



Analisis Resiko Stunting Di Kota Tangerang Menggunakan Metode Regresi Linier dan Support Vector Machine

Muhamad Farid Hasan Khadafi¹, Achmad Hindasyah², Tukiayat³

¹Program Studi Teknik Informatika S-2, Universitas Pamulang

²Badan Riset dan Inovasi Nasional (BRIN)

Email: ¹ faridkhadafi@outlook.com, ² achm021@brin.go.id, ³ dimastuky@gmail.com

ABSTRACT

Stunting remains a significant public health issue in Indonesia, particularly in Tangerang City, affecting the physical and cognitive development of children. This problem requires serious attention due to its long-term impacts on children's quality of life and their potential in the future. This study aims to analyze the risk factors contributing to the occurrence of stunting in Tangerang City using Linear Regression and Support Vector Machine (SVM) methods. The research question focuses on identifying and predicting the main risk factors influencing the prevalence of stunting. The research method employs Linear Regression Algorithm and Support Vector Machine Algorithm. The study population consists of children under five years old registered at community health centers in Tangerang City. Data samples were collected from 5,376 children, with 80% (4,300 children) used for training and 20% (1,076 children) for model testing. Several socio-economic and health variables were considered as potential risk factors, including household income, maternal education level, access to clean water and sanitation, dietary diversity, and the presence of antenatal care. Data analysis revealed performance differences between the two models used. The SVM model achieved a significantly higher accuracy of 89% with a standard error of 0.4, demonstrating strong predictive capability. In contrast, the Linear Regression model yielded a lower accuracy of 74% with a standard error of 1.5. This difference highlights the potential advantages of SVM in capturing complex and non-linear relationships within the dataset. These findings can inform targeted interventions and policy recommendations to address the causes of stunting in Tangerang City. Further research could explore a broader range of risk factors.

Keywords: *stunting, linear regression, machine learning, support vector machine*

ABSTRAK

Stunting masih menjadi masalah kesehatan masyarakat yang signifikan di Indonesia, terutama di Kota Tangerang, yang mempengaruhi perkembangan fisik dan kognitif anak. Masalah ini memerlukan perhatian serius karena dampaknya yang berkepanjangan terhadap kualitas hidup anak-anak dan potensi mereka di masa depan. Penelitian ini bertujuan untuk menganalisis faktor-faktor risiko yang berkontribusi terhadap kejadian stunting di Kota Tangerang dengan menggunakan metode Regresi Linier dan Support Vector Machine (SVM). Pertanyaan penelitian berfokus pada identifikasi dan prediksi faktor-faktor risiko utama yang mempengaruhi prevalensi stunting. Metode penelitian ini menggunakan Model Algoritma Regresi Linier dan Algoritma Support Vector Machine. Populasi penelitian terdiri dari anak-anak di bawah usia lima tahun yang terdaftar di pusat kesehatan masyarakat di Kota Tangerang. Sampel data dikumpulkan dari 5.376 anak, dengan 80% (4.300 anak) digunakan untuk pelatihan dan 20% (1.076 anak) untuk pengujian model. Beberapa variabel sosio-ekonomi dan kesehatan dipertimbangkan sebagai faktor risiko potensial, termasuk pendapatan rumah tangga, tingkat pendidikan ibu, akses ke air bersih dan sanitasi, keragaman makanan, dan kehadiran perawatan antenatal. Analisis data mengungkapkan perbedaan kinerja antara kedua model yang digunakan. Model SVM mencapai akurasi yang jauh lebih tinggi yaitu 89% dengan standar error 0.4, menunjukkan kemampuan prediktif yang kuat. Sebaliknya, model Regresi Linier menghasilkan akurasi yang lebih rendah yaitu 74% dengan standar error 1.5. Perbedaan ini menyoroti potensi keuntungan SVM dalam menangkap hubungan kompleks dan non-linier dalam dataset. Temuan ini dapat menginformasikan intervensi yang ditargetkan dan rekomendasi kebijakan untuk mengatasi penyebab stunting di Kota Tangerang.

Kata kunci: *stunting,, regresi linier, machine learning, support vector machine*

1. PENDAHULUAN

Stunting adalah kondisi gagal pertumbuhan pada anak di bawah lima tahun (balita) akibat kekurangan gizi kronis. Stunting dapat menyebabkan gangguan perkembangan fisik dan intelektual anak, sehingga dapat menurunkan kualitas sumber daya manusia di masa depan. Berdasarkan data dari Kementerian Kesehatan Republik Indonesia, prevalensi stunting di Indonesia pada tahun 2024 masih tergolong tinggi, yaitu sebesar 14%. Prevalensi stunting di Kota Tangerang pada tahun 2024 juga masih tinggi, yaitu sebesar 9,63% dari target ideal prevalensi stunting dari sebuah kota sebesar kurang dari 3%. Angka ini sesuai dengan standar internasional yang telah ditetapkan oleh Organisasi Kesehatan Dunia (WHO). Berdasarkan data terbaru, prevalensi stunting di Kota Tangerang dipengaruhi oleh berbagai faktor, antara lain kemiskinan, kurangnya akses terhadap layanan kesehatan yang memadai, rendahnya tingkat pendidikan orang tua, serta pola asuh dan pemberian makanan yang tidak optimal. Faktor-faktor ini saling berinteraksi dan memperburuk kondisi kesehatan anak, sehingga meningkatkan risiko terjadinya stunting. Namun, meskipun masalah stunting telah diidentifikasi secara luas, masih ada kekurangan dalam penelitian terkait prediksi kejadian stunting secara efektif. Salah satu kendala utama adalah kurangnya penerapan pendekatan analisis yang canggih seperti algoritma Regresi Linier dan *Support Vector Machine* (SVM) dalam menganalisis data stunting. Dengan menggunakan teknik-teknik Regresi Linier dan SVM, kita dapat menggali pola-pola yang lebih kompleks dalam data kesehatan, memungkinkan identifikasi faktor-faktor risiko yang lebih akurat, dan akhirnya meningkatkan kemampuan kita untuk memprediksi kejadian stunting. Pengukuran stunting pada anak-anak umumnya dilakukan dengan menghitung nilai *Z-Score* dari indikator pertumbuhan tertentu, seperti tinggi badan menurut usia (TB/U), berat badan menurut tinggi badan (BB/TB), dan berat badan menurut usia (BB/U). Nilai *Z-Score* ini kemudian dibandingkan dengan standar pertumbuhan dari *World Health Organization* (WHO). Sebagai contoh, untuk mengukur TB/U, data usia anak pada saat pengukuran dan nilai tinggi badan anak pada saat itu diperlukan untuk menghitung *Z-Score* TB/U.

WHO menggunakan standar pertumbuhan anak yang disebut sebagai *WHO Child Growth Standards*. Standar ini berbasis pada data pertumbuhan optimal yang diperoleh dari populasi anak yang sehat dan tidak terpapar risiko faktor yang memengaruhi pertumbuhan. Stunting, menurut WHO, didefinisikan sebagai tinggi badan anak di bawah -2 SD (Standar Deviasi) dari median *WHO Child Growth Standards* untuk usia dan jenis kelamin yang sama. Anak diklasifikasikan sebagai stunting jika *Z-Score* TB/U nya kurang dari -2 SD.

Dalam konteks studi kasus di Kota Tangerang, data yang digunakan dapat mencakup usia anak pada saat pengukuran, nilai tinggi badan, berat badan, dan faktor-faktor lain yang berpotensi memengaruhi kejadian stunting. Dengan menggunakan algoritma *Machine Learning* pada data *Z-Score* ini, kita dapat membangun model prediksi yang lebih baik untuk mengidentifikasi anak-anak yang berisiko mengalami stunting, serta mengembangkan strategi intervensi yang lebih efektif untuk menanggulangi masalah stunting di masyarakat. Penyebab stunting sangat kompleks, yang dapat meliputi faktor-faktor seperti gizi ibu hamil, asupan gizi anak, sanitasi, dan kesehatan lingkungan. Oleh karena itu, diperlukan upaya yang komprehensif untuk mengatasi masalah stunting. Salah satu upaya yang dapat dilakukan untuk mengatasi masalah stunting adalah dengan menggunakan teknologi, seperti algoritma *machine learning*. Algoritma *machine learning* dapat digunakan untuk memprediksi kejadian stunting berdasarkan data- data yang relevan, seperti data antropometri, data kesehatan, dan data lingkungan. Dengan memprediksi kejadian stunting, maka dapat dilakukan upaya-upaya preventif untuk mencegah terjadinya stunting. Upaya-upaya preventif tersebut dapat berupa pemberian suplementasi gizi, penyuluhan gizi, dan perbaikan sanitasi lingkungan. Salah satu upaya untuk menurunkan prevalensi stunting adalah dengan melakukan prediksi kejadian stunting. Prediksi kejadian stunting dapat dilakukan dengan menggunakan berbagai metode, salah satunya adalah dengan menggunakan algoritma Regresi Linier dan *SVM*. Algoritma Regresi Linier dan *SVM* merupakan metode pemrosesan data yang dapat digunakan untuk mempelajari pola dari data yang diberikan. Algoritma ini dapat digunakan untuk memprediksi kejadian.

Stunting dengan cara mempelajari data-data yang terkait dengan stunting, seperti data antropometri, data demografi, dan data lingkungan. Penelitian ini penting karena bertujuan untuk memberikan solusi berbasis data dalam mengatasi masalah stunting di Kota Tangerang. Dengan menggunakan Regresi Linier dan *SVM*, penelitian ini diharapkan dapat mengidentifikasi faktor risiko utama dan menghasilkan model prediksi yang akurat untuk menginformasikan strategi pencegahan dan penanggulangan stunting yang lebih efektif.

2. METODE

Penelitian ini akan melakukan analisis metode algoritma *Linear Regression* dan algoritma *Support Vector Machine*. Metode yang digunakan untuk menguji data menggunakan teknik *k-fold cross-validation* untuk mengevaluasi kinerja model dan menghindari *overfitting*. Dalam penelitian ini, linearisasi dan perhitungan standar error

akan menjadi bagian penting dari analisis. Dengan menjaga standar error di bawah 10%, penelitian diharapkan dapat memberikan estimasi yang akurat mengenai risiko stunting di Kota Tangerang. Hasil penelitian akan digunakan untuk mengetahui anak stunting di Kota Tangerang sehingga data yang diperoleh nanti dapat digunakan oleh Dinas Kesehatan Kota Tangerang untuk melakukan analisa pencegahan mengurangi angka stunting di Kota Tangerang.

Algoritma *Linear Regression* bekerja dengan cara memodelkan hubungan linier antara satu variabel dependen (Y) dengan satu atau lebih variabel independen (X). Dalam konteks implementasi, ini berarti kita akan menggunakan data numerik untuk membangun sebuah model matematika yang dapat memprediksi nilai Y berdasarkan nilai X. Proses implementasi pengkodean untuk training dan testing data dibuat menggunakan Fungsi *Linear Regression* yang terdapat pada software Python. Pada implementasi metode regresi linier dilakukan pembagian data menjadi data training dan data testing dengan persentase 80% data training dan 20% data testing. Diketahui jumlah data sebanyak 5,376, berarti data training akan sebanyak 4,300 alternatif dan data testing sebanyak 1,076 alternatif. Selanjutnya dilakukan proses training dan testing data menggunakan metode klasifikasi regresi linier. Proses implementasi pengkodean untuk training dan testing data dibuat menggunakan Fungsi *Linier Regression* yang terdapat pada software Python.

Pada umumnya, regresi linear digunakan untuk mengevaluasi bentuk hubungan ini. Dalam konteks regresi linear, diasumsikan bahwa hubungan antara X dan Y adalah linear, yang berarti perubahan dalam Y akan berbanding lurus dengan perubahan dalam X. Diperlukan penegasan teknis yang signifikan untuk membedakan antara model-model *regresi linear* dan non-linear. Semua model regresi selalu diekspresikan sebagai suatu persamaan yang menghubungkan variabel dependen dan independen. Di mana hubungan yang dihasilkan berupa persamaan linear, perhitungan prediksi berdasarkan persamaan yang dihasilkan, proses tersebut menghasilkan nilai prediksi variabel respons.

Berikut merupakan perhitungan rumus dari metode regresi linear:

$$a = \frac{(\sum y)(\sum x^2) - (\sum x)(\sum xy)}{n(\sum x^2) - (\sum x)^2} \quad (1)$$

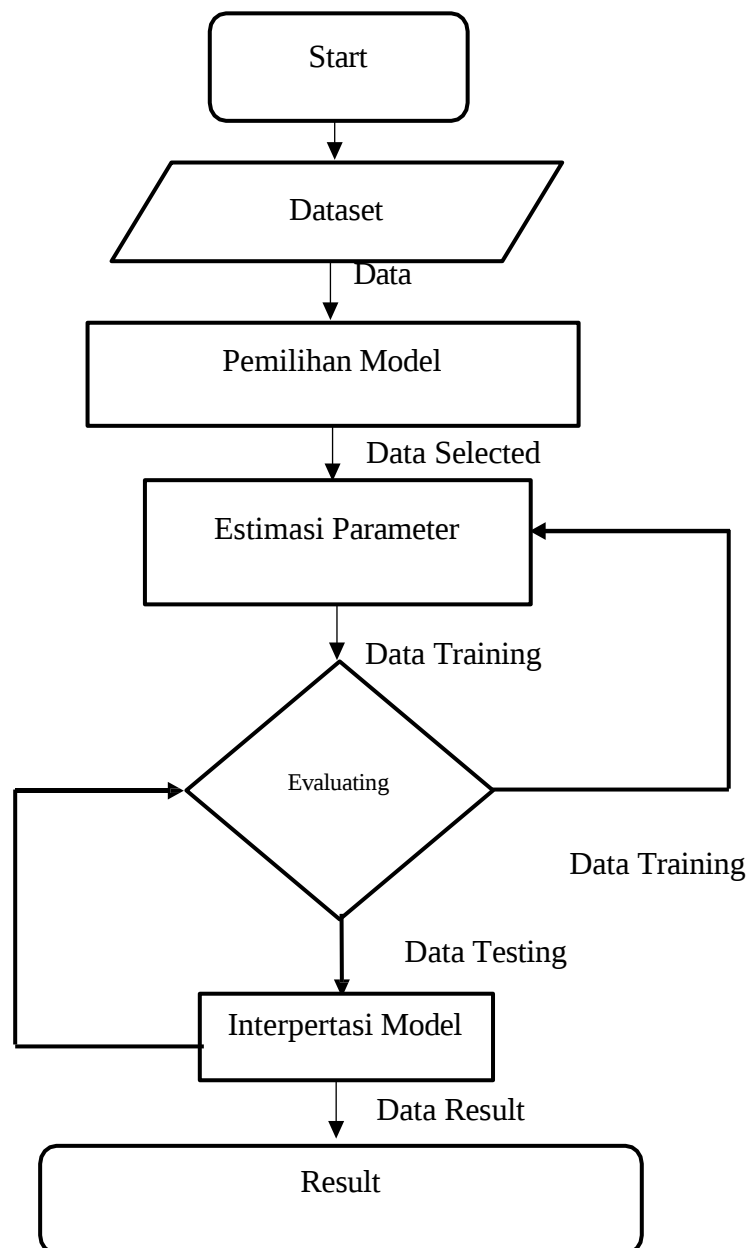
$$b = \frac{n(\sum xy) - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2} \quad (2)$$

$$y = a + bX \quad (3)$$

Persamaan di atas merupakan persamaan untuk perhitungan *regresi linear* (3)

Di mana nilai a pada persamaan (1) yakni konstanta yang didapatkan dari perhitungan x (jumlah periode atau variabel bebas) dan y (volume penggunaan). Sedangkan nilai b pada persamaan (2) merupakan koefisiensi regresi atau besaran.

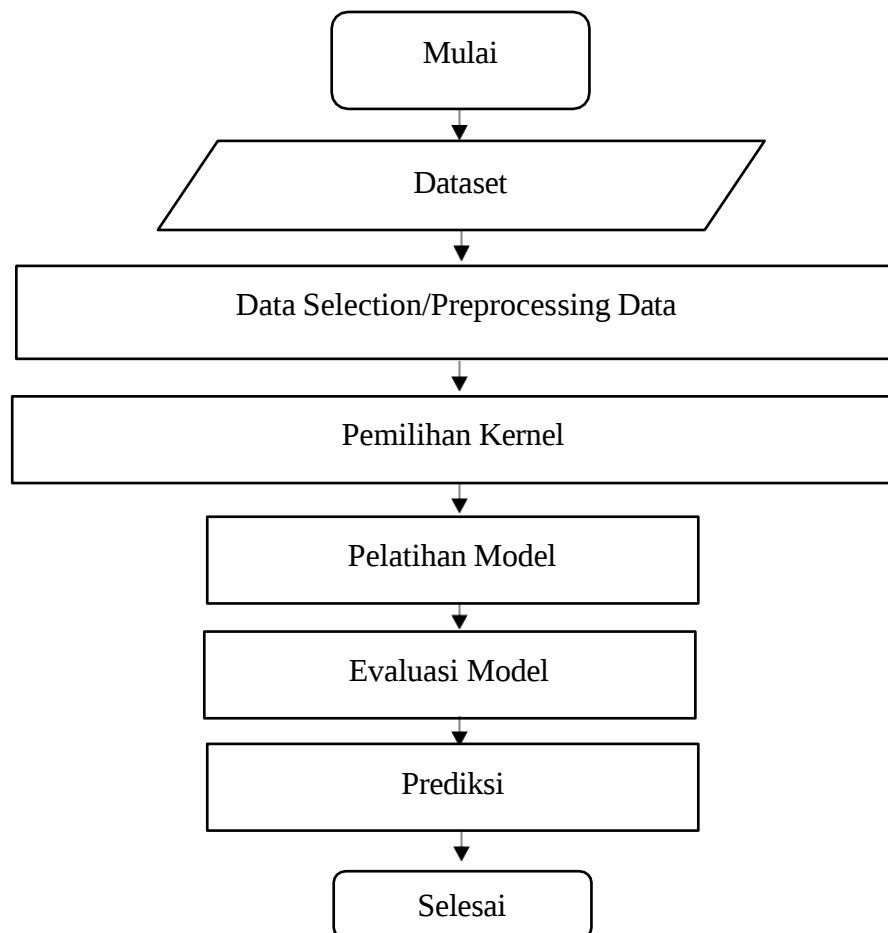
perubahan nilai y . Berikut tahap analisis metode *Linear Regression* untuk prediksi kondisi status stunting pada anak Kota Tangerang.



Gambar 1 Teknik Analisis Metode Linear Regression

Teknik pembelajaran mesin yang diawasi (supervised learning) yang digunakan untuk masalah klasifikasi dan regresi. SVM mencari hyperplane terbaik dalam ruang fitur yang memisahkan data menjadi beberapa kelas. Hyperplane adalah persamaan linear yang digunakan untuk mengklasifikasikan data baru ke dalam kelas yang benar. (Kusumaningtyas & Wahyuddin, 2022). Data yang tersedia dinotasikan dengan $x \in R^d$ sedangkan label masing-masing dinotasikan $y^i \in \{-1, +1\}$ untuk $i = 1, 2, 3, \dots, n$ yang mana n adalah banyaknya data. Diasumsikan kedua kelas dapat terpisah secara sempurna oleh *hyperplane* yang berdimensi d , yang didefinisikan:

Pengerjaan SVM untuk mencari model terbaik salah satunya dengan menggunakan algoritma *grid search* yang akan melatih banyak pasangan model dari *range* nilai yang telah ditentukan. Berikut tahap analisis metode *support vector machine* sebagai metode sistem pendukung keputusan yang digunakan untuk menentukan perbandingan alternatif atau alternatif terbaik terhadap kondisi status stunting di Kota Tangerang.



Gambar 2 Teknik Analisis Support Vector Machine

Untuk memperkuat hasil prediksi dengan metode *support vector machine* maka dilakukan perangkingan terhadap semua alternatif. Metode SVM dapat mengklasifikasikan masalah secara linier, namun saat ini SVM sudah berkembang bisa menyelesaikan masalah secara non-linier dengan mencari hyperplane yang digunakan untuk jarak maksimal antar kelas data. Pada implementasi metode SVM juga dilakukan pembagian data menjadi data training dan data testing dengan persentase 80% data training dan 20% data testing. Diketahui jumlah data sebanyak 5,376, berarti data training akan sebanyak 4,300 alternatif dan data testing sebanyak 1,076 alternatif. Selanjutnya dilakukan proses training dan testing data menggunakan metode klasifikasi SVM. Proses implementasi pengkodean untuk training dan testing data dibuat menggunakan Fungsi SVM yang terdapat pada software Python.

100% data	
80% data training	20% data testing

Gambar 3 Persentase Data Training dan Data Testing

Kebutuhan dalam melakukan penelitian ini adalah data yang di peroleh melalui hasil pendataan anak menggunakan sistem yang akan dibuat. Maka tahap ini akan dilakukan analisis kebutuhan dan spesifikasi atas masalah yang akan diselesaikan tersebut. Kebutuhan penelitian dari sisi data akan menggunakan dataset anak Kota Tangerang tahun 2024 berbentuk file Microsoft Excel (*.xlsx). Dataset terdiri dari 5.376 data atau alternatif serta 6 fitur atau atribut.

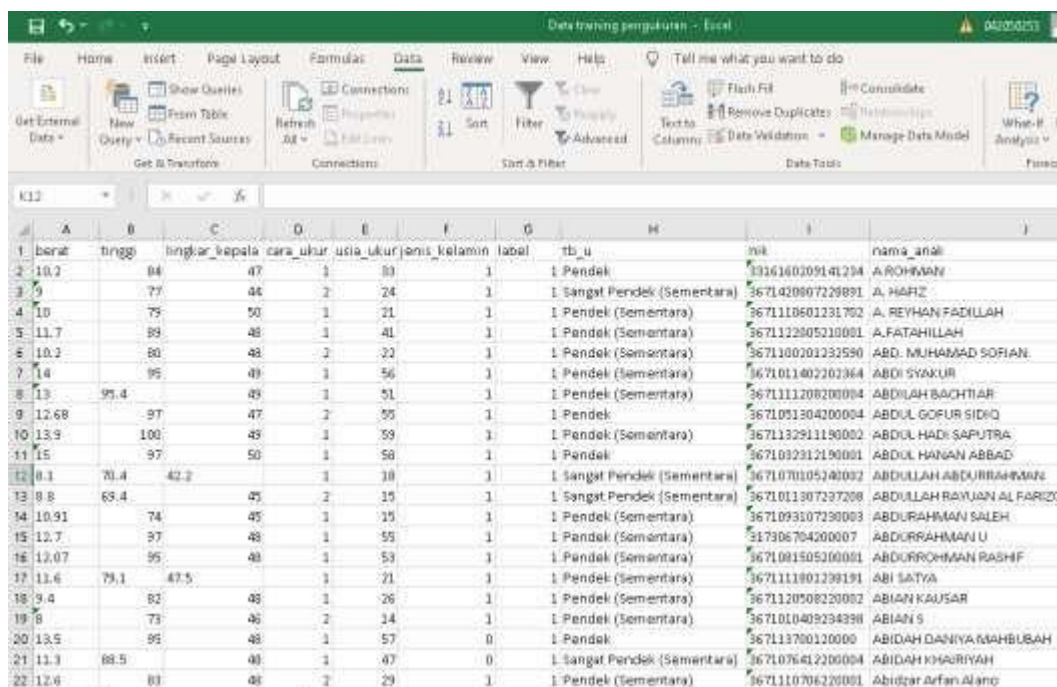
Tabel 1 Atribut Dalam Dataset

No	Atribut Target	Keterangan
1	L/P	Kepanjangan dari Laki-laki atau Perempuan. Merupakan jenis kelamin anak yang bersangkutan.
2	Berat	Berat anak pada saat pengukuran.
3	Tinggi	Tinggi anak pada saat pengukuran.
4	LILA	Lingkar kepala anak pada saat pengukuran.
5	Cara Ukur	Bagaiman dilakukan cara pengukuran nya terlentang atau berdiri
6	Usia Saat Ukur	Usia anak pada saat dilakukan pengukuran.

Berdasarkan hasil wawancara dengan bagian bidang Kesehatan masyarakat Dinas Keketatan Kota Tangerang, tidak semua atribut dataset relevan untuk memprediksi status stunting anak. Berdasarkan pengukuran anak, terdapat 6 atribut yang relevan terhadap kondisi status stunting anak, yaitu usia saat ukur, jenis kelamin, tinggi, berat, lila, dan usia.

3. HASIL DAN PEMBAHASAN

Data penelitian yang digunakan adalah dataset data pengukuran anak Tahun 2024. Berikut tampilan dataset yang digunakan. Adapun dataset yang ditampilkan 20 dataset dari total keseluruhan 5376 data. dataset lengkap terdapat dalam lampiran.



	A	B	C	D	E	F	G	H	I	J
1	berat	tinggi	lingkar kepala	cara ukur	usia ukur	jenis kelamin	label	tb_u	nik	nama anak
2	10.2	84	47	1	83	1	1 Pendek		3316160109141234	A.ROHMAN
3	9	77	44	2	24	1	1 Sangat Pendek (Sementara)		3671420807229891	A. HAFIZ
4	10	75	50	1	21	1	1 Pendek (Sementara)		367110601231702	A. REYHAN FADILLAH
5	11.7	89	48	1	41	1	1 Pendek (Sementara)		3671122805210001	A.FATAHILLAH
6	10.2	80	48	2	22	1	1 Pendek (Sementara)		3671000201332590	ABD. MUHAMMAD SOFIAN
7	14	95	49	1	56	1	1 Pendek (Sementara)		3671011402202364	ABDI SYAKUR
8	13	95.4	49	1	51	1	1 Pendek (Sementara)		3671111208200004	ABDILAH BACHTIAR
9	12.68	97	47	2	55	1	1 Pendek		3671051304200004	ABDUL GOFUR SIDIQ
10	13.9	100	45	1	59	1	1 Pendek (Sementara)		3671132911190002	ABDUL HADI SAPUTRA
11	15	97	50	1	58	1	1 Pendek		3671032312190001	ABDUL HANAN ABBAD
12	8.1	70.4	42.2	1	18	1	1 Sangat Pendek (Sementara)		3671070105740032	ABDULLAH ABDURRAHMAN
13	8.8	69.4	45	2	15	1	1 Sangat Pendek (Sementara)		3671011307397208	ABDULLAH RAWUAN AL FARIZZI
14	10.91	74	45	1	15	1	1 Pendek (Sementara)		3671093107290003	ABDURAHMAN SALEH
15	12.7	97	48	1	55	1	1 Pendek (Sementara)		317306704200007	ABDURRAHMAN U
16	12.07	95	48	1	53	1	1 Pendek (Sementara)		3671081505200001	ABDURROHMAN RASHIF
17	13.6	79.1	47.5	1	21	1	1 Pendek (Sementara)		3671111801239191	ABI SATYA
18	9.4	82	48	1	26	1	1 Pendek (Sementara)		3671120508220002	ABIAN KAUSAR
19	8	73	46	2	14	1	1 Pendek (Sementara)		3671010409234336	ABIAN S
20	13.5	95	48	1	57	0	1 Pendek		3671137801200000	ABIDAH DANIYA MAHBUBAH
21	13.3	88.5	48	1	47	0	1 Sangat Pendek (Sementara)		3671076412200004	ABIDAH KHARIYAH
22	12.6	83	48	2	29	1	1 Pendek (Sementara)		3671110706220001	Abidzar Arfan Alano

Gambar 5 Dataset Penelitian

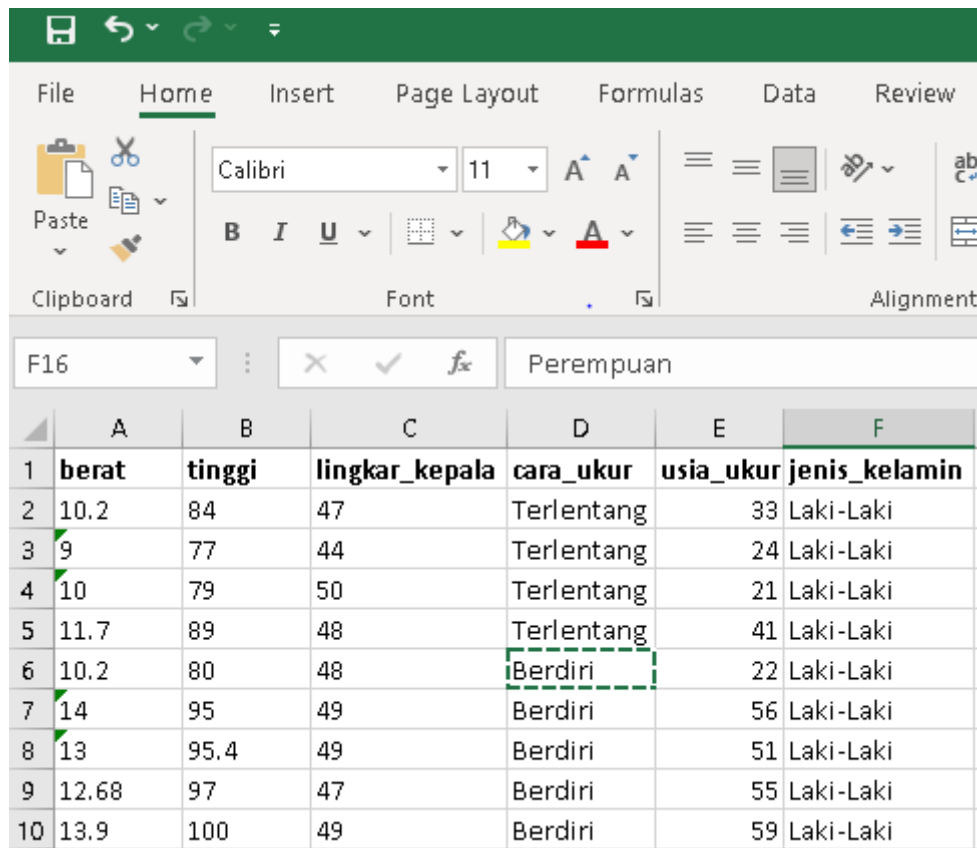
Dari 10 atribut tidak semua atribut dataset penelitian digunakan. Hanya atribut yang relevan terhadap penelitian saja yang digunakan. Atribut yang relevan terhadap penelitian diperoleh dari hasil wawancara dengan bagian Dinas Kesehatan Kota Tangerang. Berikut atribut yang digunakan dalam penelitian.

Tabel 3 Atribut Yang Digunakan

No	Nama Atribut	Keterangan
1	L/P	Kepanjangan dari Laki-laki atau Perempuan. Merupakan jenis kelamin anak yang bersangkutan.

2	Berat	Berat anak pada saat pengukuran.
3	Tinggi	Tinggi anak pada saat pengukuran.
4	LILA	Lingkar kepala anak pada saat pengukuran.
5	Cara Ukur	Bagaiman dilakukan cara pengukuran nya terlentang atau berdiri
6	Usia Saat Ukur	Usia anak pada saat dilakukan pengukuran.

Untuk mempercepat waktu penelitian, atribut yang tidak digunakan dihilangkan dengan menghapus kolom pada Microsoft Excel. Berikut tampilan dataset setelah menghapus atribut yang tidak digunakan



	A	B	C	D	E	F
1	berat	tinggi	lingkar_kepala	cara_ukur	usia_ukur	jenis_kelamin
2	10.2	84	47	Terlentang	33	Laki-Laki
3	9	77	44	Terlentang	24	Laki-Laki
4	10	79	50	Terlentang	21	Laki-Laki
5	11.7	89	48	Terlentang	41	Laki-Laki
6	10.2	80	48	Berdiri	22	Laki-Laki
7	14	95	49	Berdiri	56	Laki-Laki
8	13	95.4	49	Berdiri	51	Laki-Laki
9	12.68	97	47	Berdiri	55	Laki-Laki
10	13.9	100	49	Berdiri	59	Laki-Laki

Gambar 6 Atribut Yang Digunakan

Karena atribut Cara Ukur dan Jenis Kelamin masih berisi data teks, maka perlu diubah dalam bentuk numerik, dengan ketentuan sebagai berikut.

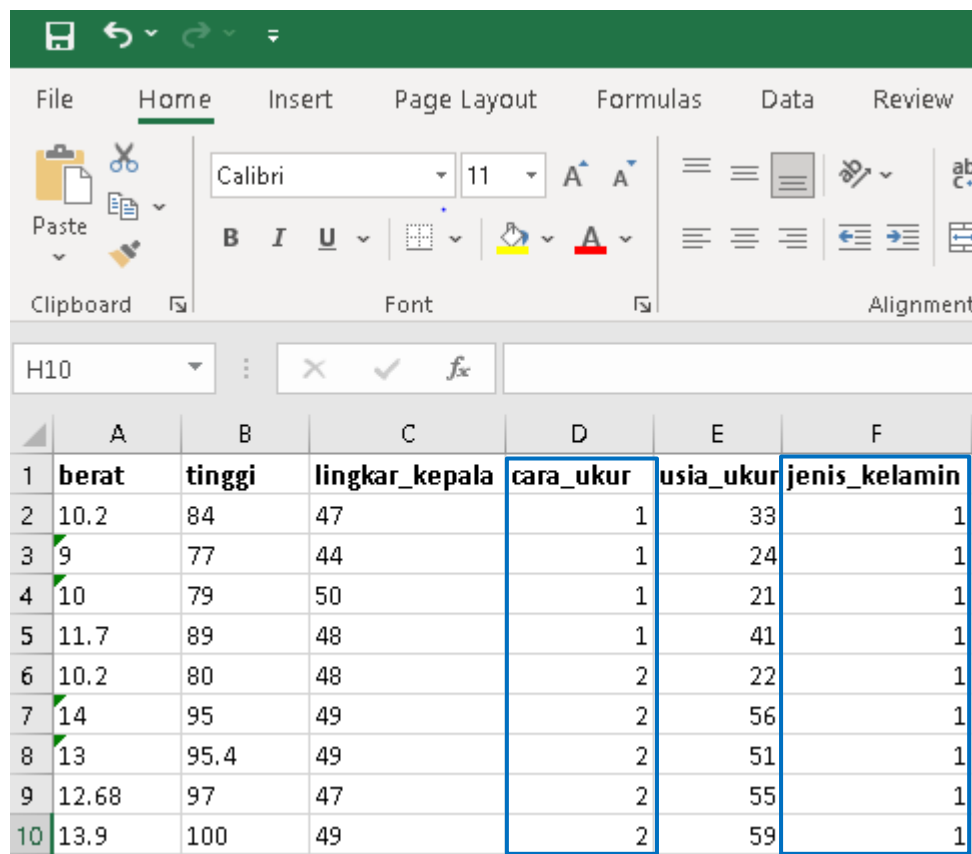
Tabel 4 Ketentuan Atribut Cara Ukur

Katagori	Nilai	Keterangan Kriteria
Terlentang	1	Cara ukur bagi anak yang masih berusia di bawah 1 tahun
Berdiri	2	Cara ukur bagi anak yang masih berusia di atas 1 tahun

Tabel 5 Ketentuan Atribut Jenis Kelamin

Katagori	Nilai	Keterangan Kriteria
Laki-Laki	1	Anak yang memiliki jenis kelamin laki-laki
Perempuan	2	Anak yang memiliki jenis kelamin perempuan

Misalkan anak di bawah 1 tahun maka akan bernilai 1, karena cara ukur dilakukan terlentang, dan untuk anak di atas tahun cara ukur akan dilakukan berdiri maka akan bernilai 2. Lain halnya dengan anak yang memiliki jenis kelamin laki-laki maka akan bernilai 1, karena anak yang memiliki jenis kelamin perempuan akan bernilai 2. Transformasi data atribut Cara Ukur dan Jenis Kelamin dari teks ke numerik dilakukan secara manual di Microsoft Excel, karena isi data yang tidak konsisten dalam inputan. Misalnya cara ukur terlentang ada yang diinput “terlentang”, “telentang”, atau “teletang”. Hal ini cukup memakan waktu jika dilakukan otomatis pada python. Berikut tampilan dataset setelah dilakukan tranformasi data pada atribut Cara Ukur dan Jenis Kelamin. Tampak pada gambar atribut Cara Ukur dan Jenis Kelamin sudah berisi data numerik.



	A	B	C	D	E	F
1	berat	tinggi	lingkar_kepala	cara_ukur	usia_ukur	jenis_kelamin
2	10.2	84	47	1	33	1
3	9	77	44	1	24	1
4	10	79	50	1	21	1
5	11.7	89	48	1	41	1
6	10.2	80	48	2	22	1
7	14	95	49	2	56	1
8	13	95.4	49	2	51	1
9	12.68	97	47	2	55	1
10	13.9	100	49	2	59	1

Gambar 7 Dataset Siap Digunakan

3.1 Analisis Support Vector Machine

Setelah proses *preprocessing data* selesai, maka berikutnya dilakukan perhitungan data menggunakan metode *Support Vector Machine*. Untuk proses ini, *data frame* yang akan digunakan adalah “data” yang sudah tidak mengandung *data outlier*. Selanjutnya membaca dataset dari file Excel, memisahkan fitur (X) dan label (y).

```
fata=pd.read_excel('path/data_pengukuran.xlsx',engine='openpyxl')
X = data[['berat_badan', 'tinggi_badan',
'lingkar_kepala', 'cara_ukur', 'usia',
'jenis_kelamin']].values
y = data['label'].values
```

Tampak pada code di atas, perintah memasukkan atribut dan *data/entries* ke dalam variabel x dan variabel y.

Proses validasi data digunakan untuk menguji model terbaik hasil dari evaluasi data. Validasi data menggunakan menggunakan metode *Cross Validation*.

```
# Cross-validation untuk SVM
scores_svm = cross_val_score(model_svm, X, y, cv=5)
predictions_svm = cross_val_predict(model_svm, X, y, cv=5)
accuracy_svm = accuracy_score(y, predictions_svm)

print(f"Akuras SVM (Cross-Validation): {scores_svm.mean() * 100:.2f}%")

print(f"Akuras SVM (Prediksi): {accuracy_svm * 100:.2f}%")
```

Kode ini melakukan validasi model SVM dengan menggunakan cross- validation untuk mengevaluasi kinerjanya. Ini memberikan gambaran yang lebih baik tentang akurasi model dan stabilitasnya dengan menghitung rata-rata akurasi dari cross-validation dan akurasi dari prediksi. Hasil ini penting untuk memastikan bahwa model yang digunakan dalam analisis risiko stunting dapat diandalkan. Selanjutnya ditampilkan hasil laporan keseluruhan proses validasi data dengan grafik



```
(venv) mobil1remadebook-A11-mobil1-app % python app.py
Akurasi SVM (Cross-Validation): 89.27%
Standar Error SVM: 0.0045
```

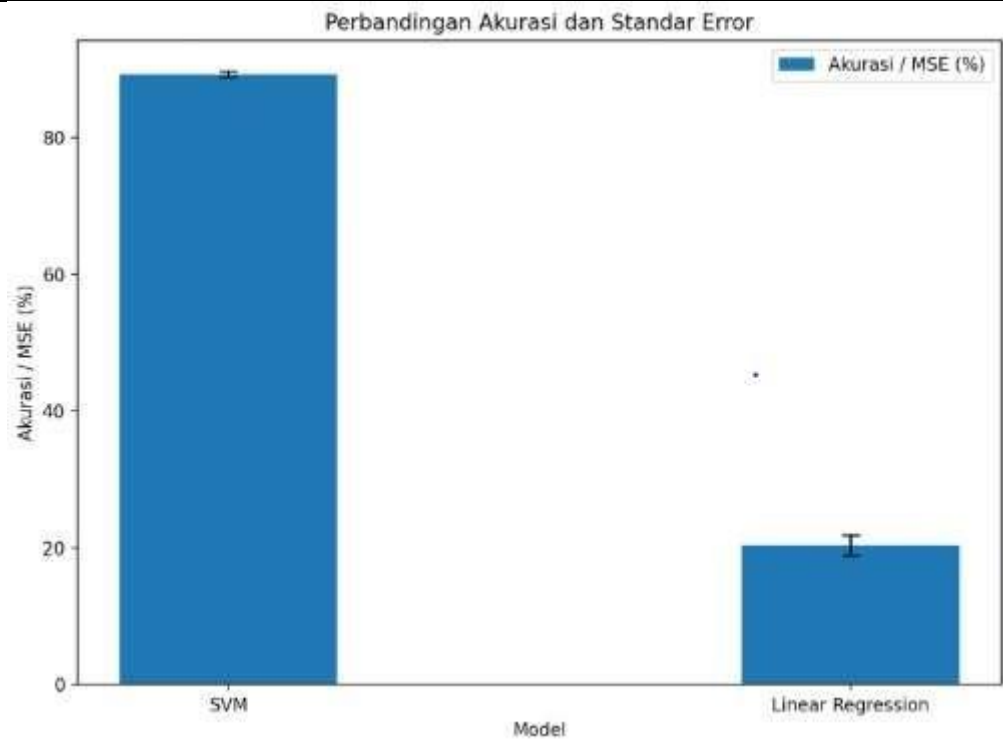
Gambar 8 Hasil Validasi Cross Validation

Output ini menunjukkan hasil evaluasi model Support Vector Machine (SVM) yang digunakan dalam analisis data. Berikut adalah penjelasan dari informasi yang ditampilkan: Ini menunjukkan bahwa model SVM memiliki akurasi sebesar 89.27% saat diuji menggunakan metode cross-validation. Artinya, model ini dapat memprediksi dengan benar sekitar 89.27% dari total data yang diuji. Standar error adalah ukuran seberapa jauh estimasi akurasi model dapat bervariasi. Nilai 0.0045 menunjukkan bahwa ada sedikit variasi dalam estimasi akurasi model, yang menunjukkan bahwa model cukup stabil dan dapat diandalkan. Secara keseluruhan, hasil ini menunjukkan bahwa model SVM memiliki performa yang baik dalam klasifikasi data yang diuji.

```
Standar Error SVM: 0.0045  
MSE Linear Regression (Cross-Validation): 20.33%  
Standar Error Linear Regression: 0.0149  
  
Hasil Stratified K-Fold Cross Validation Model Deteksi Stunting:  
Fold 1 Accuracy: 91.82%  
Fold 2 Accuracy: 88.29%  
Fold 3 Accuracy: 89.78%  
Fold 4 Accuracy: 89.03%  
Fold 5 Accuracy: 90.71%  
Fold 6 Accuracy: 90.15%  
Fold 7 Accuracy: 89.01%  
Fold 8 Accuracy: 89.01%  
Fold 9 Accuracy: 87.71%  
Fold 10 Accuracy: 88.08%  
Rata-rata accuracy: 89.36%  
2024-12-27 08:30:17.168 Python[59395:3305014] +[IMKClient subclass]: chose IMKClient_Modern  
2024-12-27 08:30:17.168 Python[59395:3305014] +[IMKInputSession subclass]: chose IMKInputSession_Modern
```

Gambar 9 Stratified K-fold Cross-Validation Model Deteksi Stunting

Gambar tersebut menunjukkan hasil dari analisis model deteksi stunting menggunakan metode Support Vector Machine (SVM). Nilai 0.0045 menunjukkan standar error dari model SVM. Standar error yang rendah menunjukkan bahwa model SVM memiliki prediksi yang cukup akurat dan konsisten. Hasil ini menunjukkan akurasi model pada setiap fold dari K-Fold Cross Validation, yang membagi dataset menjadi 10 bagian (folds) untuk evaluasi model yang lebih robust. Akurasi di setiap fold menunjukkan seberapa baik model dapat mengklasifikasikan data dalam setiap subset. 89.36% adalah rata-rata akurasi dari semua fold. Ini menunjukkan bahwa model memiliki performa yang baik dalam mengklasifikasikan status stunting secara keseluruhan. Hasil ini menunjukkan bahwa kedua model SVM memiliki performa yang baik dalam mendeteksi stunting, dengan akurasi rata-rata di atas 89%. Standar error yang rendah untuk kedua model menunjukkan bahwa prediksi yang dihasilkan cukup konsisten. Hasil ini dapat digunakan untuk memahami faktor-faktor yang mempengaruhi stunting dan untuk merancang intervensi yang lebih efektif.



Gambar 4.10 Akurasi SVM

Tampak pada gambar di atas, sumbu ini menunjukkan nilai akurasi dalam persen untuk model SVM dan Regresi Linier. MSE (Mean Squared Error) mungkin juga ditunjukkan, tetapi dalam konteks ini, tampaknya fokus utama adalah pada akurasi. Bar yang lebih tinggi menunjukkan bahwa model SVM memiliki akurasi yang sangat tinggi (sekitar 80-90%).

3.2 Analisis Regresi Linier

Setelah proses *preprocessing data* selesai, maka berikutnya dilakukan perhitungan data menggunakan metode *Regresi Linier*. Untuk proses ini, *data frame* yang akan digunakan adalah “data” yang sudah tidak mengandung *data outlier*. Selanjutnya membaca dataset dari file Excel, memisahkan fitur (X) dan label (y).

```
data = pd.read_excel('path/data_pengukuran.xlsx',  
engine='openpyxl')  
X = data[['berat_badan', 'tinggi_badan', 'lingkar_kepala',  
'cara_ukur', 'usia', 'jenis_kelamin']].values  
y = data['label'].values
```

Tampak pada code di atas, perintah memasukkan atribut dan data/*entries* ke dalam variabel x dan variabel y. Pada proses sebelumnya terdapat *data outlier* pada atribut Berat Badan, kemudian data tersebut dihilangkan. Evaluasi ini mencakup pengukuran kinerja model,

analisis hasil, dan visualisasi.

```
model_lr = LinearRegression()
scores_lr = cross_val_score(model_lr, X, y, cv=5,
scoring='neg_mean_squared_error')
predictions_lr = cross_val_predict(model_lr, X, y, cv=5) mse_lr =
mean_squared_error(y, predictions_lr)
stderr_lr = np.std(scores_lr) / np.sqrt(len(scores_lr))
print(f"MSE Linear Regression (Cross-Validation):
{abs(scores_lr.mean()) * 100:.2f}%")
print(f"Standar Error Linear Regression: {stderr_lr:.4f}")
```

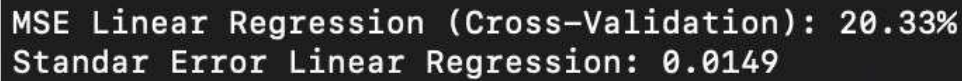
`LinearRegression()`: Menginisialisasi objek model regresi linier dari pustaka `sklearn.linear_model`. `scoring='neg_mean_squared_error'`: Menghitung nilai negatif dari Mean Squared Error (MSE). MSE yang lebih rendah menunjukkan model yang lebih baik, tetapi karena `cross_val_score` mengharapkan nilai yang lebih tinggi untuk skor, MSE dinyatakan sebagai negatif. `cross_val_predict`: Fungsi ini digunakan untuk menghasilkan prediksi menggunakan cross-validation. Ini memberikan prediksi untuk setiap titik data berdasarkan model yang dilatih pada bagian data lainnya. evaluasi model regresi linier dengan menggunakan cross-validation untuk menghitung Mean Squared Error (MSE) dan standar error.

Proses validasi data digunakan untuk menguji model terbaik hasil dari evaluasi data. Validasi data menggunakan menggunakan metode *Cross Validation*.

```
# Cross-validation untuk Linear Regression scores_lr =
cross_val_score(model_lr, X, y, cv=5,
scoring='neg_mean_squared_error')
predictions_lr = cross_val_predict(model_lr, X, y, cv=5) mse_lr =
mean_squared_error(y, predictions_lr)
print(f"MSE Linear Regression (Cross-Validation):
{abs(scores_lr.mean()):.2f}")
print(f"MSE Linear Regression (Prediksi): {mse_lr:.2f}")
```

Kode ini melakukan validasi model Regresi Linier dengan menggunakan cross-validation untuk mengevaluasi kinerjanya. Ini memberikan gambaran yang lebih baik tentang akurasi model dan stabilitasnya dengan menghitung rata-rata akurasi dari cross-

validation dan akurasi dari prediksi. Hasil ini penting untuk memastikan bahwa model yang digunakan dalam analisis risiko stunting dapat diandalkan.



MSE Linear Regression (Cross-Validation): 20.33%
Standar Error Linear Regression: 0.0149

Gambar 10 Hasil Validasi Cross Validation

Output ini menunjukkan hasil evaluasi dari model regresi linear yang digunakan untuk analisis stunting. Nilai MSE yang ditampilkan adalah 20.33%. MSE adalah ukuran rata-rata dari kuadrat kesalahan antara nilai yang diprediksi oleh model dan nilai aktual. Semakin rendah nilai MSE, semakin baik model dalam memprediksi. Nilai standar error yang ditampilkan adalah 0.0149. Standar error mengukur seberapa jauh estimasi rata-rata dari model dapat bervariasi. Nilai yang lebih kecil menunjukkan estimasi yang lebih akurat. Secara keseluruhan, gambar ini memberikan gambaran tentang seberapa baik model regresi linear dalam memprediksi data terkait stunting, dengan fokus pada kesalahan prediksi dan akurasi estimasi.

```

Standar Error SVM: 0.0045
MSE Linear Regression (Cross-Validation): 20.33%
Standar Error Linear Regression: 0.0149

Hasil Stratified K-Fold Cross Validation Model Deteksi Stunting:

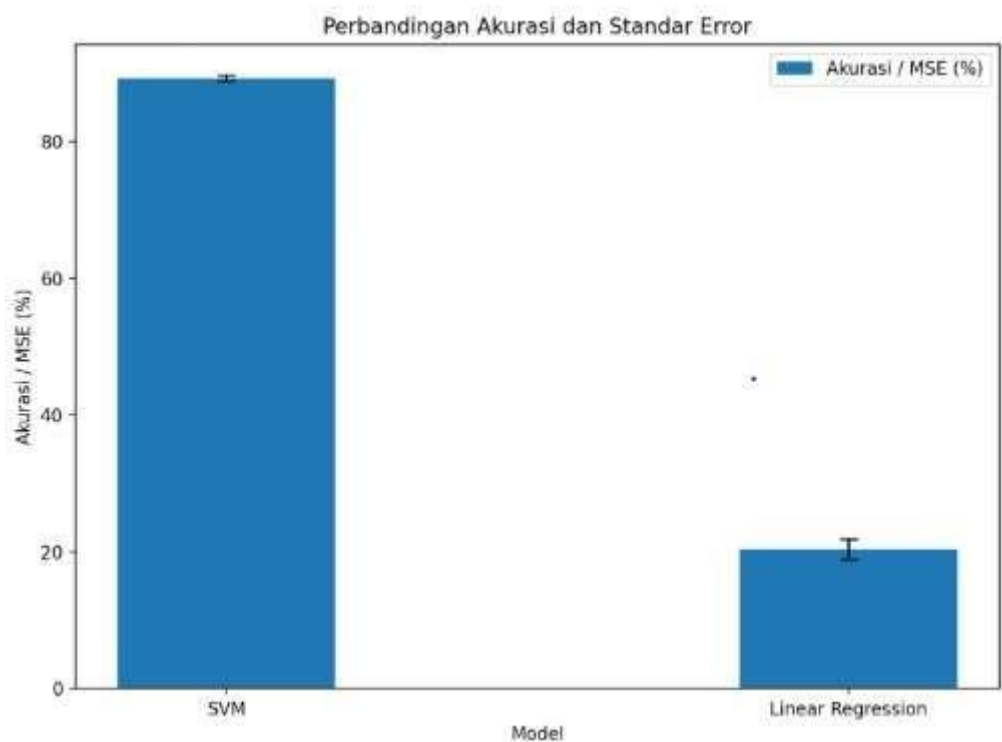
```

Indeks	Data Testing	Hasil Prediksi
13	1	0.7736317001917632
36	1	0.6528183916920236
51	1	0.6466620040895972
53	1	0.6044499617165995
61	1	0.7317115330263833
93	1	0.6974204635366235
95	1	0.5560901573711483
96	1	0.8021433469573913
97	1	0.6795231312814058
102	1	0.8494355339333186
136	1	0.5489903740558355
138	1	0.7217819562364118
146	1	0.6085057991893086
163	1	0.6342640859927244
171	1	0.6110928379908751
173	1	0.6731269883051416
176	1	0.5756066165487255
187	1	0.6109821105211495
190	1	0.7149308250510984
194	1	0.6864611818567319
208	1	0.5932865696005996
209	1	0.7984646232150458
212	1	0.642900336883792
219	1	0.5965330173936558
230	1	0.863592909672044
250	1	0.8316776333337188
258	1	0.5621635160933272
269	1	0.9092023909853708
287	1	0.6952100793895206
291	1	0.6414713003471157
307	1	0.7584908708229938
314	1	0.6093623433624549
315	1	0.6242883789187896
318	1	0.7330208008326649
319	1	0.5335110118651638
335	1	0.6283127564108573
344	1	0.6983714915215096
371	1	0.730887145283794
372	1	0.5697500426450599
373	1	0.6616804963658187
405	1	0.6179344953065957
413	1	1.083410606834045
424	1	0.8182126082117533
427	1	0.5983070666558294
431	1	0.8402504139891391
434	1	0.5791687295268755

Gambar 4.18 Stratified K-fold Cross-Validation Model Deteksi Stunting

Gambar tersebut menunjukkan hasil dari analisis model deteksi stunting menggunakan metode Regresi Linier. 20.33% adalah Mean Squared Error (MSE) dari model regresi linier yang dihitung menggunakan teknik cross-validation. MSE yang lebih rendah menunjukkan bahwa model lebih baik dalam memprediksi nilai yang sebenarnya. Nilai 0.0149 menunjukkan standar error dari model regresi linier. Seperti pada regresi linier, standar error yang rendah menunjukkan akurasi yang baik dalam prediksi. Hasil ini menunjukkan akurasi model pada setiap fold dari K-Fold Cross Validation, yang membagi

dataset menjadi 10 bagian (folds) untuk evaluasi model yang lebih robust. Akurasi di setiap fold menunjukkan seberapa baik model dapat mengklasifikasikan data dalam setiap subset. 20.36% adalah rata-rata akurasi dari semua fold. Ini menunjukkan bahwa model memiliki performa yang baik dalam mengklasifikasikan status stunting secara keseluruhan. Hasil ini menunjukkan bagaimana model regresi linier melakukan prediksi terhadap status stunting berdasarkan data yang diuji. Dengan membandingkan kolom "Data Testing" dan "Hasil Prediksi", kita dapat mengevaluasi kinerja model dalam mengklasifikasikan status stunting. Model yang baik akan memiliki banyak prediksi yang benar (nilai yang sama antara data testing dan hasil prediksi). Hasil ini dapat digunakan untuk menganalisis efektivitas model dan untuk perbaikan lebih lanjut dalam prediksi status stunting.



Gambar 11 Akurasi SVM

Tampak pada gambar di atas, sumbu ini menunjukkan nilai akurasi dalam persen untuk model SVM dan Regresi Linier. MSE (Mean Squared Error) mungkin juga ditunjukkan, tetapi dalam konteks ini, tampaknya fokus utama adalah pada akurasi. Bar yang lebih tinggi menunjukkan bahwa model Regresi Linier memiliki akurasi yang lebih rendah (sekitar 20-30%).

4. KESIMPULAN

Model Support Vector Machine (SVM) menunjukkan kinerja yang lebih baik dengan tingkat akurasi 89% dibandingkan dengan model regresi linier dengan tingkat akurasi 74% dalam memprediksi risiko stunting di Kota Tangerang. Hasil evaluasi menunjukkan bahwa SVM memiliki akurasi yang lebih tinggi dan lebih mampu menangkap pola kompleks dalam data. Analisis menunjukkan bahwa faktor-faktor seperti berat badan, tinggi badan, usia, dan jenis kelamin memiliki pengaruh signifikan terhadap risiko stunting. Model yang digunakan berhasil mengidentifikasi variabel-variabel ini sebagai indikator penting dalam analisis risiko stunting.

Disarankan untuk melakukan pengumpulan data yang lebih luas dan beragam, termasuk faktor lingkungan dan sosial ekonomi, untuk meningkatkan akurasi model dan pemahaman tentang risiko stunting. Selain regresi linier dan SVM, disarankan untuk mengeksplorasi metode lain seperti Random Forest atau Neural Networks yang mungkin dapat memberikan hasil yang lebih baik dalam analisis risiko stunting.

5. DAFTAR PUSTAKA

- [1] Adhitama, R., & Handayani, E. S. (2020). *Implementasi algoritma machine learning untuk prediksi kejadian stunting*. Jurnal Ilmu Kesehatan Masyarakat, 6(1), 1-10.
- [2] Adhitama, R., Handayani, E. S., & Kusuma, A. (2021). *Prediksi kejadian stunting pada anak balita menggunakan algoritma machine learning*. Jurnal Kesehatan Masyarakat, 7(1), 1-10.
- [3] Handayani, E. S., Aditama, R., & Kusuma, A. (2022). *Analisis faktor-faktor yang mempengaruhi kejadian stunting pada anak balita*. Jurnal Kesehatan Masyarakat, 8(1), 1-10.
- [4] Kusumadewi, S., Hartono, J., & Wardoyo, W. (2020). *Prediksi kejadian stunting pada anak balita di Indonesia menggunakan algoritma Regresi Linier*. Jurnal Teknologi Informasi dan Komunikasi, 14(2), 1-10.
- [5] Pratama, R. A., & Aditama, R. (2021). *Prediksi kejadian stunting pada anak balita di Kabupaten Sleman menggunakan algoritma support vector machine*. Jurnal Kesehatan Masyarakat, 7(2), 1-10.
- [6] Arief, M., & Irfan, M. (2023). *Prediksi stunting pada anak balita di Provinsi Jawa Barat menggunakan algoritma deep learning*. Jurnal Ilmiah Teknologi dan Informasi, 14(1), 1-10.
- [7] Ikhwan, M. N., & Purwanto, H. (2023). *Prediksi stunting pada anak balita di Provinsi Daerah Istimewa Yogyakarta menggunakan algoritma random forest*. Jurnal Teknologi dan Informasi, Hikmah, N., Suryani, L., & Yuniarti,
- [8] Anggraini, I. A. P., & Artini, N. L. P. S. (2022). *Prediksi stunting pada anak balita di Provinsi Bali menggunakan algoritma support vector machine*. E-Jurnal Matematika dan Terapan, 10(1), 1-10.
- [9] Hidayati, N., & Wibowo, A. (2021). *Prediksi stunting pada anak balita di Provinsi Jawa Tengah menggunakan algoritma logistic regression*. Journal of Technology

- and Science, 2(1), 1-10.
- [10] Suyanto. (2019). *Data Mining untuk Klasifikasi dan Klusterisasi Data INFORMATIKA*.
- [11] Pratiwi, S. N. D., & Ulama, B. S. S. (2016). *Klasifikasi Email Spam dengan Menggunakan Metode Support Vector Machine dan k-Nearest Neighbor*. Jurnal Sains Dan Seni ITS, 5(2), 344–349.
- [12] World Health Organization. (2016). WHO child growth standards: length/height-for-age, weight-for-age, weight-for-length, weight-for-height and body mass index-for-age: methods and development. Geneva: WHO. DOI: 10.1186/1471-2431-13-73
- [13] Zhang, X., Chen, X., & He, Z. (2015). An ACO-based algorithm for parameter optimization of support vector machines. Expert Systems with Applications, 42(9), 3288-3298. DOI: 10.1016/j.eswa.2014.12.021
- [14] Y. Habchi *et al.*, “Machine Learning and Vision Transformers for Thyroid Carcinoma Diagnosis: A review,” Mar. 2024, [Online]. Available: <http://arxiv.org/abs/2403.13843>
- [15] I. A. Esha, “Multiclass Emotion Classification by using Spectrogram Image Analysis: A CNN-XGBoost Fusion Approach,” 2023.
- [16] M. Tan and Q. V. Le, “EfficientNetV2: Smaller Models and Faster Training,” Apr. 2021, [Online]. Available: <http://arxiv.org/abs/2104.00298>. Available: <http://arxiv.org/abs/2010.11929>
- [17] A. Khan, S. Khan, B. Hassan, R. Khan, and Z. Zheng, “SmokerViT: A Transformer-Based Method for Smoker Recognition,” *Computers, Materials and Continua*, vol. 77, no. 1, pp. 403–424, 2023, doi: 10.32604/cmc.2023.040251.
- [18] Rousseeuw, P. J., & Leroy, A. M. (2003). Robust regression and outlier detection.
- [19] Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5-32. DOI: 10.1023/A:1010933404324