



Prediksi Penyakit Jantung Dengan Algoritma Regresi Linier

Agung Wijayadhi, Muhammad Makmun Effendi, Sugeng Budi Rahardjo

Fakultas Teknik, Teknik Informatika, Pelita Bangsa, Bekasi, Indonesia

Email : 1awijayadhi@gmail.com, 2effendiyan@pelitabangsa.ac.id, 3sugeng@pelitabangsa.ac.id

Email Penulis Korespondensi: sugeng@pelitabangsa.ac.id

Abstrak– Dalam penelitian ini, kami mengevaluasi kemampuan algoritma regresi linier untuk memprediksi risiko penyakit jantung pada individu. Kami menggunakan data dari sumber yang terpercaya dan melakukan preprocessing yang diperlukan untuk membersihkan dan menyediakan data untuk model. Hasil analisis menunjukkan bahwa algoritma regresi linier dapat digunakan dengan baik untuk memprediksi risiko penyakit jantung pada individu dengan tingkat akurasi yang cukup tinggi. Kami juga mengevaluasi beberapa faktor yang mempengaruhi risiko penyakit jantung dan menunjukkan bahwa faktor-faktor tersebut dapat diidentifikasi dan diintegrasikan ke dalam model kami untuk meningkatkan kinerjanya. Selain itu, kami juga mengevaluasi metode validasi yang digunakan untuk mengevaluasi model kami dan menunjukkan bahwa metode tersebut dapat digunakan untuk menentukan kinerja model secara obyektif. Hasil dari penelitian ini memberikan dasar yang kuat untuk mengembangkan sistem prediksi penyakit jantung yang lebih baik di masa depan. Dan hasil penelitian ini cukup sangat akurat mampu memberikan hasil yang baik dengan nilai Root Mean Squared Error: 0.379 +/- 0.000 dan Squared Error: 0.144 +/- 0.229.

Kata Kunci: Data Mining, Regresi Linier, Penyakit Jantung

Abstract– *In this study, we evaluated the ability of a linear regression algorithm to predict heart disease risk in individuals. We use data from trusted sources and perform the necessary preprocessing to clean and provide the data for the model. The results of the analysis show that the linear regression algorithm can be used well to predict the risk of heart disease in individuals with a fairly high degree of accuracy. We also evaluated several factors that influence heart disease risk and demonstrated that they could be identified and integrated into our model to improve its performance. In addition, we also evaluated the validation methods used to evaluate our models and demonstrated that they can be used to objectively determine model performance. The results from this study provide a solid foundation for developing a better heart disease prediction system in the future. And the results of this study are quite accurate enough to give good results with a Root Mean Squared Error: 0.379 +/- 0.000 and Squared Error: 0.144 +/- 0.229.*

Keywords: Data Mining, Linear Regression, Heart Disease

1. PENDAHULUAN

Penyakit jantung adalah masalah kesehatan yang sangat umum di dunia. Prediksi penyakit jantung dapat membantu dalam pencegahan dan pengobatan yang lebih efektif. Algoritma regresi linier dapat digunakan untuk memprediksi risiko penyakit jantung dengan menganalisis data klinis dan faktor risiko. Selanjutnya, dalam penelitian ini akan dijelaskan bagaimana algoritma regresi linier digunakan untuk memprediksi risiko penyakit jantung dengan menganalisis data klinis seperti tekanan darah, kolesterol, dan aktivitas fisik serta faktor risiko seperti usia, jenis kelamin, dan riwayat keluarga. Data yang digunakan dalam penelitian ini akan diambil dari sumber yang terpercaya dan valid. Analisis yang akan dilakukan akan membandingkan hasil prediksi dari algoritma regresi linier dengan metode lain seperti metode klasifikasi. Hasil dari penelitian ini diharapkan dapat memberikan informasi yang berguna bagi profesional kesehatan dalam pencegahan dan pengobatan penyakit jantung.

Selain itu, dalam penelitian ini juga akan dijelaskan bagaimana algoritma regresi linier dapat digunakan untuk mengoptimalkan faktor-faktor yang mempengaruhi risiko penyakit jantung. Hal ini akan dilakukan dengan menganalisis korelasi antara variabel-variabel yang digunakan dalam model dan mengevaluasi pengaruh masing-masing variabel terhadap hasil prediksi.

Hasil dari penelitian ini diharapkan dapat memberikan kontribusi yang signifikan dalam pengembangan metode prediksi risiko penyakit jantung yang lebih efektif dan akurat. Selain itu, dapat juga memberikan informasi yang berguna bagi individu yang memiliki risiko tinggi untuk mengambil tindakan preventif lebih awal.

Berdasarkan uraian masalah diatas, maka dalam penelitian ini digunakan algoritma regresi linear sebagai proses identifikasi data dengan judul “PREDIKSI PENYAKIT JANTUNG DENGAN ALGORITMA REGRESI LINIER”.

Penyakit jantung adalah kondisi medis yang mengacu pada masalah yang terjadi pada jantung dan pembuluh darah. Ini termasuk berbagai kondisi seperti penyakit jantung koroner, stroke, penyakit jantung bawaan, dan masalah lain yang dapat mengganggu kerja jantung dan pembuluh darah. Penyakit jantung sering disebabkan oleh faktor-faktor seperti gaya hidup tidak sehat, obesitas, stres, dan faktor risiko lainnya yang dapat mempengaruhi kesehatan jantung. Pengobatan dan pencegahan penyakit jantung sangat penting untuk memastikan kesehatan jangka panjang dan memperpanjang harapan hidup individu.[1]

Data mining adalah proses pengembangan dan aplikasi teknik untuk mengekstrak informasi dan pola dari data besar, dan menggunakannya untuk membuat keputusan yang berguna. Data mining memanfaatkan teknik statistik, machine learning, dan database untuk menganalisis data dan menemukan informasi tersembunyi yang mungkin tidak dapat ditemukan dengan cara manual. Data mining berguna untuk berbagai aplikasi, seperti pemasaran, manajemen risiko, dan pengambilan keputusan bisnis. Algoritma data mining seperti clustering, association rule mining, decision trees, dan neural networks digunakan untuk mengungkap pola dan hubungan dalam data. Data

mining membantu perusahaan dan organisasi untuk membuat keputusan yang lebih baik dengan memanfaatkan informasi yang tersembunyi dalam data.[2]

Secara sederhana Data Mining adalah proses untuk menambang atau menggali informasi yang tersembunyi dari bongkahan data besar. Inti dari proses Knowledge Discovery in Database (KDD) adalah Data mining, yang algoritmanya mengeksplor dan membangun model data. Knowledge Discovery in Database (KDD) sebagai pengorganisasian untuk proses pengidentifikasian data yang benar dan berguna untuk penemuan pola dari kumpulan data yang besar dan kompleks.[3]

Data mining adalah proses pengembangan dan aplikasi teknik untuk mengungkap pola dan informasi dari data. Berikut adalah beberapa teknik data mining yang sering digunakan: Clustering: membagi data menjadi beberapa kelompok berdasarkan kesamaan atribut. Association rule mining: menemukan hubungan antara item dalam data dengan menggunakan aturan asosiasi. Decision tree: menentukan tindakan berdasarkan suatu kondisi dan menciptakan pohon keputusan. Neural network: menggunakan jaringan saraf untuk menyelesaikan masalah dan melakukan prediksi. Naive Bayes: menggunakan teori probabilitas untuk menentukan kelas suatu data. Regression: menentukan hubungan antara variabel independen dan dependen dan membuat prediksi nilai. SVM (Support Vector Machine): membagi data menjadi dua kelas dengan memaksimalkan jarak antara kelas..[4] fungsi atau subkegiatan apa sajakah yang ada dalam data mining dalam rangka menemukan, menggali, atau menambang pengetahuan tersebut? Mengacu kepada Larose (2005), terdapat enam fungsi dalam data mining, yaitu : Fungsi deskripsi, fungsi estimasi, fungsi prediksi, fungsi klasifikasi, fungsi pengelompokkan dan fungsi asosiasi. [5]

Prediksi adalah suatu proses memperkirakan secara sistematis tentang sesuatu yang paling mungkin terjadi di masa depan berdasarkan informasi masa lalu dan sekarang yang dimiliki, agar kesalahannya (selisih antara sesuatu yang terjadi dengan hasil perkiraan) dapat diperkecil. Prediksi tidak harus memberikan jawaban secara pasti kejadian yang akan terjadi, melainkan berusaha untuk mencari jawaban sedekat mungkin yang akan terjadi.[6]

Regresi linier adalah metode statistik yang digunakan untuk memprediksi hubungan antara variabel independen (predictor/input) dan variabel dependen (respons/output) dengan membuat garis linier (linear model) sebagai model dasar. Model ini mengasumsikan hubungan antara kedua variabel tersebut adalah linier. Hasil dari regresi linier adalah sebuah equation yang dapat digunakan untuk memprediksi nilai respons/output berdasarkan nilai dari variabel independen/predictor.

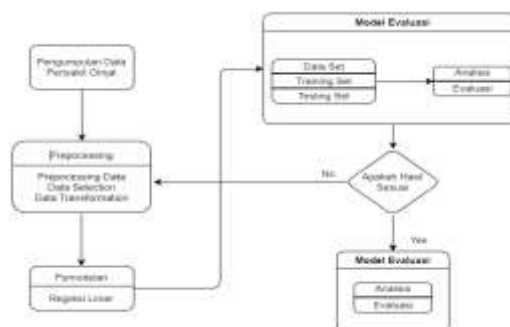
Regresi linier dapat digunakan untuk memecahkan masalah dalam banyak bidang, seperti ekonomi, teknik, dan bisnis, dengan memprediksi hubungan antara dua atau lebih variabel. Hasil dari regresi linier juga dapat digunakan untuk menentukan tingkat korelasi antara variabel dan menilai seberapa besar pengaruh suatu variabel independen terhadap variabel dependen.

Algoritma regresi linier termasuk dalam kelompok algoritma supervised learning. Algoritma regresi linier menggunakan data yang memiliki hubungan linier antara variabel dependen dan independen untuk membuat model prediksi. Algoritma ini sangat berguna dalam memprediksi nilai numerik, seperti harga saham, biaya produksi, atau jumlah penjualan. Algoritma regresi linier dapat digunakan untuk data tunggal (simple linear regression) atau data dengan lebih dari satu variabel independen (multiple linear regression). Algoritma ini bersifat sederhana dan mudah digunakan, dan memiliki hasil yang akurat jika data memenuhi syarat-syarat yang dibutuhkan, seperti linearitas, homoskedastisitas, dan tidak ada autokorelasi.[7]

2. METODOLOGI PENELITIAN

2.1 Proses Tahapan Data Mining

Dalam melakukan prediksi data penyakit jantung yang akan di uji sesuai pemodelan data yang akan digunakan agar mempermudah penelitian dan berjalan sesuai diinginkan maka dibuat alur atau tahapan dalam penelitian ini sebagai berikut:



Gambar 1. Tahapan Penelitian

2.2 Data Selection

Data selection adalah proses pemilihan data yang akan digunakan dalam pemodelan regresi linier. Tahap ini sangat penting karena kualitas data akan mempengaruhi hasil dari model. Pemilihan data yang tepat akan memastikan bahwa model regresi linier yang dibangun memiliki akurasi yang baik dan dapat menyediakan solusi yang berguna untuk masalah yang akan dicari solusinya.

Tabel 1. Data Selection

Usia	Nyeri Dada	Merokok	Kecemasan
69	1	2	2
74	2	1	1
59	1	1	1
63	2	2	2
63	1	2	1
75	1	2	1
52	2	1	1
51	2	2	2
68	2	1	2
53	2	2	2
61	2	2	2
72	1	1	1
60	2	1	1
58	2	1	1

2.3 Data Transformation

Pada Tahap ini yaitu *Data Transformation* merupakan sebuah proses mengubah format data atau atribut awal untuk proses pengujian prediksi menggunakan algoritma regresi linier dengan cara hitung manual atau dengan cara komputerasi yaitu menggunakan *tools RapidMiner*. Berikut adalah hasil data yang sudah diolah.

Tabel 2. Data Transformation

Usia	Nyeri Dada
69	1
74	2
59	1
63	2
63	1
75	1
52	2
51	2
68	2
53	2

61	2
72	1
60	2
58	2

2.4 Pemodelan

Pemodelan regresi linier melibatkan beberapa tahapan, diantaranya:

1. Analisis explorasi data: melakukan analisis terhadap data untuk memahami hubungan antara variabel independen dan dependen dan memastikan bahwa asumsi linieritas terpenuhi.
2. Pemilihan variabel independen: memilih variabel independen yang memiliki hubungan yang kuat dengan variabel dependen.
3. Persiapan data: melakukan pembersihan data dan normalisasi untuk memastikan data siap untuk model.
4. Pelatihan model: membangun model regresi linier dengan memilih model yang sesuai dan melatih model dengan data latih.
5. Evaluasi model: menilai akurasi model dengan mengukur residuals dan membandingkan hasil prediksi dengan nilai aktual.
6. Perbaikan model: melakukan perbaikan model jika diperlukan dengan memodifikasi variabel independen atau memilih model yang berbeda.
7. Implementasi: menerapkan model regresi linier pada data baru untuk membuat prediksi.

$$b = \frac{n \sum_{i=1}^n x_i y_i - \left(\sum_{i=1}^n x_i \right) \left(\sum_{i=1}^n y_i \right)}{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2} \quad (1)$$

$$a = \bar{y} - b\bar{x} \quad (2), \text{ sehingga}$$

$$a = \frac{\sum_{i=1}^n y_i}{n} - b \frac{\sum_{i=1}^n x_i}{n} \quad (3)$$

n: banyaknya pasangan data

yi: nilai perubahan tak bebas Y ke-i

xi: nilai perubahan bebas X ke-i

Menghitung persamaan regresi linearnya

$Y = a + bX$

Y: perubahan tak-bebas

a: konstanta

X: perubahan bebas

b: kemiringan

2.5 Evaluasi dan Hasil

Proses Evaluasi dilakukan dengan tujuan untuk mengetahui hasil perhitungan yang telah dianalisa dengan menggunakan metode algoritma regresi linier apakah berfungsi dengan baik atau tidak. Proses pengujian tersebut menggunakan *software tools RapidMiner*. Dengan *tools* tersebut bisa mengetahui dan melihat data apakah sesuai dengan yang diharapkan atau tidak, lalu mengetahui bagaimana variabel dependen/kriteria dapat diprediksikan melalui variabel independen atau variabel predictor secara individual.

3. HASIL DAN PEMBAHASAN

3.1 Hasil Penelitian

3.1.1 Data Uji

Penelitian yang telah diteliti ini menggunakan algoritma Regresi Linier, untuk mengidentifikasi penyakit jantung dan mendapatkan hasil *Root Mean Square Error* (RMSE) yang sangat akurat serta prediksi yang dapat digunakan dalam pengambilan keputusan pasien yang terkena penyakit atau tidak. Sumber data berasal dari situs *Kaggle.com*, dan data yang digunakan dalam penelitian ini terdiri dari atribut atau variabel seperti usia, nyeri dada dan hasil test.



3.1.2 Split Validation

Split validation adalah metode pembagian data untuk memvalidasi model machine learning dengan membagi data menjadi dua bagian, yaitu data *training* dan data *testing*. Data *training* digunakan untuk melatih model, sedangkan data *testing* digunakan untuk memvalidasi model yang telah dilatih dengan mengukur akurasi prediksinya.

Tabel 3. Dataset

Usia	Nyeri Dada	Hasil Test
69	1	1
74	2	1
59	1	0
63	2	0
63	1	0
75	1	1
52	2	1
51	2	1
68	2	0
53	2	1
61	2	1
72	1	1
60	2	0
58	2	1
69	2	0
48	1	1
75	2	1
57	2	1
68	2	1
61	1	0
44	2	1
64	1	1
21	2	0



60	2	1
72	2	1
65	1	1
61	2	1
69	1	0
53	2	1
55	1	0
57	2	0
62	2	1
56	2	0
67	2	1
59	1	0
59	2	1
60	1	1
56	1	0
56	2	1
60	2	1
68	2	1
63	1	1
77	1	1
52	2	1
70	2	1
72	2	1
62	2	1
64	2	1
70	1	1
60	1	0
56	1	1



63	2	1
54	2	1
49	2	1
57	1	1
52	1	1
63	1	1
73	1	1
47	1	1
69	2	1
70	1	1
60	1	0
70	1	1
68	1	1
74	1	1
71	2	1
56	1	1
66	2	1
76	2	1
78	2	1
68	2	1
66	2	1
67	1	1
60	2	1
61	2	1
58	2	1
76	1	1
56	2	1
67	2	1



73	2	1
58	1	0
54	2	1
62	2	1
81	1	1
56	1	1
60	1	1
66	1	1
62	1	1
62	2	1
55	2	1
62	1	1
71	1	1
52	2	1
59	1	1
48	2	1
60	1	1
61	2	1
59	2	1
64	1	1
56	2	1

3.1.3 Perhitungan Regresi Linier

Didalam perhitungan Regresi Linier, untuk mencari nilai :

X_1Y : Usia X Hasil Test

X_2Y : Nyeri Dada X Hasil Test

X_1X_2 : Usia X Nyeri Dada

X_1^2 : Usia²

X_2^2 : Nyeri Dada²

Tabel 4. Perhitungan Regresi Linier

No	Usia (X_1)	Nyeri Dada (X_2)	Hasil Test (Y)	X_1Y	X_2Y	X_1X_2	X_1^2	X_2^2
1	69	1	1	69	1	69	4761	1
2	74	2	1	74	2	148	5476	4



3	59	1	0	0	0	59	3481	1
4	63	2	0	0	0	126	3969	4
5	63	1	0	0	0	63	3969	1
6	75	1	1	75	1	75	5625	1
7	52	2	1	52	2	104	2704	4
8	51	2	1	51	2	102	2601	4
9	68	2	0	0	0	136	4624	4
10	53	2	1	53	2	106	2809	4
11	61	2	1	61	2	122	3721	4
12	72	1	1	72	1	72	5184	1
13	60	2	0	0	0	120	3600	4
14	58	2	1	58	2	116	3364	4
15	69	2	0	0	0	138	4761	4
16	48	1	1	48	1	48	2304	1
17	75	2	1	75	2	150	5625	4
18	57	2	1	57	2	114	3249	4
19	68	2	1	68	2	136	4624	4
20	61	1	0	0	0	61	3721	1
21	44	2	1	44	2	88	1936	4
22	64	1	1	64	1	64	4096	1
23	21	2	0	0	0	42	441	4
24	60	2	1	60	2	120	3600	4
25	72	2	1	72	2	144	5184	4
26	65	1	1	65	1	65	4225	1
27	61	2	1	61	2	122	3721	4
28	69	1	0	0	0	69	4761	1
29	53	2	1	53	2	106	2809	4
30	55	1	0	0	0	55	3025	1
31	57	2	0	0	0	114	3249	4
32	62	2	1	62	2	124	3844	4
33	56	2	0	0	0	112	3136	4
34	67	2	1	67	2	134	4489	4
35	59	1	0	0	0	59	3481	1
36	59	2	1	59	2	118	3481	4
37	60	1	1	60	1	60	3600	1
38	56	1	0	0	0	56	3136	1



39	56	2	1	56	2	112	3136	4
40	60	2	1	60	2	120	3600	4
41	68	2	1	68	2	136	4624	4
42	63	1	1	63	1	63	3969	1
43	77	1	1	77	1	77	5929	1
44	52	2	1	52	2	104	2704	4
45	70	2	1	70	2	140	4900	4
46	72	2	1	72	2	144	5184	4
47	62	2	1	62	2	124	3844	4
48	64	2	1	64	2	128	4096	4
49	70	1	1	70	1	70	4900	1
50	60	1	0	0	0	60	3600	1
51	56	1	1	56	1	56	3136	1
52	63	2	1	63	2	126	3969	4
53	54	2	1	54	2	108	2916	4
54	49	2	1	49	2	98	2401	4
55	57	1	1	57	1	57	3249	1
56	52	1	1	52	1	52	2704	1
57	63	1	1	63	1	63	3969	1
58	73	1	1	73	1	73	5329	1
59	47	1	1	47	1	47	2209	1
60	69	2	1	69	2	138	4761	4
61	70	1	1	70	1	70	4900	1
62	60	1	0	0	0	60	3600	1
63	70	1	1	70	1	70	4900	1
64	68	1	1	68	1	68	4624	1
65	74	1	1	74	1	74	5476	1
66	71	2	1	71	2	142	5041	4
67	56	1	1	56	1	56	3136	1
68	66	2	1	66	2	132	4356	4
69	76	2	1	76	2	152	5776	4
70	78	2	1	78	2	156	6084	4
71	68	2	1	68	2	136	4624	4
72	66	2	1	66	2	132	4356	4
73	67	1	1	67	1	67	4489	1
74	60	2	1	60	2	120	3600	4

75	61	2	1	61	2	122	3721	4
76	58	2	1	58	2	116	3364	4
77	76	1	1	76	1	76	5776	1
78	56	2	1	56	2	112	3136	4
79	67	2	1	67	2	134	4489	4
80	73	2	1	73	2	146	5329	4
81	58	1	0	0	0	58	3364	1
82	54	2	1	54	2	108	2916	4
83	62	2	1	62	2	124	3844	4
84	81	1	1	81	1	81	6561	1
85	56	1	1	56	1	56	3136	1
86	60	1	1	60	1	60	3600	1
87	66	1	1	66	1	66	4356	1
88	62	1	1	62	1	62	3844	1
89	62	2	1	62	2	124	3844	4
90	55	2	1	55	2	110	3025	4
91	62	1	1	62	1	62	3844	1
92	71	1	1	71	1	71	5041	1
93	52	2	1	52	2	104	2704	4
94	59	1	1	59	1	59	3481	1
95	48	2	1	48	2	96	2304	4
96	60	1	1	60	1	60	3600	1
97	61	2	1	61	2	122	3721	4
98	59	2	1	59	2	118	3481	4
99	64	1	1	64	1	64	4096	1
100	56	2	1	56	2	112	3136	4

Dan jika sudah terhitung semua maka akan di jumlahkan semua data dari atas sampai bawah menggunakan rumus excel yaitu (=SUM). Setelah itu di hitung kembali untuk menentukan prediksi menggunakan rumus Regresi Linier Sederhana sebagai berikut :

Ket :

Y : Variabel Dependen

X : Variabel Independen

a : Konstanta

b : Koefisien Regresi

n : Jumlah Data

$$b_1 = \frac{n \sum(x_1y) - (\sum x_1)(\sum y)}{n(\sum x_1)^2 - (\sum x_1)^2}$$

$$b_1 = \frac{100(5208) - (6202)(83)}{100(392190) - (6202)^2}$$

$$b_1 = \frac{520800 - 514766}{39219000 - 38464804}$$

$$b_1 = \frac{6034}{-34543804}$$

$$b_1 = -0,0001746768$$

Nilai b1 ini adalah hasil perhitungan dari pada Koefisien Regresi 1

$$b_2 = \frac{n \sum(x_2y) - (\sum x_2)(\sum y)}{n(\sum x_2)^2 - (\sum x_2)^2}$$

$$b_2 = \frac{100(133) - (157)(83)}{100(271) - (157)^2}$$

$$b_2 = \frac{13300 - 13031}{27100 - 24649}$$

$$b_2 = \frac{269}{2451}$$

$$b_2 = 0,109751122$$

Nilai b2 ini adalah hasil perhitungan dari pada Koefisien Regresi 2

Setelah kita mendapatkan nilai b1 dan b2 kita mencari nilai a dengan rumus sebagai berikut :

$$a = \frac{\sum y - b_1 \sum x_1 + b_2 \sum x_2}{n}$$

$$a = \frac{83 - ((-0,0001746768) * (6202)) + ((0,109751122) * (157))}{100}$$

$$a = \frac{83 - (-1,0833455136) + 17,230926154}{100}$$

$$a = \frac{83 - 16,1475806404}{100}$$

$$a = \frac{66,8524193596}{100}$$

$$a = 0,6685241936$$

Nilai a ini adalah hasil perhitungan antara Koefisien Regresi 1 dan Koefisien Regresi 2 sehingga menghasilkan nilai Konstanta.

Diambil dari 3 contoh data random untuk memprediksi apakah akurat atau tidaknya menggunakan cara hitung manual sebagai berikut :

Tabel 5. Prediksi Penyakit Jantung

No	Usia (X ₁)	Nyeri Dada (X ₂)	Hasil_Test (Y)
1	58	2	?
2	21	2	?
3	69	1	?

Menghitung Persaman Regresi Linear

$$Y = a + b_1X_1 + b_2X_2$$

$$y = 0,6685241936 + (-0,0001746768 \times 58) + (0,109751122 \times 2)$$

$$= 0,8778951832$$

$$y = 0,6685241936 + (-0,0001746768 \times 21) + (0,109751122 \times 2)$$

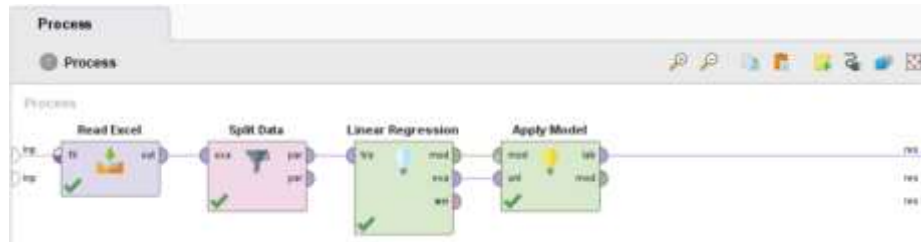
$$= 0,8843582248$$

$$y = 0,6685241936 + (-0,0001746768 \times 69) + (0,109751122 \times 1) \\ = 0,7662226164$$

3.2 Pembahasan

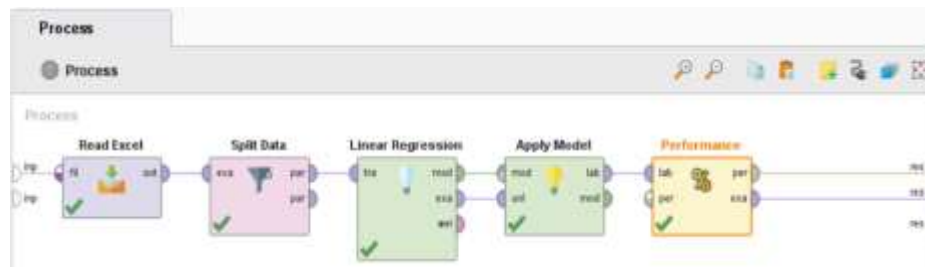
3.2.1 Proses Pengujian Data (RapidMiner)

Melakukan select *attributes* yaitu untuk mengetahui hasil prediksi dari rapidminer, hasil perhitungan manual dan hasil uji di rapidminer.



Gambar 2. Proses RapidMiner

Ketika prediksi sudah dicari langkah selanjutnya yaitu mengukur seberapa akurat hasil prediksi yang telah kita buat pada gambar 3.



Gambar 3. Proses Pencarian *Root Mean Squared Error* dan *Squared Error*

Untuk mempermudah dalam pembacaan data penyakit jantung, maka perlu di masukan tools performance untuk mencari *Root Mean Squared Error* dan *Squared Error*. Berikut ini hasilnya.

PerformanceVector

```
PerformanceVector:
root_mean_squared_error: 0.379 +/- 0.000
squared_error: 0.144 +/- 0.229
```

Gambar 4. Hasil Pengujian *Root Mean Square Error* dan *Square Error*

Langkah selanjutnya, dilakukanlah implementasi dari algoritma regresi linier dengan menggunakan *software tools RapidMiner*. Berikut adalah tahapannya :

1. Menentukan prediksi dari dataset yang telah dimasukkan kedalam RapidMiner dan menghasilkan nilai prediksi.
2. Menentukan performance untuk mengetahui tingkat akurasi dan *Root Mean Square Error* dan *Squared Error*.

3.2.2 Analisa Hasil

Berdasarkan hasil pengujian yang telah dilakukan menggunakan *tools RapidMiner* bahwa atribut (usia, nyeri dada dan hasil test) berpengaruh signifikan terhadap penelitian ini terbukti dengan menggunakan algoritma regresi linier mampu memberikan hasil yang baik dengan nilai *Root Mean Squared Error*: 0.379 +/- 0.000 dan *Squared Error*: 0.144 +/- 0.229.

4. KESIMPULAN

Berdasarkan penelitian yang dilakukan untuk menentukan prediksi penyakit jantung menggunakan metode Regresi Linier maka dapat disimpulkan bahwa :

1. Algoritma regresi linier efektif digunakan dalam memprediksi risiko penyakit jantung berdasarkan analisis data klinis dan faktor risiko.



2. Faktor-faktor yang mempengaruhi risiko penyakit jantung dapat dioptimalkan melalui analisis algoritma regresi linier, seperti mengontrol tekanan darah dan kolesterol serta meningkatkan aktivitas fisik.
3. Hasil prediksi dari algoritma regresi linier cukup akurat dibandingkan dengan metode lain dalam memprediksi risiko penyakit jantung. Proses pengujian data pada penelitian ini menggunakan algoritma regresi linear mampu memberikan hasil yang baik dengan nilai Root Mean Squared Error: 0.379 +/- 0.000 dan *Squared Error*: 0.144 +/- 0.229.

Kesimpulan ini dapat digunakan oleh profesional kesehatan dalam pencegahan dan pengobatan penyakit jantung serta memberikan informasi yang berguna bagi individu yang memiliki risiko tinggi untuk mengambil tindakan preventif lebih awal. Namun perlu diingat, hasil dari penelitian ini hanya berlaku pada data yang digunakan dalam penelitian dan tidak dapat diterapkan pada populasi lain tanpa melakukan penelitian yang sama atau serupa. Dan hasil prediksi tersebut dengan menggunakan metode Regresi Linier dapat dipakai untuk prediksi data penyakit lain atau data lain yang memang setiap bulan dan tahunnya tidak bisa terprediksi dalam hal menjaga kestabilan data tersebut.

UCAPAN TERIMAKASIH

Terima kasih kepada segenap dosen kampus, keluarga dan teman-teman yang terus mendukung saya dalam proses penelitian ini.

REFERENCES

- [1] R. Annisa, "ANALISIS KOMPARASI ALGORITMA KLASIFIKASI DATA MINING UNTUK PREDIKSI PENDERITA PENYAKIT JANTUNG," *Jurnal Teknik Informatika Kaputama (JTIK)*, vol. 3, no. 1, 2019.
- [2] Z. Nabila, A. Rahman Isnain, and Z. Abidin, "ANALISIS DATA MINING UNTUK CLUSTERING KASUS COVID-19 DI PROVINSI LAMPUNG DENGAN ALGORITMA K-MEANS," *Jurnal Teknologi dan Sistem Informasi (JTSI)*, vol. 2, no. 2, p. 100, 2021, [Online]. Available: <http://jim.teknokrat.ac.id/index.php/JTSI>
- [3] "Implementasi Data Mining Untuk Pengelempokan Buku Di Perpustakaan Yayasan Nurul Islam Indonesia Baru Dengan Metode K-Means Clustering Haryani *, Dicky Nofriansyah **, Ita Mariami *** Program Studi Sistem Informasi, STMIK Triguna Dharma ** Program Studi Sistem Informasi, STMIK Triguna Dharma," 2021, [Online]. Available: <https://ojs.trigunadharma.ac.id/index.php/jct/index>
- [4] "204-Article Text-389-1-10-20130819".
- [5] M. Pengetahuan and B. Data, "PENGANTAR DATA MINING."
- [6] D. B. Srisulistiwati¹, M. Khaerudin², S. Rejeki³, and U. Bhayangkara Jakarta, "SISTEM INFORMASI PREDIKSI PENJUALAN ALAT TULIS KANTOR DENGAN METODE FP-GROWTH (STUDI KASUS TOKO KOPERASI SEKOLAH BINA MULIA)."
- [7] P. Sari Ramadhan and N. Safitri STMIK Triguna Dharma, "Penerapan Data Mining Untuk Mengestimasi Laju Pertumbuhan Penduduk Menggunakan Metode Regresi Linier Berganda Pada BPS Deli Serdang," vol. 18, no. SAINTIKOM, pp. 55–61, 2019, [Online]. Available: <https://sirusa.bps.go.id/index.php>