Journal of Information System Research (JOSH)
Volume 6, No. 2, Januari 2025, pp 829−836
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v6i2.6277

Prediksi Penyakit Getah Bening dengan Algoritma Linear Regresi Berganda

Christofael Natalio Dondan*, Evangs Mailoa

Fakultas Teknologi Informasi, Program Studi Teknik Informatika, Universitas Kristen Satya Wacana, Salatiga
Jl. Diponegoro No.52-60, Salatiga, Kec. Sidorejo, Kota Salatiga, Jawa Tenga, Indonesia
Email : [1,*]672020171@student.uksw.edu, [2]evangs.mailoa@uksw.edu
Email Penulis Korespondensi: 672020171@student.uksw.edu
Submitted: **15/11/2024;** Accepted: **28/12/2024;** Published: **01/01/2025**

Abstrak–Penelitian ini bertujuan untuk mengevaluasi performa algoritma regresi linier berganda dalam memprediksi penyakit kelenjar getah bening dengan memanfaatkan dataset multivariat. Algoritma ini dipilih karena kemampuannya menganalisis hubungan antara variabel independen dan dependen secara kompleks, sehingga diharapkan dapat memberikan hasil prediksi yang akurat. Evaluasi model dilakukan menggunakan tiga metrik utama, yaitu Mean Absolute Error (MAE), Mean Squared Error (MSE), dan Root Mean Square Error (RMSE), untuk mengukur tingkat kesalahan prediksi dan keandalan model. Hasil penelitian menunjukkan bahwa algoritma regresi linier berganda mampu mencapai nilai MAE sebesar 0,3, MSE sebesar 0,3, dan RMSE sebesar 0,5. Nilai ini menunjukkan tingkat kesalahan prediksi yang rendah dan akurasi yang dapat diterima, sehingga algoritma ini memiliki potensi untuk diterapkan dalam membantu diagnosis penyakit kelenjar getah bening.

Kata Kunci: Linear Regression Berganda; Prediksi; Penyakit Kelenjar Getah Bening; Evaluasi Model

Abstract–This study aims to evaluate the performance of the multiple linear regression algorithm in predicting lymph node diseases by utilizing a multivariate dataset. This algorithm was chosen for its ability to analyze complex relationships between independent and dependent variables, which is expected to provide accurate prediction results. The model evaluation was conducted using three key metrics: Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Square Error (RMSE), to measure prediction error levels and model reliability. The study results indicate that the multiple linear regression algorithm achieved MAE of 0.3, MSE of 0.3, and RMSE of 0.5. These values demonstrate low prediction error and acceptable accuracy, suggesting the algorithm's potential for application in assisting the diagnosis of lymph node diseases.

Keywords: Multiple Linear Regression; Prediction; Lymph Node Disease; Model Evaluation

1. PENDAHULUAN

Kelenjar getah bening merupakan elemen kunci dalam sistem limfatik tubuh manusia, yang berperan penting dalam menjaga sistem imun atau kekebalan tubuh untuk dapat merespons infeksi dan penyakit [1]. Fungsi utama kelenjar getah bening adalah menyaring cairan limfatik serta menangkal berbagai patogen seperti bakteri, virus, dan bahkan sel-sel kanker yang berpotensi membahayakan tubuh [2][3]. Kelenjar getah bening bekerja dengan menahan dan menghancurkan mikroorganisme serta sel-sel berbahaya yang masuk, sehingga membantu tubuh dalam mempertahankan kesehatannya. Penyakit yang berkaitan dengan kelenjar getah bening dapat muncul dalam berbagai kondisi, tergantung pada penyebab serta tingkat keparahan yang dialami oleh pasien. Beberapa kondisi ini meliputi pembengkakan kelenjar akibat infeksi ringan, hingga kondisi yang lebih serius seperti limfoma atau kanker getah bening, yang memerlukan penanganan medis lebih lanjut..

Penelitian ini bertujuan untuk mengevaluasi sejauh mana algoritma regresi linier berganda dapat memprediksi penyakit pada kelenjar getah bening dengan tingkat akurasi yang optimal. Untuk mencapai tujuan ini, penelitian memanfaatkan dataset yang mencakup lebih dari dua variabel, memungkinkan penerapan analisis multivariat yang lebih kompleks. Dengan adanya berbagai variabel dalam dataset, proses prediksi dapat memperhitungkan faktor-faktor tambahan yang mempengaruhi hasil, sehingga algoritma regresi linier berganda diharapkan mampu memberikan gambaran yang lebih akurat dan relevan terhadap kondisi riil. Selain itu, penelitian ini juga akan menguji berbagai metrik evaluasi, seperti MAE, MSE, dan RMSE, untuk memberikan penilaian komprehensif mengenai kinerja algoritma dalam menangani data yang memiliki banyak variabel.

Alasan penelitian ini dilakukan adalah untuk membantu mendiagnosis penyakit kelenjar getah bening, yang mana masyarakat umum tidak tahu cara untuk mengidentifikasi tanda dan gejala yang dialami, apakah merupakan penyakit kelenjar getah bening atau bukan [4]. Rendahnya kesadaran dan pemahaman masyarakat mengenai gejala penyakit ini sering kali membuat mereka terlambat dalam mengambil tindakan pencegahan atau pengobatan yang tepat. Biasanya, masyarakat harus berkonsultasi ke dokter untuk memastikan kondisi mereka, yang dapat memakan waktu dan biaya. Dengan adanya penelitian ini, diharapkan tercipta suatu sistem prediksi berbasis machine learning yang mampu membantu mendiagnosis penyakit kelenjar getah bening secara lebih cepat dan efektif. Sistem ini tidak hanya bertujuan untuk menghemat waktu dalam proses diagnosis, tetapi juga untuk memberikan manfaat dalam memberikan pengetahuan dasar kepada masyarakat mengenai gejala yang berpotensi berhubungan dengan penyakit ini.

Pemilihan algoritma regresi linier berganda dalam penelitian ini didasarkan pada dua alasan utama. Pertama, regresi linier berganda mampu memproses lebih dari satu variabel independen atau variabel x, sehingga memungkinkan analisis yang lebih kompleks dengan mempertimbangkan berbagai faktor yang memengaruhi variabel dependen [5][6]. Kedua, koefisien dalam regresi linier berganda dapat memberikan wawasan tentang

Copyright © 2025 The Author, Page 829
This Journal is licensed under a Creative Commons Attribution 4.0 International License

Journal of Information System Research (JOSH)
Volume 6, No. 2, Januari 2025, pp 829−836
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v6i2.6277

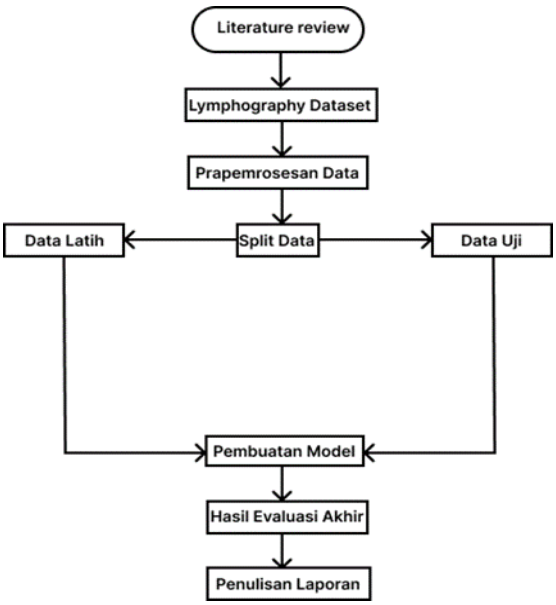
hubungan antara setiap variabel independen dengan variabel dependen, sehingga memudahkan pemahaman mengenai sejauh mana pengaruh masing-masing variabel independen terhadap hasil akhir yang diprediksi [7]. Dengan demikian, alasan-alasan ini mendukung pemilihan regresi linier berganda sebagai algoritma yang tepat untuk penelitian ini, karena selain mampu menangani data multivariat, algoritma ini juga memberikan interpretasi yang jelas mengenai kontribusi masing-masing variabel independen terhadap variabel dependen.

Beberapa penelitian terdahulu yang menggunakan algoritma regresi linier berganda untuk memprediksi penyakit menunjukkan hasil evaluasi akhir yang bervariasi. Salah satu penelitian menghasilkan nilai Root Mean Square Error (RMSE) sebesar 0,403 [8], sedangkan penelitian lainnya mencatat nilai RMSE sebesar 0,686 ± 0,000 dan Mean Squared Error (MSE) sebesar 0,471 ± 0,546 [9]. Sementara itu, penelitian lain melaporkan nilai RMSE sebesar 0,379 ± 0,000 dan MSE sebesar 0,144 ± 0,229 [10]. Perbedaan dalam hasil evaluasi ini kemungkinan dipengaruhi oleh berbagai faktor, seperti perbedaan dalam dataset yang digunakan, jumlah variabel prediktor, serta variasi dalam metode pengolahan data atau teknik prapemrosesan yang diterapkan. Analisis dari hasil penelitian-penelitian ini memberikan wawasan yang berharga dalam menilai kinerja algoritma regresi linier berganda dan menjadi acuan dalam memperkirakan tingkat akurasi yang mungkin dicapai dalam penelitian ini. Berdasarkan ketiga penelitian terdahulu, terdapat kesamaan dalam penggunaan metode prediksi penyakit dengan algoritma regresi, di mana evaluasi akhir mengandalkan metrik seperti Root Mean Square Error (RMSE) dan Mean Squared Error (MSE) untuk menilai kinerja model. Dalam penelitian ini, metode yang digunakan memiliki kesamaan dengan pendekatan terdahulu, namun terdapat perbedaan yang signifikan, yaitu penambahan Mean Absolute Error (MAE) sebagai metrik evaluasi tambahan. MAE ini ditambahkan untuk memberikan gambaran yang lebih akurat mengenai keandalan algoritma regresi linier berganda dalam memprediksi hasil. Perbedaan lainnya adalah seluruh perhitungan nilai prediksi dan evaluasi dilakukan secara manual menggunakan rumus regresi linier berganda, tanpa melibatkan perangkat lunak pengolahan data seperti SPSS atau Python. Pendekatan manual ini diambil untuk memberikan pemahaman yang lebih mendalam terhadap proses matematis yang terlibat, serta untuk menguji sejauh mana keakuratan metode yang dipilih tanpa ketergantungan pada alat bantu eksternal.

Penelitian ini mempunyai rumusan masalah sebagai berikut: Bagaimana ketepatan algoritma regresi linier berganda dalam memprediksi penyakit pada kelenjar getah bening? Bagaimana hasil evaluasi akhir dengan metrik MAE, MSE, dan RMSE yang diperoleh melalui pendekatan manual ini, dan sejauh mana metrik tersebut dapat menunjukkan kinerja dan keandalan algoritma dalam kondisi aktual? Pertanyaan ini penting untuk menguji efektivitas algoritma dalam aplikasi nyata, serta untuk menentukan tingkat akurasi dari setiap metrik yang digunakan.

2. METODOLOGI PENELITIAN

Metodologi penelitian merupakan gambaran tahapan proses yang dilalui pada penelitian. Adapaun metodologi penelitian dapat dilihat berikut:



Gambar 1. Alur Penelitian

2.1 Dataset

Data yang digunakan diambil dari UCI Machine Learning Repository dengan alamat web: https://archive.ics.uci.edu/dataset/63/lymphography. Dataset ini mencatat diagnosis pasien dengan keluhan

Copyright © 2025 The Author, Page 830
This Journal is licensed under a Creative Commons Attribution 4.0 International License

Journal of Information System Research (JOSH)
Volume 6, No. 2, Januari 2025, pp 829−836
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v6i2.6277

penyakit kelenjar getah bening. Dataset tersebut memiliki 19 variabel pendukung dengan ukuran data sebesar 3 KB pada saat diunduh.

2.2 Pra Pemrosesan data

Pada langkah selanjutnya, data yang digunakan dalam penelitian ini diproses melalui dua langkah utama: pembersihan data (cleaning data) dan pengubahan format data. Pembersihan data adalah proses yang sangat penting untuk memperbaiki atau menghapus data yang salah, duplikat, atau rusak[11][12].

2.3 Split Data

Penentuan proporsi antara data latih dan data uji merupakan langkah yang sangat penting dalam proses pengembangan algoritma. Dilakukannya split data dengan membagi dataset menjadi dua bagian bertujuan untuk memastikan bahwa model yang dikembangkan dapat belajar dari informasi data latih (training data) dan kemudian dievaluasi kinerjanya pada data uji (testing data), yang belum pernah dilihat oleh model selama pelatihan[13]. Biasanya, proporsi yang sering digunakan adalah 70:30 atau 80:20, di mana 70-80% data digunakan untuk melatih model, dan sisanya untuk pengujian.

2.4 Perhitungan Linear Regresi Berganda

Regresi linear adalah metode statistik yang digunakan untuk memodelkan hubungan antara satu variabel dependen (variabel yang ingin diprediksi) dan satu atau lebih variabel independen (variabel predictor)[14][15]. Dalam regresi linear berganda, model ini memprediksi variabel dependen berdasarkan beberapa variabel independen, memungkinkan analisis yang lebih mendalam.

$$\hat{y} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_{16} X_{16} \text{ [16]} \quad (1)$$

Dimana $\hat{y}$ adalah nilai prediksi, β_0 intercept, $\beta_1, \beta_2, \ldots, \beta_{16}$ adalah koefisien regresi untuk masing-masing variabel independen, dan $X_1, X_2, \ldots, X_{16}$ adalah nilai rata-rata dari variabel independen

2.5 Evaluasi Akhir

Dalam evaluasi akhir model, penting untuk mempertimbangkan beberapa metrik evaluasi untuk melihat tingkat akurasi algoritma[17], di antaranya adalah Root Mean Square Error (RMSE), Mean Absolute Error (MAE), dan Mean Squared Error (MSE). Ketiga metrik ini memiliki fungsi dan karakteristik unik yang membantu dalam memberikan gambaran menyeluruh tentang sejauh mana model prediksi mendekati nilai aktual.

a. MAE (Mean Absolute Error)
   MAE adalah salah satu metode yang digunakan untuk mengukur tingkat keakuratan suatu model prediksi atau peramalan[18]. MAE mengukur rata-rata kesalahan prediksi model tanpa memperhatikan arah kesalahan.

   $$\text{MAE} = \frac{\text{i}}{\text{n}} \sum |\text{error}_\text{i}| \quad (2)$$

   Dimana n adalah jumlah data yang digunakan, Error merupakan nilai selisi atau kesalahan, dan $|.|$ adalah fungsi absolut, yang menghasilkan nilai positif dari setiap perbedaan.
b. MSE (Mean Squared Error)
   MSE adalah sebuah metrik penting dalam evaluasi model regresi yang merupakan perbandingan nilai prediksi dan nilai sebenarnya[19].

   $$\text{MSE} = \frac{\text{i}}{\text{n}} \sum_{\text{i}=1}^{\text{n}} (\text{error}_\text{i})^2 \quad (3)$$

   Dimana n adalah jumlah data yang digunakan, $(\text{error}_\text{i})^2$ merupakan nilai error yang dipangkatkan 2.
c. Nilai RMSE (Root Mean Square Error)
   RMSE merupakan faktor performa regresi yang menghitung kesalahan kuadrat atau selisih antara nilai sebenarnya dengan nilai prediksi[20].

   $$\text{RMSE} = \sqrt{\text{MSE}} \quad (4)$$

   Dimana $\sqrt{\text{MSE}}$ dimana nilai MSE akan dikuadratkan.

3. HASIL DAN PEMBAHASAN

3.1 Dataset

Berdasarkan pencarian data yang dilakukan dalam penelitian ini, diperoleh dataset Lymphography dimana dataset tersebut diterbitkan oleh UCI Machine Learning Repository. Dataset yang terdiri dari 19 variabel pendukung dengan lebih dari 100 data rekaman diagnosis pasien penyakit kelenjar getah bening. Tapi dalam penelitian ini hanya mengambil sebanyak 20 data yang akan digunakan sebagai proses prediksi. Pada kolom class menggunakan

Copyright © 2025 The Author, Page 831
This Journal is licensed under a Creative Commons Attribution 4.0 International License

Journal of Information System Research (JOSH)
Volume 6, No. 2, Januari 2025, pp 829−836
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v6i2.6277

kata sebagai penanda yaitu yes yang berarti menderita penyakit kelenjar getah bening dan no tidak menderita penyakit getah bening.

Tabel 1. Dataset Mentah

No	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	x_{16}	x_{17}	x_{18}	y
1	4	2	1	1	1	1	1	2	1	2	2	2	4	8	1	1	2	2	1
2	3	2	1	1	2	2	1	2	1	3	3	2	3	4	2	2	2	2	1
3	3	2	2	2	2	2	2	2	1	4	3	3	4	8	3	2	2	7	1
4	3	1	1	1	1	2	1	2	1	3	3	4	4	4	3	1	2	6	1
5	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0
6	2	1	1	1	1	1	1	2	1	3	3	3	3	6	3	1	2	4	1
7	1	1	1	1	1	2	1	2	1	2	2	1	1	2	1	1	1	2	0
8	3	2	1	1	1	2	1	2	1	2	2	2	2	1	3	1	1	1	1
9	3	1	1	1	2	2	2	1	3	1	1	4	2	5	3	1	2	4	1
10	3	2	2	2	2	2	2	1	2	2	2	4	2	4	3	2	2	7	1
11	2	2	2	1	2	2	1	2	1	3	2	4	3	5	1	2	2	3	1
12	4	2	1	1	1	2	1	1	1	3	3	4	3	8	3	2	2	2	1
13	1	3	1	1	1	1	1	1	2	1	2	2	2	3	2	1	2	2	1
14	1	2	2	1	1	1	1	1	2	1	3	3	3	3	2	2	2	1	1
15	3	3	1	1	1	1	1	1	2	1	2	2	4	2	4	3	2	2	1
16	4	4	1	1	1	2	2	1	2	1	3	3	3	4	5	3	2	2	1
17	2	4	2	1	1	2	2	1	2	1	4	3	4	3	4	2	2	2	1
18	1	3	2	1	1	1	1	1	2	1	2	2	2	2	5	1	1	1	1
19	1	1	1	1	1	2	1	2	1	2	2	1	1	2	1	1	1	2	0
20	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0

3.2 Pra Pemrosesan Data

Salah satu hal yang harus ditentukan sebelum melakukan prediksi adalah menentukan variable independen dan variabel dependen. Dalam penelitian ini digunakan sebanyak 18 variabel independen dan 1 variabel dependen sebagai kelas atau target yang hanya digunakan sebagai tanda apakah menderita penyakit kelenjar getah bening.Sebagai contoh untuk kolom variabel pertama adalah lymphatics yang memiliki range nilai 1 sampai 4 dan kolom kedua yaitu block of affere yang memiliki range nilai 1 sampai 4. Untuk mempremudah proses matematis dan mengantisipasi kesalahan atau error karena nilai variabel dependen yang tidak menggunakan angka saat menentukan nilai prediksi akan dilakukan perubahan pada varibel dependen dengan mengubah nilainya dimana yes dan no menjadi 1 dan 0 yang berarti 1 menderita penyakit kelenjar getah bening dan 0 tidak menderita penyakit getah bening. Hasil dataset yang telah melaluli pra pemrosesan data disajikan dalam tabel 2.

Tabel 2. Dataset yang sudah siap

No	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	x_{16}	x_{17}	x_{18}	y
1	4	2	1	1	1	1	1	2	1	2	2	2	4	8	1	1	2	2	1
2	3	2	1	1	2	2	1	2	1	3	3	2	3	4	2	2	2	2	1
3	3	2	2	2	2	2	2	2	1	4	3	3	4	8	3	2	2	7	1
4	3	1	1	1	1	2	1	2	1	3	3	4	4	4	3	1	2	6	1
5	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0
6	2	1	1	1	1	1	1	2	1	3	3	3	3	6	3	1	2	4	1
7	1	1	1	1	1	2	1	2	1	2	2	1	1	2	1	1	1	2	0
8	3	2	1	1	1	2	1	2	1	2	2	2	2	1	3	1	1	1	1
9	3	1	1	1	2	2	2	1	3	1	1	4	2	5	3	1	2	4	1
10	3	2	2	2	2	2	2	1	2	2	2	4	2	4	3	2	2	7	1
11	2	2	2	1	2	2	1	2	1	3	2	4	3	5	1	2	2	3	1
12	4	2	1	1	1	2	1	1	1	3	3	4	3	8	3	2	2	2	1
13	1	3	1	1	1	1	1	1	2	1	2	2	2	3	2	1	2	2	1
14	1	2	2	1	1	1	1	1	2	1	3	3	3	3	2	2	2	1	1
15	3	3	1	1	1	1	1	1	2	1	2	2	4	2	4	3	2	2	1
16	4	4	1	1	1	2	2	1	2	1	3	3	3	4	5	3	2	2	1
17	2	4	2	1	1	2	2	1	2	1	4	3	4	3	4	2	2	2	1
18	1	3	2	1	1	1	1	1	2	1	2	2	2	2	5	1	1	1	1
19	1	1	1	1	1	2	1	2	1	2	2	1	1	2	1	1	1	2	0
20	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0

Copyright © 2025 The Author, Page 832
This Journal is licensed under a Creative Commons Attribution 4.0 International License

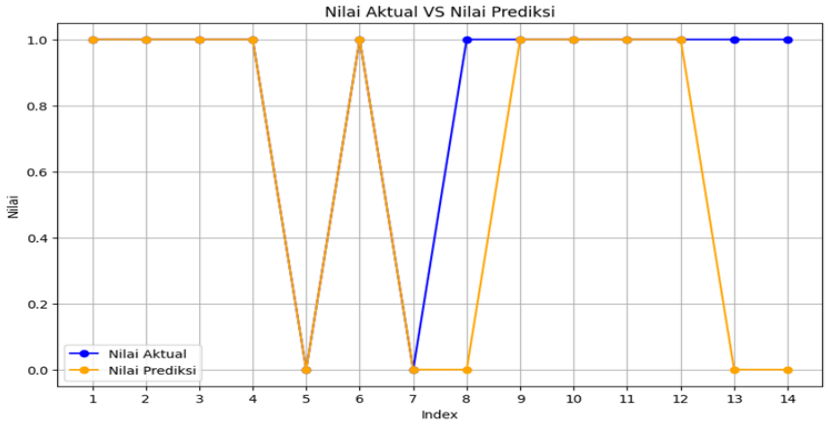
Journal of Information System Research (JOSH)
Volume 6, No. 2, Januari 2025, pp 829−836
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v6i2.6277

3.3 Split Data

Tahap selanjutnya adalah melakukan pembagian data atau split data yang mana perbandingan yang dipakai dalam penelitian ini adalah 70% data latih dan 30% data uji. Pemilihan dilakukan dengan mengambil nomor 1 sampai 14 yang merupakan data latih dan nomor 15 sampai 20 merupakan data uji. Berarti data latih yang digunakan sebanyak 14 data dan data uji yang digunakan sebanyak 6 data.

3.4 Pembuatan Model Pelatihan Linear Regresi Berganda

Pembuatan model pelatihan akan didasari oleh beberapa hal dalam penggunaan algoritmanya seperti menentukan nilai slope atau kemiringan dari suatu garis pada variabel independen. Perhitungan dilakukan dengan mencari nilai rata-rata terlebih dahulu dan dilanjutkan dengan melakukan rumus matematis untuk menentukan nilai slope.

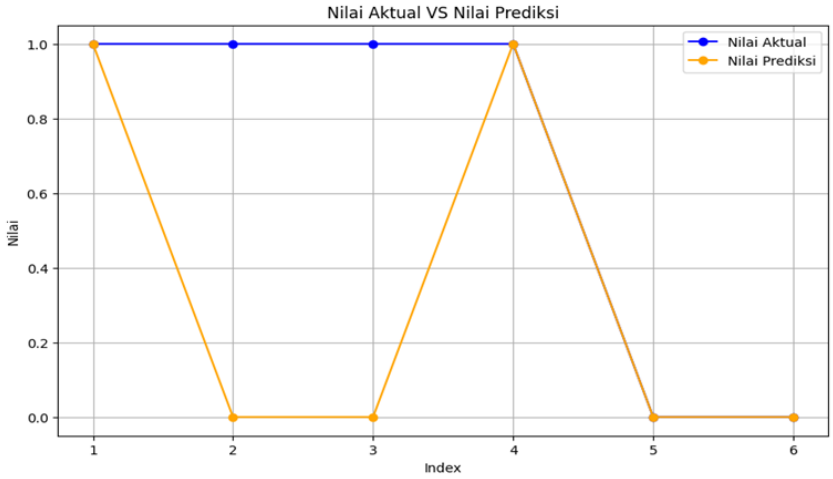


Gambar 2. Grafik Nilai Aktual vs Nilai Prediksi Data Latih

Gambar 2 menunjukkan perbandingan nilai aktual dan prediksi pada data latih. Sumbu X menggambarkan nomor pengamatan (1-14), sedangkan sumbu Y menunjukkan nilai dari 0 hingga 1. Garis biru mewakili nilai aktual yang stabil di sebagian besar pengamatan dengan nilai 1, kecuali pada pengamatan ke-5, 6, dan 13 yang turun ke 0. Sebaliknya, garis oranye untuk nilai prediksi model menunjukkan variasi yang lebih besar. Nilai prediksi stabil di sekitar 1, namun mengalami penurunan drastis pada pengamatan ke-4 hingga ke-6 sebelum kembali naik mendekati 1. Perbedaan antara garis aktual dan prediksi terlihat jelas, terutama pada pengamatan dengan fluktuasi signifikan. Ini menunjukkan bahwa model belum mampu memprediksi secara akurat, karena hasilnya sangat berbeda dari nilai aktual yang stabil. Fluktuasi ini menunjukkan bahwa model perlu ditingkatkan untuk menghasilkan prediksi yang lebih akurat. Evaluasi lebih lanjut dengan metrik seperti MAE, MSE, dan RMSE diperlukan untuk mengukur kesalahan prediksi dan menilai performa model.

3.5 Uji Algoritma Linear Regresi Berganda

Gambar 3 menggambarkan perbandingan antara nilai aktual dan nilai prediksi pada data uji.



Gambar 3. Grafik Nilai Aktual vs Nilai Prediksi Data Uji

Pada sumbu horizontal (X) terdapat enam sampel pengamatan yang diberi label dari 1 hingga 6, sementara pada sumbu vertikal (Y), nilai ditampilkan dalam rentang antara 0 hingga 1 untuk nilai aktual dan prediksi. Garis biru merepresentasikan nilai aktual, yang terlihat konsisten dan stabil pada nilai 1 di sebagian besar sampel, kecuali

Copyright © 2025 The Author, Page 833
This Journal is licensed under a Creative Commons Attribution 4.0 International License

Journal of Information System Research (JOSH)
Volume 6, No. 2, Januari 2025, pp 829−836
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v6i2.6277

pada pengamatan kedua dan ketiga, di mana nilai aktual tetap 0. Di sisi lain, garis oranye menunjukkan nilai prediksi yang tampak sangat bervariasi. Prediksi dimulai dari nilai 1 pada sampel pertama, kemudian turun drastis hingga mencapai 0 pada sampel kedua dan ketiga, sebelum naik lagi ke nilai 1 pada sampel keempat, dan kembali turun ke 0 pada sampel kelima dan keenam. Perbedaan antara garis nilai aktual dan nilai prediksi menunjukkan bahwa model memiliki beberapa ketidakcocokan dalam memprediksi data, terutama pada pengamatan kedua, ketiga, kelima, dan keenam, di mana terjadi perbedaan yang signifikan antara nilai aktual dan prediksi. Meskipun terdapat kesalahan dalam beberapa prediksi, pada sebagian besar sampel, nilai prediksi mendekati nilai aktual.

3.6 Hasil Evaluasi Akhir

Hasil evaluasi akhir dapat dilakukan jika proses pelatihan model dan uji model sudah dilakukan dengan mengambil nilai selisih pada nilai aktual dan nilai prediksi di proses uji model. Yang dimana jika keakuratan prediksi yang dihasilkan tinggi maka evaluasi akhir yang akan didapatkan akan rendah.

Tabel 3. Hasil Selisih Data Uji

Data Uji	Selisih (error)
1	0
2	1
3	1
4	0
5	0
6	0

Diliat dari hasil selisih yang baik dimana hanya ada 2 data yang mengalami nilai selisi yang cukup jauh dan yang lainnya berada di nilai 0 yang berarti hasil prediksi dan nilai aktualnya sama.

Dari hasil selisih data uji yang telah didapatkan tahap selanjutnya adalah mencari nilai MAE, MSE dan RMSE, dengan implementasi rumus berikut ini. Penentuan MAE (Mean Absolute Error) dilakukan dengan menghitung selisih antara nilai prediksi dan nilai aktual, kemudian mengambil nilai absolut dari setiap selisih untuk memastikan semua error bernilai positif. Total nilai absolut error dalam penelitian ini adalah 2. Langkah selanjutnya adalah membaginya dengan jumlah data uji (6), sehingga diperoleh MAE = 0.3. Nilai MAE yang lebih rendah menunjukkan rata-rata kesalahan prediksi yang lebih kecil, sehingga model dianggap lebih akurat.

$$MAE = |0| + |1| + |1| + |0| + |0| + |0|$$
$$MAE = 0 + 1 + 1 + 0 + 0 + 0$$
$$MAE = \frac{2}{6} = 0.3$$

MAE = 0.3

Penentuan MSE (Mean Squared Error) dimulai dengan menghitung selisih antara nilai prediksi dan aktual, lalu mengkuadratkan setiap selisih untuk menghindari nilai negatif. Total nilai kuadrat selisih dalam penelitian ini adalah 2. Langkah berikutnya adalah membagi total tersebut dengan jumlah data uji (6), sehingga diperoleh MSE = 0.3. Nilai ini menunjukkan rata-rata kuadrat kesalahan prediksi, dengan nilai MSE yang lebih rendah menandakan model prediksi yang baik.

$$MSE = 0^2 + 1^2 + 1^2 + 0^2 + 0^2 + 0^2$$
$$MSE = 0 + 1 + 1 + 0 + 0 + 0$$
$$MSE = \frac{2}{6} = 0.3$$

MSE = 0.3

RMSE (Root Mean Square Error) adalah ukuran untuk menilai perbedaan antara nilai prediksi dan nilai sebenarnya. Rumusnya adalah akar kuadrat dari MSE (Mean Squared Error), yang merupakan rata-rata dari kuadrat selisih antara nilai prediksi dan nilai aktual. RMSE memberikan deviasi standar prediksi dalam satuan yang sama dengan data, sehingga mudah diinterpretasikan. Dalam kasus ini, MSE = 0.3, sehingga RMSE = √MSE = 0.5. Nilai RMSE 0.5 menunjukkan rata-rata deviasi prediksi model dari nilai aktual.

$$RMSE = \sqrt{0.3}$$
$$RMSE = 0.5$$

Tabel 4. Regresi Linier Berganda

Linier Regresi Berganda	MAE	MSE	RMSE
Data Uji	0.3	0.3	0.5

Dalam penelitian ini, model regresi linier berganda dievaluasi melalui beberapa metrik kesalahan, yaitu Mean Absolute Error (MAE), Mean Squared Error (MSE), dan Root Mean Squared Error (RMSE). Hasil evaluasi menunjukkan bahwa pada data uji, nilai MAE dan MSE masing-masing adalah 0,3, sementara nilai RMSE sebesar 0,5. Nilai-nilai ini menunjukkan bahwa model yang digunakan memiliki performa yang cukup baik dengan tingkat kesalahan yang rendah dan berada dalam batas yang dapat diterima untuk memprediksi data baru.

This Journal is licensed under a Creative Commons Attribution 4.0 International License

Journal of Information System Research (JOSH)
Volume 6, No. 2, Januari 2025, pp 829−836
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v6i2.6277

Tabel 5. Hasil Evaluasi Akhir Penelitian Terdahulu

Linier Regresi Berganda	MAE	MSE	RMSE
Penelitian Terdahulu 1	-	0.47	0.68
Penelitian Terdahulu 2	-	0.14	0.37

Jika dibandingkan dengan penelitian sebelumnya, terdapat perbedaan hasil yang cukup mencolok. Penelitian terdahulu pertama memiliki nilai MSE sebesar 0,47 dan RMSE sebesar 0,68, menunjukkan tingkat kesalahan yang sedikit lebih tinggi dibandingkan dengan penelitian sekarang. Sementara itu, penelitian terdahulu kedua menunjukkan performa yang lebih baik dengan nilai MSE sebesar 0,14 dan RMSE sebesar 0,37, menandakan tingkat kesalahan yang lebih rendah dan akurasi yang lebih tinggi. Perbedaan hasil ini bisa disebabkan oleh beberapa faktor, termasuk jenis dataset yang digunakan serta teknik pengolahan data yang diterapkan. Selain itu, ketepatan algoritma yang digunakan dapat berpengaruh pada performa model, seperti yang terlihat pada penelitian terdahulu kedua yang berhasil menghasilkan kesalahan sangat rendah dan akurasi yang lebih baik dalam prediksi.

4. KESIMPULAN

Dari hasil penelitian ini, dapat disimpulkan bahwa algoritma regresi linier berganda menunjukkan kemampuan yang baik dalam memprediksi penyakit kelenjar getah bening dengan memanfaatkan dataset yang tersedia. Algoritma ini mampu mengolah variabel-variabel independen yang relevan dan memberikan hasil prediksi yang cukup akurat. Hal ini dibuktikan melalui nilai-nilai evaluasi seperti Mean Absolute Error (MAE), Mean Squared Error (MSE), dan Root Mean Square Error (RMSE) yang dihasilkan dalam penelitian. Nilai evaluasi akhir menunjukkan bahwa MAE sebesar 0,3, MSE sebesar 0,3, dan RMSE sebesar 0,5. Nilai-nilai ini menunjukkan performa algoritma yang cukup memadai, mengingat angka-angka tersebut jauh dari nilai 1, yang menandakan tingkat error yang rendah. Nilai MAE, MSE, dan RMSE yang berada dalam rentang yang wajar menandakan bahwa model prediksi ini memiliki tingkat akurasi yang dapat diterima. Ini menunjukkan bahwa regresi linier berganda efektif dalam memproses data yang mempunyai variabel independent lebih dari 1 untuk kasus penyakit kelenjar getah bening. Selain itu, hasil ini memperkuat keyakinan bahwa algoritma regresi linier berganda dapat digunakan sebagai alat bantu dalam mendukung diagnosis medis, terutama untuk memprediksi risiko atau keparahan kondisi berdasarkan faktor-faktor yang relevan. Algoritma ini memiliki potensi besar untuk diterapkan pada data yang lebih luas dalam rangka mendukung analisis prediktif di bidang kesehatan, karena menunjukkan akurasi yang cukup baik. Penerapan algoritma ini pada skala data yang lebih besar diharapkan dapat membantu dalam mengidentifikasi pola-pola kesehatan, memprediksi risiko penyakit, serta memberikan rekomendasi pencegahan yang lebih tepat, sehingga mampu mendukung pengambilan keputusan yang lebih efektif di sektor kesehatan.

REFERENCES

[1] S. L. Fajri and G. W. Nurcahyo, “Sistem Pakar dalam Menganalisis Penyakit Kelenjar Getah Bening Menggunakan Metode Certainty Factor,” J. Sistim Inf. dan Teknol., vol. 3, pp. 269–274, 2021, doi: 10.37034/jsisfotek.v3i4.77.
[2] J. Xu, Y. Zhang, and D. Miao, “Three-way confusion matrix for classification: A measure driven view,” Inf. Sci. (Ny)., vol. 507, pp. 772–794, 2020, doi: 10.1016/j.ins.2019.06.064.
[3] A. Ahmad, A. Ullah, K. N. Khan, and M. S. Khan, “Automated Assessment of Lymphocytes Using Machine Learning Techniques,” 2021 Int. Conf. Artif. Intell. ICAI 2021, pp. 108–112, 2021, doi: 10.1109/ICAI52203.2021.9445242.
[4] U. F. S. S. P. Cindy Saputri, Milfa Yetri, “Sistem Pakar Mendiagnosis Penyakit Kelenjar Getah Bening Dengan Menggunakan Metode Teorema Bayes,” J. Kesehat., vol. 1(2), no. 2, pp. 1–12, 2019.
[5] A. Sari and D. U. Putri, “Penerapan Regresi Linear Berganda Untuk Memprediksi Diabetes Secara Dini,” J. Tek., vol. 3, no. 1, p. 32, 2023, doi: 10.54314/teknisi.v3i1.1255.
[6] F. O. Lusiana, I. Fatma, and A. P. Windarto, “Estimasi Laju Pertumbuhan Penduduk Menggunakan Metode Regresi Linier Berganda Pada BPS Simalungun,” J. Informatics Manag. Inf. Technol., vol. 1, no. 2, pp. 79–84, 2021, doi: 10.47065/jimat.v1i2.104.
[7] M. D. Irrawati and M. Mukaramah, “Implementasi Metode Regresi Linear Berganda untuk Mengatasi Pelanggaran Asumsi Klasik,” Stud. Akuntansi, Keuangan, dan Manaj., vol. 3, no. 2, pp. 83–94, 2024, doi: 10.35912/sakman.v3i2.2743.
[8] Normah, B. Rifai, S. Vambudi, and R. Maulana, “Implementasi Data Mining Dengan Algoritma Multiple Linear Regression Untuk Memprediksi Penyakit Diabetes,” J. Tek. Komput. AMIK BSI, vol. 8, no. 2, pp. 174–180, 2022, doi: 10.31294/jtk.v4i2.
[9] M. A. Rahman Wahid, A. Nugroho, and A. Halim Anshor, “Prediksi Penyakit Kanker Paru-Paru Dengan Algoritma Regresi Linier,” Bull. Inf. Technol., vol. 4, no. 1, pp. 63–74, 2023, doi: 10.47065/bit.v4i1.501.
[10] A. Wijayadhi, M. Makmun Effendi, and S. Budi Rahardjo, “Prediksi Penyakit Jantung Dengan Algoritma Regresi Linier,” Bull. Inf. Technol., vol. 4, no. 1, pp. 15–28, 2023, doi: 10.47065/bit.v4i1.463.
[11] R. G. Wardhana, G. Wang, and F. Sibuea, “Penerapan Machine Learning Dalam Prediksi Tingkat Kasus Penyakit Di Indonesia,” J. Inf. Syst. Manag., vol. 5, no. 1, pp. 40–45, 2023, doi: 10.24076/joism.2023v5i1.1136.
[12] A. M. Majid and I. Nawangsih, “Perbandingan Metode Ensemble Untuk Meningkatkan Akurasi Algoritm Machine

This Journal is licensed under a Creative Commons Attribution 4.0 International License

Journal of Information System Research (JOSH)
Volume 6, No. 2, Januari 2025, pp 829−836
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v6i2.6277

Learning Dalam Memprediksi Penyakit Breast Cancer (Kanker Payudara),” J. SAINTIKOM (Jurnal Sains Manaj. Inform. dan Komputer), vol. 23, no. 1, p. 97, 2024, doi: 10.53513/jis.v23i1.9563.

[13] F. Akbar and Rahmaddeni, “Jurnal Politeknik Caltex Riau Komparasi Algoritma Machine Learning untuk Memprediksi Penyakit Alzheimer,” J. Komput. Terap., vol. 8, no. 2, pp. 236–245, 2022, [Online]. Available: https://jurnal.pcr.ac.id/index.php/jkt/

[14] F. H. Hamdanah and D. Fitrianah, “Analisis Performansi Algoritma Linear Regression dengan Generalized Linear Model untuk Prediksi Penjualan pada Usaha Mikra, Kecil, dan Menengah,” J. Nas. Pendidik. Tek. Inform., vol. 10, no. 1, p. 23, 2021, doi: 10.23887/janapati.v10i1.31035.

[15] A. Eka, A. Juarna, T. Informatika, F. T. Industri, and U. Gunadarma, “Prediksi Pro duksi Daging Sapi Nasional dengan Meto de Regresi Linier dan Regresi Polinomial,” J. Ilm. Komputasi, vol. 20, no. 2, pp. 209–215, 2021, doi: 10.32409/jikstik.20.2.2722.

[16] S. Rahmatullah, E. H. Juningsih, and S. Rachmawati, “Prediksi nilai akademik peserta didik di masa pandemi covid-19 dengan regresi linier berganda,” JISAMAR (Journal Inf. Syst. Applied, Manag. Account. Res., vol. 7, no. 1, pp. 112–123, 2023, doi: 10.52362/jisamar.v7i1.1012.

[17] N. R. Muntiari and K. H. Hanif, “Klasifikasi Penyakit Kanker Payudara Menggunakan Perbandingan Algoritma Machine Learning,” J. Ilmu Komput. dan Teknol., vol. 3, no. 1, pp. 1–6, 2022, doi: 10.35960/ikomti.v3i1.766.

[18] M. Aditya Pratama, M. Munawaroh, W. Joko Pranoto, P. Studi Teknik Informatika, F. Sains dan Teknologi, and U. Muhammadiyah Kalimantan Timur, “Perbandingan Performa Algoritma Linear Regresi dan Random Forest untuk Prediksi Harga Bawang Merah di Kota Samarinda,” J. Ilmu Tek., vol. 1, no. 2, pp. 172–182, 2024, [Online]. Available: https://doi.org/10.62017/tektonik

[19] Y. Syakir et al., “Analisis Marketplace Shopee Untuk Memprediksi Penjualan dengan Algoritma Regresi Linier,” J. Sains Komput. Inform. (J-SAKTI, vol. 6, no. 2, pp. 904–915, 2022.

[20] S. Zahara and Sugianto, “Peramalan Data Indeks Harga Konsumen Berbasis Time Series Multivariate Menggunakan Deep Learning,” J. RESTI (Rekayasa Sist. dan Teknol. Informasi), vol. 5, no. 1, pp. 24–30, 2021, doi: 10.29207/resti.v5i1.2562.

Copyright © 2025 The Author, Page 836
This Journal is licensed under a Creative Commons Attribution 4.0 International License