Jurnal Sains dan Edukasi Sains
Vol.8, No.2, Agustus 2025: 133-142
https://doi.org/10.24246/juses.v8i2p133-142

Perbandingan Kinerja Metode *Support Vector Regression* dan Metode Regresi Linier Berganda dalam Memprediksi BMI pada Dataset ASTHMA

Titus Antonius David Kurniawan[1], Adi Setiawan[2,*], Faldy Tita[1]

[1]Program Studi Matematika, Fakultas Sains dan Matematika, Universitas Kristen Satya Wacana, Salatiga, Jawa Tengah 50711
[2]Program Studi Magister Sains Data, Fakultas Sains dan Matematika, Universitas Kristen Satya Wacana, Salatiga, Jawa Tengah 50711

*email korespondensi: adi.setiawan@uksw.edu

Received: 10 Juli 2024; ***Revised***: 16 Februari 2025; ***Accepted***: 6 Mei 2025; ***Published***: 1 Agustus 2025

ABSTRAK

Body Mass Index (BMI) merupakan metode untuk mengevaluasi apakah seseorang memiliki berat badan ideal, yang berperan penting dalam mengidentifikasi risiko kesehatan seperti penyakit jantung dan diabetes. Dalam penelitian ini, BMI digunakan sebagai variabel prediktor untuk menentukan kemungkinan seseorang menderita asma, yaitu penyakit kronis yang memengaruhi saluran pernapasan. Untuk memprediksi BMI berdasarkan variabel-variabel lain dalam dataset Asthma, penelitian ini membandingkan dua metode regresi, yaitu regresi linier berganda dan Support Vector Regression (SVR). Evaluasi akurasi model dilakukan menggunakan Mean Absolute Percentage Error (MAPE), yang mengukur kesalahan prediksi dalam bentuk persentase, di mana nilai MAPE yang lebih rendah menunjukkan tingkat akurasi prediksi yang lebih baik. Data uji dibagi menjadi empat skenario, yaitu 10%, 20%, 30%, dan 40% dari keseluruhan data. Hasil perhitungan menunjukkan bahwa metode regresi linier berganda menghasilkan nilai MAPE sebesar 15,42%; 15,44%; 15,51%; dan 15,63% secara berturut-turut. Sementara itu, metode SVR menghasilkan nilai MAPE sebesar 15,77%; 15,74%; 15,77%; dan 15,81%. Berdasarkan hasil tersebut, regresi linier berganda terbukti memberikan prediksi yang lebih akurat dibandingkan SVR dalam konteks dataset Asthma. Dengan demikian, regresi linier berganda lebih efektif dalam memodelkan BMI dibandingkan SVR dan dapat menjadi pertimbangan penting dalam pengembangan model prediktif untuk mendukung pengambilan keputusan terkait risiko penyakit pernapasan seperti asma.

Kata-kata kunci: *Asthma; BMI; MAPE; Regresi linier berganda; Support Vector Regression*

PENDAHULUAN

Penyakit inflamasi saluran pernapasan yang berhubungan dengan hiperresponsif saluran napas yang menyebabkan penyempitan saluran napas berlebihan yang dipicu oleh virus, alergen, dan aktivitas fisik yang menyebabkan bersin, kesulitan bernapas, sesak napas, dan batuk yang dapat berbeda seiring dengan waktu dan intensitasnya biasa disebut asma (Permatasari & Yanti, 2020). Asma ditandai dengan mengi, sesak napas, dada terasa berat dan batuk terutama pada malam hari, disertai hiperinflamasi dan hiperventilasi pernapasan (Fitri & Kartikasari, 2021).

Body Mass Index (BMI) adalah metode pengukuran yang memperbandingkan berat badan dengan tinggi badan seseorang. BMI digunakan menentukan apakah tubuh seseorang termasuk dalam kategori berat kurang, normal, obesitas ringan, obesitas sedang, dan obesitas berat. Selain itu, BMI berguna dalam mengelola berat badan agar mencapai tingkat normal yang sesuai dengan tinggi badan. BMI memiliki rumus yang diperlihatkan pada persamaan (1) (Rismawan & Kusumadewi, 2008).

$$BMI = \frac{Berat\ badan\ (kg)}{\left(tinggi\ badan(m)\right)^2} \qquad (1)$$

Dengan pengelompokan BMI sebagai berikut:

> 30	: obesitas berat,
$25 - 29{,}9$	: obesitas sedang,
$23 - 24{,}9$	: obesitas ringan,
$18{,}5 - 22{,}9$	: normal, dan
$< 18{,}5$	: berat kurang.

Penelitian sebelumnya mengenai perbandingan metode *support vector regression* dengan regresi linier berganda yang dilakukan oleh (Ginting et al., 2021) tentang memprediksi stok obat dengan metode *support vector machine* dan regresi linier berganda, hasil penelitian tersebut menyatakan bahwa metode regresi linier berganda lebih baik digunakan untuk memprediksi stok obat. Dibuktikan dengan nilai *Mean Squared Error* (MSE) pada regresi linier berganda lebih kecil dibandingkan MSE SVR.

Metode regresi merupakan pendekatan statistik yang digunakan untuk memahami hubungan antara satu atau lebih variabel bebas dengan variabel terikat, berfungsi juga untuk mengukur atau memprediksi nilai variabel terikat berdasarkan variabel bebas tersebut. Metode regresi ini mempunyai beberapa jenis seperti regresi linier sederhana, metode linier berganda, regresi gamma, regresi *poisson, support vector regression*, dan masih banyak lagi. Tujuan penelitian ini adalah membandingkan kinerja dari metode regresi linier berganda dengan metode *support vector regression* untuk memprediksi BMI dalam data set Asthma. Kedua metode akan dievaluasi menggunakan RMSE, MAE, MAPE, dan koefisien determinasi untuk mengetahui metode mana yang lebih baik.

DASAR TEORI

Regresi

Regresi merupakan analisis yang merepresentasikan ketergantungan variabel terikat (*dependent*) terhadap satu atau lebih variabel bebas (*independent*), yang bertujuan memprediksi dan memperkirakan rata-rata populasi atau nilai rata-rata dari variabel terikat berdasarkan nilai variabel bebas yang diketahui. Fokus utama dalam regresi adalah menjelaskan dan menilai hubungan antara variabel-variabel tersebut (Yang, 2017; Basri, 2018).

Support Vector Regression (SVR)

Support Vector Regression atau biasa disingkat SVR adalah modifikasi dari metode *Support Vector Machine* dimana SVR digunakan untuk masalah regresi. Secara umum konsep regresi fungsi linier dapat dituliskan seperti pada persamaan (2) (Prakoso, 2019):

$$f(x) = \langle w, x \rangle + b \quad dengan \quad w \in R^d, \quad x \in R^d, \quad b \in R \tag{2}$$

Misalkan terdapat n set *data training*, (x_i, y_i) dengan $i = 1,2,\dots,n$. Dalam hal ini $x_i \in R^d$ merupakan vektor fitur pada data ke-*i* dan $y_i \in R$ merupakan nilai target yang sesuai sehingga fungsinya dapat dituliskan sebagai :

$$y_i = w.x_i + b \tag{3}$$

dengan koefisien dari *w* dan *b* berfungsi sebagai parameter yang mengoptimalkan model linear $f(x) = \langle w, x \rangle + b$ untuk menghasilkan prediksi akurat. Dalam konteks pembelajaran mesin, akurasi model diukur melalui fungsi risiko, yang mencerminkan selisih antara nilai prediksi dan sebenarnya. Karena distribusi data sebenarnya tidak diketahui, fungsi risiko tersebut didekati melalui risiko empiris berdasarkan data pelatihan. Sehingga tugas pembelajaran dikonversi menjadi masalah optimasi $w, \xi_i, \; dan \; \xi_i^*$ yang meminimalkan fungsi objektif, diperlihatkan pada persamaan (4)-(7).

$$\left(\frac{1}{2}||w||^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*)\right) \tag{4}$$

dengan kendala,

$$y_i \leq f(x_i) + \varepsilon + \xi_i \tag{5}$$

$$y_i \geq f(x_i) - \varepsilon - \xi_i \tag{6}$$

$$\xi_i, \xi_i^* \geq 0 \tag{7}$$

Dalam hal ini, ‖w‖ diminimumkan dengan tujuan membuat fungsi f setipis (*flat*) mungkin. Nilai tawar (*trade off*) antara ketipisan fungsi f dan batas atas deviasi lebih dari ε yang masih bisa ditoleransi disebut Konstanta C (Cost) > 0, sedangkan ξ_i , ξ_i^* merupakan variabel slack positif, digunakan untuk mengatasi masalah pembatas yang tidak layak (infeasible constraint) dalam masalah optimasi (Purnama, 2020).

Regresi Linear Berganda

Regresi linier berganda merupakan suatu metode pemodelan yang digunakan untuk menggambarkan hubungan fungsional antara satu variabel terikat dengan dua atau lebih variabel bebas (Masruroh & Mauladi, 2020). Model regresi ini dapat diformulasikan sebagaimana yang dijelaskan oleh (Niu et al., 2019) direpresentasikan pada persamaan (8).

$$Y = a + b_1X_1 + b_2X_2 + \cdots + b_nX_n \tag{8}$$

Dengan :
Y = variabel tak bebas (nilai variabel yang akan diprediksi),
a = konstanta,
$b_1, b_2, \dots, b_n$ = nilai koefisien regresi,
$X_1, X_2, \dots, X_n$ = variabel bebas.

Pada kasus regresi, metode yang baik dapat ditentukan dengan menghitung kesalahan prediksinya (Sivana Ihzaniah et al., 2023). Terdapat berbagai cara yang dapat digunakan untuk menghitung kesalahan prediksi, di antaranya :

1. ***Mean Absolute Error*** **(MAE)**

   Mean Absolute Error (MAE) merupakan metode yang digunakan untuk mengevaluasi seberapa akurat sebuah model peramalan. Nilai MAE semakin kecil menunjukkan bahwa tingkat akurasi sebuah model peramalan semakin tinggi. Hasil MAE menghitung rata-rata selisih absolut antara nilai yang sebenarnya dengan nilai yang diprediksi oleh model, menggambarkan seberapa besar perbedaan absolut antara nilai sebenarnya dengan prediksi model (Suryanto, 2019). MAE dirumuskan pada persamaan (9).

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|f_i - y_i| \tag{9}$$

   Dengan :
   f_i = nilai hasil peramalan (prediksi) ke-i, $(i = 1,2, \dots, n)$,
   y_i = nilai sebenarnya (aktual) ke-i, $(i = 1,2, \dots, n)$,
   n = ukuran sampel.

2. ***Root Mean Squared Error*** **(RMSE)**

   Root Mean Squared Error (RMSE) merupakan akar dari jumlah kuadrat kesalahan yang dirata ratakan, juga dapat dipakai suatu ukuran kesalahan yang dihasilkan dari suatu model prediksi (Nurani et al., 2023). Nilai yang semakin rendah menandakan kinerja model yang Semakin baik. Nilai RMSE dapat dicari dengan persamaan (10) (Mostafa & Amano, 2019).

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(f_i - y_i)^2}{n}} \tag{10}$$

   Dengan :
   f_i = nilai hasil peramalan (prediksi) ke-i, $(i = 1,2, \dots, n)$,
   y_i = nilai sebenarnya (aktual) ke-i, $(i = 1,2, \dots, n)$,
   n = ukuran sampel.

3. ***Mean Absolute Precent Error*** **(MAPE)**

   Mean absolute percent error (MAPE) atau kesalahan persentase absolut rata-rata merupakan ukuran yang menunjukkan tingkat kesalahan relative dengan menyajikan dalam bentuk persentase. MAPE

menyatakan persentase kesalahan hasil peramalan dengan permintaan aktual dalam jangka waktu tertentu. Nilai ini mengindikasikan seberapa besar kesalahan prediksi, apakah terlalu tinggi atau terlalu rendah dibandingkan data aktual (Arumningsih & Darsyah, 2018; Nurfadilah et al ., 2022). Nilai MAPE dihitung menggunakan persamaan (11).

$$MAPE = \frac{100}{n}\sum_{i=1}^{n}\frac{|y_i - f_i|}{y_i} \tag{11}$$

Dengan :
f_i = nilai hasil peramalan (prediksi) ke-i, $(i = 1,2,\dots,n)$,
y_i = nilai sebenarnya (aktual) ke-i, $(i = 1,2,\dots,n)$,
n = ukuran sampel.

Penggunaan nilai MAPE pada suatu evaluasi hasil prediksi dapat menghindari pengukuran akurasi antara besaran nilai aktual dengan nilai prediksi, dimana kriteria nilai MAPE dikatakan sangat baik jika persentase nilai MAPE tidak lebih dari 10%, baik bila nilai MAPE antara 10% - 20%, cukup bila persentase MAPE antara 20% sampai dengan 50%, dan dikatakan buruk jika persentase MAPE-nya lebih dari 50% (Putro et al., 2018).

4. Koefisien Determinasi

Koefisien Determinasi merupakan hasil yang menampilkan seberapa besar variabel bebas dalam mempengaruhi variabel terikatnya. Nilai yang berkisaran antara 0 sampai 1 yang dapat mengidentifikasi besaran kombinasi variabel bebas yang secara bersamaan mempengaruhi nilai variabel dependennya. Nilai antara 1 sampai dengan 0 dibagi dalam beberapa kategori. Hubungan sangat kuat bila bernilai 0,8 sampai dengan 1; kuat 0,6 sampai 0,79; cukup kuat 0,4 sampai 0,59; lemah 0,2 sampai 0,39; dan sangat lemah bila di interval 0 sampai 0,19. Rumus Koefisien Determinasi dijabarkan pada persamaan (12) (Chicco et al., 2021).

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{12}$$

Dengan :
$\hat{y}_i$ = nilai hasil peramalan (prediksi) ke-i, $(i = 1,2,\dots,n)$,
y_i = nilai sebenarnya (aktual) ke-i, $(i = 1,2,\dots,n)$,
$\bar{y}$ = rata-rata,
n = ukuran sampel.

Contoh Sederhana

Sebagai ilustrasi sederhana, digunakan sebagian data dari *dataset* Asthma dengan variabel dependen y adalah *Body Mass Index* (BMI). Sementara itu, 2 variabel independen yang digunakan adalah x_1 status merokok (*smoke*) dan x_2 yaitu nilai fitur genetik rs4490198 yang telah melalui proses skoring. Fitur rs4490198 merupakan salah satu *Single Nucleotide Polymorphism* (SNP) yang relevan dalam studi genetik dan telah diolah sebelumnya agar dapat digunakan sebagai variabel numerik dalam model regresi dengan pemisalan sebagai berikut :

$y = (20, 27, 33, 29, 41, 24)$,
$x_1 = (1, 0, 0, 1, 0, 0)$,
$x_2 = (1, 1, 0{,}5, 0{,}5, 0, 0)$.

Metode regresi linier berganda menggunakan aplikasi R dengan fungsi 'lm', sehingga didapatkan regresinya yaitu $y = 33{,}846 - 4{,}154x_1 - 6{,}923x_2$ yang mengakibatkan diperoleh-nya $\hat{y}$ berturut-turut adalah 22,770; 26,923; 30,385; 26,231; 33,846; 33,846 dan rata-rata sebesar 29. Selanjutnya dicari nilai MAE, RMSE, MAPE, dan koefisien determinasi sebagai berikut :

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|f_i - y_i|$$

$$= \frac{1}{6}(|22{,}770 - 20| + |26{,}923 - 27| + |30{,}385 - 33| + |26{,}231 - 29| + |33{,}846 - 41| + |33{,}846 - 24|) = 4{,}205,$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(f_i - y_i)^2}{n}}$$

$$= \sqrt{\frac{((22{,}770 - 20)^2 + (26{,}923 - 27)^2 + (30{,}385 - 33)^2 + (26{,}231 - 29)^2 + (33{,}846 - 41)^2 + (33{,}846 - 24)^2)}{6}}$$

$$= \sqrt{\frac{7{,}669 + 0{,}006 + 6{,}840 + 7{,}669 + 51{,}177 + 96{,}947}{6}} = 28{,}385,$$

$$MAPE = \frac{100}{n}\sum_{i=1}^{n}\frac{|y_i - f_i|}{y_i}$$

$$= \frac{100}{6}(0{,}138 + 0{,}003 + 0{,}079 + 0{,}095 + 0{,}174 + 0{,}410) = 15{,}013$$

$$R^2(y, \hat{y}) = 1 - \frac{\sum_{i=1}^{n}(y_i - f_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

$$= 1 - \frac{(20 - 22{,}77)^2 + (27 - 26{,}923)^2 + (33 - 30{,}385)^2 + (29 - 26{,}231)^2 + (41 - 33{,}846)^2 + (24 - 33{,}846)^2}{(20 - 29)^2 + (27 - 29)^2 + (33 - 29)^2 + (29 - 29)^2 + (41 - 29)^2 + (24 - 29)^2}$$

$$= 1 - 0{,}631$$

$$= 0{,}369$$

Model *Support Vector Regression* (SVR) digunakan untuk memprediksi nilai indeks massa tubuh (*Body Mass Indeks*/BMI) berdasarkan fitur yang tersedia. Implementasi dilakukan menggunakan *package* e1071 dalam perangkat lunak R dengan fungsi svm. Hasil prediksi ($\hat{y}$) yang diperoleh dari model secara berturut-turut adalah 22,513; 27,737; 32,261; 28,267; 30,827; dan 30,827. Terhadap evaluasi performa model, diperoleh dari ke-4 metrik yaitu MAE sebesar 3,620; RMSE sebesar 5,132; MAPE sebesar 12,220%; dan koefisien determinasi sebesar 0,415.

Data dan Skoring

Penelitian ini menggunakan data asma dan data ini didapat dari paket program "SNPassoc" dengan nama data "Asthma". Data Asthma terdiri dari *country, gender, BMI, Smoke, casecontrol*, dan kode genetik *Single Nucleotide Polymorphism* (SNP). Kode-kode genetik dibedakan oleh nukleotida tunggal yang disusun menjadi rangkaian *Deoxyribonucleic acid* (DNA). Fitur BMI untuk penelitian ini menjadi variabel terikat, sedangkan fitur lain seperti negara, jenis kelamin, casecontrol, dan kode genetik menjadi variabel bebas.

Data harus diproses dengan membersihkan data dan mengkonversi data, sehingga dapat dilakukan simulasi. Data diproses menggunakan pengkodean machine learning, data dibagi menjadi 2 bagian, bagian pertama merupakan data *training* dan data kedua merupakan data *testing*. Data *training* digunakan untuk memperoleh parameter dalam model sedangkan data testing (data uji) digunakan untuk menentukan kinerja metode yang digunakan.

Penilaian dilakukan dilakukan dengan memberikan skor terhadap setiap fitur. Negara diberi skor 1 sampai dengan 10 secara urut Australia, Belgia, Estonia, Perancis, Jerman, Norwegia, Spanyol, Swedia, Swiss, dan Inggris, sedangkan untuk fitur jenis kelamin Perempuan diberi skor 0 dan laki-laki diberi skor 1. Skoring kode genetik GG diberi skor 0, AG/GA diberi nilai 0,5 serta GG diberi nilai 1.

HASIL DAN DISKUSI

Statistik Deskriptif

Tabel 1 menunjukkan statistik deskriptif untuk fitur-fitur data Asthma seperti *age, body mass index*, *smoke* dan *case control*, dimana untuk fitur age dengan nilai minimal yaitu 26,44 dan maksimal 56,47 dengan rata-rata 42,91 dan nilai tengah 43,47. BMI nilai minimal 16,17 dan maksimal 69,20, nilai tengah 24,93 dan rata-rata 25,53. *Smoke* mempunyai nilai minimal 0 dan nilai maksimal 1 dengan rata-rata 0,3. *Case control* mempunyai nilai minimal 0 dan nilai maksimal 1 dengan rata-rata 0,22 sedangkan untuk *country* dan *gender* tidak bisa dihitung karena data masih dalam bentuk tulisan.

Tabel 1. Statistik Deskriptif Data Asthma

Statistik	*Age*	BMI	*Smoke*	*Case control*
Min	26,44	16,71	0,00	0,00
Kuartil 1	36,96	22,66	0,00	0,00
Median	43,47	24,93	0,00	0,00
Mean	42,91	25,53	0,30	0,22
Kuartil 3	48,68	27,69	1,00	0,00
Max	56,47	69,20	1,00	1,00

Setelah diskoring nilai-nilai statistika deskriptif dari *country* dan *gender* sudah bisa dihitung yang dapat dilihat pada **Tabel 2.** Nilai untuk minimal dan maksimal sama namun ada sedikit perbedaan pada nilai rata-rata dan nilai tengahnya, dan sudah termasuk fitur *country* dan *gender*. Nilai nilai statistika deskriptif untuk smoke dan casecontrol sama dengan nilai sebelum dilakukan skoring. *Country* nilai minimal 1 dan maksimal 10 dengan rata-rata 42,91 dan nilai tengah 43,47. Fitur *gender* nilai minimal yaitu 0 dan maksimal 1 dengan nilai tengah 1 dan rata-rata 0,5. Fitur *age* dan BMI nilai minimal dan maksimal sama dengan data sebelum diskoring namun ada perbedaan pada nilai tengah dan rata-rata, untuk *age* nilai tengahnya 43,51 dengan rata-rata 43,00. Sedangkan untuk BMI nilai tengahnya menjadi 24,78 dengan rata-rata 25,41.

Tabel 2. Statistika Deskriptif Data Asthma Setelah Skoring

Statistik	*Country*	*Gender*	*Age*	BMI	*Smoke*	*Case control*
Min	1,00	0,00	26,44	16,71	0,00	0,00
Kuartil 1	5,00	0,00	36,84	22,58	0,00	0,00
Median	7,00	1,00	43,51	24,78	0,00	0,00
Mean	6,74	0,50	43,00	25,41	0,30	0,22
Kuartil 3	8,00	1,00	48,84	27,45	1,00	0,00
Max	10,00	1,00	56,47	69,20	1,00	1,00

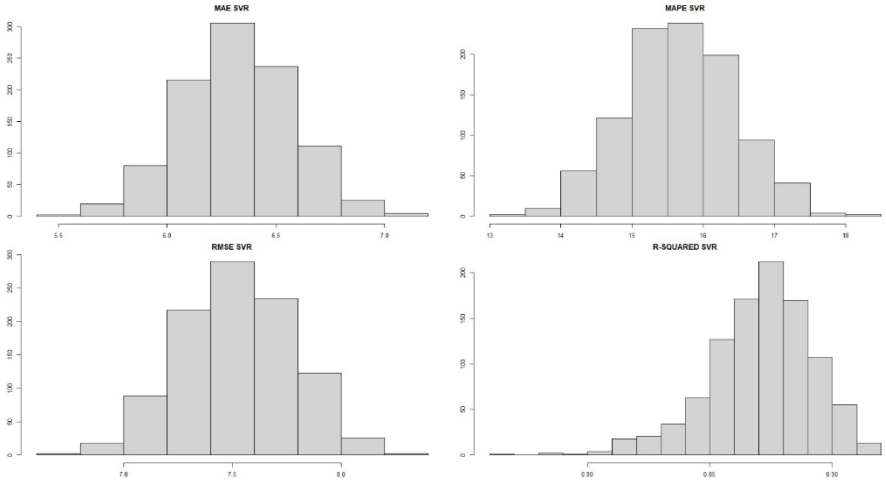
Support Vector Regression

Data Asthma pada *software* R Studio memiliki 1578 baris dengan 57 kolom yang menunjukkan kategori data. Dari dataset asma dilakukan regresi dengan metode *Support Vector Regression* dengan menggunakan *packages* 'e1071'. Nilai NA dihilangkan terlebih dahulu, setelah itu diperoleh data yang memiliki 1076 baris dan kolom 57. Kemudian diubah ke dalam data numerik, untuk menentukan nilai RMSE, MAE, MAPE, dan koefisien determinasi serta dilakukan pengulangan sebanyak 1000 kali dengan metode *Support Vector Regression* dan membagi 2 data, yaitu 80% merupakan data latih dan 20% data uji. Hasil nilai-nilai MAE, MAPE, RMSE, dan koefisien determinasi dapat dilihat pada **Tabel 3** sedangkan diagram histogram dapat dilihat pada **Gambar 1.** Di samping itu, juga dilakukan 1000 kali pengulangan dengan data uji 10%; 20%; 30%; 40% untuk memprediksi peubah respon atau target data BMI (*Body Mass Index*) dan diperoleh hasil mean RMSE, MAPE, MAE, dan koefisien determinasi yang dinyatakan pada **Tabel 4.**

Tabel 3. Hasil Perhitungan 1000 kali metode SVR dengan proporsi data uji 20%

Statistik	MAE	MAPE	RMSE	Koefisien Determinasi
Min	5,724	13,83	6,870	0,7726
Kuartil 1	6,192	15,23	7,378	0,8585
Median	6,355	15,73	7,546	0,8752
Mean	6,350	15,74	7,548	0,8712
Kuartil 3	6,501	16,23	7,710	0,8863
Max	7,045	17,78	8,323	0,9221

Ilustrasi yang ditampilkan pada **Gambar 1** merupakan histogram dari metode *Support Vector Regression* dengan data uji 20%, terlihat pada histogram MAPE, MAE, dan RMSE cenderung simetris dan untuk histogram dari koefisien determinasi cenderung miring ke kiri.



Gambar 1. Histogram Perhitungan 1000 kali metode SVR dengan proporsi data uji 20%

Tabel 4 menunjukkan hasil perhitungan menggunakan metode support vector regression, dilakukan sebanyak 1000 kali iterasi dengan variasi proporsi data uji sebesar 10%; 20%; 30%; dan 40% dimana terlihat bahwa nilai MAE cenderung menurun seiring bertambahnya proporsi data uji dari 10% hingga 30%. Namun pada proporsi 40%, terjadi sedikit peningkatan nilai MAE, menunjukan adanya fluktuasi performa model pada proporsi data tertentu. Nilai MAPE dan nilai RMSE turun dari proporsi data uji 10% menjadi 20% namun setelah proporsi data uji 20% nilai MAPE dan nilai RMSE semakin besar yang mengindikasikan bahwa peningkatan jumlah data uji tidak selalu menjamin penurunan kesalahan prediksi *relative* dan absolut. Sebaliknya, nilai koefisien determinasi menunjukan peningkatan konsisten seiring bertambahnya proporsi data uji. Hal ini berarti model semakin mampu menjelaskan variasi data aktual dengan prediksi yang dihasilkan. Semakin tinggi nilai koefisien determinasi, maka semakin baik pula kemampuan model dalam menjelaskan hubungan antara variabel *input* dan *output* sehingga akurasi model dikatakan lebih baik.

Tabel 4. Hasil Perhitungan Support Vector Regression sebanyak 1000 kali

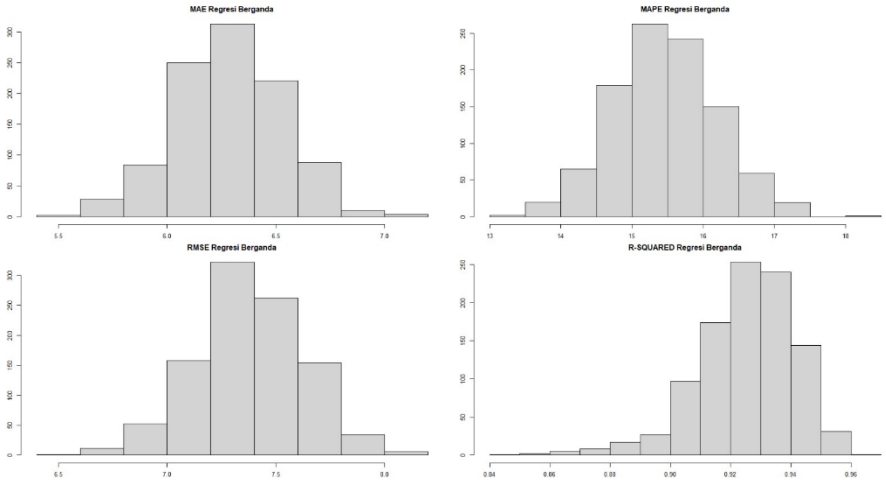
Proporsi Data Uji	MAE	MAPE	RMSE	Koefisien Determinasi
10%	6,364	15,77	7,563	0,8643
20%	6,350	15,74	7,548	0,8712
30%	6,356	15,77	7,557	0,8731
40%	6,371	15,81	7,568	0.8758

Regresi Linier Berganda

Pada penelitian ini digunakan metode regresi linier berganda sebagai pembanding metode *Support Vector Regression*. Penentuan nilai RMSE, MAE, MAPE, dan koefisien determinasi rata-rata mempergunakan 1000 kali perulangan dengan metode regresi linier berganda dan membagi menjadi 2 data yaitu 80% merupakan data latih dan 20% data uji. Hasil nilai-nilai MAE, MAPE, RMSE, dan koefisien determinasi terdapat pada **Tabel 5** dan hasil histogram dapat dilihat pada **Gambar 2.** Simulasi dilakukan dengan 1000 kali pengulangan dan data uji 10%, 20%, 30%, dan 40% dengan memprediksi peubah respon atau target data BMI (*Body Mass Index*). Diperoleh hasil mean RMSE, MAPE, MAE, dan koefisien determinasi yang dinyatakan pada **Tabel 6**. **Gambar 2** merupakan histogram dari Regresi Linear Berganda (RLB) dengan data uji 20% terlihat pada histogram MAPE dan MAE cenderung simetris dan untuk koefisien determinasi dan RMSE histogram cenderung miring ke kiri.

Tabel 5. Hasil Perhitungan 1000 kali metode Regresi Linier Berganda pada proporsi data uji 20%

Statistik	MAE	MAPE	RMSE	Koefisien Determinasi
Min	5,586	13,27	6,744	0,8743
Kuartil 1	6,124	14,98	7,238	0,9200
Median	6,279	15,43	7,374	0,9302
Mean	6,284	15,44	7,380	0,9288
Kuartil 3	6,441	15,89	7,526	0,9396
Max	6,930	17,55	8,093	0,9641



Gambar 2. Hasil Perhitungan 1000 kali metode Regresi Linier Berganda pada data uji 20%

Tabel 6 menyajikan hasil perhitungan menggunakan metode regresi linier berganda, dimana nilai MAE, MAPE, dan RMSE semakin besar proporsi uji datanya semakin besar nilainya. Nilai koefisien determinasi cenderung menurun ketika proporsi data uji semakin besar. Berdasarkan pada **Tabel 4** dan **Tabel 6** dapat dilakukan perbandingan metode *Support Vector Regression* dengan metode regresi linier berganda berdasarkan simulasi yang dilakukan 1000 kali menggunakan proporsi data uji 10%, 20%, 30%, dan 40% didapatkan hasil perbandingan menunjukkan nilai MAE pada metode SVR dengan proporsi 20% sebesar 6,350 dimana nilai tersebut lebih tinggi dibandingkan dengan nilai MAE regresi linier berganda yakni 6,284. Nilai MAPE pada Metode SVR dengan proporsi 20% sebesar 15,74% dimana nilai ini lebih besar dibandingkan MAPE regresi linier dengan proporsi 20% yakni 15,44%. Nilai RMSE pada metode SVR dengan proporsi 20% sebesar 7,548 dimana nilai ini lebih tinggi dibandingkan dengan nilai RMSE regresi linier berganda dengan proporsi 20% yakni 7,380. Sedangkan nilai koefisien determinasi metode SVR dengan proporsi 20% sebesar 0,8712 nilai ini lebih rendah dibandingkan dengan koefisien determinasi metode regresi linier berganda dengan proporsi 20% yakni 0,9288. Hal itu berarti bahwa, dalam kasus ini metode regresi linier berganda lebih baik digunakan daripada *support vector regression*.

Tabel 6. Hasil Perhitungan Regresi Linier Berganda sebanyak 1000 kali

Proporsi Data Uji	MAE	MAPE	RMSE	Koefisien Determinasi
10%	6,278	15,42	7,361	0,9302
20%	6,284	15,44	7,380	0,9288
30%	6,308	15,51	7,424	0,9199
40%	6,356	15,63	7,492	0,9058

Penelitian mengenai perbandingan kinerja metode SVR dengan Metode regresi linier berganda telah dilakukan (Delbari et al., 2019) yang menghasilkan bahwa metode *Support Vector Regression* lebih baik digunakan daripada metode regresi linier berganda untuk memperkirakan suhu diatas permukaan tanah. Berbeda dengan penelitian tersebut pada penelitian untuk memprediksi BMI lebih baik menggunakan metode regresi linier berganda dibandingkan dengan metode SVR.

Penelitian yang lain dilakukan oleh (Hoffman et al., 2016) yang berjudul "*Use of multivariate linear regression and support vector regression to predict functional outcome after surgery for cervical spondylotic myelopathy*". Hasil penelitian menjelaskan bahwa metode *Support Vector Regression* lebih akurat dari metode regresi linier berganda, dibuktikan dengan hasil koefisien determinasi, dimana koefisien determinasi pada SVR yaitu 0,932 lebih besar dari koefisien determinasi pada metode regresi linier berganda yaitu 0,452. Penelitian lain tentang perbandingan metode SVR dengan regresi linier berganda telah diteliti oleh (Obi, 2020) yang berjudul "*A Comparative Study of the Multiple Regression and Support Vector Regression*" menjelaskan bahwa metode *support vector regression* lebih baik dan lebih akurat dibandingkan dengan metode regresi linier berganda.

KESIMPULAN

Penelitian ini bertujuan untuk membandingkan keakuratan prediksi dua metode regresi yaitu *Support Vector regression* (SVR) dan metode regresi linear berganda. Pada kasus data ASTHMA ditunjukkan bahwa metode regresi linier berganda lebih baik dibandingkan dengan SVR. Hasil dari metode ini, nilai MAPE regresi linier berganda berturut-turut sebesar 15,42%; 15,44%; 15,51%; dan 15,63% sebanyak 1000 kali percobaan dengan proporsi data uji 10%, 20%, 30%, dan 40% sedangkan nilai MAPE untuk metode SVR sebesar 15,77%; 15,74%; 15,77%; dan 15,81% untuk 1000 kali percobaan. Penelitian yang telah dilakukan ini juga dapat dikembangkan dengan metode-metode lain dalam *machine learning*, seperti regresi logistik. Selain mengganti metode juga dapat dikembangkan dengan mengubah data yang digunakan.

DAFTAR PUSTAKA

Arumningsih, L. D., & Darsyah, M. Y. (2018). Peramalan Indeks Harga Konsumen Kota Malang Tahun 2014–2016 dengan Menggunakan Metode Moving Average dan Exponential Smoothing Holt-Winter. *Prosiding Seminar Nasional …*, *1*(January 2018), 310–315.

Basri, H. (2018). Pemodelan Regresi Berganda Untuk Data Dalam Studi Kecerdasan Emosional. *Didaktika Jurnal Kependidikan*, *12*(12), 103–116.

Chicco, D., Warrens, M. J., & Jurman, G. (2021). The coefficient ofdetermination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Computer Science*, *00*(00), 1–24. https://doi.org/10.7717/peerj-cs.623

Delbari, M., Sharifazari, S., & Mohammadi, E. (2019). Modeling daily soil temperature over diverse climate conditions in Iran—a comparison of multiple linear regression and support vector regression techniques. *Theoretical and Applied Climatology*, *135*(3–4), 991–1001. https://doi.org/10.1007/s00704-018-2370-3

Ginting, L. M., Sigiro, M. M., Manurung, E. D., & Sinurat, J. J. P. (2021). Perbandingan Metode Algoritma Support Vector Regression dan Multiple Linear Regression Untuk Memprediksi Stok Obat. *Journal of Applied Technology and Informatics Indonesia*, *1*(2), 29–34. https://doi.org/10.54074/jati.v1i2.36

Hoffman, H., Lee, S. I., Garst, J. H., Lu, D. S., Li, C. H., Nagasawa, D. T., Ghalehsari, N., Jahanforouz, N., Razaghy, M., Ghavamrezaii, A., Paak, B. H., Wu, I., Sarrafzadeh, M., Lu, C., Angeles, L., Angeles, L., Angeles, L., Angeles, L., Angeles, L., … Angeles, L. (2016). *Use of Multivariate Linear Regression and Support Vector Regression to Predict Functional Outcome After Surgery for Cervical Spondylotic Myelopathy*. *22*(9), 1444–1449. https://doi.org/10.1016/j.jocn.2015.04.002.Use

Masruroh, M., & Mauladi, K. F. (2020). Penerapan Metode Regresi Linear Berganda Dalam Sistem Prediksi Nilai Ujian Nasional Siswa Smp. *Jurnal Teknika*, *12*(1), 1. https://doi.org/10.30736/jt.v12i1.393

Mostafa, S. M., & Amano, H. (2019). Effect of clustering data in improving machine learning model accuracy. *Journal of Theoretical and Applied Information Technology*, *97*(21), 2973–2981.

Niu, W. J., Feng, Z. K., Feng, B. F., Min, Y. W., Cheng, C. T., & Zhou, J. Z. (2019). Comparison of multiple linear regression, artificial neural network, extreme learning machine, and support vector machine in deriving operation rule of hydropower reservoir. *Water (Switzerland)*, *11*(1). https://doi.org/10.3390/w11010088

Nurfadilah, A., Budi, W., Kurniati, E., Suhaedi, D., Matematika, P. S., Bandung, U. I., & Statistik, B. P. (2022). Penerapan Metode Moving Average untuk Prediksi Indeks Harga Konsumen. *… : Jurnal Teori Dan …*, *21*(1), 19–25.

Obi, J. C. (2020). A Comparative Study of the Multiple Regression and Support Vector Regression. *COOU Journal of Physical S Ciences 3(1), 3(1)*(May), 465–473.

Prakoso, B. H. (2019). Implementasi Support Vector Regression pada Prediksi Inflasi Indeks Harga Konsumen. *MATRIK : Jurnal Manajemen, Teknik Informatika Dan Rekayasa Komputer*, *19*(1), 155–162. https://doi.org/10.30812/matrik.v19i1.511

Purnama, D. I. (2020). Peramalan Jumlah Penumpang Datang Melalui Transportasi Udara Di Sulawesi Tengah Menggunakan Support Vector Regression (SVR). *Jurnal Ilmiah Matematika Dan Terapan*, *17*(1). https://doi.org/10.22487/2540766x.2020.v17.i1.15186

Putro, B., Furqon, M. T., & Wijoyo, S. H. (2018). Prediksi Jumlah Kebutuhan Pemakaian Air Menggunakan Metode Exponential Smoothing. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, *2*(11), 4679–4686.

Sivana Ihzaniah, L., Setiawan, A., & Wulan Wijaya, R. N. (2023). Perbandingan Kinerja Metode Regresi K-Nearest Neighbor Dan Metode Regresi Linear Berganda Pada Data Boston Housing. *JAMBURA Journal Of Probability And Statistics*, *4*(1), 18–29.

Suryanto, A. A. (2019). Penerapan Metode Mean Absolute Error (Mea) Dalam Algoritma Regresi Linear Untuk Prediksi Produksi Padi. *Saintekbu*, *11*(1), 78–83. https://doi.org/10.32764/saintekbu.v11i1.298

Yang, Q. (2017). *Regression* (L. Schintler A & C. McNeely L (eds.)). Springer. https://doi.org/10.1007/978-3-319-32001-4