

# Penerapan Data Mining Untuk Menganalisis Data Faktor Resiko Penyakit Jantung Menggunakan Metode Logistic Regression

Exzaraja Indiandra Scandea<sup>1</sup>, Muhammad Aqsha Rizki Sugiarto<sup>2\*</sup>, Fany Lestari<sup>3</sup>, Dwi Hartanti<sup>4</sup>

<sup>1,2,3,4</sup>Informatika

Universitas Duta Bangsa Surakarta

<sup>1</sup>202030028@mhs.udb.ac.id, <sup>2\*</sup>202030248@mhs.udb.ac.id, <sup>3</sup>202020964@mhs.udb.ac.id, <sup>4</sup>dwi\_hartanti@udb.ac.id

**Abstrak**— Penyakit jantung adalah kondisi ketika jantung mengalami gangguan. Sakit jantung tidak hanya disebabkan oleh satu gejala saja, melainkan beragam gejalanya. Penyakit jantung juga bisa menyerang siapa saja tanpa pandang usia, laki-laki maupun perempuan. Terutama bagi orang yang memiliki gaya hidup tidak sehat. Oleh karena itu, penting bagi kita untuk mengetahui penyebab penyakit jantung agar kita dapat melakukan pencegahan. Seiring berkembangnya teknologi, semakin mudah kita untuk menganalisis penyebab penyakit jantung. Untuk mengetahui lebih dalam faktor penyebab penyakit jantung, kita akan melakukan analisis dengan menggunakan logistic regression.

**Kata kunci**— Penyakit Jantung, teknologi, logistic regression.

**Abstract**— Heart disease is a condition when the heart is disturbed. Heart disease is not only caused by one symptom, but a variety of symptoms. Heart disease can also affect anyone regardless of age, male or female. Especially for people who have an unhealthy lifestyle. Therefore, it is important for us to know the causes of heart disease so that we can take prevention. As technology develops, it becomes easier for us to analyze the causes of heart disease. To find out more about the causes of heart disease, we will do an analysis using logistic regression.

**Keywords**— Heart disease, technology, logistic regression.

## I. PENDAHULUAN

Penyakit Jantung Koroner (PJK) merupakan penyakit yang menyerang jantung. Organ tersebut memiliki fungsi memompa darah ke seluruh tubuh. Kelainan pada organ tersebut dapat menyebabkan penyempitan pembuluh darah arteri yang mengalirkan darah ke otot jantung, sehingga mengakibatkan berkurangnya suplai oksigen dan nutrisi untuk menggerakkan jantung secara optimal. Penyempitan pembuluh darah tersebut disebabkan oleh pengendapan kalsium dan endapan lemak berwarna kuning yang dikenal dengan aterosklerosis (Soeharto, 2001).

Penyakit jantung atau penyakit kardiovaskular merupakan penyakit yang relatif umum. Penyakit ini terjadi karena adanya disfungsi jantung. Kondisi kronis ini dianggap sebagai penyebab kematian yang cukup umum di seluruh dunia untuk pria dan wanita dari semua ras.

Fungsi jantung adalah memompa darah nan kaya oksigen ke semua organ tubuh, lalu darah organ yang kekurangan oksigen kembali ke jantung, atrium di sebelah kanan diteruskan ke ventrikel kanan. Di antara atrium kanan dengan ventrikel kanan Memiliki katup trikuspid untuk mencegah darah mengalir kembali ke atrium kanan, jadi saat ventrikel berkontraksi, kemudian darah akan

diangkut melalui katup arteri pulmonalis ke paru-paru untuk diisi dengan oksigen.

Setelah darah diisi oksigen oleh paru-paru, kemudian diangkut ke atrium kiri melalui vena pulmonalis. Kemudian atrium sebelah kiri akan berkontraksi dan darah kaya oksigen akan mengalir ke ventrikel kiri melalui katup mitral. Kemudian ventrikel kiri berkontraksi dan darah akan mengalir semua organ tubuh.

Semakin majunya teknologi, penyakit semakin mudah untuk dideteksi. Adanya banyak cara atau klasifikasi yang bisa digunakan, salah satunya adalah logistic regression.

Logistic Regression adalah jenis analisis statistik yang biasa digunakan analisis data untuk pemodelan prediktif. Dalam metode analisis ini, variabel dependennya terbatas atau kategoris, bisa A atau B (regresi biner) atau beberapa opsi hingga A, B, C atau D (regresi polinomial). Jenis analisis statistik ini digunakan dalam perangkat lunak statistik untuk memahami hubungan antara variabel dependen dan satu atau lebih variabel independen dengan memperkirakan probabilitas. Jenis analisis ini dapat membantu untuk memprediksi probabilitas.

## II. METODOLOGI PENELITIAN

### A. Studi Pendahuluan

Studi pendahuluan ini bertujuan untuk memperoleh pemahaman awal tentang penggunaan data mining dan metode regresi logistik dalam menganalisis faktor risiko penyakit jantung. Dalam studi ini, metode regresi logistik digunakan untuk memodelkan hubungan antara faktor risiko tertentu dan keberadaan penyakit jantung.

Dalam tahap awal studi pendahuluan, telah dilakukan eksplorasi data untuk mengumpulkan dan menyiapkan data faktor risiko penyakit jantung. Data tersebut mencakup informasi seperti usia, jenis kelamin, riwayat merokok, tekanan darah, kadar kolesterol, dan faktor risiko lainnya yang relevan.

Setelah itu, analisis regresi logistik diterapkan untuk mengidentifikasi faktor risiko yang signifikan dalam pengembangan penyakit jantung. Metode regresi logistik memungkinkan pengukuran tingkat hubungan antara setiap faktor risiko dengan kejadian penyakit jantung. Koefisien regresi yang dihasilkan dapat memberikan indikasi sejauh mana setiap faktor risiko berkontribusi terhadap risiko penyakit jantung.

Studi pendahuluan ini juga mempertimbangkan pra pemrosesan data yang melibatkan pembersihan data, penggabungan data, dan transformasi data jika diperlukan. Langkah-langkah ini diperlukan untuk memastikan data yang digunakan dalam analisis regresi logistik akurat dan dapat diandalkan.

Melalui penerapan metode regresi logistik pada data faktor risiko penyakit jantung, diharapkan studi ini dapat memberikan wawasan yang lebih dalam tentang faktor-faktor yang berkontribusi terhadap risiko penyakit jantung. Informasi ini dapat berguna dalam pengembangan strategi pencegahan, deteksi dini, dan pengelolaan penyakit jantung.

Dalam kesimpulannya, studi pendahuluan penerapan data mining menggunakan metode regresi logistik untuk menganalisis data faktor risiko penyakit jantung menunjukkan potensi dalam mengidentifikasi faktor risiko yang signifikan dan memberikan wawasan yang berharga bagi penelitian lebih lanjut tentang penyakit jantung.

### B. Perumusan Masalah

Perumusan masalah penerapan data mining untuk menganalisis data faktor risiko penyakit

jantung menggunakan metode regresi logistik dapat dirumuskan sebagai berikut:

1. Bagaimana penerapan data mining dapat membantu mengidentifikasi faktor risiko yang signifikan dalam pengembangan penyakit jantung?
2. Bagaimana data faktor risiko penyakit jantung dapat dikumpulkan, dipersiapkan, dan diproses sebelum dianalisis menggunakan metode regresi logistik?
3. Apa manfaat dari menganalisis data faktor risiko penyakit jantung menggunakan metode regresi logistik dalam pengembangan strategi pencegahan, deteksi dini, dan pengelolaan penyakit jantung?

### C. Pengumpulan Data dan Pengolahan Data

- Tentukan variabel yang relevan untuk menganalisis faktor risiko penyakit jantung, seperti usia, jenis kelamin, riwayat keluarga, kebiasaan merokok, tekanan darah, kolesterol, diabetes, dan lain sebagainya.
- Peroleh data yang berkaitan dengan variabel-variabel tersebut, misalnya melalui survei, studi kohort, atau basis data medis.
- Lakukan pembersihan data dengan menghapus atau menangani nilai yang hilang, outliers, atau noise yang dapat mempengaruhi hasil analisis.
- Transformasikan data jika diperlukan, seperti mengubah variabel kategorikal menjadi variabel dummy atau melakukan normalisasi pada variabel numerik.

### D. Analisis

- Kumpulkan data yang relevan tentang faktor risiko penyakit jantung, seperti data demografis, riwayat medis, kebiasaan hidup, dan hasil tes medis terkait.
- Lakukan preprocessing data dengan membersihkan data yang hilang atau rusak, menangani nilai yang hilang atau outliers, dan melakukan transformasi data jika diperlukan.

### E. Kesimpulan

Model Logistic Regression yang dilatih dengan menggunakan data faktor risiko dapat

digunakan untuk memprediksi keberadaan penyakit jantung pada individu berdasarkan variabel prediktor yang diukur. Hal ini memungkinkan identifikasi dini dan pencegahan yang lebih efektif terhadap penyakit jantung.

### III. HASIL DAN PEMBAHASAN

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
	sex	age	cp	trtbps	chol	fbs	restecg	thalachh	exng	oldpeak	slp	caa	thall	output
2	1	66	0	160	228	0	0	138	0	2,3	2	0	1	1
3	0	71	0	132	149	0	1	125	0	1,6	1	0	2	1
4	1	64	3	170	227	0	0	155	0	0,6	1	0	3	1
5	0	66	2	146	278	0	0	152	0	0	1	1	2	1
6	0	39	2	138	220	0	1	152	0	0	1	0	2	1
7	0	58	0	130	197	0	1	131	0	0,6	1	0	2	1
8	1	47	2	130	253	0	1	179	0	0	2	0	2	1
9	1	35	1	122	192	0	1	174	0	0	2	0	2	1
10	1	58	1	125	220	0	1	144	0	0,4	1	4	3	1
11	1	56	1	130	223	0	0	163	0	0	2	0	3	1
12	1	56	1	120	240	0	1	169	0	0	1	1	2	1
13	0	55	1	132	342	0	1	166	0	1,2	2	0	2	1
14	1	41	1	120	157	0	1	182	0	0	2	0	2	1
15	1	38	2	138	175	0	1	173	0	0	2	4	2	1
16	1	38	2	138	175	0	1	173	0	0	2	4	2	1
17	1	67	0	160	286	0	0	108	1	1,5	1	3	2	0
18	1	67	0	120	229	0	0	129	1	2,6	1	2	3	0
19	0	62	0	140	268	0	0	160	0	3,6	0	2	2	0
20	1	63	0	130	254	0	0	147	0	1,4	1	1	3	0
21	1	53	0	140	203	1	0	155	1	3,1	0	0	3	0
22	1	56	2	130	256	1	0	142	1	0,6	1	1	1	0
23	1	48	1	110	229	0	1	168	0	1	0	0	3	0
24	1	58	1	120	284	0	0	160	0	1,8	1	0	2	0
25	1	58	2	132	224	0	0	173	0	3,2	2	2	3	0
26	1	60	0	130	206	0	0	132	1	2,4	1	2	3	0

Gambar 1 Data Set Penyakit Jantung

Data ini terdiri dari 14 atribut yang akan digunakan dalam proses pengolahan data serta terdapat 2 label yang meliputi 0 = lebih sedikit kemungkinan terkena penyakit serangan jantung, 1 = lebih tinggi kemungkinan terkena penyakit serangan jantung. Berikut ini penjelasan dari ke 14 atribut yang ada :

No	Nama	Keterangan
1	Sex	1= Laki-Laki, 0= Perempuan
2	Age	Umur Pasien
3	Cp	Jenis Nyeri Dada (typical angina, atypical angina, non-anginal pain, dan asymptomatic). (0 = Gejala umum nyeri dada dengan kemungkinan penyumbatan arteri koroner adalah Typical Angina. 1 = Gejala tidak rinci, kemungkinan penyumbatan lebih rendah adalah Atypical Angina. 2 = Rasa menusuk dan sakit dalam jangka waktu panjang atau pendek adalah Non-anginal pain. 3 = Tidak menunjukkan gejala penyakit adalah Asymptomatic)
4	trtbps	Dalam satuan mm/Hg tekanan darah pasien ketika dalam kondisi istirahat
5	chol	Serum kolesterol dalam satuan mg/dl.
6	fbs	Besar gula darah dalam satuan mg/dl, lebih besar atau kurang dari 120mg/dl. (0 = kurang dari 120mg/dl, 1 = lebih dari 120mg/dl).
7	restecg	Hasil elektrokardiografi istirahat. (0 = Normal, 1 = gelombang ST meningkat / menurun lebih dari 0,5 mV, 2 = Ventricular kiri mengalami hipertropi).
8	thalachh	Detak Jantung Maksimum.
9	exng	Nyeri dada akibat olahraga (0 = tidak nyeri, 1 = nyeri).
10	oldpeak	Besar segmen ST dari olahraga relatif terhadap kondisi istirahat
11	slp	Besar kemiringan segmen ST pada kondisi latihan pucak atau maksimum. (0 = downsloping, 1 = flat, 2 = upsloping)
12	caa	Jumlah pembuluh darah utama (0-3).
13	thall	Status jantung yang dibagi menjadi 4 diantaranya, 0 = tidak diketahui, 1 = cacat tetap, 2 = normal, 3 = cacat reversibel.
14	output	Terindikasi penyakit serangan jantung atau tidak. (0 = lebih kecil kemungkinan terkena serangan jantung, 1 = lebih besar kemungkinan terkena serangan jantung).

Gambar 2 Keterangan Atribut

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
	sex	age	cp	trtbps	chol	fbs	restecg	thalachh	exng	oldpeak	slp	caa	thall	output	0	1	0	1	0	1	0
2	1	66	0	160	228	0	0	138	0	2,3	2	0	1	1	0	1	0	0	0	0	0
3	0	71	0	132	149	0	1	125	0	1,6	1	0	2	1	1	0	0	0	0	0	0
4	1	64	3	170	227	0	0	155	0	0,6	1	0	3	1	1	0	0	0	0	0	0
5	0	66	2	146	278	0	0	152	0	0	1	1	2	1	1	0	0	0	0	0	0
6	0	39	2	138	220	0	1	152	0	0	1	0	2	1	1	0	0	0	0	0	0
7	0	58	0	130	197	0	1	131	0	0,6	1	0	2	1	1	0	0	0	0	0	0
8	1	47	2	130	253	0	1	179	0	0	2	0	2	1	1	0	0	0	0	0	0
9	1	35	1	122	192	0	1	174	0	0	2	0	2	1	1	0	0	0	0	0	0
10	1	58	1	125	220	0	1	144	0	0,4	1	4	3	1	1	0	0	0	0	0	0
11	1	56	1	130	223	0	0	163	0	0	2	0	3	1	1	0	0	0	0	0	0
12	1	56	1	120	240	0	1	169	0	0	1	1	2	1	1	0	0	0	0	0	0
13	0	55	1	132	342	0	1	166	0	1,2	2	0	2	1	1	0	0	0	0	0	0
14	1	41	1	120	157	0	1	182	0	0	2	0	2	1	1	0	0	0	0	0	0
15	1	38	2	138	175	0	1	173	0	0	2	4	2	1	1	0	0	0	0	0	0
16	1	38	2	138	175	0	1	173	0	0	2	4	2	1	1	0	0	0	0	0	0
17	1	67	0	160	286	0	0	108	1	1,5	1	3	2	0	1	0	0	0	0	0	0
18	1	67	0	120	229	0	0	129	1	2,6	1	2	3	0	1	0	0	0	0	0	0
19	0	62	0	140	268	0	0	160	0	3,6	0	2	2	0	1	0	0	0	0	0	0
20	1	63	0	130	254	0	0	147	0	1,4	1	1	3	0	1	0	0	0	0	0	0
21	1	53	0	140	203	1	0	155	1	3,1	0	0	3	0	1	0	0	0	0	0	0
22	1	56	2	130	256	1	0	142	1	0,6	1	1	1	0	1	0	0	0	0	0	0
23	1	48	1	110	229	0	1	168	0	1	0	0	3	0	1	0	0	0	0	0	0
24	1	58	1	120	284	0	0	160	0	1,8	1	0	2	0	1	0	0	0	0	0	0
25	1	58	2	132	224	0	0	173	0	3,2	2	2	3	0	1	0	0	0	0	0	0
26	1	60	0	130	206	0	0	132	1	2,4	1	2	3	0	1	0	0	0	0	0	0

Gambar 3 Decision Variabel

Pada gambar 3 merupakan langkah pertama dari logistic regression yaitu menambahkan table Decision Variabel dengan nilai 0,001 sesuai dengan jumlah atribut yang ada pada data dimulai dari 0.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
	sex	age	cp	trtbps	chol	fbs	restecg	thalachh	exng	oldpeak	slp	caa	thall	output	0	1	0	1	0	1	0
2	1	66	0	160	228	0	0	138	0	2,3	2	0	1	1	0	1	0	0	0	0	0
3	0	71	0	132	149	0	1	125	0	1,6	1	0	2	1	1	0	0	0	0	0	0
4	1	64	3	170	227	0	0	155	0	0,6	1	0	3	1	1	0	0	0	0	0	0
5	0	66	2	146	278	0	0	152	0	0	1	1	2	1	1	0	0	0	0	0	0
6	0	39	2	138	220	0	1	152	0	0	1	0	2	1	1	0	0	0	0	0	0
7	0	58	0	130	197	0	1	131	0	0,6	1	0	2	1	1	0	0	0	0	0	0
8	1	47	2	130	253	0	1	179	0	0	2	0	2	1	1	0	0	0	0	0	0
9	1	35	1	122	192	0	1	174	0	0	2	0	2	1	1	0	0	0	0	0	0
10	1	58	1	125	220	0	1	144	0	0,4	1	4	3	1	1	0	0	0	0	0	0
11	1	56	1	130	223	0	0	163	0	0	2	0	3	1	1	0	0	0	0	0	0
12	1	56	1	120	240	0	1	169	0	0	1	1	2	1	1	0	0	0	0	0	0
13	0	55	1	132	342	0	1	166	0	1,2	2	0	2	1	1	0	0	0	0	0	0
14	1	41	1	120	157	0	1	182	0	0	2	0	2	1	1	0	0	0	0	0	0
15	1	38	2	138	175	0	1	173	0	0	2	4	2	1	1	0	0	0	0	0	0
16	1	38	2	138	175	0	1	173	0	0	2	4	2	1	1	0	0	0	0	0	0
17	1	67	0	160	286	0	0	108	1	1,5	1	3	2	0	1	0	0	0	0	0	0
18	1	67	0	120	229	0	0	129	1	2,6	1	2	3	0	1	0	0	0	0	0	0
19	0	62	0	140	268	0	0	160	0	3,6	0	2	2	0	1	0	0	0	0	0	0
20	1	63	0	130	254	0	0	147	0	1,4	1	1	3	0	1	0	0	0	0	0	0
21	1	53	0	140	203	1	0	155	1	3,1	0	0	3	0	1	0	0	0	0	0	0
22	1	56	2	130	256	1	0	142	1	0,6	1	1	1	0	1	0	0	0	0	0	0
23	1	48	1	110	229	0	1	168	0	1	0	0	3	0	1	0	0	0	0	0	0
24	1	58	1	120	284	0	0	160	0	1,8	1	0	2	0	1	0	0	0	0	0	0
25	1	58	2	132	224	0	0	173	0	3,2	2	2	3	0	1	0	0	0	0	0	0
26	1	60	0	130	206	0	0	132	1	2,4	1	2	3	0	1	0	0	0	0	0	0

Gambar 4 Logit

Langkah kedua adalah menentukan nilai logit dengan rumus :  

$$= \$U\$2 + \$U\$3 * B2 + \$U\$4 * C2 + \$U\$5 * D2 + \$U\$6 * E2 + \$U\$7 * F2 + \$U\$8 * G2 + \$U\$9 * H2 + \$U\$10 * I2 + \$U\$11 * J2 +$$

$$1 * J2 + \$U\$12 * K2 + \$U\$13 * L2 + \$U\$14 * M2 + \$U\$15 * N2$$

Gambar 5 e^Logit

Langkah ketiga adalah menentukan  $e^{Logit}$  dengan rumus =EXP(O2)

Gambar 6 Probability

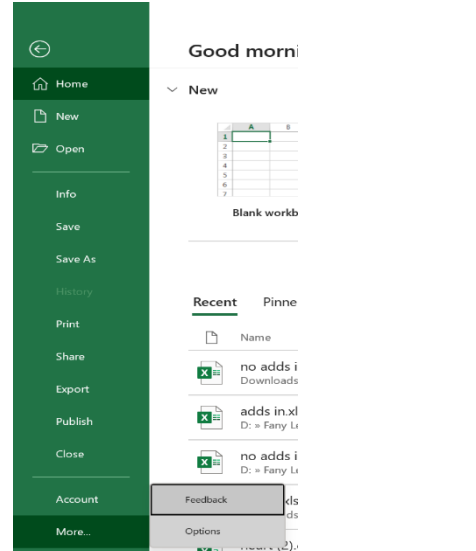
Langkah keempat adalah menentukan probability dengan menggunakan rumus =P2/(1+P2)

Gambar 7 Log Likelihood

Langkah kelima adalah menentukan log likelihood dengan rumus  $=(A2 * LN(Q2)) + ((1 - A2) * LN(1 - Q2))$ . Nilai log-likelihood dalam model regresi mengukur kebaikan model. Nilai kemungkinan log yang lebih tinggi menunjukkan kecocokan yang lebih baik, menghitung kemungkinan log dengan mencari logaritma natural dari probabilitas yang dihitung dengan parameter tertentu.

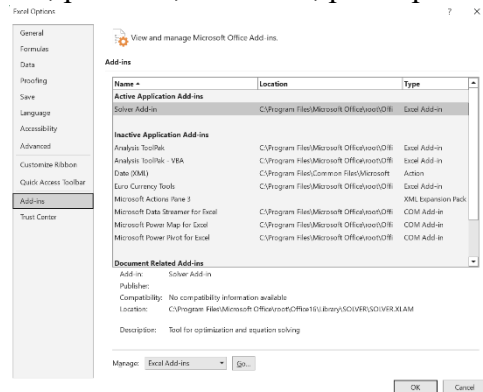
Gambar 8 Jumlah Log Likelihood

Selanjutnya adalah menjumlahkan log likelihood dengan rumus =SUM(R2:R26) maka akan didapatkan hasil yaitu -14,6633278



Gambar 9 Options

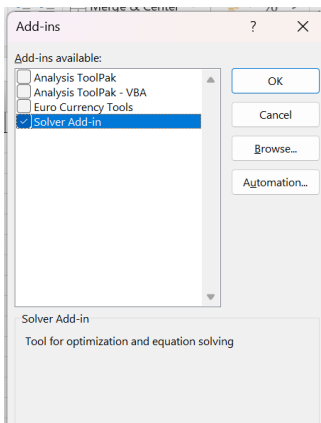
Setelah itu, pilih file, klik more, pilih options.



Gambar 220 Add-ins

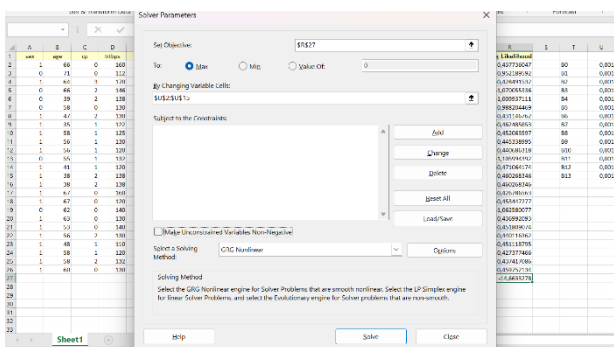
Pada kotak dialog Excel options, klik pada tab Add-ins. Pilih go.

Pilih Keep Solver Solution, lalu klik Ok, untuk menyimpan hasil solver.



Gambar 11 Solver Add- ins

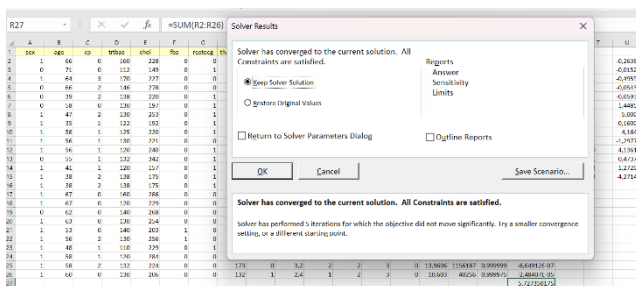
Pastikan opsi Add-in Solver dicentang. Klik OK untuk melanjutkan.



Gambar 12 Solver

Di tab Data, klik Analisis Data. Di kotak teks, masukkan sel yang menghitung total semua nilai kemungkinan log. Tujuannya adalah untuk memaksimalkan nilai. pilih nilai placeholder sebagai rentang yang akan diubah.

Pastikan bahwa 'Make Unconstrained Variables Non-Negative' tidak dipilih. Pilih opsi GRG Nonlinier sebagai metode penyelesaian. Klik Solve untuk melanjutkan.



Gambar 13 Solver Result

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	1	66	0	160	228	0	0	118	0	213	0	0	0	0	0	0	0	0	0	0	0
2	0	75	0	110	149	0	0	115	0	146	1	0	0	0	0	0	0	0	0	0	0
3	0	66	3	170	217	0	0	115	0	0	0	1	0	0	0	0	0	0	0	0	0
4	0	66	2	146	278	0	0	115	0	0	1	1	2	1	0	0	0	0	0	0	0
5	0	39	2	138	220	0	1	132	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0	39	2	138	220	0	1	132	0	0	0	0	0	0	0	0	0	0	0	0	0
7	0	47	2	130	197	0	1	131	0	0	0	0	0	0	0	0	0	0	0	0	0
8	1	47	2	130	203	0	1	144	0	0	0	0	0	0	0	0	0	0	0	0	0
9	1	35	1	122	192	0	1	174	0	0	0	0	0	0	0	0	0	0	0	0	0
10	1	56	1	135	220	0	1	144	0	0	0	0	0	0	0	0	0	0	0	0	0
11	1	56	1	130	221	0	0	143	0	0	0	0	0	0	0	0	0	0	0	0	0
12	1	42	1	120	157	0	1	142	0	0	0	0	0	0	0	0	0	0	0	0	0
13	0	55	1	132	242	0	1	146	0	1	2	0	0	0	0	0	0	0	0	0	0
14	1	42	1	120	157	0	1	142	0	0	0	0	0	0	0	0	0	0	0	0	0
15	1	38	2	138	175	0	1	173	0	0	0	0	0	0	0	0	0	0	0	0	0
16	1	38	2	138	175	0	1	173	0	0	0	0	0	0	0	0	0	0	0	0	0
17	1	47	0	120	229	0	0	129	1	2	6	1	2	3	0	0	0	0	0	0	0
18	1	47	0	120	229	0	0	129	1	2	6	1	2	3	0	0	0	0	0	0	0
19	0	62	0	140	248	0	0	160	0	3	6	0	2	2	0	0	0	0	0	0	0
20	1	40	0	130	254	0	0	147	0	3	4	1	1	1	0	0	0	0	0	0	0
21	1	52	0	140	205	1	0	155	1	3	1	0	0	0	0	0	0	0	0	0	0
22	1	56	2	130	216	1	0	142	1	0	6	1	1	1	0	0	0	0	0	0	0
23	1	48	1	130	219	0	1	148	0	1	0	0	0	0	0	0	0	0	0	0	0
24	1	58	1	130	244	0	0	160	0	1	8	1	0	0	0	0	0	0	0	0	0
25	1	58	2	130	214	0	0	175	0	3	2	2	2	3	0	0	0	0	0	0	0
26	1	60	0	130	206	0	0	152	1	2	4	1	2	3	0	0	0	0	0	0	0

Gambar 14 Hasil

Berdasarkan hasil tersebut, maka didapatkan faktor resiko penyakit jantung tertinggi adalah slp yaitu Besar kemiringan segmen ST pada kondisi latihan pucak atau maksimum dengan hasil 4,164596. Hal itu bisa kita lihat pada table Decision Variabel.

#### IV. KESIMPULAN

Metode regresi logistik adalah pendekatan statistik yang efektif untuk memodelkan hubungan antara variabel independen (faktor risiko penyakit jantung) dan variabel dependen (keberadaan atau tidaknya penyakit jantung).

Dalam penerapan data mining, regresi logistik dapat digunakan untuk mengidentifikasi faktor risiko yang paling signifikan dalam mengembangkan penyakit jantung. Melalui analisis regresi logistik, kita dapat mengukur tingkat hubungan antara setiap faktor risiko dengan kejadian penyakit jantung.

Dengan menggunakan data mining, seperti teknik eksplorasi data, pra pemrosesan data, dan pemilihan fitur, kita dapat mempersiapkan data yang tepat untuk dianalisis menggunakan metode regresi logistik. Hal ini melibatkan pembersihan data, penggabungan data dari berbagai sumber, dan transformasi data jika diperlukan.

Setelah melakukan analisis regresi logistik, kita dapat mendapatkan koefisien regresi yang mengindikasikan sejauh mana setiap faktor risiko berkontribusi terhadap risiko penyakit jantung. Koefisien positif menunjukkan peningkatan risiko, sementara koefisien negatif menunjukkan pengurangan risiko.

Selain itu, dengan menggunakan metode regresi logistik, kita juga dapat memprediksi probabilitas

individu terkena penyakit jantung berdasarkan faktor risiko yang dimilikinya. Ini dapat memberikan informasi yang berharga dalam menentukan tindakan pencegahan atau perawatan yang tepat untuk individu tersebut.

Dalam kesimpulannya, penerapan data mining dengan menggunakan metode regresi logistik dapat membantu dalam menganalisis faktor risiko penyakit jantung dengan mengidentifikasi faktor risiko yang signifikan dan memprediksi risiko individu. Hal ini dapat berguna dalam pengembangan strategi pencegahan, deteksi dini, dan pengelolaan penyakit jantung.

#### REFERENSI

- [1] Abdhul, Y. (2023). Studi Pendahuluan: Definisi, Tujuan, Cara Membuat dan Contoh - Deepublish Store. Diakses 27 Juni 2023, dari <https://deepublishstore.com/blog/studi-pendahuluan/>
- [2] Mengenal Penyakit Jantung Koroner dan Pencegahannya – Dinas Kesehatan Provinsi Kalimantan Barat. (2023). Diakses 27 Juni 2023, dari <https://dinkes.kalbarprov.go.id/mengenal-penyakit-jantung-koroner-dan-pencegahannya/>
- [3] Menor, Deion. (2023) Cara Melakukan Regresi Logistik di Excel. Diakses dari <https://sheetaki.com/logistic-regression-excel/>
- [4] Algorit.ma. (2022). 3 Tipe Logistic Regression yang Wajib Diketahui Data Analyst. Diakses dari <https://algorit.ma/blog/logistic-regression-adalah-2022/>
- [5] Vincentmichael089. (2019). Machine Learning : Mengenal Logistic Regression. Diakses dari <https://vincentmichael089.medium.com/machine-learning-2-logistic-regression-96b3d4e7b603>
- [6] Saretta, Radius, Irene. (2021). Data Mining : Pengertian, Fungsi, Penerapan, dan Metode Pengambilannya. Diakses dari <https://www.cermati.com/artikel/data-mining-pengertian-fungsi-penerapan-dan-metode-pengambilannya>
- [7] Kunci. Apa itu Data Mining ? Tujuan, Metode, dan Contohnya. Diakses dari <https://www.kunci.com/posts/apa-itu-data-mining>
- [8] Opendata.jabar. Dataset – Open Data Jabar. Diakses dari <https://opendata.jabarprov.go.id/id/dataset/jumlah-pemotongan-domba-berdasarkan-jenis-kelamin-dan-kabupatenkota-di-jawa-barat>
- [9] Wikipedia. (2021). Regresi logistik. Diakses dari [https://id.wikipedia.org/wiki/Regresi\\_logistik](https://id.wikipedia.org/wiki/Regresi_logistik)
- [10] Hendayana, Rachmat. Penerapan Metode Regresi Logistik Dalam Menganalisis. Diakses dari <https://media.neliti.com/media/publications/31101-penerapan-metode-regresi-logistik-dalam-a796735e.pdf>
- [11] Amazon. Apa itu Regresi Logistik ?. Diakses dari <https://aws.amazon.com/id/what-is/logistic-regression/#:~:text=Regresi%20logistik%20adalah%20teknik%20analisis,terbatas%2C%20seperti%20ya%20atau%20tidak.>