MENDETEKSI PENYAKIT JANTUNG MENGGUNAKAN MACHINE LEARNING DENGAN ALGORITMA LOGISTIC REGRESSION

Jefri Junifer Pangaribuan[1*)], Henry Tanjaya[2)], Kenichi[3)]

[1]Fakultas Ilmu Komputer, Universitas Pelita Harapan
E-mail: jefri.pangaribuan@uph.edu[1)]
[2]Fakultas Ilmu Komputer, Universitas Pelita Harapan
E-mail: ht6411@student.uph.edu[2)]
[3]Fakultas Ilmu Komputer, Universitas Pelita Harapan
E-mail: kk2848@student.uph.edu[3)]

**Penulis Korespondensi*

Abstract – Heart Disease or also called cardiovascular disease is one of the dangerous diseases that can cause death. As technology develops and the popularity of machine learning technology increases, machine learning technology can be used to help detect heart disease using patient data. There are various methods that can be used to diagnose whether a person has heart disease. This research implements the use of an algorithm called logistic regression, where the algorithm uses logistic functions to produce binary or zero and one as a classification determination. After the experiment is carried out with a logistic regression algorithm the results have different advantages over other methods based on the confusion matrix analysis model. In training data, logistic regression methods has the highest sensitivity value is 88.54% compared to other methods. In testing data, logistic regression method has the highest specificity value of 87.50% compared to other methods.

***Keywords**: Heart Disease, Cardiovascular Disease, Machine Learning, Logistic Regression*

Abstrak – Penyakit Jantung atau disebut juga penyakit kardiovaskular merupakan salah salah satu penyakit berbahaya yang dapat menyebabkan kematian. Seiring berkembangnya teknologi dan peningkatan popularitas teknologi machine learning, teknologi machine learning tersebut dapat digunakan untuk membantu mendeteksi penyakit jantung dengan menggunakan data pasien. Terdapat berbagai jenis metode yang dapat digunakan untuk mendiagnosa apakah seseorang terkena penyakit jantung atau tidak. Penelitian ini mengimplementasikan penggunaan algoritma yaitu logistic regresi, dimana algoritma tersebut memakai fungsi logistik untuk menghasilkan binary atau nol dan satu sebagai penentuan klasifikasi. Setelah eksperimen dilakukan dengan algoritma logistik regresi memberikan hasil yang memiliki keunggulan yang berbeda-beda terhadap metode lainnya berdasarkan model analisa confusion matrix. Pada data training, metode logistik regresi mempunyai nilai sensitivity yang paling tinggi yaitu 88.54% dibanding metode lainnya. Pada data testing, metode logistik regresi mempunyai nilai kekhususan yang paling tinggi yaitu 87.50% dibanding metode lainnya.

Diterima <07072021>, Revisi <09072021>, Diterima untuk publikasi <26072021>.

ISSN : 2528-5114

Kata Kunci: *Penyakit Jantung, Penyakit Kardiovaskular, Machine Learning, Logistik Regresi*

PENDAHULUAN

Penyakit jantung atau dikenal juga sebagai penyakit kardiovaskular adalah semua penyakit yang terjadi akibat adanya gangguan fungsi jantung. Penyakit jantung merupakan hasil dari penumpukan plak di dalam arteri koroner, yang menghambat aliran darah ke jantung serta meningkatkan risiko serangan jantung dan komplikasi lainnya [1]. Penyakit jantung merupakan salah satu penyebab kematian tertinggi di Indonesia pada tahun 2017 menurut Kementerian Kesehatan Indonesia [2].

Tentunya dengan kemajuan teknologi angka tersebut dapat diminimalisir. Salah satu teknologi yang ada sekarang yaitu *machine learning* yang dapat digunakan untuk mendeteksi sebuah penyakit. Seperti penelitian yang pernah dilakukan untuk mendiagnosis penyakit Diabetes Melitus dengan menggunakan *Extreme Learning Machine* [3], maupun penelitian yang dilakukan untuk mendiagnosis penyakit jantung koroner dengan membandingkan algoritma C4.5 dan *Extreme Learning Machine* [4].

Machine learning sendiri bukanlah sebuah teknologi yang memiliki keakuratan 100% untuk melakukan analisa data dan mendapat kesimpulan berdasarkan analisa data tersebut. Namun keakuratan yang dihasilkan cukup efektif sehingga *machine learning* telah terbukti membantu di bidang kesehatan [5].

Ada banyak algoritma klasifikasi yang dapat dipakai di *machine learning*, tetapi dalam kasus ini penulis menggunakan algoritma *Logistic Regression*. *Logistic Regression* adalah analisis regresi yang tepat untuk dilakukan ketika variabel dependen adalah biner (dua kemungkinan). *Logistic Regression* digunakan untuk menggambarkan data dan untuk menjelaskan hubungan antara satu variabel biner dependen dan satu atau lebih variabel independen nominal, ordinal, interval atau rasio tingkat. Penulis juga akan menggunakan *dataset* yang diambil dari *Kaggle* [6].

STUDI LITERATUR

Penyakit Jantung

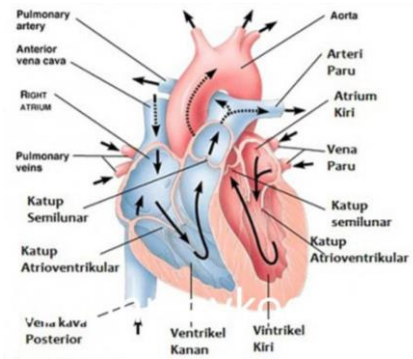
Penyakit jantung merupakan keadaan dimana jantung mengalami masalah atau gangguan. Bentuk gangguan jantung bermacam-macam, mulai dari pembuluh darah jantung, irama jantung, katup jantung, atau gangguan akibat bawaan lahir.

Anatomi jantung terdiri dari 4 ruangan, dua ruang bagian atas merupakan atrium (serambi) kiri dan kanan, sedangkan ruang bagian bawah merupakan ventrikel (bilik) kiri dan kanan. Antara ruang kanan dan kiri dipisahkan oleh otot yang berfungsi untuk mencegah tercampurnya darah yang miskin oksigen dengan yang kaya akan oksigen.

Fungsi jantung adalah untuk memompa darah yang kaya dengan oksigen ke seluruh organ tubuh, lalu darah yang miskin dengan oksigen dari organ kembali ke jantung, atrium kanan lalu diteruskan ke ventrikel kanan. Diantara atrium kanan dengan ventrikel kanan terdapat katup trikuspid yang berfungsi untuk mencegah darah kembali ke atrium kanan, lalu ketika ventrikel berkontraksi, maka darah akan dibawa melalui katup pulmonal dan arteri pulmonal ke paru paru untuk diisi dengan oksigen.

Setelah darah telah diisi oksigen oleh paru paru, maka akan dibawa ke atrium

kiri melalui vena pulmonal. Lalu atrium kiri akan berkontraksi dan darah yang kaya akan oksigen akan mengalir ke ventrikel kiri melalui katup mitral, dimana fungsi katup mitral adalah mencegah dara untuk kembali ke atrium kiri. Kemudian ventrikel kiri akan berkontraksi dan darah akan mengalir ke seluruh organ tubuh. Gambar 1 menunjukkan anatomi jantung yang terdiri dari 4 ruang serta penghubung antar ruang [7].



Gambar 1. Anatomi jantung

Machine Learning

Machine Learning merupakan sebuah metode berbasis komputer yang tidak perlu diatur dahulu oleh manusia dan dapat belajar dengan bantuan data dan akan semakin pintar seiring dengan banyaknya data yang telah diolah (belajar melalui pengalaman). Metode ini sering digunakan dalam menyelesaikan kasus klasifikasi dan *clustering* dan biasanya digunakan untuk menangani data dalam skala besar atau *big data.*

Di dalam *machine learning*, proses pengolahan datanya terbagi menjadi *data training* dan *data testing* dimana *data training* bertujuan untuk melatih metode atau algoritma yang digunakan dalam *machine learning* sedangkan *data testing* bertujuan untuk mengetahui tingkat kinerja dari algoritma yang telah dilatih tersebut ketika digunakan pada data yang belum pernah ditemui saat melakukan proses *data mining.*

Machine learning secara umum jika dilihat dari teknik pembelajarannya terbagi menjadi *unsupervised learning* dan *supervised learning*. *Supervised learning* merupakan teknik di dalam *machine learning* dalam mengolah dataset yang telah memiliki label data untuk dipelajari sehingga label data tersebut mampu diidentifikasi oleh mesin saat melakukan proses klasifikasi maupun prediksi sedangkan *unsupervised learning* merupakan teknik di dalam *machine learning* yang memperbolehkan mesin itu sendiri dalam menarik kesimpulan atas informasi yang diperlukan dan biasanya dilakukan pada data tanpa *label* [8].

Logistik Regresi

Model logistik regresi adalah suatu model statistik yang digunakan untuk mengetahui pengaruh variabel prediktor (X) terhadap variabel respon (Y) dengan variabel responnya berupa data dikotomi yaitu bernilai 1 menyatakan bahwa variabel respon memiliki kriteria yang ditentukan dan 0 menyatakan bahwa variabel respon tidak memiliki kriteria yang ditentukan. Menurut Hosmer dan Lemeshow [9], persamaan model regresi logistik adalah:

$$\pi(x_i) = \frac{e^{\beta_0+\beta_1+x_{1i}}}{1+e^{\beta_0+\beta_1+x_{1i}}} \quad (1)$$

Dengan fungsi logit $g(x_i)$ yaitu:

$$g(x_i) = \ln\left[\frac{\pi(x_i)}{1-\pi(x_i)}\right] = \beta_0 + \beta_1 x_{1i} \quad (2)$$

Dari persamaan (1) dan (2) dapat disederhanakan menjadi:

$$\pi(x_i) = \frac{e^{g(x_i)}}{1+e^{g(x_i)}} \quad (3)$$

Fungsi Sigmoid

Persamaan model yang digunakan untuk logistik regresi yang dapat disebut juga fungsi sigmoid adalah sebagai berikut

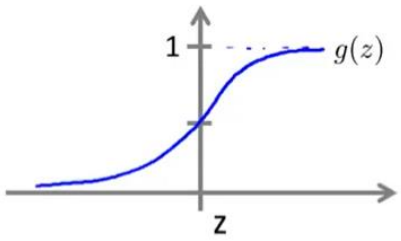
$$h_\theta = g(\theta^T x) \quad (4)$$

$$z = \theta^T x \quad (5)$$

$$g(z) = \frac{1}{1+e^{-z}} \quad (6)$$

$$h_\theta(x) = \frac{1}{1+e^{-\theta^T x}} \quad (7)$$

$h_\theta(x)$ akan memberikan hasil diantara 0 dan 1 , contohnya $h_\theta(x)$ memberikan nilai 0.7 dimana memberikan hasil probabilitas 70% terhadap output dari 1 dan untuk probabilitas hasil 0 adalah 30%. Fungsi dengan kurva yang berbentuk huruf S. Untuk setiap nilai x yang dipetakan ke dalam interval 0 sampai 1 dinamakan fungsi sigmoid biner, sedangkan output yang memiliki rentang antara -1 sampai dengan 1 disebut sigmoid tan. Berikut gambar 2 adalah gambar dari persamaan sigmoid yang berbentuk huruf S.



Gambar 2. Fungsi Sigmoid

Misalnya, untuk memprediksi

- *y = 1, maka nilai dari* $h_\theta(x) \geq 0.5$, *dimana nilai dari* $z \geq 0$
- *y = 0, maka nilai dari* $h_\theta(x) < 0.5$, *nilai dari* $z < 0$,

Dari penjelasan di atas, dapat disimpulkan bahwa ketika memprediksi hasil dari sebuah persamaan, 0 ataupun 1, sama seperti memprediksi y = 1 ketika nilai dari $\theta^T x \geq 0$ dan sebaliknya memprediksi nilai y = 0, $\theta^T x < 0$.

METODOLOGI

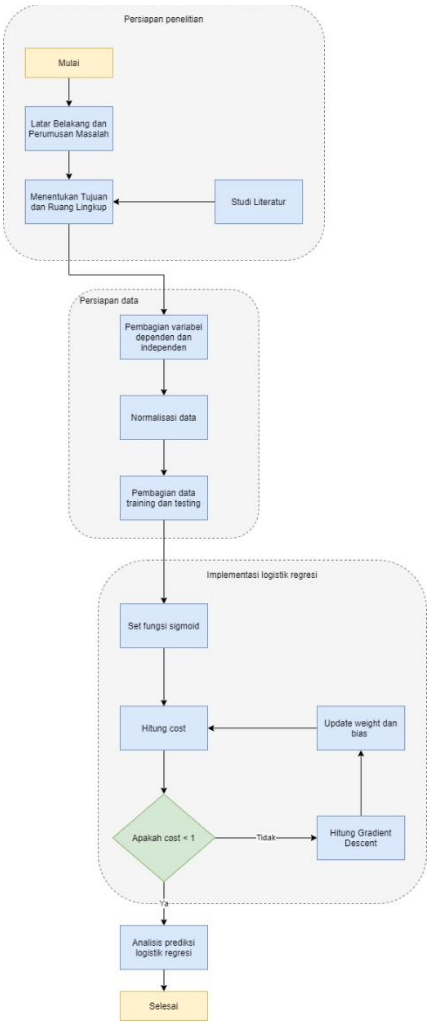
Tahapan Penelitian

Tahap awal dalam melakukan persiapan penelitian yaitu menentukan latar belakang dan tujuan penelitian serta mendefinisikan batasan dan ruang lingkup, serta studi literatur. Studi Literatur mempunyai bagian penting dalam pelaksanaan penelitian karena dapat membantu meningkatkan pemahaman akan komponen komponen yang terlibat dalam penelitian, yaitu cara kerja prediksi algoritma Logistik Regresi serta faktor faktor yang dapat mengakibatkan apakah seseorang memiliki penyakit jantung.

Tahapan kedua dalam langkah penelitian yaitu melakukan persiapan data. Dimana dalam persiapan data, data yang akan digunakan dilakukan pembagian variabel menjadi variabel dependen dan variabel independen. Dilanjutkan dengan normalisasi data agar data rentang antara 0 dan 1 untuk memudahkan perhitungan. Dan selanjutnya dilakukan pembagian data menjadi data *training* dan data *testing*

Tahapan ketiga yaitu implementasi algoritma Logistik Regresi untuk mendiagnosa apakah seseorang memiliki penyakit jantung atau tidak. Proses dari implementasi terdiri dar mencari *cost* atau bisa disebut juga sebagai *mean error* menggunakan *cost function* disebut juga *forward propagation*, lalu melihat apakah *cost* sudah sesuai dengan yang diinginkan, jika tidak, maka dapat menggunakan gradient descent atau bisa disebut juga backward propagation untuk menghitung bobot dan melakukan update terhadap bobot maupun *bias* pada variabelnya dan hitung ulang *cost*-nya menggunakan *cost function*. Jika sudah sesuai maka dapat dilanjutkan ke tahap berikutnya.

Dan langkah terakhir yaitu melakukan analisa terhadap hasil implementasi logistik regresi dan menarik kesimpulan serta memberikan saran dari hasil penelitian yang telah diperoleh. Analisa menggunakan confusion matrix untuk mengevaluasi keakuratan model logistik regresi. Gambar 3 menunjukkan *flowchart* proses dari penelitian yang dilakukan.



Gambar 3. Tahapan Penelitian

Metode Pengumpulan Data

Dataset didapatkan melalui *repository dataset* yang berada pada *website* Kaggle.com [6]. Kaggle merupakan tempat kerkumpulan *data scientist* dimana tersedia banyak *dataset* yang dapat diambil dan tersedia banyak perlombaan yang berkaitan dengan *Machine Learning.*

Dataset yang digunakan berjudul *Heart Disease UCI* dimana *dataset* tersebut didapat dari *Cleaveland Heart Disease Database* dan dipublikasikan pada *Kaggle*. *Dataset* tersebut sudah digunakan oleh para peneliti *machine learning* saat ini, dimana *dataset* tersebut mengandung 76 atribusi yang memiliki pengaruh pada penyakit jantung yang diidap seseorang, tetapi berdasarkan eksperimen yang biasanya dijalankan, biasanya menggunakan 14 atribusi dan terdapat 303 data pasien seperti yang ditunjukkan pada tabel 1. Konklusi terdapatnya penyakit jantung pada kolom *target* dengan nilai 0 atau 1.

Tabel 1. Penjelasan Label pada Dataset

Singkatan Label	Makna Label
Age	Umur
Sex	Jenis Kelamin
CP (Chest Pain)	Rasa sakit dada
Trestbps	Tekanan darah saat istirahat (mmHg)
Chol	Kolestrol (mg/dl)
Fbs	Gula darah puasa (>120mg/dl) (1=ya;0=tidak)
Restecg	Hasil elektrographic saat istirahat
Thalac	Detak jantung maksimal
Exang	Latihan yang diinduksi angina (1=ya;0=tidak)
Oldpeak	Depresi yang diinduksi oleh latihan relatif
Slope	Kemiringan puncak ST Segmen
Ca	Jumlah pembuluh darah yang berwarna setelah diwarnai flourosopy
Thal	Tipe kerusakan pembuluh darah, 3=normal;6=cacat tetap;7=cacat sementara
Target	1 atau 0

Penelitian Dengan Logistik Regresi

Diagnosis penyakit jantung dengan logistik regresi menggunakan *script*

phython sebagai *tool*-nya dan proses dari implementasi logistik regresi dapat dijelaskan sebagai berikut:

a. *Pembagian variabel dependen dan independen*

Pembagian variabel menjadi variabel independen yaitu kolom lain selain target dan variabel dependen yaitu target, adalah sebagai berikut, variable independen yaitu *Age, Sex, Chest Pain, Trestbps, Chol, Fbs, Restecg, Thalac, Exang, Oldpeak, Slope, Ca, Thal,* sedangkan variable dependen yaitu *Target.*

b. *Normalisasi data*

Data yang akan dimasukkan ke dalam logistik regresi dinormalisasi sehingga mempunyai nilai dengan rentang 0 dan 1. Berikut ini merupakan rumus yang digunakan untuk normalisasi data

$$X_{changed} = \frac{X - X_{min}}{X_{max} - X_{min}}$$

Dimana:

X_{min} = nilai minimum pada data set sebelum normalisasi

X_{max} = nilai maksimum pada data set sebelum normalisasi

$X_{changed}$ = nilai data setelah normalisasi

X = nilai data asli sebelum normalisasi

c. *Pembagian data training dan testing*

Data akan dibagi untuk *training* dan *testing*. Data sebesar 80 % dari *dataset* akan dipilih secara acak untuk dijadikan data *training* dan 20 % dari *dataset* akan digunakan sebagai *testing*.

d. *Menentukan fungsi*

e. *Fungsi sigmoid*

Untuk setiap nilai x yang dipetakan ke dalam interval 0 sampai 1 dinamakan fungsi sigmoid biner, sedangkan output yang memiliki rentang antara -1 sampai dengan 1 disebut sigmoid tan.

f. *Forward propagation*

Forward propagation digunakan untuk menghitung *cost function* dan y *(output). Cost* atau disebut juga *mean error,* mengukur perbedaan antara 2 variabel kontiniu. Berikut merupakan persamaan dengan *forward propagation.*

$$J(\theta) = -\frac{1}{m}\sum_{i=1}^{m}\left[y^{(i)}\, log\left(h_\theta\left(x^{(i)}\right)\right) - \left(1 - y^{(i)}\right) log\,\left(1 - h_\theta\left(x^{(i)}\right)\right)\right]$$

g. *Backward propagation*

Backward Propagation digunakan untuk menghitung *gradient descent. Gradient descent* merupakan algoritma optimisasi untuk meminimalkan beberapa fungsi secara iteratif. Penggunaan *gradient descent* digunakan untuk memperbaharui parameter bobot pada model regresi logistik. Gambar 3.3 menunjukkan proses gradient descent dimana terlihat gradient membantu mengarahkan *weight* ke titik optimal. Berikut merupakan persamaan dari *backward propagation:*

Repeat {

$$\theta_j := \theta_j - \frac{a}{m}\sum_{i=1}^{m}\left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right) x_j^{(i)}$$

}

h. *Menentukan learning rate*

Learning rate ditentukan menggunakan intuisi karena *learning rate* bersifat fleksibel, apabila

learning rate terlalu rendah maka akan memerlukan banyak waktu dan juga biaya tenaga *gpu* untuk mendapatkan *cost function* yang optimal. Sebaliknya apabila *cost function* terlalu tinggi, maka *cost function* tidak akan pernah mencapai titik optimal karena akan melewati *cost function* optimal tersebut [10].

i. Mencari cost function yang optimal

Dengan menggunakan *data training* dan *forward propagation* dan *backward propagation* dalam iterasi yang cukup banyak akan mendapatkan *cost function* yang optimal. Setelah mendapatkan *cost function* kita dapat menyimpulkan *weight* (beban) pada setiap variabel untuk dijadikan model pada *machine learning.*

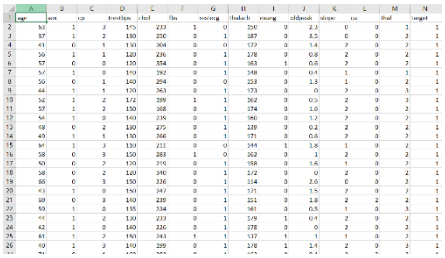
j. Testing model

Setelah mendapatkan model, kita dapat melakukan *testing* dengan memasukkan data *testing* kedalam model dan mendapatkan hasilnya. Hasil tersebut dapat diolah dalam *confusion matrix* untuk mendapatkan analisa terhadap keakuratan dalam prediksi model tersebut.

HASIL DAN PEMBAHASAN

Hasil Pengumpulan Data

Berdasarkan pencarian data, diperoleh *dataset Heart Disease UCI* dimana *dataset* tersebut didapat dari *Cleaveland Heart Disease Database* dan dipublikasikan pada *Kaggle*. *Dataset* tersebut sudah digunakan oleh para peneliti *Machine Learning* saat ini, dimana *dataset* tersebut mengandung 76 atribusi yang memiliki pengaruh pada penyakit jantung yang diidap seseorang, tetapi berdasarkan eksperimen yang dijalankan, menggunakan 14 atribusi dan terdapat 303 data pasien. Konklusi terdapatnya penyakit jantung pada kolom *target* dengan nilai 0 atau 1. Gambar 4 menunjukkan informasi nilai *dataset* disediakan dengan format sebagai berikut:



Gambar 4. *Dataset* Pasien Penyakit Jantung

Prediksi Dengan Logistik Regresi

Salah satu hal yang perlu ditentukan dalam prediksi menggunakan logistik regresi adalah jumlah *input.* Penelitian ini akan menggunakan 13 buah variabel *input.* Pada kolom 14, khusus kolom *target,* tidak termasuk pada variabel *input* melainkan hanya penanda apakah seseorang mengidap penyakit jantung atau tidak.

Sebelum melakukan penelitian dilakukan pembagian variabel menjadi variabel independen yaitu kolom selain *target* dan variable dependen yaitu *target*. Selanjutnya dilakukan normalisasi terlebih dahulu agar data memiliki rentang antara 0 dan 1.

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal
0	0.708333	1.0	1.000000	0.481132	0.244292	1.0	0.0	0.603053	0.0	0.370968	0.0	0.0	0.333333
1	0.166667	1.0	0.666667	0.339623	0.283105	0.0	0.5	0.885496	0.0	0.564516	0.0	0.0	0.666667
2	0.250000	0.0	0.333333	0.339623	0.178082	0.0	0.0	0.770992	0.0	0.225806	1.0	0.0	0.666667
3	0.562500	1.0	0.333333	0.245283	0.251142	0.0	0.5	0.816794	0.0	0.129032	1.0	0.0	0.666667
4	0.583333	0.0	0.000000	0.245283	0.520548	0.0	0.5	0.702290	1.0	0.096774	1.0	0.0	0.666667

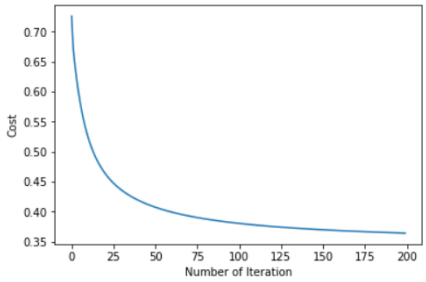
Gambar 5. Sampel Data yang Dinormalisasi

Tahap selanjutnya adalah membagi data *training* dan *testing* sebesar 80% dari *dataset* akan dipilih secara acak untuk dijadikan data *training* dan 20% dari *dataset* akan digunakan sebagai data *testing.* Data *training* yang sudah diacak akan dilakukan proses *transpose matrix.*

Gambar 6. Data Setelah *Transpose*

Implementasi Logistik Regresi

Setelah dilakukan iterasi sebesar 200 kali dan *learning rate* sebesar 1, maka pemilihan iterasi menggunakan intuisi dengan berdasarkan pada besaran penurunan *cost*, dimana penentuan sebesar perubahan *cost* minimal 0.005. Maka dari itu, bobot *weight* yang akan digunakan adalah pada iterasi ke-76 dikarenakan setelah iterasi ke-76, perubahan *cost* sudah tidak terlalu signifikan. Sebagai gambaran, perubahan *cost* dari iterasi ke-1 hingga iterasi ke-76 sebesar 0.33572448 sedangkan pada iterasi ke-76 hingga iterasi ke-152, perubahan *cost* hanya sebesar 0.0204791, dan terus terjadi penurunan *cost* hingga iterasi ke-200. Gambar 7 merupakan hasil pemetaan antara *cost* dengan iterasi.



Gambar 7. Grafik Pemetaan *Cost* dengan Iterasi

Testing Logistik Regresi

Berdasarkan bobot *output* yang didapat dari proses *training*, maka tahap selanjutnya adalah melakukan prediksi dengan algoritma logistik regresi. Data yang digunakan adalah data *testing* sebesar 20% dari total data sehingga didapatkan sebesar 62 data. Tabel 2 menunjukkan hasil bobot pada setiap variabel.

Tabel 2. Bobot *Output* Hasil *Training* Untuk Masing-Masing Variabel

Age	-0.34399865
Sex	-1.31237031
CP (Chest Pain)	1.7729785
Trestbps	-0.26222051
Chol	-0.0371197
Fbs	-0.169084
Restecg	0.38723275
Thalac	1.33400028
Exang	-1.1463211
Oldpeak	-1.04004649
Slope	1.15955381
Ca	-1.85961429
Thal	-0.67790772
Target	

Analisa Hasil Prediksi Logistik Regresi

Hasil analisa prediksi logistik regresi dapat dilakukan pada proses *training* dan proses *testing*. Keakuratan prediksi logistik regresi dapat dianalisa dari hasil prediksi *label target* seseorang mengidap penyakit jantung. Dari proses logistik regresi, digunakan metode *confusion matrix* dimana terdapat masing masing proses *training* dan *testing* seperti pada tabel 2.

Tabel 3. Hasil Proses Data *Training* & *Testing*

Proses	TP	TN	FP	FN
Training Regresi Logistik	116	86	25	15
Testing Regresi Logistik	31	21	6	1

Berdasarkan hasil dari tabel *confusion matrix* di atas, dapat dihitung *sensitivity, specificity, accuracy* dan *precision* pada masing masing proses *training* dan *testing* logistik regresi. Tabel 3 menunjukkan hasil analisa model prediksi menggunakan *confusius matrix* terhadap data *testing* dan data *training.*

Tabel 4. Hasil *Confusion Matrix* Data *Training* dan *Testing*

Proses	*Specificity*	*Sensitivity*	*Precision*	*Training Accuracy*
Training Regresi Logistik	85.14 %	88.54 %	82.26 %	83.47%
Testing Regresi Logistik	87.50 %	91.17 %	83.78 %	85.25%

Pada penelitian sebelumnya telah dilakukan perbandingan antara algorima C.50, SVM, KNN, dan *neural network* untuk mendeteksi penyakit jantung pada *dataset* yang sama [11]. Berdasarkan hasil dari data *training*, terlihat bahwa algoritma KNN lebih unggul dibanding dengan algoritma lainnya dan algoritma logistik regresi lebih rendah tingkat akurasi, presisi dan spesifikasi dibanding algoritma yang lain.

Tabel 5. Perbandingan Algoritma pada Data *Training*

Algoritma	*Specificity*	*Sensitivity*	*Precision*	*Accuracy*
C.50	89.62 %	84.61 %	85.71 %	87.50 %
SVM	84.90 %	79.48 %	79.48 %	82.61 %
KNN	91.50 %	79.48 %	87.32 %	86.41 %
Neural Network	91.50 %	78.20 %	87.14 %	85.87 %
Logistik Regresi	85.14 %	88.54 %	82.26 %	83.47 %

Berdasarkan hasil dari data *testing*, terlihat bahwa algoritma C.50 lebih unggul dibanding dengan algoritma lainnya dan algoritma *neural network* lebih rendah tingkat akurasi, sensitivitas, presisi, dan spesifikasi dibanding dengan algoritma lain, sedangkan algoritma logistik regresi hanya unggul pada spesifikasi saja.

Tabel 6. Perbandingan Algoritma pada *Data Testing*

Algoritma	*Specificity*	*Sensitivity*	*Precision*	*Accuracy*
C.50	90.90 %	95.23 %	90.90 %	93.02 %
SVM	90.90 %	80.95 %	89.47 %	86.05 %
KNN	88.63 %	88.09 %	88.09 %	88.37 %
Neural Network	86.36 %	73.80 %	83.78 %	80.23 %
Logistik Regresi	87.50 %	91.17 %	83.78 %	85.25 %

KESIMPULAN

Dari hasil percobaan dan analisa dari prediksi penyakit jantung dengan menggunakan logistik regresi maka dapat diambil kesimpulan sebagai berikut:

- Tahapan dalam implementasi logistik regresi adalah menentukan variabel dependen dan independen, normalisasi data, pembagian data *training* dan *testing*, mendefinisikan fungsi s*igmoid, forward propagation* dan *backward propagation*, menentukan *learning rate* dan jumlah iterasi, menentukan cost function yang optimal dan melakukan *testing* pada data.
- Jumlah iterasi dan *learning rate* dalam penelitian menggunakan logistik regresi akan menentukan hasil akurasi penelitian. Dimana iterasi dan inisiasi *learning rate* menggunakan intuisi yang didasari pada perubahan *cost* pada setiap iterasi.
- Diagnosis penyakit jantung dengan menggunakan logistik regresi memiliki keunggulan yang berbeda beda terhadap metode lainnya pada model analisa *confusius matrix*.
- Semakin rendah *cost* dalam logistik regresi, semakin bagus hasil akurasi yang dihasilkan di dalam penelitian ini.

DAFTAR PUSTAKA

[1] "Informasi umum seputar penyakit jantung," Hellosehat, 2019. [Online]. Available:https://hellosehat.com/pusat-kesehatan/serangan-jantung/informasi-umum-seputar-penyakit-jantung/.

[2] "Kementerian Kesehatan Republik Indonesia," 2017. [Online]. Available: http://www.depkes.go.id/article/view/17073100005/penyakit-jantung-penyebab-kematian-tertinggi-kemenkes-ingatkan-cerdik-.html.

[3] J. J. Pangaribuan, "Mendiagnosis Penyakit Diabetes Melitus Dengan Menggunakan Metode Extreme Learning Machine," *Journal Information System Development (ISD)*, vol. 1, pp. 32-40, 2016.

[4] J. J. Pangaribuan, C. Tedja and S. Wibowo, "Perbandingan Metode Algoritma C4.5 Dan Extreme Learning Machine Untuk Mendiagnosis Penyakit Jantung Koroner," *Jurnal Informatics Engineering Research and Technology (IERT)*, vol. 1, pp. 9-15, 2019.

[5] Springer Nature Publishing, "Nature Biomedical Engineering," Nature Research, 2019. [Online]. Available: https://www.nature.com/collections/zbkpvddmhm.

[6] Kaggle, "Kaggle," 2019. [Online]. Available: https://www.kaggle.com/ronitf/heart-disease-uci.

[7] Fatmawati, Identifikasi Karakteristik Penyakit Jantung Koroner di Polikninik RSU Bahteramas Provinsi Sulawesi Utara, 2017.

[8] E. Retnoningsih and R. Pramudita, "Mengenal Machine Learning Dengan Teknik Supervised dan Unsupervised Learning Menggunakan Python," *Bina Insani ICT Journal*, vol. 7, pp. 156-165, 2020.

[9] N. I. Mardini, L. Marlena and E. Azhar, "Regresi Logistik Pada Model Problem Based Learning Berbantu Software Cabri 3D," *Jurnal Mercumatika: Jurnal Penelitian Matematika dan Pendidikan Matematika*, vol. 4, pp. 64-70, 2019.

[10] H. Zulkifli, "Understanding Learning Rates and How It Improves Performance in Deep Learning," Towards Data Science, 22 1 2018. [Online]. Available: https://towardsdatascience.com/understanding-learning-rates-and-how-it-improves-performance-in-deep-learning-d0d4059c1c10.

[11] M. Abdar, S. R. N. Kalhori, T. Sutikno, M. I. I. Subroto and G. Arji, "Comparing Performance of Data Mining Algorithms in Prediction Heart Diseases," *International Journal of Electrical and Computer Engineering*, 2015.