
IMPLEMENTASI METODE REGRESI LINEAR DALAM MEMPREDIKSI PENYAKIT ANEMIA SECARA DINI

Anggi Mei Sarah^{1*}, Bambang Kurniadi², Endang Warsini³

^{1,2,3}Sistem Informasi, STMIK Royal Kisaran

email: *anggisarah@gmail.com

Abstract: *At this time, people are still not fully aware that they have anemia or not, so many people are not vigilant. The cost is quite expensive and the place to check for anemia is also quite far away. in checking blood health conditions. The aim of this research is to predict whether someone suffers from anemia by using the regression method. In the regression method there are 7 more methods including (1) Linear Regression, (2) Support Vector Regression - Linear, (3) Support Vector Regression - RBF, (4) Decision Tree Regression, (5) Random Forest Regressor, (6) Gradient Boosting Regression, (7) NLP Regressor applied in this study. Based on the 7 regression methods, we will look for the 1 best method with the best accuracy value that will be used in the deploy process to predict someone's anemia. Decision Tree Regression is the best method among other regression methods. Where the Decision Tree Regression method produces the best accuracy values in 4 accuracy tests with ratios of 90:10, 80:20, 70:30 and 60:40, where each of the ratios produces an accuracy rate of 99% with the best estimated value, namely MAE of 0. on the MSE of 0. on the RMSE of 0. and also on the R2-Score the results are 1.*

Keywords: *anemia, prediction, regression.*

Abstrak: Pada masa ini masyarakat masih belum sepenuhnya menyadari bahwa ia menderita anemia atau tidak sehingga banyak masyarakat tidak waspada.. Biaya yang cukup mahal dan tempat untuk memeriksa anemia juga memiliki jarak yang cukup jauh, tentunya akan menambah biaya lagi untuk mencari kendaraan sehingga banyak masyarakat yang lalai dalam memeriksa kondisi kesehatan darah. Tujuan dari penelitian ini adalah untuk memprediksi apakah seseorang menderita penyakit anemia dengan menggunakan metode regresi. Pada metode regresi terdapat 7 metode lagi meliputi (1) *Linear Regression*, (2) *Support Vector Regression – Linear*, (3) *Support Vector Regression – RBF*, (4) *Decision Tree Regression*, (5) *Random Forest Regressor*, (6) *Gradient Boosting Regression*, (7) *NLP Regressor* yang diterapkan dalam penelitian ini. Berdasarkan 7 metode regresi tersebut akan dicari 1 metode terbaik dengan nilai akurasi paling terbaik yang akan digunakan dalam proses pengolahan deploy untuk memprediksi penyakit anemia seseorang. *Decision Tree Regression* menjadi metode terbaik diantara metode regresi lainnya. Dimana metode *Decision Tree Regression* menghasilkan nilai akurasi terbaik dalam 4 pengujian akurasi dengan rasio 90:10, 80:20, 70:30 dan 60:40, dimana masing-masing semua rasio menghasilkan tingkat akurasi sebesar 99% dengan nilai estimasi terbaik yaitu MAE sebesar 0. pada MSE sebesar 0. pada RMSE sebesar 0. dan serta pada R2-Score hasil sebesar 1.

Kata Kunci: anemia, prediksi, regresi.

PENDAHULUAN

Kemajuan teknologi saat ini, hampir setiap bidang pekerjaan membutuhkan komputer sebagai alat untuk menyelesaikan pekerjaan. Kemajuan teknologi membantu dalam munculnya inovasi-inovasi baru untuk penyajian informasi yang membantu kebutuhan informasi. Hingga saat ini kemajuan teknologi juga sudah masuk kedalam bidang kesehatan khususnya Anemia.

Anemia adalah suatu keadaan dimana jumlah sel darah merah atau kadar hemoglobin (Hb) di dalam darah lebih rendah daripada nilai normal. Data WHO tahun 2015 menunjukkan lebih dari 20% atau 2 miliar orang di dunia berstatus anemia. Prevalensi anemia diperkirakan 9 persen di negara-negara maju, dan 43 persen di negara berkembang. Prevalensi anemia pada WUS sebesar 30 persen dan pada wanita hamil sebesar 42 persen. Banyaknya orang yang menderita anemia, penerapan teknologi membantu untuk memberikan solusi. Penerapan teknologi untuk anemia digunakan untuk mendiagnosis seseorang apakah kemungkinan menderita anemia.

Dalam mendiagnosis seseorang apakah kemungkinan menderita anemia, tentunya dokter akan melakukan pemeriksaan laboratorium untuk pasien. Pemeriksaan laboratorium untuk diagnosis anemia dilakukan secara bertahap. Pemeriksaan berikutnya dilakukan dengan memperhatikan hasil pemeriksaan terdahulu sehingga lebih terarah dan efisien. Pemeriksaan yang dilakukan meliputi tes penyaring, pemeriksaan darah seri anemia, pemeriksaan sumsum tulang dan pemeriksaan khusus sesuai jenis anemia. Tentu hal ini akan memiliki

waktu yang lama untuk memberikan hasil pemeriksaan apakah pasien tersebut menderita anemia atau tidak. Selain itu biaya yang cukup mahal, membuat masyarakat akan berfikir ulang untuk melakukan pemeriksaan anemia. Tempat untuk memeriksa anemia juga memiliki jarak yang cukup jauh, untuk masyarakat yang tidak memiliki akses untuk pergi, tentunya akan menambah biaya lagi untuk mencari kendaraan sewa. Saat ini masyarakat juga masih belum sepenuhnya paham dan menyadari apakah ia menderita anemia sehingga banyak masyarakat tidak waspada. Apalagi anemia bisa menyebabkan kelelahan, dan jika anemia bertambah berat bisa menyebabkan stroke hingga seragan jantung.

Dari permasalahan di atas maka peneliti tertarik melakukan proses penelitian untuk mendapatkan solusi memprediksi apakah seseorang kemungkinan menderita anemia, yang dapat membantu masyarakat mendapat peringatan dini untuk menjaga kesehatan agar tidak menderita anemia dan segera melakukan pengobatan jika hasilnya menderita anemia. Metode yang digunakan pada penelitian ini adalah metode Regresi Linear. Melalui pembuatan sistem/aplikasi berbasis web untuk memprediksi anemia, maka para masyarakat mendapatkan peringatan dini dan menjaga kesehatan agar tidak mengalami anemia.

Metode Regresi Linear merupakan sebuah metode statistik yang melakukan prediksi menggunakan pengembangan hubungan matematis antara variabel, yaitu variabel dependen (Y) dengan variabel independen (X). Variabel dependen merupakan variabel

akibat atau variabel yang dipengaruhi, sedangkan variabel independen merupakan variabel sebab atau variabel yang mempengaruhi. Prediksi terhadap nilai variabel dependen dapat dilakukan jika variabel independennya diketahui [1].

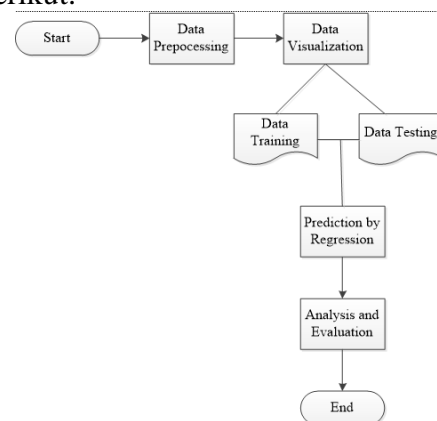
Pada penerapannya dalam penelitian di bidang kesehatan, metoderegresi linear bergandajuga telah banyak digunakan salah satunya oleh Hariyadi Deny, et., al (2021) dengan judul “Identifikasi Tingkat Resiko Penyakit Stroke Menggunakan Algoritma Regresi Linear Berganda”. Pada penelitian tersebut regresi linear berganda digunakan sebagai metode untuk memprediksi tingkat penyakit risiko jantung [2].

Oleh karena itu, penelitian ini bertujuan mendapatkan hasil apakah seorang apakah seseorang kemungkinan menderita anemia dan mengetahui metode manakah yang menghasilkan nilai akurasi yang lebih akurat dengan membandingkan 7 model yang terdapat dalam regresi diantaranya adalah (1) *Linear Regression*, (2) *Support Vector Regression – Linear*, (3) *Support Vector Regression – RBF*, (4) *Decision Tree Regression*, (5) *Random Forest Regressor*, (6) *Gradient Boosting Regression*, (7) *NLP Regressor*. Serta untuk hasil pengujiannya menggunakan 4 model pengukuran nilai estimasi yaitu *Mean Absolute Error (MAE)*, *Mean Squared Error (MSE)*, *Root Mean Square Error (RMSE)*, dan *R2-Score*.

METODE RISET

Dalam *machine learning*, penelitian ini dikategorikan sebagai penelitian *supervised learning* yaitu sebuah teknik model dimana data yang

akan diproses memiliki label/target/class dengan tujuan untuk mengetahui hubungan kausalitas antara variabel bebas (variabel independen) dan variabel yang menjadi target/labelnya (variabel y) [3]. Oleh karena itu, variabel yang terdapat dalam penelitian ini terdiri dari variabel independen atau variabel x meliputi *Hemoglobin*, *Mean Corpuscular Hemoglobin*, *Mean Corpuscular Hemoglobin Concentration*, dan *Mean Corpuscular Volume*. Serta yang menjadi variabel dependennya adalah Result (merupakan hasil seseorang menderita anemia atau tidak). Adapun tahapan yang akan dilakukan dalam proses *data mining* adalah sebagai berikut:



Gambar 1. Tahapan *Data Mining*

Pada Gambar 1 dapat dilihat bahwa tahapan pertama yang akan dilakukan adalah *Data Preprocessing* dimana dalam proses tersebut terdapat *data cleaning* untuk melakukan *handling missing* pada data, *data selection* untuk menyeleksi data berdasarkan kolom yang akan diteliti dan *data transformation*. Pada Tahapan kedua adalah melakukan *data visualization* yaitu untuk menyajikan representasi dari sebaran data yang digunakan. Tahapan selanjutnya adalah

melakukan pemisahan data set menjadi 2 bagian yaitu *data training* dan *data testing*. Setelah itu, memasuki tahapan prediksi menggunakan regresi yang meliputi 7 model regresi dan menguji data tersebut dengan data testing. Kemudian tahapan yang terakhir adalah tahapan analisis dan evaluasi data dengan menggunakan 4 metode evaluasi yaitu MAE, MSE, RMSE dan R2-Score, tujuannya adalah untuk mendapatkan hasil akurasi dari

proses mendiagnosis penyakit anemia sehingga diperoleh metode terbaik dengan hasil evaluasi yang terbaik pula.

1. Data Riset

Untuk data riset yang akan digunakan merupakan sebuah dataset *Anemia* yang berasal dari *kaggle.com*. Data tersebut berjumlah sebanyak 1441 baris data. Sampel dataset yang akan digambarkan dalam penelitian ini menggunakan 5 data teratas seperti yang terdapat dalam Tabel 1 berikut :

Tabel 1. Sampel Dataset

Hemoglobin	Mean Corpuscular Hemoglobin	Mean Corpuscular Hemoglobin Concentration	Mean Corpuscular Volume	Anemia
14.9	22.7	29.1	83.7	0
15.9	25.4	28.3	72.0	0
9.0	21.5	29.6	71.2	1
14.9	16.0	31.4	87.5	0
11.6	22.3	30.9	74.5	1

2. Pembagian Data

Pembagian data dalam penelitian ini menggunakan *data training* dan *data testing* dengan rasio sebesar 90:10 Dimana *data training* memiliki kisaran data sebanyak 90 dari dataset serta *data testing* sebanyak 10 dari dataset. Algoritma yang menggunakan rasio 90:10 memiliki performa yang lebih baik dibandingkan dengan algoritma lainnya dimana tingkat akurasi dapat diperoleh sebesar 99,40% [4]. Adapun rasio 80:20, 70:30, dan 60:40 yang digunakan dalam penelitian ini sehingga rasio pembagian data lebih bervariasi dimana tujuan yang diharapkan adalah agar memperoleh model regresi manakah yang memiliki hasil estimasi terbaik.

3. Estimasi dengan 7 Model Regresi

Regresi merupakan alat yang paling populer yang digunakan untuk memperoleh hasil evaluasi pengatur sebuah variabel x (variabel bebas) dan variabel y (variabel terikat). Dalam proses prediksi menggunakan 7 model regresi dengan tujuan untuk membandingkan model regresi terbaik yang akan digunakan. 7 model regresi tersebut yakni :

3.1 Linear Regression

Linear regression atau regresi linier terbagi atas 2 yaitu *simple linear regression* dan *multiple linear regression*. Menurut penelitian bahwa regresi linear sederhana adalah model yang menganalisis hubungan antara 1 variabel prediktor dan 1 variabel respon.

Dalam penelitian tersebut juga menyatakan bahwa prediksi menggunakan regresi linear sederhana memperoleh hasil prediksi dengan kategori baik karena telah menunjukkan hasil prediksi terbaik selama tahun 2021 pada periode bulan januari [5]. Adapun persamaan dalam regresi linear sederhana beriku ini :

$$Y = a + bx$$

Dimana :

Y : variabel dependen

a : titik perpotongan garis koordinat y

b : koefisien variabel independen

x : variabel independen

Kemudian *multiple linear regression* atau regresi linear berganda menurut [6] merupakan suatu model analisis yang menggambarkan hubungan variabel respon (Y) dengan variabel – variabel yang mempengaruhinya (x) dimana variabel (x) tersebut berjumlah lebih dari satu. Persamaan yang digunakan dalam regresi linear berganda adalah:

$$Y = b_0 + b_1X_1 + \dots + b_nX_n \quad (1)$$

Dimana :

Y : variabel respon

X : variabel prediktor

b₀ : intercept

b : koefisien

3.2 Support Vector Regression Linear

Merupakan salah satu model regresi yang digunakan untuk memperbaiki *overfitting* dan memiliki nilai akurasi yang baik [7]. *Overfitting* adalah kondisi dimana data saat diolah/*training* mendekati hasil prediksi hampir sempurna [8]. Persamaan yang terdapat dalam model *support vector regression* sebagai berikut ini :

$$f(x) = w^t(p(x)) + b$$

Dimana :

f(x) : fungsi regresi

W^t : vektor bobot dimensi l

(p(x)): titik dalam *space* F, hasil dari pemetaan x pada input *space*

b : merupakan bias

3.3 Support Vector Regression – RBF (Radial Basis Function)

Adapun persamaan yang terdapat dalam *support vector regression* dengan fungsi kernel Gaussian-RBF yaitu :

$$K(x_i, x) = \exp\left(-\frac{1}{2a^2} \|x - x_i\|^2\right) \quad (2)$$

3.4 Decision Tree Regression

Persamaan *decision tree regression* terlebih dahulu mencari nilai entropy seperti dalam persamaan berikut ini :

$$entro(S) = \sum_{i=1}^m -p(w_i|S) \cdot \log_2(w_i|S) \quad (3)$$

Dimana :

S : Himpunan Kasus

M : Total kelas data

(w_i|S) : Proporsi kelas ke-i dalam semua data latih yang diproses di node S.

Kemudian, setelah diperoleh nilai entropy maka selanjutnya kita mencari nilai gain nya sebagai berikut

$$Gai(S, J) = Entropy(S) - \sum_{i=j}^n p(v_i|S) * E(S_i)$$

Dimana :

S : Himpunan kasus

J : Fitur

n : Banyak kelas dalam *node* (akar)

(v_i|S) : Proporsi nilai v muncul pada kelas dalam *node* (akar)

(S_i) : Entropi komposisi nilai v dari kelas ke-j dalam data ke-i node

3.5 Random Forest Regressor

Menurut penelitian [9] *random forest* merupakan model algoritma pada regresi dengan tehnik yang mengkombinasikan prediksi dari beberapa pemecahan dalam *machine learning* guna mendapatkan prediksi dengan akurasi lebih baik daripada mode tunggal. Dalam penelitian tersebut juga menyimpulkan bahwa *random forest regression* lebih akurat dibandingkan dengan *linear regression* dimana rating akurasi pada *random forest* diperoleh sebesar 97.7% dengan menggunakan perhitungan nilai RMSE dan MAPE. Persamaan *random forest regressor* seperti yang terdapat dibawah ini :

$$\hat{y}_i = \frac{1}{N_{\text{tree}}} \sum_{n=1}^{N_{\text{tree}}} \hat{y}_n \quad (5)$$

Dimana :

\hat{y}_i : hasil prediksi

N_{tree} : total jumlah pohon

\hat{y}_n : hasil prediksi pohon ke-n

3.6 Gradient Boosting Regression

Merupakan bagian dari algoritma ensemble yang menggunakan peningkatan akurasi sebuah nilai. Model *gradient boosting regression* mampu mengatasi pola yang kompleks. Struktur data dari *gradient boosting regression* adalah pohon keputusan (*decision tree*). Adapun persamaan *gradient boosting regression* adalah sebagai berikut ini :

$$r_{im} = - \left[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)} \right]_{F(x)=F_{m-1}(x)} \quad \text{for } i=1, \dots, n$$

3.7 Natural Language Processing (NLP) Regressor

Adalah kemampuan komputer dalam mengolah bahasa manusia baik secara lisan maupun tulisan dengan tujuan membangun sebuah aplikasi yang digunakan untuk memfasilitasi komunikasi melalui bahasa alami

4. Pengukuran Akurasi

Data yang telah diolah menggunakan 7 model regresi kemudian diukur nilai akurasinya. Pengolahan data yang bersifat estimasi yang masih mengandung unsur ketidakpastian maka diperlukan untuk memasukkan nilai *error* sebagai pengukuran akurasi dari estimasi tersebut. Nilai *Error* merupakan selisih antara nilai pengamatan yang sebenarnya dengan nilai prediksi. Untuk estimasi yang diukur menggunakan MAE, MSE, dan RMSE nilai yang paling baol adalah nilai yang paling kecil, semakin kecil nilai yang dihasilkan berarti hasil estimasi semakin bagus. Berbeda dengan *R2-Score* nilai koefisien determinan yang mendekati 1 mengartikan bahwa variabel independennya memberikan semua informasi yang diperlukan dalam mengestimasi variabel dependennya [10].

HASIL DAN PEMBAHASAN

Pada pembahasan dalam penelitian ini akan dilakukan visualisasi data agar data dapat lebih mudah untuk diamati serta dapat mengetahui hubungan antara tiap variabel x terhadap variabel y. Selanjutnya, data akan diuji dengan 4 bentuk pembagian data dengan rasio yang berbeda-beda yang dimana didalamnya akan dilakukan pengujian data menggunakan 7 model regresi dan diukur dengan 4 model nilai evaluasi. Hal tersebut bertujuan untuk menemukan model regresi manakah yang paling baik dengan menghasilkan nilai akurasi yang paling akurat untuk digunakandalam memprediksi seseorang apakah menderita anmia atau tidak.

a. Pengujian dengan 7 Regresi

Dalam penelitian ini menggunakan metode regresi yang akan dilakukan perbandingan dari 7 metode regresi agar memperoleh model dengan hasil akurasi terbaik untuk digunakan dalam proses memprediksi seseorang apakah menderita *anemia* atau tidak.

Penelitian ini menggunakan pembagian data untuk diuji sebanyak 4 bentuk diantaranya pembagian rasio pertama yaitu 90:10, kedua 80:20,

ketiga 70:30 dan rasio keempat yaitu 60:40. nantinya akan diketahui model regresi yang terbaik dalam melakukan proses memprediksi seseorang apakah menderita anemia atau tidak.

b. Pengujian Pada Rasio Data 90:10

Pengujian data kedua menggunakan pembagian data dengan rasio 80:20. Terdapat sebanyak 80% *data training* dan 20% *data testing* dari dataset. diperoleh hasil berikut.

Tabel 2. Pengujian Akurasi Data Rasio 90:10

MODEL	MAE	MSE	RMSE	R2-SCORE
LinearRegression	0.24166387	0.0927601	0.30456544	0.62668102
SVR-Linear	0.24019935	0.0955814	0.30916241	0.61532655
SVR-RBF	0.23780095	0.09100236	0.30166598	0.63375516
DecisionTree	0.	0.	0.	1.
RandomForest	0.0078125	0.0015625	0.03952847	0.99371162
GradientBoosting	0.08414339	0.02153946	0.14676329	0.91331305
MLP-Regressor	0.2512619	0.10367496	0.32198597	0.58275347

Ketentuan yang digunakan dalam melihat hasil memprediksi anemia yang baik adalah untuk nilai pengukuran menggunakan MAE, MSE dan RMSE nilai yang paling kecil yang dihasilkan merupakan nilai dengan hasil yang paling bagus. Sedangkan untuk R2-Score nilai angka 1 (satu) merupakan nilai dengan estimasi yang paling bagus. Berdasarkan tabel diatas, pada model pengukuran MAE nilai yang paling bagus dihasilkan oleh metode *Decision Tree Regression* dengan nilai sebesar 0. MSE sebesar 0. pada RMSE sebesar 0. dan R2-Score dengan hasil sebesar 1. Berdasarkan pengamatan tabel dan

pemaparan diatas, pengujian akurasi dengan rasio data 90:10 menghasilkan *Decision Tree Regression* sebagai metode regresi yang paling baik karena metode tersebut paling banyak menghasilkan nilai terbaik pada 4 model pengukuran yaitu MAE, MSE, RMSE dan R2-Score.

c. Pengujian Pada Rasio Data 80:20

Pengujian data kedua menggunakan pembagian data dengan rasio 80:20. Terdapat sebanyak 80% *data training* dan 20% *data testing* dari dataset.

Tabel 3. Pengujian Akurasi Data Rasio 80:20

MODEL	MAE	MSE	RMSE	R2-SCORE
LinearRegression	0.24068319	0.09126450	0.30210014	0.62515653
SVR-Linear	0.23718242	0.09253993	0.30420377	0.61991802
SVR-RBF	0.23662406	0.08978162	0.29963582	0.63124701
DecisionTree	0.	0.	0.	1.
RandomForest	0.0139738	0.00419214	0.06474674	0.98278195
GradientBoosting	0.08894949	0.0263573	0.16234932	0.89174473
MLP-Regressor	0.25445036	0.10504389	0.32410476	0.5685615

Berdasarkan tabel diatas, diperoleh metode *Decision Tree Regression* sebagai metode regresi terbaik diantara 7 metode regresi lainnya. Hal tersebut dapat kita lihat bahwa nilai estimasi *Decision Tree Regression* yang paling bagus pada 4 model pengukuran yaitu nilai sebesar 0. pada MAE, 0. pada MSE, 0. pada RMSE dan R2-SCORE nilai akurasi sebesar sebesar 1. Maka, dapat diketahui pada pengujian akurasi

data 80:20 menghasilkan metode *Decision Tree Regression* sebagai metode dengan nilai akurasi yang paling baik.

d. Pengujian Pada Rasio Data 70:30

Pengujian data ketiga dengan menggunakan rasio data 70:30 yaitu 70% *data training* dan 30% *data testing* menggunakan 7 model regresi diperoleh hasil sebagai berikut :

Tabel 4. Pengujian Akurasi Data Rasio 70:30

MODEL	MAE	MSE	RMSE	R2-SCORE
LinearRegression	0.23888422	0.08821371	0.29700793	0.62515653
SVR-Linear	0.23797199	0.08975595	0.29959297	0.63418899
SVR-RBF	0.23969109	0.08948881	0.2991468	0.63527775
DecisionTree	0.	0.	0.	1.
RandomForest	0.01727575	0.00584718	0.07646683	0.97616914
GradientBoosting	0.08563743	0.02359887	0.15361923	0.90382002
MLP-Regressor	0.331471	0.16135451	0.40168956	0.34238059

Berdasarkan tabel diatas, *Decision Tree Regression* menjadi metode regresi yang paling banyak muncul dengan menghasilkan nilai prediksi yang paling

baik diantara nilai yang dihasilkan oleh metode lainnya. *Decision Tree Regression* menghasilkan 0. dengan nilai pengukuran MAE, 0. dengan MSE,

0, RMSE dan 1 R2-Score. Dengan begitu *Decision Tree* pada pengujian data pada rasio 70:30 ini metode paling baik

e. Pengujian Pada Rasio 60:40

Pengujian data keempat diperoleh hasil sebagai berikut ini :

Tabel 5. Pengujian Akurasi Data Rasio 60:40

MODEL	MAE	MSE	RMSE	R2-SCORE
LinearRegression	0.24910268	0.09314994	0.30520475	0.61420038
SVR-Linear	0.25015722	0.09645112	0.3105658	0.60052787
SVR-RBF	0.25255812	0.09629609	0.31031612	0.60116993
DecisionTree	0.	0.	0.	1.
RandomForest	0.02369942	0.00728324	0.08534188	0.96983498
GradientBoosting	0.08646946	0.02396207	0.15479687	0.90075616
MLP-Regressor	0.28373996	0.12676511	0.35604089	0.47497622

Pada pengujian akurasi data rasio 60:40, metode *Decision Tree Regression* menjadi metode regresi terbaik dengan paling banyak menghasilkan nilai yang terbaik daripada metode lainnya. Nilai akurasi yang dihasilkan pada MAE sebesar 0., pada MSE sebesar 0., pada RMSE 0. , dan R2-SCORE nilai akurasi sebesar 1.

f. Regresi Terbaik

Setelah melakukan pengujian akurasi dengan 4 bentuk rasio data, kemudian diperoleh metode terbaik yang paling banyak menghasilkan nilai estimasi terbaik yaitu metode *Decision Tree Regression*, yang memiliki tingkat akurasi sebesar 99% sebagaimana dapat dilihat pada tabel dibawah ini :

Tabel 6. Metode Terbaik

Metode Regresi	90:10	80:20	70:30	60:40
<i>Decision Tree Regression</i>	99%	99%	99%	99%

KESIMPULAN

Berdasarkan hasil pengolahan data menggunakan 7 metode regresi menggunakan *software jupyter notebook, visual studio code* dan *python* sebagai bahasa pemrogramannya serta pengujiannya menggunakan 4 rasio pengukuran nilai akurasi maka metode terbaik yang menghasilkan nilai akurasi

paling bagus yaitu *Decision Tree Regression* dengan tingkat akurasinya sebesar 99%. Dengan demikian metode regresi yaitu *Decision Tree Regression* digunakan sebagai metode regresi yang akan diimplementasikan dalam pembuatan *deploy* aplikasi untuk memprediksi apakah seseorang menderita anemia atau tidak.

DAFTAR PUSTAKA

- [1] E. Triyanto, H. Sismoro, And A. D. Laksito, "Implementasi Algoritma Regresi Linear Berganda Untuk Memprediksi Produksi Padi Di Kabupaten Bantul," *Rabit J. Teknol. Dan Sist. Inf. Univrab*, Vol. 4, No. 2, Pp. 66–75, 2019, Doi: 10.36341/Rabit.V4i2.666.
- [2] D. Haryadi, D. Marini Umi Atmaja, A. Rahman Hakim, And N. Suwaryo, "Identifikasi Tingkat Resiko Penyakit Stroke Menggunakan Algoritma Regresi Linear Berganda," *Deny Haryadi, Sntem*, Vol. 1, No. November, Pp. 1198–1207, 2021.
- [3] R. Artikel, N. Luh, P. Chandra, R. A. Rahman, R. Venyutzky, And N. A. Rakhmawati, "Analisis Klasifikasi Sentimen Terhadap Sekolah Daring Pada Twitter Menggunakan Supervised Machine Learning," Vol. 7, No. April, Pp. 47–58, 2021.
- [4] A. I. Sang, E. Sutoyo, And I. Darmawan, "Analisis Data Mining Untuk Klasifikasi Data Kualitas Udara Dki Jakarta Menggunakan Algoritma Decision Tree Dan Support Vector Machine Data Mining Analysis For Classification Of Air Quality Data Dki Jakarta Using Decision Tree Algorithm And Support Vector," Vol. 8, No. 5, Pp. 8954–8963, 2021.
- [5] M. M. Syaifulloh, "Prediksi Indeks Standar Pencemaran Udara Di Kota Surabaya Berdasarkan Konsentrasi Gas Karbon Monoksida," Vol. 2, No. November, 2021.
- [6] A. Prasetyo, "Prediksi Produksi Kelapa Sawit Menggunakan Metode Regresi Linier Berganda," Vol. 6, No. 2, Pp. 76–80, 2021.
- [7] A. Aulia *Et Al.*, "Prediksi Harga Emas dengan menggunakan Algoritma Support Vector Regression (Svr) Dan linear regression (Lr)," Vol. 8, No. 5, 2022, Doi: 10.5281/Zenodo.6408864.
- [8] N. D. Maulana, B. D. Setiawan, And C. Dewi, "Implementasi Metode Support Vector Regression (Svr) Dalam Peramalan Penjualan Roti (Studi Kasus : Harum Bakery)," Vol. 3, No. 3, Pp. 2986–2995, 2019.
- [9] S. Fachid And A. Triayudi, "Perbandingan Algoritma Regresi Linier Dan Regresi Random Forest Dalam Memprediksi Kasus Positif Covid-19," *J. Media Inform. Budidarma*, Vol. 6, No. 1, Pp. 68–73, 2022, Doi: 10.30865/Mib.V6i1.3492.
- [10] "Memahami Koefisien Determinasi Dalam Regresi Linear – Accounting."