

Analysis of Stunting Prevalence in Indonesia: A Comparative Study Using K-Means Clustering, Support Vector Machine, and *Decision Tree* Algorithms

Analisis Prevalensi Stunting di Indonesia: Studi Perbandingan Menggunakan Metode K-Means Clustering, Support Vector Machine, dan Algoritma *Decision Tree*

Albertus Christian Wahyu, Evangeline Suciadi, Justin Stephen, Nurfajriah Oktaviani,
Rosa Aulia Wangi,

¹²³⁴⁵ Information System, Faculty of Engineering & Information
Universitas Multimedia Nusantara, Tangerang Banten 15810, Indonesia.

¹Albertus.Christian@student.umn.ac.id,

Abstrak – Stunting, kondisi gagal pertumbuhan pada anak akibat kekurangan gizi selama 1000 hari pertama kehidupan, menjadi ancaman serius bagi Indonesia dengan prevalensi 21,6% pada tahun 2022 dan target penurunan sebesar 14% dalam RPJMN 2020-2024. Penelitian ini menghadapi kompleksitas stunting dengan menerapkan metode machine learning inovatif, termasuk K-Means Clustering, Support Vector Machine (SVM), dan Decision Tree. Faktor prediktor dipilih berdasarkan kelompok sasaran program penurunan stunting dan disesuaikan untuk mencerminkan pendekatan multisektoral. Data yang digunakan berasal dari Badan Pusat Statistik dan SSGI 2022. Hasil analisis menunjukkan variabel "BBLR" berkorelasi positif signifikan dengan prevalensi stunting, sementara variabel seperti "Persalinan FASYANKES," "K4," "KF Lengkap," "mCPR," dan "A 6-11" berkorelasi negatif signifikan. K-Means Clustering mengidentifikasi pola regional, memberikan wawasan untuk intervensi yang lebih terfokus, namun tanpa perbedaan yang nyata antara kluster. SVM mencapai tingkat akurasi 88%, terutama dengan kernel linear, sementara Decision Tree mencatat keakuratan sempurna 100%, sehingga pada penelitian ini model *Decision Tree* memberikan performa yang lebih baik.

Kata Kunci: *Decision Tree, Indonesia, K-means, Stunting, Support Vector Machine*

I. PENDAHULUAN

A. Latar Belakang dan Tujuan

Stunting, yang merupakan kondisi gagal pertumbuhan pada anak akibat kekurangan gizi selama 1000 hari pertama kehidupan, menjadi isu kritis yang mengancam masa depan Indonesia [1]. Dampak jangka panjang dari stunting meliputi penurunan kognitif, hambatan psikomotorik, kesulitan dalam prestasi akademik dan olahraga, serta risiko penyakit degeneratif di masa dewasa [2]. Angka prevalensi stunting di Indonesia mencapai

21,6% dan target penurunan yang ditetapkan pada Rencana Pembangunan Jangka Menengah Nasional (RPJMN) 2020-2024 sebesar 14% membuat upaya percepatan dalam penanganan stunting menjadi sangat penting [3].

Sebagai langkah mengatasi kompleksitas faktor yang memengaruhi stunting, penelitian ini mengadopsi pendekatan inovatif dengan memanfaatkan tiga metode machine learning, meliputi K-Means Clustering, Support Vector Machine (SVM), dan Decision Tree. Faktor prediktor yang dipilih didasarkan pada kelompok sasaran program penurunan stunting dan telah disesuaikan untuk mencerminkan pendekatan multisektoral. K-Means Clustering digunakan untuk mengidentifikasi pola regional terkait stunting, memberikan wawasan penting untuk strategi intervensi yang lebih terfokus. SVM dan Decision Tree, di sisi lain, dipilih untuk analisis klasifikasi dan penelitian lebih lanjut terhadap faktor-faktor stunting di tiap provinsi. Dasar pemilihan metode tersebut didukung oleh literatur sebelumnya. K-Means Clustering, sebagai algoritma pembelajaran tanpa pengawasan, telah terbukti efektif dalam mengidentifikasi pola regional yang serupa [4]. SVM menunjukkan performa akurasi yang superior dalam beberapa studi klasifikasi [5] [6] [7], sementara Decision Tree ditemukan lebih unggul dalam beberapa penelitian lain [8] [9] [10] [11] [12] [13]. Oleh karena itu, penelitian ini akan membandingkan performa kedua metode tersebut untuk menentukan pendekatan terbaik.

Menggabungkan data dari Badan Pusat Statistik dan SSGI 2022, tujuan utama penelitian ini adalah menyediakan wawasan berbasis bukti yang dapat digunakan dalam merumuskan strategi kesehatan masyarakat yang efektif. Melalui kombinasi keunggulan prediktif Decision Tree, keandalan SVM, dan kemampuan K-Means Clustering, penelitian ini diharapkan dapat memberikan kontribusi signifikan

dalam upaya mengurangi prevalensi stunting di Indonesia hingga tahun 2024.

B. Studi Literatur

a) Algoritma K-Means

Algoritma k-means merupakan pendekatan tanpa supervisi yang digunakan untuk mengkategorikan data tak berlabel menjadi sejumlah kluster. Proses implementasi algoritma k-means mencakup langkah-langkah berikut:

- Memilih jumlah kluster (k) yang diperlukan untuk mengelompokkan data.
- Mengatur pusat kluster dengan cara tertentu, bisa berupa inisialisasi acak.
- Mengklasifikasikan setiap data ke kluster yang paling dekat berdasarkan perhitungan jarak antara data dan pusat kluster. Dalam hal ini, jarak diukur dengan metode *Euclidean Distance*.
- Mengukur hubungan suatu data dengan kluster tertentu dengan membandingkan jarak data tersebut dengan pusat kluster. Ini melibatkan penghitungan jarak antara setiap data dan setiap pusat kluster menggunakan rumus Euclidean Distance, dengan variabel seperti $D(i, j)$ yang mewakili jarak antara data ke i dan pusat kluster j seperti berikut.

$$D(i, j) = \sqrt{(X_{1i} - Y_{1j})^2 + (X_{2i} - Y_{2j})^2 + \dots + (X_{ki} - Y_{kj})^2}$$

Dengan keterangan:

$D(i, j)$: jarak dari data ke- i ke pusat kluster ke- j

X_{ki} : data ke- i di atribut ke- k

Y_{kj} : titik pusat ke- j di atribut ke- k

- Memodifikasi pusat kluster dengan menghitung ulang rata-rata dari semua data di kluster tersebut. Cara ini bisa berupa perhitungan rata-rata atau pilihan medoid kluster.
- Melakukan iterasi untuk setiap data dengan mempertimbangkan pusat kluster yang telah diperbarui. Jika ada perubahan pada pusat kluster, langkah sebelumnya diulang. Iterasi ini terus berlanjut hingga pusat kluster stabil, menandakan selesainya proses pengelompokan.

Beberapa teknik evaluasi dapat digunakan untuk menentukan jumlah kluster yang paling sesuai. Pada penelitian ini, digunakan skor siluet dan Indeks Calinski-Harabasz. Skor siluet berfungsi sebagai indikator evaluasi yang mengukur sejauh mana setiap sampel dikelompokkan dengan benar oleh algoritma seperti k-means. Rentang skor ini bervariasi antara -1 hingga 1, semakin mendekati 1, menandakan sampel berada jauh dari kelompok lain; skor 0 mengindikasikan bahwa sampel berada di antara batas keputusan kelompok; dan skor negatif mengimplikasikan potensi kesalahan dalam pengelompokan [14]. Adapun Indeks Calinski-Harabasz (CHI) juga merupakan alat evaluasi yang berharga dalam menilai efisiensi algoritma pengelompokan. CHI, yang kadang-kadang disebut sebagai *Variance Ratio Criterion* (VRC),

mengevaluasi kualitas pengelompokan dengan membandingkan variasi antar kelompok dengan variasi dalam kelompok. Dengan kata lain, semakin tinggi nilai CHI, semakin efisien dan berkualitas pengelompokan yang dihasilkan oleh algoritma [15].

b) Algoritma Support Vector Machine

Algoritma Support Vector Machine (SVM) merupakan salah satu pendekatan dalam pembelajaran mesin yang digunakan untuk klasifikasi data, khususnya untuk dua kelas. Dalam prakteknya, SVM berupaya menemukan sebuah hyperplane yang memisahkan data dengan seoptimal mungkin. Hyperplane ini, yang juga disebut sebagai pengajar, mencari posisi terbaik untuk memisahkan data, sehingga margin atau jarak antara pengajar dengan data terdekat dari masing-masing kelas maksimal. Walaupun SVM awalnya dirancang untuk data linier, namun dengan penggunaan fungsi kernel, data dapat diperluas ke dimensi yang lebih tinggi, memungkinkan SVM untuk menangani data yang tidak linier. Selain klasifikasi, SVM juga dapat digunakan dalam tugas regresi. Berikut adalah beberapa jenis fungsi kernel dalam SVM:

- **Linear Kernel:** Fungsi kernel linier merupakan pendekatan yang optimal ketika data yang diberikan memiliki karakteristik linier yang jelas. Kernel ini mengimplementasikan pengajar linier tanpa memerlukan transformasi data tambahan, dengan asumsi data dapat dipisahkan secara linier. Melalui pendekatan ini, SVM menunjukkan efisiensi yang signifikan dalam mengklasifikasikan data yang menunjukkan linieritas [16].
- **Polynomial Kernel:** Diterapkan pada situasi di mana data tidak memungkinkan pemisahan dengan model linier. Kernel polinomial mentransformasikan data ke dimensi yang lebih tinggi, menghasilkan pengajar dengan sifat polinomial. Keunggulan utamanya terletak pada fleksibilitasnya untuk menyesuaikan derajat polinomial sesuai dengan kompleksitas data yang ada [17].
- **Radial Basis Function (RBF) Kernel:** Relevan ketika data menunjukkan sifat non-linier yang signifikan. Melalui transformasi data ke ruang dimensi yang lebih tinggi, kernel RBF menghasilkan pengajar yang mampu memodelkan karakteristik non-linier dengan presisi. Kemampuannya dalam menangkap nuansa pola non-linier menjadikannya pilihan yang strategis untuk analisis data yang kompleks. [17].
- **Sigmoid Kernel:** Direkomendasikan untuk data yang menunjukkan sifat non-linier yang tidak dapat diakomodasi oleh model linier atau polinomial. Kernel sigmoid memodifikasi data ke dimensi yang lebih tinggi, menghasilkan pengajar dengan sifat sigmoidal. Pendekatan ini khususnya efektif dalam menangani struktur data yang menunjukkan hubungan non-linier yang rumit

[17].

Penting untuk menekankan bahwa seleksi kernel yang paling sesuai harus didasarkan pada analisis mendalam mengenai sifat intrinsik dan distribusi data yang akan diproses [18].

c) Algoritma Decision Tree

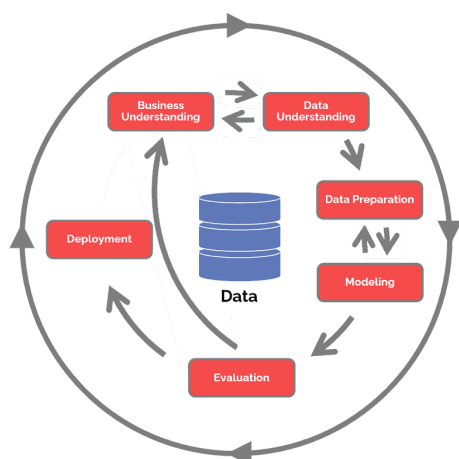
Decision Tree digunakan untuk membantu dalam memprediksi atau mengklasifikasikan data. Alat ini bekerja dengan mengambil sekelompok data dan mencari pola atau aturan di dalamnya untuk membantu dalam pengambilan keputusan. Langkah membangun pohon keputusan adalah sebagai berikut:

- Mengambil semua data awal dan menyimpannya dalam satu kelompok utama.
- Mencari tahu data atau variabel apa yang paling berpengaruh dalam menentukan hasil, biasanya menggunakan metode Attribute Selection Measure.
- Membagi kelompok utama tersebut berdasarkan variabel yang telah ditemukan, sehingga didapatkan kelompok-kelompok yang lebih kecil namun lebih spesifik.
- Pada setiap kelompok yang baru dibuat, langkah 2 dan 3 perlu diulangi. Proses ini akan terus berlanjut hingga program tidak bisa lagi membagi data menjadi kelompok yang lebih kecil. Kelompok-kelompok yang terakhir ini disebut sebagai "simpul daun".

Keputusan di setiap langkah dibuat berdasarkan metrik yang disebut gain information. Metrik ini mengukur seberapa baik sebuah variabel dapat membantu dalam memprediksi hasil. Semakin tinggi Gain Informasi, semakin baik variabel tersebut [10].

II. METODOLOGI

Proses yang dilakukan oleh peneliti dapat diilustrasikan secara rinci melalui Gambar 1.



Gambar 1. proses CRISP-DM

Pada sub-bab selanjutnya, akan diuraikan secara mendalam penjelasan mengenai setiap tahapan yang telah dilakukan, memberikan pemahaman yang komprehensif mengenai metodologi yang diterapkan dalam penelitian ini.

A. Business Understanding

Data yang diolah harus dimengerti terlebih dahulu tujuan bisnisnya untuk apa. Dalam permasalahan ini, dilakukan identifikasi faktor-faktor yang berkontribusi pada prevalensi stunting dan memberikan rekomendasi untuk intervensi yang efektif.

B. Data Understanding

Penelitian ini memanfaatkan data sekunder yang bersumber dari publikasi Badan Pusat Statistik tahun 2022. Data ini dapat diperoleh dan diverifikasi melalui laman resmi Badan Pusat Statistik di <https://www.bps.go.id/>. Dataset yang telah disusun dan dikumpulkan disajikan dalam format file Excel. Selanjutnya, data tersebut telah diimpor dan diintegrasikan ke dalam lingkungan Jupyter Notebook untuk proses analisis lebih lanjut.

```
#Input Data
src = pd.read_excel("C:/Users/asus/Downloads/Stunt dataset.xlsx")
src.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 520 entries, 0 to 519
Data columns (total 23 columns):
#   Column                                     Non-Null Count  Dtype
---  -
0   Kabupaten/Kota Prov Indonesia            520 non-null    object
1   Prevalensi Stunting (TB/U) %             520 non-null    float64
2   K4                                         520 non-null    float64
3   Persalinan FASYANKES                     520 non-null    float64
4   KF Lengkap                              520 non-null    float64
5   Vit A Ibu                                520 non-null    float64
6   bumil TTD                                520 non-null    float64
7   BBLR                                       520 non-null    float64
8   IMD                                        520 non-null    float64
9   ASI                                       520 non-null    float64
10  CPKB                                      520 non-null    float64
11  IDL                                       520 non-null    float64
12  A 611                                    520 non-null    float64
13  A 1259                                   520 non-null    float64
14  A 659                                    520 non-null    float64
15  mCPR                                      520 non-null    float64
16  Air Minum Layak                         520 non-null    float64
17  Sanitasi Layak                          520 non-null    float64
18  IKP                                       520 non-null    float64
19  BPNT 40%                                520 non-null    float64
20  KKS 40%                                  520 non-null    float64
21  APK PAUD                                520 non-null    float64
22  UMK                                       520 non-null    float64
```

Gambar 2. Informasi Variabel dalam Dataset

Dataset yang digunakan dalam penelitian ini mencakup 23 kolom variabel dengan sebagian besar diantaranya menggunakan unit persentase, kecuali untuk kolom 'UMK' yang menggunakan satuan rupiah. Untuk informasi yang lebih rinci mengenai setiap variabel, dapat ditemukan dalam Tabel 1. Secara teknis, mayoritas kolom memiliki tipe data float64 yang mengindikasikan tipe data numerik, namun perlu dicatat bahwa kolom 'Kabupaten/Kota Prov Indonesia' memiliki tipe data object yang menunjukkan tipe data string.

Tabel 1. Identifikasi Variabel

Nama Kolom	Definisi
Prevalensi Stunting (TB/U) %	Persentase balita dengan kondisi gizi yang dikategorikan sebagai sangat pendek dan pendek berdasarkan indeks tinggi badan menurut umur (TB/U) dengan Z score kurang dari -2 standar deviasi di setiap Kabupaten/Kota di

	Indonesia pada tahun 2022.
K4	Persentase ibu hamil yang menjalani pemeriksaan kehamilan minimal 4 (empat) kali selama periode kehamilan. Pemeriksaan ini terdiri dari satu kali pemeriksaan pada trimester pertama, satu kali pemeriksaan pada trimester kedua, dan dua kali pemeriksaan pada trimester ketiga di setiap Kabupaten/Kota di Indonesia pada tahun 2022.
Persalinan FASYANKES	Persentase persalinan yang dilaksanakan di FASYANKES di setiap Kabupaten/Kota di Indonesia pada tahun 2022.
KF lengkap	Persentase ibu bersalin/nifas yang menerima kunjungan nifas lengkap hingga 42 hari setelah melahirkan di setiap Kabupaten/Kota Indonesia pada tahun 2022.
Vit A Ibu	Persentase ibu nifas yang menerima suplemen Vitamin A di setiap Kabupaten/Kota Indonesia pada tahun 2022
Bumil TTD (90 Tablet)	Persentase ibu hamil yang memperoleh 90 tablet Tablet Tambah Darah (TTD) selama masa kehamilan di setiap Kabupaten/Kota Indonesia pada tahun 2022.
BBLR (Berat Badan Lahir Rendah)	Persentase bayi dengan BBLR tiap Kabupaten/Kota Indonesia 2022
IMD (Inisiasi Menyusui Dini)	Persentase bayi baru lahir yang mendapat IMD di tiap Kabupaten/Kota Indonesia 2022
ASI (Air Susu Ibu)	Persentase bayi usia < 6 bulan yang diberi ASI eksklusif di tiap Kabupaten/Kota Indonesia 2022
CPKB (Cakupan Pelayanan Kesehatan Bayi)	Persentase bayi laki-laki dan perempuan yang menerima layanan kesehatan neonatal esensial, mencakup tindakan resusitasi, pencegahan hipotermia, praktik pemberian ASI awal dan eksklusif, perawatan mata, tali pusat, kulit, imunisasi, pemberian vitamin K, manajemen terpadu balita muda (MTBM), dan penyuluhan bagi ibunya mengenai perawatan neonatus di rumah dengan menggunakan buku KIA, di setiap Kabupaten/Kota di Indonesia pada tahun 2022.
IDL (Imunisasi Dasar Lengkap)	Persentase bayi yang telah menerima satu dosis vaksinasi Hepatitis B, satu dosis vaksinasi BCG, tiga dosis vaksinasi DPT-HB/DPT-HB-Hib, empat dosis vaksinasi polio, dan satu dosis vaksinasi campak rubella di setiap Kabupaten/Kota Indonesia pada tahun 2022.
A 6-11	Persentase bayi usia 6-11 bulan yang menerima suplemen vitamin A di setiap Kabupaten/Kota di Indonesia pada tahun 2022.
A 12-59	Persentase bayi usia 12-59 bulan yang

	menerima suplemen vitamin A di setiap Kabupaten/Kota di Indonesia pada tahun 2022.
A 6-59	Persentase bayi usia 6-59 bulan yang menerima suplemen vitamin A di setiap Kabupaten/Kota di Indonesia pada tahun 2022.
mCPR (Peserta KB Aktif Metode Modern)	Persentase peserta program KB baru dan yang telah lama masih dalam keadaan aktif menggunakan kontrasepsi secara terus-menerus dengan metode modern seperti kondom, suntik, pil, AKDR, MOW, MOP, Implan, dan MAL untuk tujuan menunda, menjarangkan kehamilan, atau mengakhiri kesuburan di setiap Kabupaten/Kota di Indonesia pada tahun 2022.
Air Minum Layak	Persentase rumah tangga menurut Kabupaten/Kota di Indonesia pada tahun 2022 yang memiliki akses layak terhadap air minum.
Sanitasi Layak	Persentase rumah tangga menurut Kabupaten/Kota di Indonesia pada tahun 2022 yang memiliki akses sanitasi layak
IKP (Indeks Ketahanan Pangan)	Persentase IKP terdiri dari tiga aspek utama ketahanan pangan, yaitu ketersediaan, keterjangkauan, dan pemanfaatan pangan.
BNPT 40%	Persentase rumah tangga dalam kelompok kuintil 1 dan 2 (penduduk pada kelompok pengeluaran 40% terendah) di Indonesia diuraikan berdasarkan kabupaten/kota dan pembelian/penerimaan beras Raskin/Bantuan Pangan Non Tunai (BPNT) selama empat bulan terakhir pada tahun 2022
KKS 40%	Persentase rumah tangga kelompok kuintil 1 dan 2 (penduduk pada kelompok pengeluaran 40% terbawah) penerima kartu perlindungan sosial (KPS)/kartu keluarga sejahtera (KKKS) di Indonesia terperinci menurut Kabupaten/Kota Tahun 2022.
APK PAUD	Persentase Angka Partisipasi Kasar Penduduk Usia 3-6 Tahun Dirinci Menurut Kabupaten/Kota tahun 2022
UMK (Upah Minimum Kabupaten/Kota)	Upah terendah (termasuk tunjangan teratur tetapi tidak termasuk upah lembur) yang dibayarkan kepada karyawan (per jenis jabatan/pekerjaan)

C. Data Preparation

a) preparasi dan eksplorasi data

Pada tahapan ini, proses pra-pemrosesan data dilakukan dengan tujuan untuk mempersiapkan data sebelum dianalisis menggunakan metode K-Means, Support Vector Machine, dan Decision Tree. Tahapan awal dalam proses ini adalah melakukan pengecekan missing value, mengingat bahwa ketiga algoritma yang digunakan tidak dapat menangani data yang

memiliki missing value. Gambaran yang lebih rinci mengenai proses pengecekan tersebut dapat dilihat pada ilustrasi yang disajikan dalam Gambar 3.

```
#Cek missing value
src.isna().sum()

Kabupaten/Kota Prov Indonesia    0
Prevalensi Stunting (TB/U) %      0
K4                                0
Persalinan FASYANKES              0
KF Lengkap                       0
Vit A Ibu                        0
bumil TTD                        0
BBLR                             0
IMD                              0
ASI                              0
CPKB                             0
IDL                              0
A 611                            0
A 1259                           0
A 659                            0
mCPR                             0
Air Minum Layak                  0
Sanitasi Layak                   0
IKP                              0
BPNT 40%                        0
KKS 40%                          0
APK PAUD                         0
UMK                              0
dtype: int64
```

Gambar 3. Pengecekan Missing Value

Berdasarkan hasil pengecekan, dapat disimpulkan bahwa dataset yang digunakan tidak mengandung missing value. Selanjutnya, tahapan berikutnya adalah melakukan pengecekan terhadap adanya outlier dengan fokus pada evaluasi variabel non-numerik dalam dataset. Proses identifikasi outlier ini mengacu pada metode nilai interquartile, yang rinciannya dapat diperiksa dan dipahami melalui ilustrasi yang diberikan dalam Gambar 4.

```
#cek outlier
Q1 = srcnokab.quantile(q=.25)
Q3 = srcnokab.quantile(q=.75)
IQR = Q3-Q1

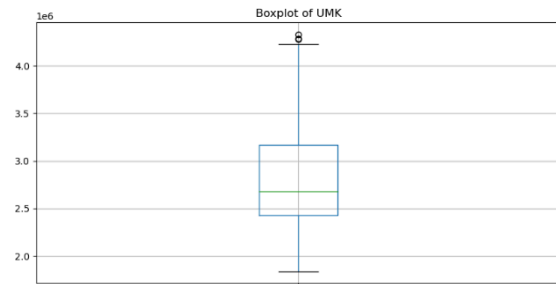
data_iqr = srcnokab[~((srcnokab < (Q1-1.5*IQR)) | (srcnokab > (Q3+1.5*IQR)))]
data_iqr.shape

print("Dimensi dataset awal", srcnokab.shape)
print("Dimensi dataset setelah pengecekan outlier", data_iqr.shape)
```

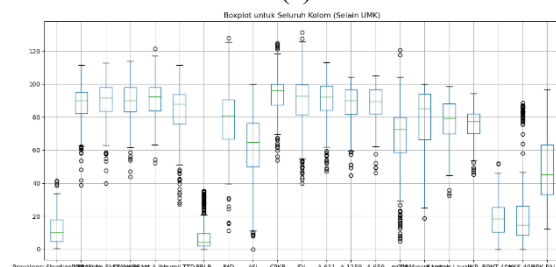
Dimensi dataset awal (520, 22)
Dimensi dataset setelah pengecekan outlier (274, 22)

Gambar 4. Pengecekan Outlier Menggunakan Interquartile

Gambar 4 mengindikasikan bahwa ukuran awal dataset adalah 520 baris dan 22 kolom. Pasca eliminasi outlier, sejumlah baris telah dieliminasi, mengurangi ukuran dataset menjadi 274 baris, meskipun jumlah kolomnya tidak berubah. Langkah selanjutnya dalam memeriksa outlier diteruskan dengan memanfaatkan boxplot, seperti yang tergambar pada Gambar 5.



(a)

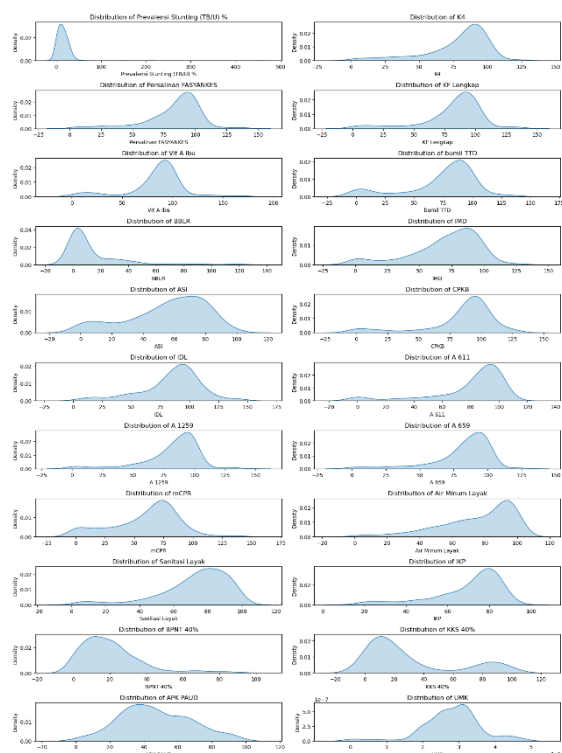


(b)

Gambar 5. Boxplot Variabel Numerik

Outlier ditemukan pada setiap variabel seperti yang ditunjukkan di Gambar 5. Menghilangkan outlier tidak akan dipertimbangkan, karena dapat mengurangi informasi krusial dari dataset. Oleh sebab itu, dalam penelitian ini, penanganan outlier akan memanfaatkan teknik yang resisten terhadap dampak nilai outlier dan mempergunakan transformasi data. Teknik transformasi yang dipilih akan menyesuaikan dengan karakteristik persebaran setiap variabel. Sebagai langkah pertama, plot Kernel Density Estimation (KDE) akan disiapkan untuk mendalami pola distribusi data lebih detail.

```
Skewness of Prevalensi Stunting (TB/U) %: 13.017
Skewness of K4: -1.401
Skewness of Persalinan FASYANKES: -1.307
Skewness of KF Lengkap: -1.381
Skewness of Vit A Ibu: -0.849
Skewness of bumil TTD: -1.131
Skewness of BBLR: 2.633
Skewness of IMD: -1.027
Skewness of ASI: -0.681
Skewness of CPKB: -1.633
Skewness of IDL: -0.878
Skewness of A 611: -1.836
Skewness of A 1259: -1.513
Skewness of A 659: -1.733
Skewness of mCPR: -0.428
Skewness of Air Minum Layak: -1.1
Skewness of Sanitasi Layak: -1.383
Skewness of IKP: -1.382
Skewness of BPNT 40%: 1.64
Skewness of KKS 40%: 1.09
Skewness of APK PAUD: 0.242
Skewness of UMK: -0.638
```



Gambar 6. Plot Kernel Density Estimation (KDE)

Dari evaluasi nilai skewness pada masing-masing variabel, dapat disimpulkan bahwa "Prevalensi Stunting (TB/U) %" (Skewness: 13.017), "BPNT 40%" (Skewness: 1.64), serta "KKS 40%" (Skewness: 1.09) menunjukkan kecenderungan skewness positif yang berarti. Skewness dengan angka positif yang signifikan menggambarkan distribusi data dengan ekor panjang menuju nilai yang lebih tinggi, sedangkan sebagian besar data berada pada nilai yang lebih rendah. Sementara itu, variabel "K4" dengan Skewness sebesar -1.401 mengindikasikan adanya kemiringan distribusi data ke sisi kanan, dimana mayoritas data berkumpul pada nilai-nilai yang lebih tinggi. Nilai skewness ini mencerminkan plot dari setiap variabel.

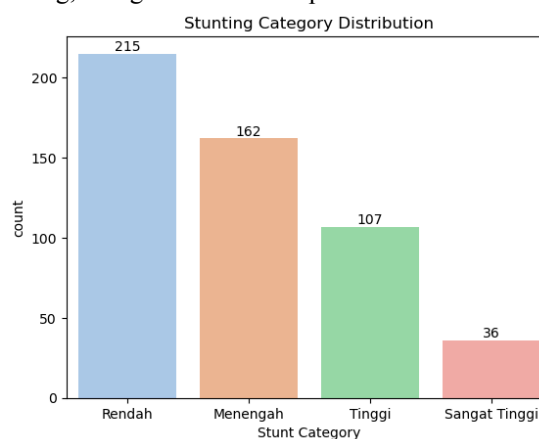
Guna mengurangi distribusi yang tidak simetris, dilakukan standarisasi pada data. Hal ini dilakukan dengan tujuan membuat distribusi data lebih simetris dan mengurangi pengaruh dari nilai-nilai yang berada di luar kisaran normal. Adanya proses standarisasi diarahkan untuk meningkatkan kemampuan interpretasi hasil analisis dan memastikan keandalan model statistik yang akan digunakan selanjutnya.

```
from sklearn import preprocessing
srcz = preprocessing.scale(srcnokab)
srcz

array([[ 0.43378231,  0.32265872,  0.2393182 , ..., -0.65467682,
         0.70923903, -1.06641861],
       [ 0.09986639,  0.70091003,  0.64587553, ..., -0.11416977,
        -0.05339914, -0.60080731],
       [-0.05678553, -0.04709256, -0.09803787, ...,  0.03831383,
        1.0150661 , -0.99347532],
       ...,
       [-0.23404954, -2.77560203, -1.27445906, ...,  2.16618955,
        -1.53477918, -0.02977162],
       [-0.11037697,  1.89941419, -0.49594504, ..., -0.75200336,
        0.0088275 ,  0.58191388],
       [ 0.30186491, -1.33059702,  1.45034002, ...,  1.7813729 ,
        -0.71473813, -0.31955558]])
```

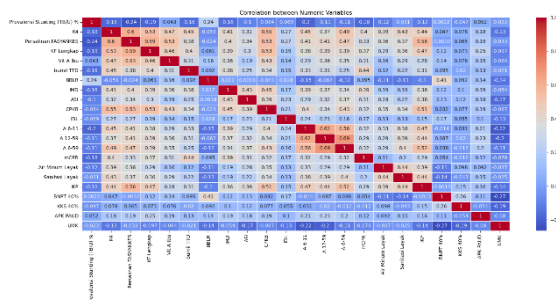
Gambar 7. Data yang Telah Distandarisasi

Selanjutnya, variabel tambahan dibuat yang berisi data ordinal untuk keperluan analisis klasifikasi. Proses ini esensial karena dalam dataset yang ada, tidak ada variabel kategorik yang tersedia. Pendekatan pembuatan variabel tambahan mengacu pada rekomendasi WHO untuk mengklasifikasikan prevalensi stunting, memanfaatkan empat dari lima kategori yang telah ditentukan, yaitu rendah, menengah, tinggi, dan sangat tinggi [19]. Ilustrasi dari proses pengelompokan ini dapat dilihat melalui grafik batang, sebagaimana terlihat pada Gambar 8.



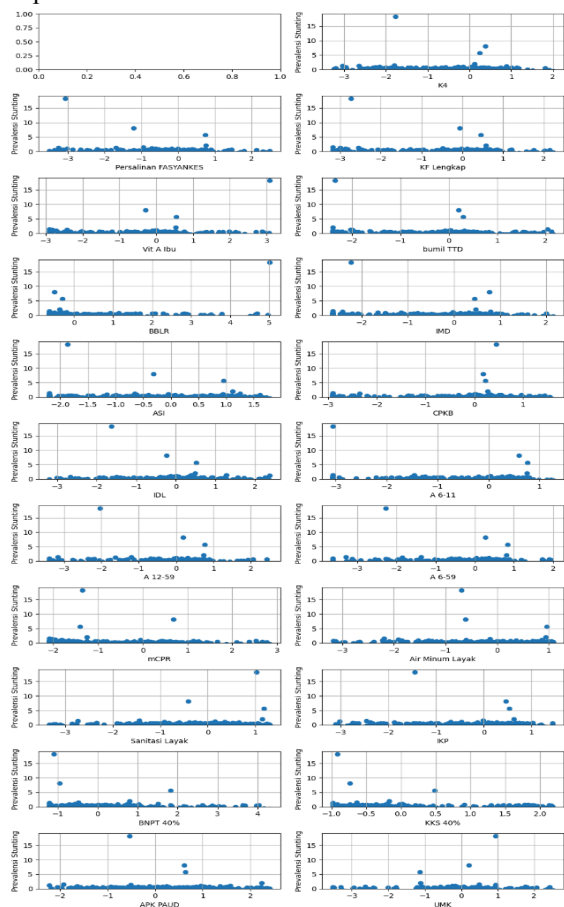
Gambar 8. Sebaran Kabupaten/Kota Berdasarkan Kategori Prevalensi Stunting

Pada kategori prevalensi stunting yang sangat tinggi, terdapat 36 Kabupaten/Kota, misalnya Kabupaten Bondowoso. Sementara itu, Kabupaten/Kota yang diklasifikasikan sebagai tinggi berjumlah 107, sedangkan yang tergolong menengah sebanyak 162, dan yang terkategori rendah mencapai 215 Kabupaten/Kota. Jika diamati dari diagram ini, dapat diketahui bahwa sebagian besar Kabupaten/Kota di Indonesia sudah menerapkan kebijakan penanganan stunting dengan baik yang ditandai dengan banyaknya daerah yang masuk ke dalam kategori rendah. Langkah berikutnya melibatkan proses seleksi variabel yang memiliki pengaruh paling signifikan, berdasarkan nilai korelasi dengan prevalensi stunting di tiap Kabupaten/Kota sebagai variabel respon.



Gambar 9. Korelasi Variabel Numerik Terhadap Variabel Respon

Variabel prediktor yang mempengaruhi persentase "Stunting Prevalence (TB/U)" dipilih berdasarkan korelasi yang signifikan dengan variabel respons tersebut. Variabel prediktor yang menunjukkan korelasi positif yang signifikan melibatkan "BBLR", sementara variabel yang memiliki korelasi negatif yang signifikan mencakup "Persalinan FASYANKES", "K4", "KF Lengkap", "mCPR", dan "A 6-11". Seleksi variabel dilakukan dengan mempertimbangkan korelasi mendekati nilai $\pm 0,5$ dan tidak kurang dari $\pm 0,2$, menandakan adanya keterkaitan yang kuat. Variabel-variabel lain diabaikan karena memiliki korelasi mendekati nol, menunjukkan ketidakcukupan hubungan untuk menjelaskan variabel respons.



Gambar 10. Grafik Sebaran Variabel Numerik Terhadap Prevalensi Stunting

Dalam ilustrasi dari Gambar 10, terlihat bahwa titik-titik data pada variabel yang memiliki korelasi mendekati $\pm 0,5$ dan minimal $\pm 0,2$ menunjukkan pola distribusi tertentu. Sebaliknya, titik-titik pada variabel dengan korelasi di luar interval tersebut menyebar secara acak. Dari analisis ini, dapat disimpulkan bahwa variabel "BBLR" memiliki hubungan positif yang signifikan dengan prevalensi stunting. Di sisi lain, variabel "Persalinan FASYANKES", "K4", "KF Lengkap", "mCPR", dan "A 6-11" menunjukkan hubungan negatif yang signifikan dengan prevalensi stunting di berbagai Kabupaten/Kota.

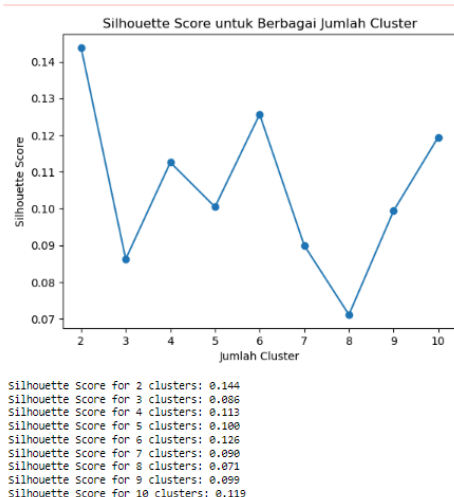
Langkah berikutnya melibatkan pembagian data menjadi dua kelompok, yakni data pelatihan dan data pengujian. Tujuan dari langkah ini adalah untuk mengevaluasi performa model yang telah dikembangkan. Dalam konteks penelitian ini, diterapkan perbandingan 80:20, yang mengindikasikan bahwa 80% dari dataset akan digunakan sebagai data pelatihan, sementara 20% sisanya akan dijadikan sebagai data pengujian. Pembagian dataset ini diharapkan dapat memberikan latihan yang memadai bagi model machine learning, dengan harapan dapat meningkatkan akurasi hasil penelitian.

D. Modeling

a) K-means Clustering

Teknik analisis kluster digunakan dalam data mining untuk eksplorasi serta klasifikasi data berdasarkan kesamaan ciri-ciri di antara observasi [20]. Melalui pendekatan ini, observasi yang berada dalam satu kelompok menunjukkan ciri-ciri yang lebih mirip dibandingkan dengan observasi di kelompok lainnya. Penelitian ini berupaya mengkategorikan Kabupaten/Kota di Indonesia berdasarkan kesamaan variabel prediktor dalam mempengaruhi variabel respon. Penentuan jumlah kluster didasarkan pada nilai Silhouette maksimal, menggambarkan tingkat homogenitas dan pemisahan yang optimal di antara kluster-klusternya.

Pemilihan fitur dilakukan berdasarkan korelasi antar variabel. Dari analisis korelasi, dipilih enam fitur yang memiliki korelasi paling signifikan dengan prevalensi stunting. Fitur-fitur ini adalah fitur ke-1, ke-2, ke-3, ke-6, ke-11, dan ke-14 dari dataset asli. Fitur-fitur ini disimpan dalam variabel X, sementara prevalensi stunting (kolom pertama) disimpan dalam variabel Y.



Gambar 11. Penentuan Jumlah Klaster

Pemilihan Jumlah Cluster Optimal digunakan Untuk menentukan jumlah cluster yang optimal, dilakukan iterasi dari 2 hingga 10 klaster. Pada setiap iterasi, dihitung Silhouette Score yang mengukur seberapa mirip data dalam satu klaster dan seberapa berbeda dengan data di klaster lain. Hasil dari iterasi ini digambarkan dalam grafik, di mana nilai Silhouette Score tertinggi menunjukkan jumlah klaster yang optimal yang terdapat dalam Gambar 11.

b) Support Vector Machine (SVM)

Data yang digunakan dalam analisis ini berasal dari dataset yang telah dipecah menjadi data latih (xc_train, yc_train) dan data uji (xc_test, yc_test). Dataset ini mencakup berbagai fitur yang relevan untuk klasifikasi dan label kelas untuk setiap contoh.

Untuk melakukan klasifikasi, digunakan Support Vector Machine (SVM) dengan berbagai jenis kernel: linear, RBF (Radial Basis Function), sigmoid, dan polynomial. Setiap model dilatih menggunakan data latih dan kemudian dievaluasi menggunakan data uji.

Hasil prediksi dibandingkan dengan label sebenarnya menggunakan laporan klasifikasi (classification_report) yang mencakup metrik seperti presisi, recall, dan F1-score. Selain itu, ditampilkan juga Confusion Matrix untuk melihat distribusi prediksi yang benar dan salah.

Maka Metodologi ini melibatkan pelatihan dan evaluasi model SVM dengan berbagai kernel untuk menentukan performa terbaik dalam klasifikasi data. Dengan menggunakan laporan klasifikasi dan ROC Curve, kita dapat membandingkan kinerja model secara komprehensif dan memilih model yang paling sesuai untuk data kita.

c) Decision Tree

Data yang digunakan dalam analisis ini berasal dari dataset yang berisi berbagai variabel yang relevan untuk klasifikasi serta label kelas untuk setiap sampel. Dataset ini dibagi menjadi data latih (xc_train, yc_train) dan data uji (xc_test, yc_test).

Untuk melakukan klasifikasi, digunakan model Decision Tree dengan parameter maksimal kedalaman pohon (max_depth) sebesar 4 dan minimal sampel pada tiap daun (min_samples_leaf) sebesar 2. Model ini dilatih menggunakan data latih dan kemudian dievaluasi menggunakan data uji.

Hasil prediksi dibandingkan dengan label sebenarnya menggunakan laporan klasifikasi (classification_report) yang mencakup metrik seperti presisi, recall, dan F1-score. Selain itu, ditampilkan juga Confusion Matrix untuk melihat distribusi prediksi yang benar dan salah.

Maka Metodologi ini melibatkan pelatihan dan evaluasi model Decision Tree untuk klasifikasi data multi-kelas. Dengan menggunakan laporan klasifikasi dan ROC Curve, kita dapat mengevaluasi kinerja model secara komprehensif dan memahami kemampuan model dalam membedakan antara kelas-kelas yang berbeda.

E. Evaluation

a) K-means Clustering

Pada tahap validasi dan evaluasi model klusterisasi menggunakan K-Means, dua metrik utama digunakan untuk menentukan jumlah klaster yang optimal, yaitu Silhouette Score dan Calinski-Harabasz Index. Silhouette Score mengukur seberapa mirip setiap titik data dengan klaster mereka sendiri dibandingkan dengan klaster lain, dengan nilai tertinggi menunjukkan klusterisasi terbaik. Evaluasi dilakukan untuk jumlah klaster antara 2 hingga 10, dan hasilnya menunjukkan bahwa jumlah klaster optimal adalah 2, karena nilai Silhouette Score tertinggi tercapai pada jumlah klaster tersebut. Calinski-Harabasz Index, yang mengukur rasio antara dispersi antara klaster dengan dispersi dalam klaster, juga mendukung keputusan ini dengan nilai tertinggi pada 2 klaster. Visualisasi klaster dan analisis rata-rata prevalensi stunting di setiap klaster memperkuat temuan ini, dimana klaster 0 memiliki rata-rata prevalensi stunting yang lebih tinggi dibandingkan klaster 1. Kombinasi dari kedua metrik ini memastikan bahwa pemilihan 2 klaster adalah yang paling sesuai untuk data yang digunakan.

b) Support Vector Machine (SVM)

Pada tahap validasi dan evaluasi model Support Vector Machine (SVM) untuk klasifikasi multikelas, dilakukan eksperimen dengan berbagai kernel, yaitu linear, RBF, sigmoid, dan polynomial. Model SVM dengan kernel linear pertama kali diimplementasikan, dimana hasil prediksi dibandingkan dengan nilai

aktual menggunakan confusion matrix dan metrik evaluasi seperti precision, recall, dan f1-score yang dihasilkan oleh classification_report. Confusion matrix divisualisasikan untuk memberikan gambaran jelas tentang performa model.

c) *Decision Tree*

Pada tahap validasi dan evaluasi model Decision Tree untuk klasifikasi multikelas, model dibangun dengan parameter max_depth 4 dan min_samples_leaf 2. Model dilatih menggunakan data pelatihan dan kemudian diuji menggunakan data uji, dengan hasil prediksi dibandingkan terhadap nilai aktual melalui confusion matrix dan metrik evaluasi seperti precision, recall, dan f1-score yang diperoleh dari classification_report. Confusion matrix divisualisasikan untuk memberikan gambaran jelas mengenai performa model

F. Deployment

Dalam studi kasus ini, deployment dilakukan dengan mengintegrasikan model K-means clustering, Support Vector Machine (SVM), dan Decision Tree ke dalam sebuah sistem aplikasi berbasis web yang dirancang untuk mendukung pengambilan keputusan dalam penanganan stunting. Aplikasi ini memungkinkan pengguna, seperti petugas kesehatan dan pembuat kebijakan, untuk memasukkan data baru dan memperoleh hasil analisis secara real-time. Model K-means clustering digunakan untuk mengidentifikasi pola regional terkait stunting, memberikan wawasan untuk strategi intervensi yang lebih terfokus. Sementara itu, model SVM dan Decision Tree digunakan untuk mengklasifikasikan data baru berdasarkan faktor-faktor yang relevan, membantu dalam penelitian lebih lanjut dan evaluasi efektivitas intervensi. Implementasi ini melibatkan pengembangan backend menggunakan framework seperti Flask atau Django, serta frontend yang user-friendly untuk memudahkan interaksi pengguna. Pengujian dan validasi model dilakukan secara berkala untuk memastikan akurasi dan performa tetap optimal dalam lingkungan produksi, mendukung target penurunan prevalensi stunting sesuai RPJMN 2020-2024.

III. HASIL DAN PEMBAHASAN

A. Analisa Masalah

1. Perbandingan Penggunaan Algoritma

Pendekatan inovatif dalam penelitian ini menggunakan tiga metode machine learning: K-Means Clustering, Support Vector Machine (SVM), dan Decision Tree. K-Means Clustering digunakan untuk mengidentifikasi pola regional terkait stunting, memberikan wawasan penting untuk strategi intervensi yang lebih terfokus. SVM dan Decision Tree dipilih untuk analisis klasifikasi faktor-faktor stunting di tiap provinsi. Keputusan ini didasarkan

pada keunggulan masing-masing metode yang terbukti dalam literatur sebelumnya. Meskipun SVM menunjukkan performa akurasi yang superior dalam beberapa studi klasifikasi, Decision Tree ditemukan lebih unggul dalam penelitian lain. Dengan membandingkan performa kedua metode ini, penelitian ini bertujuan untuk menentukan pendekatan terbaik dalam mengatasi stunting di Indonesia.

2. Pemahaman Pola Stunting di Tingkat Kabupaten

Analisis menggunakan pendekatan multisektoral untuk mengidentifikasi faktor-faktor yang mempengaruhi stunting di tiap provinsi. Hasilnya dapat memberikan pemahaman mendalam tentang pola stunting yang berbeda-beda di setiap wilayah. Informasi ini penting untuk merumuskan strategi intervensi yang sesuai dengan kondisi lokal, memastikan efektivitas program penurunan stunting secara holistik. Dengan demikian, penelitian ini diharapkan dapat memberikan panduan yang lebih tepat dalam alokasi sumber daya dan implementasi program kesehatan masyarakat di tingkat kabupaten.

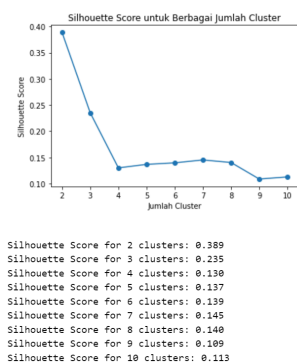
3. Analisis Faktor Penyebab Stunting di Indonesia

Penelitian ini mencermati faktor-faktor yang mempengaruhi stunting di Indonesia, yang meliputi faktor gizi, sanitasi, akses kesehatan, dan determinan sosio-ekonomi lainnya. Dengan menganalisis korelasi antara faktor-faktor ini dan prevalensi stunting di setiap provinsi, penelitian ini dapat mengidentifikasi faktor-faktor yang paling berperan dalam masalah stunting di tiap wilayah. Informasi ini penting untuk merancang strategi intervensi yang lebih efektif dan terarah, serta meningkatkan koordinasi lintas sektor dalam upaya penanganan stunting.

B. Hasil Pemodelan

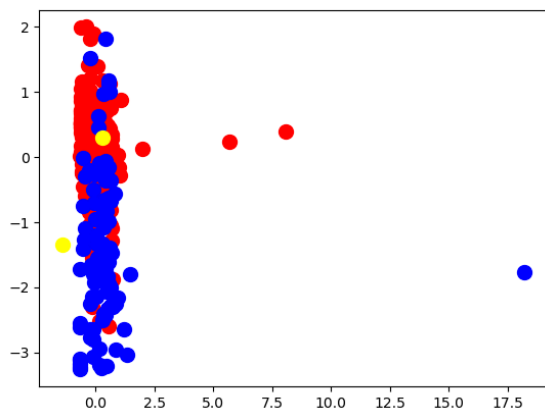
a) K means Clustering

Teknik analisis kluster digunakan dalam data mining untuk eksplorasi serta klasifikasi data berdasarkan kesamaan ciri-ciri di antara observasi [20]. Melalui pendekatan ini, observasi yang berada dalam satu kelompok menunjukkan ciri-ciri yang lebih mirip dibandingkan dengan observasi di kelompok lainnya. Penelitian ini berupaya mengkategorikan Kabupaten/Kota di Indonesia berdasarkan kesamaan variabel prediktor dalam mempengaruhi variabel respon. Penentuan jumlah kluster didasarkan pada nilai Silhouette maksimal, menggambarkan tingkat homogenitas dan pemisahan yang optimal di antara kluster-klasternya.



Gambar 12. Penentuan Jumlah Kluster

Pada Gambar 12, terlihat bahwa nilai Silhouette mencapai puncaknya pada jumlah kluster sama dengan dua, dengan nilai Silhouette sebesar 0,389. Oleh karena itu, Kabupaten/Kota di Indonesia akan dikelompokkan menjadi dua. Informasi lebih lanjut mengenai hasil pengelompokkan dapat diakses pada visualisasi yang disajikan pada Gambar 12.



Gambar 12. Pengelompokkan Kabupaten/Kota Menjadi Dua Kluster

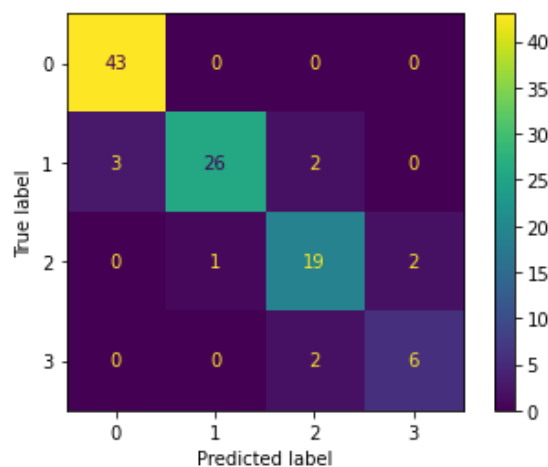
Berdasarkan hasil analisis dua kluster yang dihasilkan dari metode K-Means Clustering, terdapat perbedaan karakteristik yang signifikan antara Kluster 0 dan Kluster 1 dalam hal Prevalensi Stunting dan indikator kesehatan masyarakat lainnya. Kluster 0 memiliki Prevalensi Stunting (TB/U) % lebih rendah, sekitar 13.55%, dibandingkan dengan Kluster 1 yang memiliki tingkat Prevalensi Stunting sebesar 25.20%. Selain itu, Kluster 0 juga menunjukkan angka yang lebih baik pada sebagian besar variabel indikator kesehatan, seperti K4, Persalinan FASYANKES, KF Lengkap, Vit A Ibu, bumil TTD, BBLR, IMD, ASI, CPKB, dan lainnya, dengan nilai rata-rata yang relatif lebih tinggi. Di sisi lain, Kluster 1 menunjukkan angka yang cenderung lebih rendah pada sejumlah indikator tersebut. Kesimpulan ini

memberikan gambaran bahwa Kluster 0 dapat diidentifikasi sebagai kelompok dengan kondisi kesehatan masyarakat yang lebih baik, sementara Kluster 1 cenderung memiliki tingkat prevalensi stunting yang lebih tinggi dan kondisi kesehatan masyarakat yang lebih menantang. Hasil ini dapat menjadi dasar untuk pengembangan strategi intervensi kesehatan yang lebih terfokus dan responsif di masing-masing kluster guna meningkatkan kondisi kesehatan masyarakat secara keseluruhan.

Namun demikian, terlihat bahwa tidak ada pengelompokkan terpisah dari anggota kluster 0 dan kluster 1. Hal ini menunjukkan bahwa karakteristik kedua kluster tidak berbeda jauh, sehingga penggunaan K-means pada penelitian ini dapat dianggap tidak diperlukan

b) Support Vector Machine (SVM)

Hasil dan pembahasan Support Vector Machine menunjukkan performa model dalam mengklasifikasikan Kabupaten/Kota ke dalam berbagai kategori prevalensi stunting. Model berhasil mengkategorikan sejumlah Kabupaten/Kota secara akurat ke dalam kategori rendah, menengah, tinggi, dan sangat tinggi.



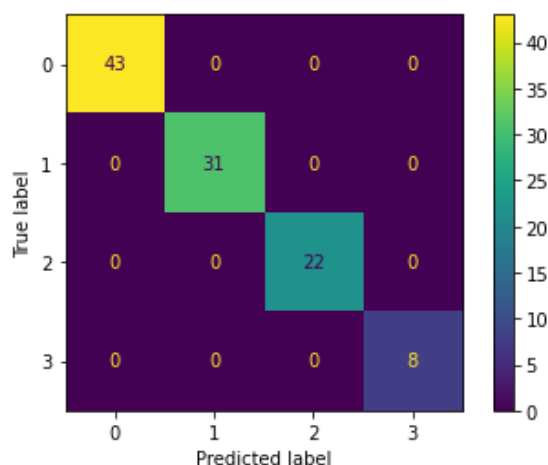
Gambar 14. Confusion Matrix SVM Kernel Linear

Pada Gambar 14 model berhasil mengkategorikan 43 Kabupaten/Kota ke dalam kategori rendah, 26 Kabupaten/Kota ke kategori menengah, 19 Kabupaten/Kota ke kategori tinggi, serta 6 Kabupaten/Kota ke kategori sangat tinggi secara akurat. Namun, ada beberapa kesalahan dalam klasifikasi, yaitu 3 Kabupaten/Kota yang sebenarnya berada di kategori menengah namun dikategorikan sebagai rendah, 1 Kabupaten/Kota yang seharusnya berada di kategori tinggi namun diklasifikasikan sebagai menengah, 2 Kabupaten/Kota yang seharusnya berada di kategori sangat tinggi namun dikategorikan sebagai tinggi, 2 Kabupaten/Kota yang sebenarnya berada di kategori menengah namun

diklasifikasikan sebagai tinggi, dan 2 Kabupaten/Kota yang seharusnya berada di kategori tinggi namun dikategorikan sebagai sangat tinggi.

c) Decision Tree

Hasil dan pembahasan Decision Tree menunjukkan performa model dalam mengklasifikasikan Kabupaten/Kota ke dalam berbagai kategori prevalensi stunting. Model berhasil mengkategorikan sejumlah Kabupaten/Kota secara akurat ke dalam kategori rendah, menengah, tinggi, dan sangat tinggi.



Gambar 15. Confusion Matrix Decision Tree

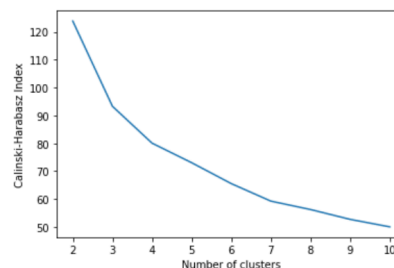
Pada Gambar 15 Dari total 104 Kabupaten/Kota yang dianalisis, model berhasil mengkategorikan sejumlah Kabupaten/Kota secara akurat ke dalam kategori rendah, menengah, tinggi, dan sangat tinggi. Lebih khusus lagi, dari Gambar 14, dapat diamati bahwa model berhasil mengklasifikasikan 43 Kabupaten/Kota ke dalam kategori rendah, 31 Kabupaten/Kota ke kategori menengah, 22 Kabupaten/Kota ke kategori tinggi, serta 8 Kabupaten/Kota ke kategori sangat tinggi dengan tingkat akurasi yang memuaskan. Hal ini menunjukkan kemampuan model Decision Tree dalam memahami dan membedakan pola-pola yang mendasari prevalensi stunting di tiap wilayah, memberikan wawasan yang berharga untuk upaya-upaya intervensi dan kebijakan yang ditujukan untuk mengatasi masalah stunting di tingkat Kabupaten/Kota.

C. Hasil Validasi dan Evaluasi Model

a) K-Means Clustering

analisis evaluasi kinerja pengelompokan Kabupaten/Kota di Indonesia berdasarkan variabel prediktor prevalensi stunting menggunakan indeks Calinski-Harabasz.

Number of clusters: 2, Calinski-Harabasz Index: 123.85
 Number of clusters: 3, Calinski-Harabasz Index: 93.30
 Number of clusters: 4, Calinski-Harabasz Index: 80.07
 Number of clusters: 5, Calinski-Harabasz Index: 73.06
 Number of clusters: 6, Calinski-Harabasz Index: 65.56
 Number of clusters: 7, Calinski-Harabasz Index: 59.27
 Number of clusters: 8, Calinski-Harabasz Index: 56.26
 Number of clusters: 9, Calinski-Harabasz Index: 52.73
 Number of clusters: 10, Calinski-Harabasz Index: 50.04



Gambar 16. Indeks Calinski-Harabasz

Evaluasi kinerja pengelompokan Kabupaten/Kota di Indonesia berdasarkan variabel prediktor prevalensi stunting dengan menggunakan indeks Calinski-Harabasz menemukan bahwa jumlah optimal kluster adalah dua. Kualitas kinerja yang lebih baik tercermin dari nilai indeks yang tinggi. Perlu ditekankan bahwa diperlukan setidaknya dua cluster untuk mengelompokkan observasi, sehingga temuan ini mendukung pandangan peneliti bahwa penerapan metode pengelompokan k-means mungkin tidak diperlukan dalam konteks penelitian ini.

b) Support Vector Machine (SVM)

Pembuatan model SVM telah dilakukan empat kali dengan menggunakan kernel linear, rbf, sigmoid, dan poly. Detail evaluasi kinerja masing-masing model dapat dirujuk pada Tabel 2.

	precision	recall	f1-score	support
1	0.78	0.91	0.84	43
2	0.43	0.39	0.41	31
3	0.38	0.41	0.39	22
4	0.00	0.00	0.00	8
accuracy			0.58	104
macro avg	0.40	0.43	0.41	104
weighted avg	0.53	0.58	0.55	104
(kernel rbf)				
	precision	recall	f1-score	support
1	0.93	1.00	0.97	43
2	0.96	0.84	0.90	31
3	0.83	0.86	0.84	22
4	0.75	0.75	0.75	8
accuracy			0.90	104
macro avg	0.87	0.86	0.86	104
weighted avg	0.91	0.90	0.90	104
(kernel linear)				
	precision	recall	f1-score	support
1	0.71	0.81	0.76	43
2	0.32	0.39	0.35	31
3	0.44	0.32	0.37	22
4	1.00	0.25	0.40	8
accuracy			0.54	104
macro avg	0.62	0.44	0.47	104
weighted avg	0.56	0.54	0.53	104
(kernel sigmoid)				
	precision	recall	f1-score	support
1	0.49	0.98	0.66	43
2	0.36	0.13	0.19	31
3	0.68	0.14	0.22	22
4	0.00	0.00	0.00	8
accuracy			0.47	104
macro avg	0.36	0.31	0.37	104
weighted avg	0.44	0.47	0.38	104
(kernel poly)				

Tabel 3. Perbandingan Kebaikan Model SVM

Setelah melakukan analisis terhadap kualitas model SVM berdasarkan berbagai kernel yang digunakan, ditemukan bahwa model yang memanfaatkan kernel

linear menunjukkan performa terbaik dengan akurasi sebesar 90%, sedangkan penggunaan kernel rbf hanya menghasilkan akurasi sebesar 58%, kernel sigmoid sebesar 54%, dan kernel poly sebesar 47%. Hasil ini berbeda dengan temuan sebelumnya yang menyarankan bahwa kernel rbf lebih efektif [21]. Kemungkinan perbedaan ini terjadi karena struktur data yang memiliki kecenderungan linear, sehingga kernel linear menjadi opsi yang paling tepat. Kernel linear memang dirancang khusus untuk data yang dapat dipisahkan dengan garis lurus. Se jauh ini, kernel RBF lebih direkomendasikan untuk data yang memerlukan representasi di dimensi yang lebih tinggi agar dapat dipisahkan dengan lebih efisien.

Pada aspek precision, model menunjukkan kinerja sangat baik dengan mencapai 93% untuk klasifikasi tingkat rendah dan 96% untuk klasifikasi tingkat menengah. Meski demikian, recall pada kategori sangat tinggi merupakan yang paling rendah, yaitu sebesar 75%. Namun, F1-score model secara keseluruhan mencapai 90%, menandakan adanya keseimbangan optimal antara precision dan recall. Secara keseluruhan, model mampu dengan efektif mengidentifikasi kasus stunting tingkat rendah, namun ada potensi untuk meningkatkan deteksi kasus pada tingkat menengah. Dengan demikian, dapat disimpulkan bahwa model ini efektif dalam mengenali kasus stunting pada tingkat rendah dan menunjukkan performa yang memadai untuk tingkat menengah.

c) Decision Tree

Evaluasi hasil klasifikasi berdasarkan kategori stunting WHO (World Health Organization) menunjukkan bahwa model Decision Tree bekerja dengan baik.

	precision	recall	f1-score	support
1	1.00	1.00	1.00	43
2	1.00	1.00	1.00	31
3	1.00	1.00	1.00	22
4	1.00	1.00	1.00	8
accuracy			1.00	104
macro avg	1.00	1.00	1.00	104
weighted avg	1.00	1.00	1.00	104

Gambar 18. Kebaikan Model Decision Tree

Hasil evaluasi model *Decision Tree* menunjukkan kinerja yang sangat memuaskan dalam klasifikasi faktor-faktor stunting di Indonesia ke dalam empat kelas berbeda, yaitu rendah, menengah, tinggi, dan sangat tinggi. Dari metrik *precision*, *recall*, dan *f1-score* yang masing-masing mencapai nilai 1.00 untuk setiap kelas, hal ini menandakan bahwa model memiliki kemampuan yang sempurna dalam mengidentifikasi dan membedakan antara kelas-kelas tersebut tanpa kesalahan.

Secara rinci, presisi dan *recall* untuk semua kelas mencapai nilai 1.00. Ini artinya, semua data yang masuk ke setiap kelas berhasil diidentifikasi dengan

benar tanpa kesalahan. Selanjutnya, *F1-score* juga mencapai nilai 1.00 untuk setiap kelas, menunjukkan bahwa model ini baik dalam mengenali serta mengingat data dengan sangat baik.

D. Pembahasan Hasil yang Didapatkan

a) Perbandingan Penggunaan Algoritma

Dari hasil evaluasi kinerja menggunakan algoritma Support Vector Machine (SVM) dan Decision Tree, terlihat bahwa keduanya menunjukkan performa yang baik dalam mengklasifikasikan faktor-faktor stunting di Indonesia ke dalam empat kelas yang berbeda: rendah, menengah, tinggi, dan sangat tinggi. Namun, terdapat perbedaan signifikan dalam akurasi dan kemampuan kedua algoritma dalam mengklasifikasikan kelas tersebut.

Pertama, SVM yang menggunakan kernel linear menunjukkan performa terbaik dengan akurasi sebesar 90%. Meskipun demikian, ada penurunan performa secara signifikan saat menggunakan kernel lainnya seperti rbf, sigmoid, dan poly, yang masing-masing hanya mencapai akurasi sekitar 58%, 54%, dan 47%. Meskipun nilai akurasi yang rendah pada kernel lainnya bisa disebabkan oleh struktur data yang cenderung linear, hasil tersebut menunjukkan bahwa SVM mungkin tidak cocok untuk dataset yang kompleks.

Sementara itu, Decision Tree menunjukkan kinerja yang sangat memuaskan dengan presisi, recall, dan F1-score mencapai nilai 1.00 untuk setiap kelas, menandakan kemampuan model dalam mengidentifikasi dan membedakan antara kelas-kelas tersebut tanpa kesalahan. Evaluasi berdasarkan ROC Curve juga menunjukkan kinerja yang sangat baik dengan nilai area untuk setiap kategori mendekati sempurna.

Dengan demikian, meskipun SVM dengan kernel linear memiliki akurasi yang tinggi, Decision Tree menunjukkan keunggulan dalam kemampuan klasifikasi dan interpretasi data faktor-faktor stunting di Indonesia. Keandalan Decision Tree dalam mengklasifikasikan faktor-faktor stunting di Indonesia ke dalam empat kelas dengan akurasi 100% dan tidak adanya kesalahan dalam proses klasifikasi menegaskan bahwa model ini lebih dapat diandalkan dalam analisis dan interpretasi data stunting dibandingkan SVM. Oleh karena itu, dalam konteks penelitian ini, penggunaan Decision Tree lebih disarankan untuk analisis lebih lanjut terhadap faktor-faktor stunting di Indonesia.

b) Pemahaman Pola Stunting di Tingkat Kabupaten

Hasil dan pembahasan Decision Tree menunjukkan bahwa model secara efektif mengklasifikasikan

Kabupaten/Kota ke dalam berbagai kategori prevalensi stunting dengan akurasi yang memuaskan. Dari total 104 Kabupaten/Kota yang dianalisis, model mampu mengkategorikan secara tepat ke dalam kategori rendah, menengah, tinggi, dan sangat tinggi menunjukkan bahwa model mampu mengklasifikasikan 43 Kabupaten/Kota ke dalam kategori rendah, 31 Kabupaten/Kota ke kategori menengah, 22 Kabupaten/Kota ke kategori tinggi, serta 8 Kabupaten/Kota ke kategori sangat tinggi dengan tingkat akurasi yang memuaskan.

Pentingnya hasil ini terletak pada kemampuan model Decision Tree dalam memahami dan membedakan pola-pola yang mendasari prevalensi stunting di setiap wilayah. Dengan memperoleh wawasan yang mendalam mengenai faktor-faktor yang memengaruhi tingkat stunting di tingkat Kabupaten/Kota, para pembuat kebijakan dan pelaku intervensi kesehatan dapat merancang strategi yang lebih terarah dan efektif dalam menanggulangi masalah stunting. Oleh karena itu, hasil ini memberikan kontribusi yang berharga dalam upaya pencegahan dan penanggulangan stunting di tingkat lokal, memungkinkan adopsi kebijakan yang lebih tepat sasaran dan solusi intervensi yang lebih efisien.

c) Analisis Faktor Penyebab Stunting di Indonesia

Hasil analisis kami menunjukkan bahwa terdapat beberapa variabel yang memiliki korelasi signifikan dengan prevalensi stunting di Indonesia. Variabel "BBLR" atau Berat Badan Lahir Rendah, secara positif berkorelasi signifikan dengan prevalensi stunting. Ini menunjukkan bahwa bayi dengan berat badan lahir rendah memiliki risiko lebih tinggi untuk mengalami stunting pada tahap perkembangan selanjutnya.

Di sisi lain, variabel seperti "Persalinan FASYANKES", "K4", "KF Lengkap", "mCPR", dan "A 6-11" memiliki korelasi negatif signifikan dengan prevalensi stunting. Hal ini menunjukkan bahwa peningkatan akses dan kualitas pelayanan kesehatan maternal, seperti persalinan di fasilitas kesehatan yang memadai, cakupan imunisasi ibu hamil (K4), cakupan keluarga yang mendapatkan Kunjungan Neonatal Pertama (KF Lengkap), cakupan penggunaan alat kontrasepsi (mCPR), dan asupan makanan tambahan bagi balita usia 6-11 bulan dapat membantu mengurangi risiko stunting.

Dengan demikian, hasil analisis ini menyoroti pentingnya peningkatan akses dan kualitas pelayanan kesehatan maternal dan perawatan bayi baru lahir, serta pentingnya promosi pola makan yang sehat pada balita usia 6-11 bulan sebagai strategi utama dalam upaya pencegahan stunting di Indonesia.

IV. SIMPULAN DAN SARAN

A. Simpulan

Penelitian ini menilai keberhasilan berbagai teknik analisis data, seperti K-Means, Support Vector Machine (SVM), dan Decision Tree, dalam mengidentifikasi tingkat stunting di 520 Kabupaten/Kota di Indonesia. Hasil pemilihan faktor prediktor yang signifikan menunjukkan bahwa variabel "BBLR" memiliki hubungan positif yang signifikan dengan prevalensi stunting. Sebaliknya, variabel "Persalinan FASYANKES", "K4", "KF Lengkap", "mCPR", dan "A 6-11" menunjukkan hubungan negatif yang signifikan dengan prevalensi stunting di berbagai Kabupaten/Kota.

Dari hasil analisis, K-Means tidak menghasilkan perbedaan yang nyata antara kluster yang dibentuk, sehingga analisisnya kurang mendalam. Sebaliknya, SVM mencapai tingkat keakuratan 88%, terutama saat menggunakan kernel linear. Decision Tree, di sisi lain, mencatat keakuratan sempurna 100%, dan secara umum, metode Decision Tree mengungguli SVM dalam pengklasifikasian. Penggunaan teknik K-Means berhasil mengelompokkan 520 Kabupaten/Kota menjadi dua kluster. Kluster pertama menunjukkan angka prevalensi stunting yang lebih rendah dibandingkan kluster kedua. Namun, metode pengelompokan ini belum optimal karena belum ada kriteria yang tegas membedakan kedua kluster tersebut.

B. Saran

Implementasi Program Efektif: Disarankan kepada pemerintah Indonesia untuk terus mengimplementasikan program-program yang telah terbukti efektif dalam mengurangi prevalensi stunting, khususnya di Kabupaten/Kota yang masuk dalam kluster 1.

Eksplorasi Teknik Lain: Untuk penelitian mendatang, sebaiknya eksplorasi lebih lanjut dilakukan terhadap teknik pengelompokan atau algoritma machine learning alternatif untuk mendapatkan interpretasi yang lebih mendalam.

Pengembangan Rencana Intervensi: Hasil dari penelitian ini dapat menjadi pondasi bagi pengembangan rencana intervensi kesehatan yang lebih spesifik dan adaptif di setiap daerah, sehingga penanganan stunting dapat lebih tepat sasaran dan efektif.

REFERENSI

- [1] Kementerian Kesehatan Republik Indonesia, "Warta KESMAS," 2023.
- [2] H. Dasman, "Empat Dampak Stunting Bagi Anak dan Negara Indonesia," *The Conversation*, 2019.
- [3] Kementrian PPN/Bappenas, "Peraturan Presiden Nomor 18 Tahun 2020," 2020. [Online]. Available: <https://jdih.bappenas.go.id/peraturan/detailperaturan/1037>.
- [4] D. H. Manongga, S. P. Manongga and H. Hendry, "PREDICTING A STUNTING PREVALENCE," *International Journal of Innovative Computing, Information, and Control*, vol. 19, 2023.
- [5] D. Septyha, K. Rahayu, S. Rabbani, V. Fitria, Rahmaddeni, Y. Irawan and R. Hayami, "Implementasi Algoritma Decision Tree dan Support Vector Machine untuk Klasifikasi Penyakit Kanker Paru," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 3, pp. 15-19, 2023.
- [6] S. Syahrial, R. Ilham, Z. F. Asikin and S. S. I. Nurdin, "Stunting Classification in Children's Measurement Data Using Machine Learning Models," *Journal La Multiapp*, 2022.
- [7] D. Widyawati, A. Faradibah and P. L. L. Belluano, "Comparison Analysis of Classification Model Performance in Lung Cancer Prediction Using Decision Tree, Naive Bayes, and Support Vector Machine," *Indonesian Journal of Data and Science*, vol. 4, 2023.
- [8] G. V. S. Charan and N. S. Kumar, "Analysis and Comparison for Innovative Prediction Technique of COVID-19 using Decision Tree Algorithm over the Support Vector Machine Algorithm with Improved Accuracy," *Cardiometry*, 2022.
- [9] O. N. Chilyabanyama, R. Chilengi, M. Simuyandi, C. C. Chisenga, M. Chirwa, K. Hamusonde, R. K. Saroj, N. T. Iqbal, I. Ngaruye and S. Bosomprah, "Performance of Machine Learning Classifiers in Classifying Stunting among Under-Five Children in Zambia," *Children*, 2022.
- [10] K. N. S. S. Raj and K. Thinakaran, "Prediction of Heart Disease using Decision Tree over Logistic Regression using Machine Learning with Improved Accuracy," *Cardiometry*, no. 25, pp. 1514-1519, 2022.
- [11] A. Nugroho, H. L. H. S. Warnars, F. L. Gaol and T. Matsuo, "Trend of Stunting Weight for Infants and Toddlers Using Decision Tree," *IAENG: International Journal of Applied Mathematics*, 2022.
- [12] M. Ula, A. F. Ulva, Mauliza, M. A. Ali and Y. R. Said, "APPLICATION OF MACHINE LEARNING IN DETERMINING THE CLASSIFICATION OF CHILDREN'S NUTRITION WITH DECISION TREE," *JUTIF: Jurnal Teknik Informatika*, vol. 3, 2022.
- [13] A. Byna, "Comparative Analysis of Machine Learning Algorithms for Classification about Stunting Genesis," pp. 74-84, 2019.
- [14] Scikit Learn, "Selecting the number of clusters with silhouette analysis on KMeans clustering," 2023. [Online]. Available: https://scikit-learn.org/stable/auto_examples/cluster/plo_t_kmeans_silhouette_analysis.html. [Accessed 2023].
- [15] D. Abdullah, Susilo, A. S. Ahmar, R. Rusli and R. Hidayat, ["https://link.springer.com/article/10.1007/s11135-021-01176-w,"](https://link.springer.com/article/10.1007/s11135-021-01176-w) [Online]. Available: <https://link.springer.com/article/10.1007/s11135-021-01176-w>.
- [16] N. Pratiwi and Y. Setyawan, "ANALISIS AKURASI DARI PERBEDAAN FUNGSI KERNEL DAN COST PADA SUPPORT VECTOR MACHINE STUDI KASUS KLASIFIKASI CURAH HUJAN DI JAKARTA," *Journal of Fundamental Mathematics and Applications (JFMA)*, vol. 4, 2021.
- [17] H. Harafani and R. S. Wahono, "Optimasi Parameter pada Support Vector Machine Berbasis Algoritma Genetika Untuk Estimasi Kebakaran Hutan," *Journal of Intelligent Systems*, vol. 1, 2015.
- [18] MathWorks, "Support Vector Machine," 2023. [Online]. Available: <https://www.mathworks.com/discovery/support-vector-machine.html>.
- [19] World Health Organization, "Nutrition Landscape Information System (NLIS)," 2018. [Online]. Available: <https://apps.who.int/nutrition/landscape/help.aspx?menu=0&helpid=391&lang=EN>.
- [20] Columbia Mailman School of Public Health, 2023. [Online]. Available: <https://www.publichealth.columbia.edu/research/population-health-methods/k-means-cluster-analysis>.
- [21] A. Z. Praghakusma and N. Charibaldi, "Komparasi Fungsi Kernel Metode Support Vector Machine untuk Analisis Sentimen Instagram dan Twitter (Studi Kasus: Komisi Pemberantasan Korupsi)," *Jurnal Sarjana Teknik Informatika*, vol. 9, pp. 33-42, 2021.
- [22] Badan Pusat Statistik, "Indeks Khusus Penanganan Stunting Menurut Provinsi," 2022.

PEMBAGIAN TUGAS KELOMPOK 1 :**Rosa Aulia Wangi -00000069798**

- Menulis bagian metodologi artikel jurnal
- Membuat bagian data modelling dan explorasi data pada file ipynb

Albertus Christian Wahyu -00000068921

- Menulis bagian kesimpulan dan saran artikel jurnal
- Membuat bagian data modelling pada file ipynb

Evangeline Suciadi - 00000068887

- Menulis hasil dan pembahasan artikel jurnal
- Membuat bagian preparasi data dan evaluasi pada file ipynb

Nurfajriah Oktaviani - 00000068637

- Menulis hasil dan pembahasan artikel jurnal
- Membuat bagian data understanding dan evaluasi pada file ipynb

Justin Stephen -00000072126

- Membuat bagian pendahuluan pada artikel jurnal
- Membuat bagian abstrak artikel jurnal