

PERBANDINGAN MODEL REGRESI LINIER DENGAN LINEAR SUPPORT VECTOR REGRESSION DALAM MEMPREDIKSI KASUS STUNTING DI JAWA BARAT

Ryan Fadhilah Faizal Hakim¹

1. Universitas Koperasi Indonesia
Jl. Jatinangor KM. 20, 5, Cibeusi, Sumedang, Kabupaten Sumedang, Jawa Barat 45363
Email : ryanhakimikopin@gmail.com.

ABSTRAK

Stunting merupakan masalah kesehatan yang serius di Indonesia, khususnya di Jawa Barat, yang dapat berdampak pada pertumbuhan fisik dan perkembangan kognitif anak. Penelitian ini bertujuan untuk membandingkan keakuratan model regresi linier dengan metode Ordinary Least Squares (OLS) dan model Linear Support Vector Regression (Linear SVR) dalam memprediksi jumlah kasus stunting di Jawa Barat. Data yang digunakan adalah data sekunder dari Dinas Kesehatan Jawa Barat, yang mencakup jumlah balita stunting dan jumlah balita bergizi buruk dari tahun 2019 hingga 2023. Analisis dilakukan dengan menggunakan bahasa pemrograman Python dan beberapa library terkait. Hasil penelitian menunjukkan bahwa nilai Adjusted R-Squared (Adj. R-Squared) dari kedua model adalah sama, yaitu sebesar 0.795 atau 79.5%. Ini menunjukkan bahwa kedua model memiliki kemampuan yang setara dalam menjelaskan variabilitas data jumlah balita stunting berdasarkan variabel jumlah balita bergizi buruk. Model regresi linier memberikan interpretasi koefisien yang langsung, sedangkan model Linear SVR menawarkan fleksibilitas yang lebih besar dalam mengatur margin dan kesalahan. Kesimpulan dari penelitian ini adalah bahwa kedua model dapat digunakan untuk memprediksi jumlah kasus stunting dengan keakuratan yang sama, namun pemilihan model terbaik harus mempertimbangkan karakteristik data dan kebutuhan interpretasi.

Kata kunci: *Stunting, Regresi Linier, Linear SVR, Prediksi.*

1. PENDAHULUAN

Stunting atau kondisi gagal tumbuh kembang anak menjadi salah satu permasalahan yang masih dihadapi Indonesia. Kondisi ini menyebabkan masalah serius bagi pertumbuhan dan perkembangan dari segi fisik serta kemampuan kognitif pada anak. Terhambatnya tumbuh kembang anak dapat mengakibatkan penurunan kualitas sumber daya manusia yang terbentuk di masa mendatang. Berbagai upaya dilakukan pemerintah untuk menekan kasus *stunting* sebagai program nasional untuk mempersiapkan generasi emas 2045.

Masalah *stunting* yang berkepanjangan dapat menyebabkan penurunan kualitas pertumbuhan anak-anak seperti gangguan fisik dan perkembangan mental, kekebalan tubuh rendah, gangguan nutrisi dan kesehatan, prestasi akademik rendah, serta produktivitas dan ekonomi dalam jangka panjang.

Berdasarkan data Survei Status Gizi Indonesia (SSGI) Kementerian Kesehatan, prevalensi balita *stunting* di Jawa Barat mencapai 20,2% pada tahun 2022. Angka tersebut menurun sebesar 4,3% dari prevalensi tahun 2021 yaitu 24,5%. Prevalensi *stunting* di Jawa Barat tahun 2022 lebih rendah dari prevalensi *stunting* nasional yaitu 24,4%. Di tahun 2024, Pemerintah Jawa Barat berupaya menurunkan angka prevalensi *stunting* menjadi 14%.

Variabel yang berpengaruh terhadap *stunting* menurut Jurnal Ilmiah Kebidanan Universitas dr. Soebandi Jember, Vol. 9, No. 2 (2022) yaitu status gizi yang kurang. Pada penelitian ini dibuatkan model regresi linier dengan metode *Ordinary Least Squares* (OLS) dan model *Linear Support Vector Regression* dengan tujuan untuk memprediksi kasus *stunting* di Jawa Barat. Kedua model ini selain digunakan sebagai model prediksi, tetapi juga akan dilihat perbandingan dari keakuratan model tersebut dalam memprediksi.

2. TINJAUAN PUSTAKA

2.1. Stunting

Menurut Pasal 1 ayat (1) Peraturan Presiden Nomor 72 Tahun 2021 tentang Percepatan Penurunan, *stunting* adalah gangguan pertumbuhan dan perkembangan anak akibat kekurangan gizi kronis dan infeksi berulang, yang ditandai dengan panjang atau tinggi badannya berada di bawah standar yang ditetapkan oleh menteri yang menyelenggarakan urusan pemerintahan di bidang kesehatan.

2.2. Regresi

Analisis regresi merupakan suatu proses statistik untuk mengestimasi hubungan antara variabel-variabel, yakni berupa teknik-teknik memodelkan dan melakukan analisis beberapa variabel atas dasar hubungan antara satu variabel terikat (dependen) dan satu atau lebih variabel bebas (independen) (Armstrong, 2012:689).

Persamaan umum regresi adalah sebagai berikut:

$$\hat{Y} = a + bx \quad (1)$$

2.3. Ordinary Least Squares (OLS)

Ordinary Least Squares atau dikenal juga dengan *Least Squares Regression* adalah metode meminimalkan jumlah kuadrat residu yang digunakan untuk menyesuaikan (fit) model regresi.

Persamaannya adalah sebagai berikut:

$$RSS = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \quad (2)$$

$$RSS = \sum_{i=1}^n (Y_i - \hat{b}_0 - \hat{b}_1 X_i)^2 \quad (3)$$

2.4. R-Squared

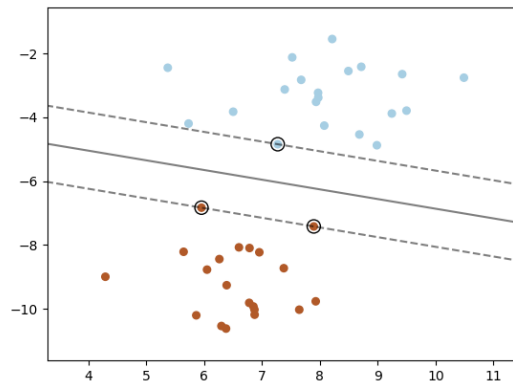
R-Squared (R^2) adalah sebuah ukuran statistik yang digunakan dalam regresi untuk mengevaluasi seberapa baik model yang dibuat dengan data observasi.

Persamaannya adalah sebagai berikut:

$$R^2 = 1 - \left(\frac{SS_{res}}{SS_{tot}} \right) \quad (4)$$

2.5. Support Vector Machine (SVM)

Support Vector Machine (SVM) adalah algoritma *supervised machine learning* yang mengklasifikasikan data dengan menemukan sebuah garis optimal atau *hyperplane* yang memaksimalkan jarak antara setiap kelas dalam ruang berdimensi N.



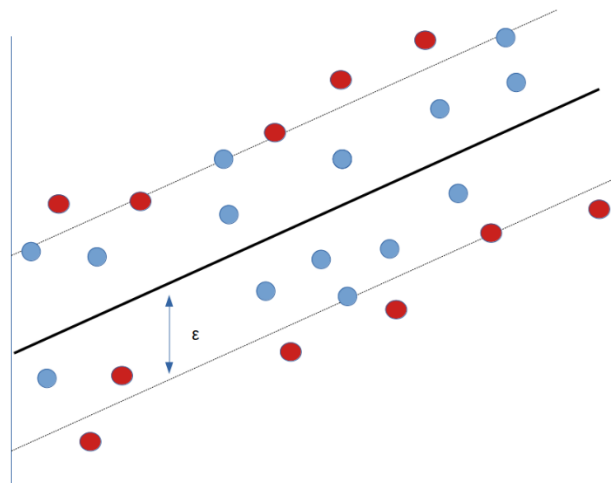
Gambar 1 Support Vector Machines

Persamaannya SVM adalah sebagai berikut:

$$\min \tau(W) = \frac{1}{2} \|W\|^2 \quad (5)$$

2.6. Support Vector Regression (SVR)

Support Vector Regression (SVR) adalah ekstensi dari SVM yang digunakan untuk permasalahan regresi. SVR menggunakan prinsip yang sama dengan SVM tetapi berfokus pada prediksi yang bersifat kontinu.



Gambar 2 Support Vector Regression

Persamaannya adalah sebagai berikut:

$$f(x) = \langle w, x \rangle + b \quad (6)$$

2.7. Linear Support Vector Regression (Linear SVR)

Linear Support Vector Regression (Linear SVR) adalah versi dari SVR yang hanya menggunakan kernel *linear*. Tujuan utama dari *Linear SVR* adalah untuk menemukan *hyperplane* yang meminimalkan deviasi (epsilon) dari data, yang disebut sebagai *epsilon-insensitive tube*.

Berikut merupakan persamaannya:

$$f(x) = \langle w, x \rangle + b \quad (7)$$

$$\min_{w,b,\xi,\xi^*} \frac{1}{2} \|W\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) \quad (8)$$

Terdapat beberapa parameter dalam model ini di antaranya:

Parameter C yang mengontrol *trade-off* antara margin dan kesalahan *training*. Parameter *Loss Function* digunakan untuk menggambarkan seberapa jauh prediksi model dari nilai sebenarnya. Dan parameter Epsilon digunakan untuk menentukan lebar rentang di sekitar *hyperplane* yang dibuat oleh model.

2.8. Grid Search

Grid Search adalah metode untuk menemukan kombinasi terbaik dari nilai parameter sehingga dapat menghasilkan performa model yang optimal.

3. METODE PENELITIAN

3.1. Sumber Data

Data yang digunakan dalam penelitian adalah data sekunder yang didapatkan dari Dinas Kesehatan Jawa Barat. Data sekunder yang digunakan yaitu data jumlah *stunting* berdasarkan kabupaten/kota di Jawa Barat tahun 2019 sampai 2023 dan data jumlah balita berdasarkan kategori balita gizi buruk di Jawa Barat tahun 2019 sampai 2023.

3.2. Variabel Penelitian

Variabel Penelitian yang digunakan dalam penelitian ini disajikan dalam Tabel 1.

Tabel 1.
Variabel Penelitian

Variabel	Keterangan	Penjelasan	Skala Data
Y	Jumlah Balita <i>Stunting</i>	Jumlah balita <i>stunting</i> yang akan diprediksi	Diskrit
X	Jumlah Balita Bergizi Buruk	Jumlah balita kekurangan gizi yang menjadi variabel independen	Diskrit

3.3. Alat Analisis

Dalam penelitian ini, bahasa pemrograman Python sebagai alat utama untuk mengolah data dan melakukan analisis yang dipadukan dengan penggunaan library seperti Pandas yang digunakan untuk mengolah data tabel, Numpy yang digunakan untuk perhitungan matematis, StatsModels yang digunakan untuk membuat model regresi linier, Matplotlib yang digunakan untuk membuat grafik, dan Scikit Learn yang digunakan untuk membuat model *Linear SVR*.

3.4. Langkah Analisis

1. Melakukan penggabungan *datasets* jumlah balita *stunting* tahun 2019 sampai 2023 dan jumlah balita bergizi buruk tahun 2019 sampai 2023 dari data Dinas Kesehatan Jawa Barat.
2. Membuat model regresi linier dengan jumlah balita *stunting* sebagai variabel dependen dan jumlah balita bergizi buruk sebagai variabel independen dengan metode *Ordinary Least Squared* (OLS).
3. Menguji koefisien determinasi (R^2) pada model regresi linier.
4. Melakukan visualisasi model regresi dengan *scatterplot*.
5. Melakukan *Grid Search* untuk menemukan parameter terbaik untuk model *Linear SVR*.
6. Membuat model *Linear SVR* dengan jumlah balita *stunting* sebagai variabel dependen dan jumlah balita bergizi buruk sebagai variabel Independen.
7. Menghitung koefisien determinasi (R^2) pada model *Linear SVR*.
8. Melakukan visualisasi model *Linear SVR* dengan *scatterplot*.
9. Membandingkan model regresi dengan model *Linear SVR*.
10. Menginterpretasikan kedua model yang telah dibuat.
11. Menarik kesimpulan dari kedua model yang telah dibuat.

4. HASIL PENELITIAN DAN PEMBAHASAN

Datasets yang telah didapatkan diolah dengan menggunakan bahasa pemrograman Python serta dilakukan penggabungan *datasets* dengan menggunakan library Pandas dan mengambil variabel jumlah balita *stunting* sebagai variabel dependen dan jumlah balita bergizi buruk sebagai variabel independen. Variabel tersebut digunakan untuk membuat model regresi linier dengan metode *Ordinary Least Squares* (OLS) yang tersedia pada library Statsmodels sehingga didapatkan hasil Tabel 2.

Tabel 2
Model Regresi Linier

Variabel	Koefisien
Konstanta	-78,7844
Jumlah Balita Bergizi Buruk (X)	1,6811

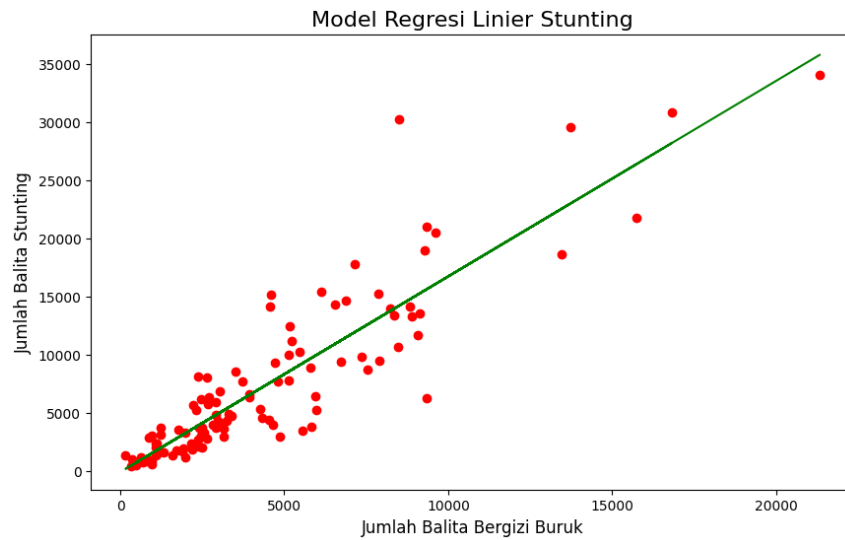
Dari Tabel 2 didapatkan model persamaan regresi sebagai berikut:

$$\hat{Y} = -78,7844 + 1,6811X \quad (9)$$

Tabel 3
Hasil R-Squared dari Model Regresi Linier

Sumber	R-Squared	Adj. R-Squared
Regresi	0,797	0,795

Berdasarkan Tabel 3, model regresi linier memperoleh nilai *R-Squared* sebesar 79,7% dengan nilai *Adj. R-Squared* (*R-Squared* yang disesuaikan) yaitu sebesar 79,5%. Dari nilai tersebut, model dapat menjelaskan prediksi *stunting* dengan persentase keakuratan sebesar 79,5%.



Gambar 3 Scatterplot Model Regresi Linier Stunting

Berdasarkan Gambar 3, *Scatterplot* menunjukkan bahwa terdapat hubungan positif antara jumlah balita bergizi buruk (X) dengan jumlah balita *stunting* (Y).

```
param_grid = {
    'C': [0.1, 1, 10, 100],
    'loss': ['epsilon_insensitive', 'squared_epsilon_insensitive'],
    'epsilon': [0.01, 0.1, 0.5, 1],
}

grid_search = GridSearchCV(model, param_grid, cv=5)
grid_search.fit(X, Y)

best_params = grid_search.best_params_
best_model = grid_search.best_estimator_
```

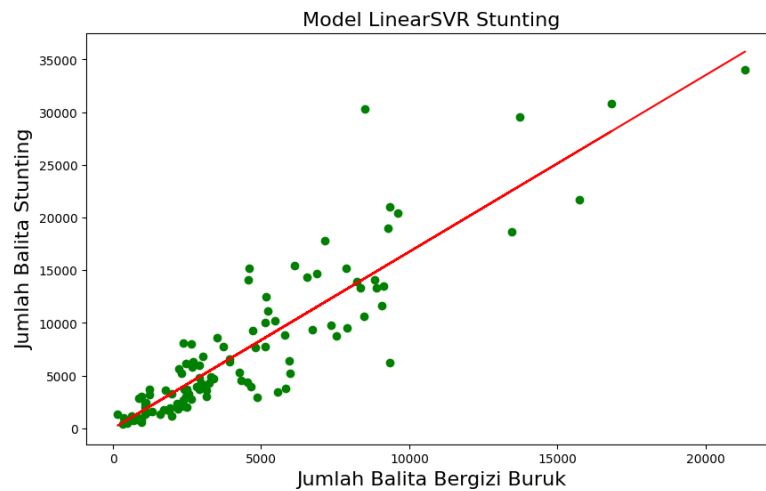
Gambar 4 Sintak Grid Search

Berdasarkan penggunaan sintak dalam Gambar 4, parameter terbaik yang didapatkan yaitu $C = 1$; $\text{loss} = \text{'epsilon_insensitive'}$ dan $\text{epsilon} = 0.01$.

```
model = LinearSVR(C=1, loss='epsilon_insensitive', epsilon=0.01)
model.fit(X, Y)
```

Gambar 5 Sintak Model Linear SVR

Pembuatan model Linear SVR dilakukan dengan sintak pada Gambar 5. Model dibuat menggunakan parameter terbaik yang didapatkan dari metode *Grid Search* pada Gambar 4.



Gambar 6 Scatterplot Model LinearSVR Stunting

Tabel 4
Hasil R-Squared dari Model Linear SVR

Sumber	R-Squared	Adj. R-Squared
Linear SVR	0,796	0,795

Berdasarkan Tabel 4, nilai *R-Squared* yang didapatkan dari model *Linear SVR* yaitu sebesar 79,6% dengan nilai dari *Adj. R-Squared* (*R-Squared* yang disesuaikan) yaitu sebesar 79,5%. Dengan ini, model dapat menjelaskan prediksi *stunting* yaitu 79,5%.

4.1. Interpretasi Perbandingan Model Regresi Linier dengan Model Linear SVR

Nilai *Adjusted R-Squared* (*Adj. R-Squared*) berdasarkan Tabel 3 dan Tabel 4 untuk kedua model sama yaitu sebesar 0.795 atau 79.5%. Hal ini menunjukkan bahwa kedua model, baik regresi linier maupun Linear SVR, memiliki kemampuan yang sama dalam menjelaskan variabilitas data jumlah balita *stunting* di Jawa Barat berdasarkan variabel jumlah balita bergizi buruk.

Model regresi linier memberikan persamaan 9, di mana konstanta sebesar -78,7844 dan koefisien untuk jumlah balita bergizi buruk (X) sebesar 1,6811 menunjukkan bahwa setiap peningkatan satu unit balita bergizi buruk akan meningkatkan jumlah balita *stunting* sebesar 1,6811 unit, dengan asumsi faktor lain tetap konstan.

Sebaliknya, model *Linear SVR* menggunakan prinsip yang berbeda dalam mengatur margin dan kesalahan. Model ini dirancang untuk menemukan *hyperplane* yang meminimalkan deviasi (*epsilon*) dari data yang disebut sebagai *epsilon-insensitive tube*. Hal ini membuat interpretasi koefisien menjadi tidak langsung seperti dalam regresi linier.

Kedua model memiliki nilai *Adj. R-Squared* yang sama, menunjukkan keandalan yang serupa dalam memprediksi jumlah balita *stunting* berdasarkan jumlah balita bergizi buruk. Namun, cara kedua model mencapai hasil tersebut berbeda. Regresi linier mengandalkan minimalisasi jumlah kuadrat residu, sedangkan *Linear SVR* mengoptimalkan margin dan deviasi *epsilon* untuk membuat prediksi.

Visualisasi *scatterplot* dari kedua model menunjukkan bahwa ada hubungan positif antara jumlah balita bergizi buruk dengan jumlah balita *stunting*. Pada *scatterplot* model regresi linier, hubungan ini digambarkan dengan garis regresi yang cenderung naik. Sementara itu, *scatterplot* model *Linear SVR* juga menunjukkan tren positif, tetapi dengan margin yang lebih fleksibel yang diatur oleh parameter *epsilon* dan C.

5. KESIMPULAN

Dari hasil analisis di atas, dapat disimpulkan bahwa kedua model memiliki kinerja yang sama baik dalam menjelaskan variabilitas data jumlah balita *stunting* di Jawa Barat. Namun, perbedaan utama terletak pada pendekatan yang digunakan oleh masing-masing model. Regresi linier menawarkan interpretasi koefisien yang langsung dan sederhana, sementara *Linear SVR* menawarkan fleksibilitas yang lebih besar dalam mengatur margin dan kesalahan, yang mungkin lebih cocok untuk data yang memiliki *noise* atau *outlier* yang signifikan.

Dengan demikian, pemilihan model terbaik harus mempertimbangkan karakteristik data dan kebutuhan interpretasi. Jika interpretasi langsung dari koefisien penting, regresi linier mungkin lebih cocok. Namun, jika fleksibilitas dalam mengatur margin dan mengatasi *outlier* lebih diutamakan, *Linear SVR* bisa menjadi pilihan yang lebih baik.

6. DAFTAR PUSTAKA

- [1]. Amirudin, M., & Wowor, A. D. (2023). Analisis Perbandingan Klasifikasi Balita Beresiko Stunting Menggunakan Metode Support Vector Machine dan Decision Tree. In Proceedings of the National Conference on Electrical Engineering, Informatics, Industrial Technology, and Creative Media (CENTIVE 2023) (Vol. 3, No. 1). Satya Wacana Christian University & Universitas Kristen Satya Wacana.
- [2]. Armstrong, Scott J. (2012). Illusion in Regression Analysis, International Journal Forecasting, Vol. 28, 689-693
- [3]. Andriyani, S. Y., Lydia, M. S., & Efendi, S. (2023). Optimization of Support Vector Machine Algorithm Using Stunting Data Classification. Prisma Sains Jurnal Pengkajian Ilmu dan Pembelajaran Matematika dan IPA IKIP Mataram, 11(1), 164.
- [4]. Badan Penelitian dan Pengembangan Kesehatan Kementerian Kesehatan Republik Indonesia. (2018). Riset Kesehatan Dasar (Riskesdas) 2018. Jakarta: Kementerian Kesehatan Republik Indonesia.
- [5]. Nababan, J. F. (2021). Klasifikasi penderita stunting dengan metode Support Vector Machine (Studi Kasus: Lima Puskesmas di Kota Bandar Lampung) (Skripsi). Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lampung, Bandar Lampung.
- [6]. Pratista, S., Nazir, A., Iskandar, I., Budianita, E., & Afrianty, I. (2023). Perbandingan Teknik Prediksi Pemakaian Obat Menggunakan Algoritma Simple Linear Regression dan Support Vector Regression. Sains dan Teknologi, Teknik Informatika, Universitas Islam Negeri Sultan Syarif Kasim, Pekanbaru, Indonesia.
- [7]. Romadhani, G. K. (2021). Sistem Prediksi Risiko Stunting Dengan Metode Support Vector Machine (SVM) Pada Anak Di Indonesia. Undergraduate thesis, Institut Teknologi Sepuluh Nopember.
- [8]. Syafika, V. A. N., & Karisma, R. D. L. N. (2023). On implementation of Support Vector Machine (SVM) in determining the classification of stunting-specific intervention index in Indonesia. Aplikasi Statistika, Seminar Nasional Official Statistics 2023, 1. Universitas Islam Negeri Maulana Malik Ibrahim.
- [9]. Yuningsih. (2022). Hubungan Status Gizi dengan Stunting pada Balita. Oksitosin: Jurnal Ilmiah Kebidanan, 9(2), 102-109. Program Studi Kebidanan, Fakultas Ilmu Kesehatan, Universitas dr. Soebandi Jember.