari penelitian ini yang berpengaruh yaitu tekanan darah tinggi, kolesterol tinggi dan denyut jantung tinggi menyebabkan kemungkinana tinggi serangan jantung itu terjadi, dan dari data diatas dapat juga kita lihat usia 40-60 memiliki resiko yang besar untuk terkena serangan jantung tersebut.

DAFTAR PUSTAKA

- [1] A. B. Wibisono and A. Fahrurrozi, "Perbandingan Algoritma Klasifikasi Dalam Pengklasifikasian Data Penyakit Jantung Koroner," *J. Ilm. Teknol. dan Rekayasa*, vol. 24, no. 3, pp. 161–170, 2019, doi: 10.35760/tr.2019.v24i3.2393.
- [2] L. F. Tampubolon, A. Ginting, and F. E. Saragi Turnip, "Gambaran Faktor yang Mempengaruhi Kejadian Penyakit Jantung Koroner (PJK) di Pusat Jantung Terpadu (PJT)," *J. Ilm. Permas J. Ilm. STIKES Kendal*, vol. 13, no. 3, pp. 1043–1052, 2023, doi: 10.32583/pskm.v13i3.1077.
- [3] A. P. Silalahi and H. G. Simanullang, "Supervised Learning Metode K-Nearest Neighbor Untuk Prediksi Diabetes Pada Wanita," *METHOMIKA J. Manaj. Inform. dan Komputerisasi Akunt.*, vol. 7, no. 1, pp. 144–149, 2023, doi: 10.46880/jmika.vol7no1.pp144-149.
- [4] Elen Riswana Safila Putri, Fahriza Novianti, Yasirah Rezqita Aisyah Yasmin, and Dian Candra Rini Novitasari, "Prediksi Kasus Aktif Kumulatif Covid-19 Di Indonesia Menggunakan Model Regresi Linier Berganda," *Transform. J. Pendidik. Mat. dan Mat.*, vol. 5, no. 2, pp. 567–577, 2021, doi: 10.36526/tr.v5i2.1231.
- [5] A. R. Syamsuri, C. W. Anggraini, D. P. Pratiwi, S. Afriliana, and Maissy, "PENGARUH KUALITAS PRODUK DAN HARGA TERHADAP KEPUASAN PELANGGAN UD ADLI DI DESA SUKAJADI KECAMATAN PERBAUNGAN," *Bisnis Mhs.*, pp. 215–224, 2021.