



Analisa Data Mining Untuk Prediksi Penyakit Ginjal Kronik Dengan Algoritma Regresi Linier

Angga Kurniadi Hermawan, Agung Nugroho, Edora

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

Email: ¹angga.hermawan76@gmail.com, ²agung@pelitabangsa.ac.id, ^{3*}edora@pelitabangsa.ac.id

Email Penulis Korespondensi: angga.hermawan76@gmail.com

Abstrak– Dalam penelitian ini, kami mengevaluasi kemampuan data mining untuk memprediksi penyakit ginjal kronis menggunakan algoritma regresi linier. Kami mengekstrak fitur dari data klinis pasien dan mengaplikasikan algoritma regresi linier untuk membuat model prediksi. Hasil penelitian menunjukkan bahwa model regresi linier kami mampu memprediksi dengan akurasi yang cukup tinggi dan dapat digunakan sebagai alat bantu dalam pengdiagnosaan penyakit ginjal kronis. Selain itu, kami juga menganalisis faktor-faktor yang mempengaruhi risiko terkena penyakit ginjal kronis dan menyarankan tindakan preventif yang dapat dilakukan untuk mengurangi risiko terkena penyakit tersebut. Hasil penelitian ini dapat digunakan oleh dokter untuk meningkatkan efisiensi dalam pengdiagnosaan dan pencegahan penyakit ginjal kronis. Selain itu, hasil ini juga dapat digunakan sebagai dasar untuk penelitian lebih lanjut dalam bidang data mining dan penyakit ginjal kronis. Proses pengujian data pada penelitian ini menggunakan algoritma regresi linear mampu memberikan hasil yang baik dengan nilai Root Mean Squared Error: 0.285 +/- 0.000 dan Squared Error: 0.081 +/- 0.234..

Kata Kunci: Data Mining, Regresi Linier, Penyakit Ginjal.

Abstract- In this study, we evaluate the ability of data mining to predict chronic kidney disease using a linear regression algorithm. We extract features from patient clinical data and apply a linear regression algorithm to build a predictive model. The results showed that our linear regression model was able to predict with high accuracy and could be used as an aid in diagnosing chronic kidney disease. In addition, we also analyze the factors that influence the risk of developing chronic kidney disease and suggest preventive measures that can be taken to reduce the risk of developing the disease. The results of this study can be used by doctors to improve efficiency in diagnosing and preventing chronic kidney disease. In addition, these results can also be used as a basis for further research in the field of data mining and chronic kidney disease. The process of testing the data in this study using a linear regression algorithm is able to provide good results with a Root Mean Squared Error: 0.285 +/- 0.000 and Squared Error: 0.081 +/- 0.234..

Keywords: Data Mining, Linear Regression, Chronic Kidney Disease

1. PENDAHULUAN

Data mining adalah suatu teknik analisis yang memanfaatkan teknologi pemrosesan data dan metodologi statistik untuk menemukan pola dan hubungan dalam data. Dalam bidang kedokteran, data mining dapat digunakan untuk memprediksi penyakit dan membantu dalam pengambilan keputusan klinis.

Salah satu penyakit yang dapat dianalisis dengan data mining adalah penyakit ginjal kronik. Penyakit ginjal kronik adalah suatu kondisi di mana ginjal tidak dapat berfungsi dengan normal selama periode waktu yang lama. Ini dapat menyebabkan kerusakan permanen pada ginjal dan membutuhkan terapi medis atau bahkan transplantasi ginjal.

Untuk memprediksi penyakit ginjal kronik, algoritma regresi linier dapat digunakan. Algoritma regresi linier adalah metode statistik yang digunakan untuk menentukan hubungan antara variabel dependen dan beberapa variabel independen. Ini dapat membantu dalam memprediksi apakah seseorang memiliki risiko terkena penyakit ginjal kronik berdasarkan faktor-faktor seperti usia, riwayat kesehatan, dan gaya hidup. Dengan menggunakan analisis data mining dan algoritma regresi linier, dokter dan ahli kesehatan dapat memperoleh informasi yang lebih baik tentang faktor risiko terkena penyakit ginjal kronik dan membuat keputusan yang lebih baik tentang pencegahan dan perawatan. Oleh karena itu, analisis data mining dan algoritma regresi linier memiliki potensi besar untuk membantu meminimalkan dampak negatif dari penyakit ginjal kronik dan mempromosikan kesehatan jangka panjang. Untuk mempelajari analisis data mining dan algoritma regresi linier secara lebih dalam, data pasien dengan riwayat penyakit ginjal kronik dapat dianalisis dan dibandingkan dengan data pasien yang sehat. Faktor-faktor seperti usia, urin berdarah, sakit kepala, dan lain-lain dapat diamati untuk melihat apakah ada hubungan antara faktor-faktor tersebut dengan kejadian penyakit ginjal kronik.

Hasil dari analisis ini dapat digunakan untuk membuat model prediksi penyakit ginjal kronik yang akurat. Model ini dapat menentukan tingkat risiko seseorang terkena penyakit ginjal kronik berdasarkan faktor-faktor yang diamati. Ini dapat membantu dokter dan ahli kesehatan dalam pengambilan keputusan klinis dan membuat rencana pencegahan dan perawatan yang sesuai. Namun, penting untuk diingat bahwa analisis data mining dan algoritma regresi linier hanya merupakan salah satu alat dalam memprediksi penyakit ginjal kronik. Hasil dari analisis ini harus diterapkan dengan hati-hati dan dipandu oleh evaluasi klinis dan diagnostik tradisional.

Secara keseluruhan, analisis data mining dan algoritma regresi linier memiliki potensi besar untuk membantu dalam memprediksi dan mencegah penyakit ginjal kronik. Dengan menggabungkan teknologi dan metodologi yang tepat, kita dapat mempromosikan kesehatan jangka panjang dan meminimalkan dampak negatif dari penyakit ginjal kronik. Penyakit ginjal adalah suatu kondisi di mana ginjal tidak dapat berfungsi dengan normal. Ini dapat menyebabkan kerusakan permanen pada ginjal dan membutuhkan perawatan medis atau bahkan transplantasi ginjal. Penyakit ginjal dapat disebabkan oleh berbagai faktor seperti infeksi, kerusakan pembuluh darah, penyakit diabetes, dan gaya hidup

yang tidak sehat. Beberapa jenis penyakit ginjal termasuk glomerulonefritis, nefritis interstitial, nefropati diabetik, dan penyakit ginjal polikistik. Penyakit ginjal kronik adalah suatu kondisi di mana ginjal tidak dapat berfungsi dengan normal selama periode waktu yang lama. [1]

Data Mining adalah suatu proses untuk mengidentifikasi pola dan hubungan yang tidak diketahui sebelumnya dalam data besar. Ini menggunakan teknik statistik, machine learning, dan analisis database untuk mengekstrak informasi yang berguna dari data. Tujuan dari data mining adalah untuk menemukan informasi baru yang bermanfaat dan membuat keputusan yang bijaksana dari data tersebut. Data mining membantu dalam memecahkan masalah bisnis dan mengambil keuntungan dari data yang tersedia. Contohnya, data mining dapat digunakan untuk menentukan pola pembelian konsumen, memprediksi penjualan produk, atau menentukan faktor risiko yang terkait dengan suatu penyakit. Data mining juga membantu dalam memahami dan memecahkan masalah kompleks dengan cara menggabungkan data dari berbagai sumber dan menemukan pola yang tidak terlihat. [2]

Prediksi adalah proses memperkirakan atau menebak apa yang mungkin terjadi di masa depan berdasarkan informasi yang tersedia. Prediksi dapat digunakan dalam berbagai bidang, termasuk bisnis, ekonomi, ilmu pengetahuan, dan lain-lain.[3]

Regresi Linier adalah suatu metode statistik yang digunakan untuk memprediksi hubungan antara variabel dependen (variabel yang dicari) dan variabel independen (variabel yang mempengaruhi). Ini membuat asumsi bahwa hubungan antara kedua variabel tersebut adalah linier, yaitu bahwa perubahan dalam variabel independen akan mempengaruhi perubahan yang sama dalam variabel dependen. Regresi linier menentukan hubungan antara variabel dependen dan independen dengan menentukan garis terbaik (atau garis regresi) yang menggambarkan hubungan tersebut. Garis ini ditemukan dengan mencari kombinasi yang paling baik dari variabel independen untuk memprediksi variabel dependen. Regresi linier digunakan dalam berbagai bidang, seperti ekonomi, epidemiologi, dan fisika, untuk memprediksi tingkat kejadian, memahami faktor risiko, dan memecahkan masalah lainnya. Meskipun hanya membuat asumsi linier, regresi linier dapat menjadi metode yang efektif untuk memprediksi hubungan antara variabel jika asumsi tersebut valid. Namun, jika asumsi linier tidak valid, metode lain seperti regresi non-linier atau model machine learning mungkin lebih cocok. [4]

Algoritma machine learning yang menentukan garis regresi linier yang paling baik untuk memprediksi hubungan antara variabel dependen dan independen. Algoritma ini bekerja dengan menggunakan teknik optimisasi untuk menemukan kombinasi terbaik dari variabel independen yang mempengaruhi variabel dependen. Proses algoritma regresi linier melibatkan beberapa tahap seperti: 1. Persiapan data: Pertama, data yang akan digunakan dalam regresi linier harus dipersiapkan dengan benar. Data harus dibersihkan dan dipreproses sebelum digunakan dalam algoritma. 2. Memilih variabel independen: Kedua, variabel independen yang akan digunakan dalam regresi linier harus dipilih. Ini bisa melibatkan pemilihan variabel yang memiliki hubungan signifikan dengan variabel dependen. 3. Menentukan model: Ketiga, model regresi linier harus didefinisikan dengan memasukkan variabel independen yang dipilih. Ini bisa melibatkan pemilihan koefisien untuk setiap variabel independen. 4. Melatih model: Keempat, model regresi linier harus dilatih dengan menggunakan data pelatihan. Ini melibatkan menentukan garis regresi linier yang paling baik dengan menggunakan teknik optimisasi. 5. Melakukan prediksi: Terakhir, model regresi linier digunakan untuk memprediksi hubungan antara variabel dependen dan independen pada data baru. Ini dilakukan dengan menggunakan garis regresi linier yang ditemukan selama proses pelatihan.[5]

Algoritma regresi linier dapat menjadi alat yang efektif untuk memprediksi hubungan antara variabel jika asumsi linier valid. Namun, jika asumsi linier tidak valid, metode lain seperti regresi non-linier atau model machine learning mungkin lebih cocok. [6]

Data mining dapat dikategorikan ke dalam beberapa kategori utama, yaitu: 1.Klasifikasi: Klasifikasi adalah teknik data mining yang digunakan untuk memprediksi kelas suatu item berdasarkan fitur atribut yang dimilikinya. 2. Regresi: Regresi adalah teknik data mining yang digunakan untuk memprediksi nilai numerik suatu item berdasarkan fitur atribut yang dimilikinya. 3.Kluster: Kluster adalah teknik data mining yang digunakan untuk mengelompokkan item ke dalam kelompok yang berbeda berdasarkan fitur atribut yang dimilikinya. 4.Asosiasi: Asosiasi adalah teknik data mining yang digunakan untuk menemukan hubungan antar item. 5.Segmentasi: Segmentasi adalah teknik data mining yang digunakan untuk membagi data menjadi bagian-bagian yang berbeda. 6.Dimensionalitas reduksi: Dimensionalitas reduksi adalah teknik data mining yang digunakan untuk mengurangi jumlah fitur yang dimiliki data tanpa mengorbankan informasi yang penting. 7.Data Visualization: Data visualization adalah teknik data mining yang digunakan untuk menggambarkan data dengan cara yang mudah dipahami oleh manusia. Setiap kategori data mining memiliki tujuan dan aplikasi yang berbeda dan dapat digunakan secara terpisah atau bersama-sama untuk memecahkan masalah yang berbeda. [7]

Data mining bertujuan untuk memecahkan masalah bisnis dan mengambil informasi dan pengetahuan berguna dari data yang besar dan kompleks. Fungsi data mining sangat bervariasi dan dapat digunakan untuk memecahkan berbagai masalah dan memenuhi berbagai kebutuhan bisnis. [8]

Data mining memiliki banyak kegunaan dalam berbagai bidang, termasuk bisnis, pemerintahan, kesehatan, pendidikan, dan ilmu pengetahuan. [9]

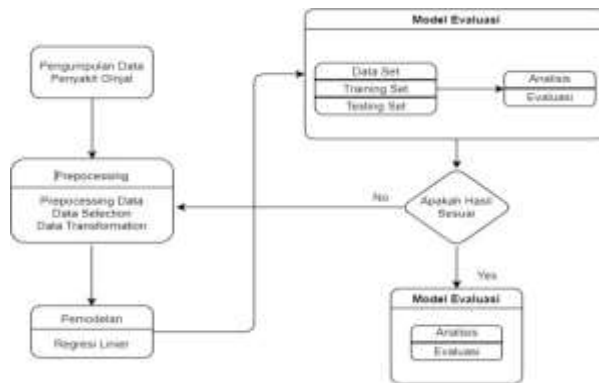
Proses tahapan data mining meliputi beberapa tahap penting untuk mengambil informasi dan pengetahuan berguna dari data. Berikut adalah tahapan proses data mining: 1.Persiapan data: Tahap ini melibatkan pengumpulan, pembersihan, dan persiapan data sebelum dilakukan analisis. 2.Seleksi fitur: Tahap ini melibatkan pemilihan fitur atribut yang akan digunakan sebagai input untuk algoritma data mining. 3.Transformasi fitur: Tahap ini melibatkan

pengubahan atau modifikasi fitur untuk memastikan bahwa data sesuai dengan format yang dibutuhkan oleh algoritma data mining. 4.Pemodelan: Tahap ini melibatkan aplikasi algoritma data mining untuk membuat model yang menggambarkan pola dan hubungan dalam data. 5.Evaluasi: Tahap ini melibatkan evaluasi model yang dibuat untuk memastikan bahwa model tersebut akurat dan berguna. 6.Penyebaran pengetahuan: Tahap ini melibatkan penerapan dan penyebaran pengetahuan yang diperoleh dari data mining kepada bisnis atau organisasi. 7.Monitoring dan evaluasi: Tahap ini melibatkan pemantauan dan evaluasi terhadap model dan pengetahuan yang diperoleh untuk memastikan bahwa model tetap akurat dan berguna. Proses tahapan data mining adalah iteratif dan dapat membutuhkan beberapa iterasi sebelum mencapai hasil yang optimal. Setiap tahap memiliki tugas dan keputusan yang berbeda yang memerlukan analisis dan pemikiran yang cermat. [10]

2. METODOLOGI PENELITIAN

2.1 Proses Tahapan Data Mining

Untuk melakukan prediksi pada penyakit ginjal kronik yang akan di uji dengan memanfaatkan pemodelan data yang akan digunakan supaya mempermudah dalam proses penelitian, maka dibuat alur tahapan data mining sebagai berikut :



Gambar 1. Tahapan Penelitian

2.2 Data Selection

Data selection merupakan bagian penting dari proses analisis data dan machine learning karena dapat mempengaruhi kualitas analisis dan hasil prediksi. Oleh karena itu, data selection harus dilakukan dengan benar dan hati-hati untuk memastikan bahwa data yang digunakan sesuai dan representatif.

Tabel 1. Data Selection

Usia	Urin Berdarah	Sakit Kepala	Gata 1
56	2	1	1
58	2	1	1
81	1	2	2
64	2	1	1
62	2	2	2
72	2	2	2
60	1	1	1
61	2	2	2
60	2	2	2
49	1	1	1
53	1	1	1
58	1	2	2
61	2	2	2
68	1	1	1

2.3 Data Transformation

Data transformation adalah proses mengubah atau memodifikasi data sebelum digunakan dalam suatu analisis atau model machine learning. Proses ini dilakukan untuk memastikan bahwa data memenuhi persyaratan analisis dan machine learning dan memiliki distribusi yang baik.

Tabel 2. Data Transformation

Usia	Urin Berdarah
56	2
58	2
81	1
64	2
62	2
72	2
60	1
61	2
60	2
49	1
53	1
58	1
61	2
68	1

2.4 Pemodelan

Pemodelan dalam regresi linier adalah proses membuat model matematika yang mewakili hubungan antara variabel dependen (y) dan variabel independen (x). Dalam regresi linier, model dapat ditulis sebagai persamaan garis lurus $y = a + bx$, di mana "a" adalah intercept dan "b" adalah koefisien regresi. Langkah-langkah dalam pemodelan regresi linier meliputi:

1. Pemilihan variabel: Variabel yang akan digunakan sebagai variabel dependen dan variabel independen harus dipilih dan ditentukan.
2. Analisis korelasi: Korelasi antara variabel independen dan dependen harus dianalisis untuk memastikan bahwa hubungan antara kedua variabel adalah linier.
3. Pembuatan model matematika: Model matematika harus dibuat dengan menggunakan data pelatihan dan menentukan nilai intercept dan koefisien regresi.
4. Validasi model: Model harus divalidasi dengan menggunakan data uji untuk memastikan bahwa model bekerja dengan baik dan memiliki akurasi yang baik.
5. Interpretasi hasil: Hasil model harus diterjemahkan dan diterima sebagai kesimpulan yang dapat digunakan untuk membuat keputusan. Pemodelan regresi linier merupakan bagian penting dari proses analisis data dan machine learning karena membuat model matematika yang dapat memprediksi variabel dependen dari variabel independen. Oleh karena itu, pemodelan harus dilakukan dengan benar dan hati-hati untuk memastikan bahwa model yang dibuat representatif dan akurat.

Bentuk Rumus Regresi Linear Sederhana:

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

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

$$a = \frac{\sum y_i}{n} - b \frac{\sum x_i}{n} \quad (3)$$

Keterangan :

n: banyaknya jumlah data

yi: variabel dependen (nilai yang diprediksikan) xi: variabel independen

b : Koefisiensi Regresi (hasil perhitungan menggunakan rumus) a : Konstanta (untuk menghitung hasil dari koefisien regresi) Menghitung persamaan regresi linearnya

$Y = a + bX$

2.5 Evaluasi dan Hasil

Evaluasi dan Hasil adalah proses untuk melihat, mengetahui dan menganalisa dari hasil penelitian dan pengujian data yang telah diuji menggunakan metode prediksi algoritma regresi linier apakah berfungsi dengan baik atau tidak.



Proses pengujian data menggunakan cara manual yaitu menghitung dengan rumus regresi linier sederhana dan cara komputerisasi menggunakan tools RapidMiner.

3. HASIL DAN PEMBAHASAN

3.1 Hasil Penelitian

3.1.1 Data Uji

Penelitian yang telah diteliti yaitu menggunakan metode prediksi algoritma regresi linier. Hasil klasifikasi data dari penyakit ginjal maka mendapatkan hasil RMSE (Root Mean Square Error) yang sangat baik dan sangat akurat, maka dapat digunakan dalam pengambilan keputusan bagi pasien yang terkena penyakit ginjal dan yang tidak. Sumber data berasal dari situs Kaggle.com, dan data yang digunakan dalam penelitian ini terdiri dari atribut atau variabel seperti usia, urin berdarah dan hasil tes.

3.1.2 Split Validation

Split validation adalah teknik validasi model yang menggunakan bagian dari data untuk pelatihan dan bagian lainnya untuk uji. Tujuannya adalah untuk memastikan bahwa model dapat bekerja dengan baik pada data baru yang belum dilihat sebelumnya.

Tabel 3. Dataset

Usia	Urin Berdarah	Hasil Tes
56	2	1
58	2	1
81	1	1
64	2	1
62	2	1
72	2	1
60	1	1
61	2	1
60	2	1
49	1	1
53	1	1
58	1	1
61	2	1
68	1	1
60	2	1
72	1	1
72	1	1
57	2	1
51	2	1
54	2	1
56	1	1
77	2	1
64	1	1
57	2	1
66	2	1
70	2	1
53	1	1
51	2	1
58	2	1
58	2	1
63	1	0
51	2	1



61	1	1
61	2	1
76	2	1
71	2	1
69	1	1
56	2	1
67	1	1
54	2	1
63	1	1
47	2	1
62	2	1
65	2	1
63	2	1
64	1	1
65	2	1
51	1	1
56	1	1
70	2	1
58	2	1
67	2	1
62	1	1
74	1	1
69	2	0
64	2	0
75	2	1
47	2	0
57	2	0
56	1	1
68	1	0
55	1	1
62	2	1
73	2	1
68	2	1
75	1	1
63	1	1
61	1	1
62	1	1
44	1	1
56	2	1
54	1	1
57	1	0
56	1	1
69	1	1
72	1	1
59	2	1
70	1	1
64	2	1
61	2	1
72	2	1



63	2	1
74	2	1
71	1	1
71	2	0
72	2	1
77	2	1
72	1	1
55	2	1
65	2	1
67	2	1
69	1	1
55	2	1
51	2	1
64	1	1
63	1	1
69	1	1
64	1	1
59	1	1
73	2	1

3.1.3 Perhitungan Regresi Linier

Didalam perhitungan Regresi Linier, untuk mencari nilai : X_1Y : Usia X Hasil Test

X_2Y : Urin Berdarah X Hasil Test X_1X_2 : Usia X Urin Berdarah

X_{12} : Usia²

X_{22} : Urin Brdarah²

Tabel 4. Perhitungan Regresi Linier

Usia (X_1)	Urin Berdarah (X_2)	Hasil Tes (Y)	X_1 Y	X_2 Y	X_1 X_2	$(X_1)^2$	$(X_2)^2$
56	2	1	56	2	11	3136	4
58	2	1	58	2	11	3364	4
81	1	1	81	1	81	6561	1
64	2	1	64	2	12	4096	4
62	2	1	62	2	12	3844	4
72	2	1	72	2	14	5184	4
60	1	1	60	1	60	3600	1
61	2	1	61	2	12	3721	4
60	2	1	60	2	12	3600	4
49	1	1	49	1	49	2401	1
53	1	1	53	1	53	2809	1
58	1	1	58	1	58	3364	1
61	2	1	61	2	12	3721	4
68	1	1	68	1	68	4624	1
60	2	1	60	2	12	3600	4
72	1	1	72	1	72	5184	1
72	1	1	72	1	72	5184	1
57	2	1	57	2	11	3249	4



BULLETIN OF INFORMATION TECHNOLOGY (BIT)

Vol 4, No 1, Maret 2023, Hal 37 - 48

ISSN 2722-0524 (media online)

DOI [10.47065/bit.v3i1.475](https://doi.org/10.47065/bit.v3i1.475)

<https://journal.fkpt.org/index.php/BIT>

51	2	1	51	2	10	2601	4
					2		
54	2	1	54	2	10	2916	4
					8		
56	1	1	56	1	56	3136	1
77	2	1	77	2	15	5929	4
					4		
64	1	1	64	1	64	4096	1
57	2	1	57	2	11	3249	4
					4		
66	2	1	66	2	13	4356	4
					2		
70	2	1	70	2	14	4900	4
					0		
53	1	1	53	1	53	2809	1
51	2	1	51	2	10	2601	4
					2		
58	2	1	58	2	11	3364	4
					6		
58	2	1	58	2	11	3364	4
					6		
63	1	0	0	0	63	3969	1
51	2	1	51	2	10	2601	4
					2		
61	1	1	61	1	61	3721	1
61	2	1	61	2	12	3721	4
					2		
76	2	1	76	2	15	5776	4
					2		
71	2	1	71	2	14	5041	4
					2		
69	1	1	69	1	69	4761	1
56	2	1	56	2	11	3136	4
					2		
67	1	1	67	1	67	4489	1
54	2	1	54	2	10	2916	4
					8		
63	1	1	63	1	63	3969	1
47	2	1	47	2	94	2209	4
62	2	1	62	2	12	3844	4
					4		
65	2	1	65	2	13	4225	4
					0		
63	2	1	63	2	12	3969	4
					6		
64	1	1	64	1	64	4096	1
65	2	1	65	2	13	4225	4
					0		
51	1	1	51	1	51	2601	1
56	1	1	56	1	56	3136	1
70	2	1	70	2	14	4900	4
					0		
58	2	1	58	2	11	3364	4
					6		
67	2	1	67	2	13	4489	4
					4		
62	1	1	62	1	62	3844	1
74	1	1	74	1	74	5476	1
69	2	0	0	0	13	4761	4
					8		
64	2	0	0	0	12	4096	4
					8		



BULLETIN OF INFORMATION TECHNOLOGY (BIT)

Vol 4, No 1, Maret 2023, Hal 37 - 48

ISSN 2722-0524 (media online)

DOI [10.47065/bit.v3i1.475](https://doi.org/10.47065/bit.v3i1.475)

<https://journal.fkpt.org/index.php/BIT>

75	2	1	75	2	15 0	5625	4
47	2	0	0	0	94	2209	4
57	2	0	0	0	11 4	3249	4
56	1	1	56	1	56	3136	1
68	1	0	0	0	68	4624	1
55	1	1	55	1	55	3025	1
62	2	1	62	2	12 4	3844	4
73	2	1	73	2	14 6	5329	4
68	2	1	68	2	13 6	4624	4
75	1	1	75	1	75	5625	1
63	1	1	63	1	63	3969	1
61	1	1	61	1	61	3721	1
62	1	1	62	1	62	3844	1
44	1	1	44	1	44	1936	1
56	2	1	56	2	11 2	3136	4
54	1	1	54	1	54	2916	1
57	1	0	0	0	57	3249	1
56	1	1	56	1	56	3136	1
69	1	1	69	1	69	4761	1
72	1	1	72	1	72	5184	1
59	2	1	59	2	11 8	3481	4
70	1	1	70	1	70	4900	1
64	2	1	64	2	12 8	4096	4
61	2	1	61	2	12 2	3721	4
72	2	1	72	2	14 4	5184	4
63	2	1	63	2	12 6	3969	4
74	2	1	74	2	14 8	5476	4
71	1	1	71	1	71	5041	1
71	2	0	0	0	14 2	5041	4
72	2	1	72	2	14 4	5184	4
77	2	1	77	2	15 4	5929	4
72	1	1	72	1	72	5184	1
55	2	1	55	2	11 0	3025	4
65	2	1	65	2	13 0	4225	4
67	2	1	67	2	13 4	4489	4
69	1	1	69	1	69	4761	1
55	2	1	55	2	11 0	3025	4
51	2	1	51	2	10 2	2601	4
64	1	1	64	1	64	4096	1
63	1	1	63	1	63	3969	1



69	1	1	69	1	69	4761	1
64	1	1	64	1	64	4096	1
59	1	1	59	1	59	3481	1
73	2	1	73	2	14	5329	4
62	157	92	57	144	6	4001	271
78			82		98	04	
					47		

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

$$b_1 = \frac{100(5782) - (6278)(92)}{100(400104) - (6278)^2}$$

$$b_1 = \frac{578200 - 577576}{40010400 - 39413284}$$

$$b_1 = \frac{624}{597116}$$

$$b_1 = 0,0010450230775$$

Setelah mendapatkan nilai b1, selanjutnya mencari nilai b2

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

$$b_2 = \frac{100(144) - (157)(92)}{100(271) - (157)^2}$$

$$b_2 = \frac{14400 - 14444}{27100 - 24649}$$

$$b_2 = \frac{-44}{2451}$$

$$b_2 = -0,01795185638514$$

Nilai a ini adalah hasil perhitungan antara b1 dan b2 sehingga menghasilkan nilai Konstanta (a).

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

Tabel 5. Kolom data random prediksi penyakit ginjal

No	Usia (X₁)	Nyeri Dada (X₂)	Hasil Tes (Y)
1	73	2	?
2	59	1	?
3	64	1	?

Menghitung Persaman Regresi Linear $Y = a + b_1X_1 + b_2X_2$

$$y = 0,79257786571339 + (0,0010450230775x \ 73) + (-0,01795185638514x \ 2)$$

$$= 0,83296083760061$$

$$y = 0,79257786571339 + (0,0010450230775x \ 59) + (-0,01795185638514x \ 1)$$

$$= 0,83628237090075$$

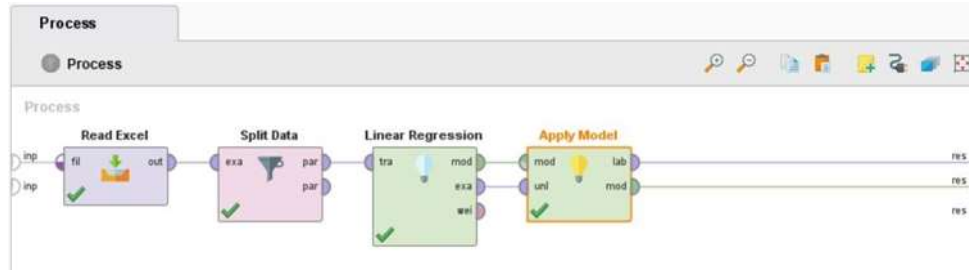
$$y = 0,79257786571339 + (0,0010450230775x \ 64) + (-0,01795185638514x \ 1)$$

$$= 0,8415074862882$$

3.2 Pembahasan

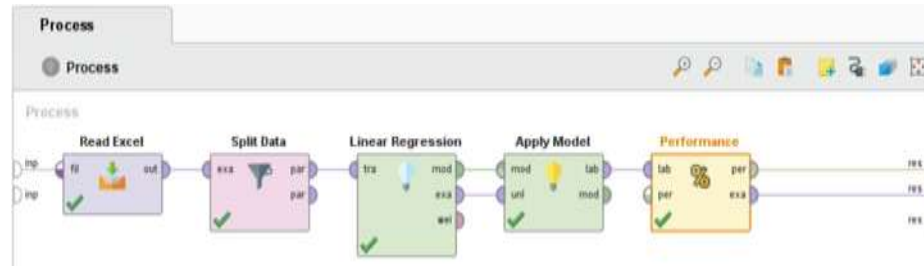
3.2.1 Proses Pengujian Data (RapidMiner)

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 ketemu, langkah selanjutnya mencari tingkat akurasi yang ada menggunakan Performance pada gambar 3.



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

Untuk mempermudah dalam pembacaan akurasi data penyakit ginjal, maka perlu di masukan tools performance untuk mencari RMSE (Root Mean Squared Error) dan Squared Error.

PerformanceVector

```
PerformanceVector:
root_mean_squared_error: 0.285 +/- 0.000
squared_error: 0.081 +/- 0.234
```

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

3.2.2 Analisa Hasil

Berdasarkan hasil pengujian yang telah dilakukan menggunakan tools RapidMiner bahwa atribut (usia, urin berdarah dan hasil tes) berpengaruh signifikan terhadap penelitian ini terbukti dengan menggunakan algoritma regresi linier mampu memberikan hasil yang sangat baik dan sangat akurat dengan nilai Root Mean Squared Error: 0.285 +/- 0.000 dan Squared Error: 0.081 +/- 0.234.

4. KESIMPULAN

Algoritma regresi linier efektif digunakan dalam memprediksi risiko penyakit ginjal berdasarkan analisis dari dataset yang ada. Kemampuan algoritma regresi linier sangat baik dalam memprediksi penderita penyakit ginjal dengan memanfaatkan data atribut (usia, urin berdarah dan hasil tes). Hasil prediksi dari algoritma regresi linier sangat baik dan akurat dibandingkan dengan metode lain dalam memprediksi risiko penyakit ginjal. Proses pengujian data pada penelitian ini menggunakan algoritma regresi linear mampu memberikan hasil yang baik dengan nilai Root Mean Squared Error: 0.285 +/- 0.000 dan Squared Error: 0.081 +/- 0.234.

UCAPAN TERIMA KASIH

Terima kasih kepada segenap jajaran pihak kampus, bapak, ibu beserta keluarga yang lainnya, kekasih yang senantiasa selalu mendampingi dan teman-teman yang terus mensupport saya dalam proses pengerjaan penelitian ini.

REFERENCES

- [1] K. Jurusan and K. Tangerang, "Hubungan Kepatuhan Menjalani Hemodialisis Dan Dukungan Keluarga Dengan Kualitas Hidup Pasien Gagal Ginjal Kronik Di Ruang Hemodialisa Rumah Sakit Umum Kabupaten Tangerang," 2018.
- [2] C. Zai and T. Komputer, "IMPLEMENTASI DATA MINING SEBAGAI PENGOLAHAN DATA."
- [3] A. Fahreza, "2022 25 Penerapan Data Mining dengan Metode Single Moving Average dalam Pengolahan Data Penerimaan Siswa Baru."



- [4] K. Puteri and A. Silvanie, "MACHINE LEARNING UNTUK MODEL PREDIKSI HARGA SEMBAKO DENGAN METODE REGRESI LINIER BERGANDA 1)," 2020. [Online]. Available: www.data.jakarta.go.id.
- [5] T. Akhir, "Analisa Algoritma Regresi Linear dan Decision Tree Dalam Prediksi Penjualan Produk (Studi Kasus: Lookma Boutique)." [Online]. Available: <https://lib.mercubuana.ac.id/>
- [6] F. H. Hamdanah and D. Fitrianah, "Analisis Performansi Algoritma Linear Regression dengan Generalized Linear Model untuk Prediksi Penjualan pada Usaha Mikra, Kecil, dan Menengah," *Jurnal Nasional Pendidikan Teknik Informatika (JANAPATI)*, vol. 10, no. 1, p. 23, Apr. 2021, doi: 10.23887/janapati.v10i1.31035. "204-Article Text-389-1-10-20130819".
- [7] "KLASIFIKASI PENYAKIT JANTUNG MENGGUNAKAN PARTICLE SWARM OPTIMIZATION DAN ALGORITMA K-NEAREST NEIGHBORS."
- [8] E. Widodo, "PELITA TEKNOLOGI Prediksi Penjurusan IPA, IPS dan BAHASA dengan Menggunakan Machine Learning Abstrak Informasi Artikel," *Jurnal Pelita Teknologi*, vol. 15, no. 1, pp. 37–48.
- [9] K. Erwansyah, "J-SISKO TECH Jurnal Teknologi Sistem Informasi dan Sistem Komputer TGD Implementasi Data Mining Untuk Menganalisa Hubungan Data Penjualan Produk Bahan Kimia Terhadap Persediaan Stok Barang Menggunakan Algoritma FP (Frequent Pattern) Growth Pada PT. Grand Multi Chemicals," n, vol. 30,no. 2, pp. 30–40, 2019.