

# DIAGNOSA STUNTING BERDASARKAN GEJALA MEDIS MENGUNAKAN ALGORITMA NAIVE BAYES, SVM DAN K-NN

Aris Subadi<sup>1</sup>, Kusri<sup>2</sup>

<sup>1,2</sup> Magister Informatika, Universitas AMIKOM Yogyakarta  
<sup>1</sup>aris.subadi@students.amikom.ac.id., <sup>2</sup>kusri@amikom.ac.id

---

## Abstrak

Stunting merupakan masalah kesehatan global yang signifikan, mempengaruhi pertumbuhan dan perkembangan anak-anak di seluruh dunia, termasuk Indonesia. Penelitian ini bertujuan untuk mengimplementasikan dan mengevaluasi kinerja algoritma machine learning dalam mendiagnosa stunting berdasarkan gejala medis. Tiga algoritma utama yang digunakan adalah Naive Bayes, Support Vector Machine (SVM), dan K-Nearest Neighbor (K-NN). Dataset yang digunakan dalam penelitian ini terdiri dari data gejala medis yang relevan dengan stunting, yang dikumpulkan dari berbagai sumber terpercaya. Data tersebut diproses melalui tahap pembersihan, normalisasi, dan pembagian menjadi set pelatihan dan set pengujian. Setelah data diproses, setiap algoritma machine learning dilatih menggunakan set pelatihan dan kemudian diuji menggunakan set pengujian untuk mengevaluasi kinerjanya. Hasil evaluasi menunjukkan bahwa algoritma Naive Bayes memberikan skor akurasi tertinggi sebesar 90%, diikuti oleh SVM dengan akurasi 86,56%, dan K-NN dengan akurasi 79,43%. Analisis hasil menunjukkan bahwa Naive Bayes adalah algoritma yang paling efektif untuk diagnosa stunting berdasarkan gejala medis, karena memberikan akurasi tertinggi dan efisiensi dalam komputasi. SVM juga menunjukkan kinerja yang kuat dalam menangani data berdimensi tinggi, sementara K-NN memberikan hasil yang memuaskan meskipun dengan akurasi yang lebih rendah. Kesimpulan penelitian ini adalah bahwa Naive Bayes dapat menjadi pilihan utama untuk implementasi dalam sistem diagnosa stunting, dengan mempertimbangkan keunggulan dalam akurasi dan efisiensi. Penelitian ini memberikan kontribusi signifikan dalam bidang kesehatan dengan menyediakan model prediksi yang efektif untuk diagnosa stunting, serta memberikan rekomendasi praktis bagi tenaga medis dalam memanfaatkan teknologi machine learning. Validasi eksternal diusulkan sebagai langkah lanjutan untuk memastikan generalisasi model.

**Kata kunci:** stunting, machine learning, naive bayes, support vector machine, k-nearest neighbor, diagnosa medis.

---

## 1. Pendahuluan

Isu stunting menjadi salah satu fokus perhatian utama di dunia pada beberapa tahun terakhir ini. Menurut data yang dikumpulkan World Health Organization (WHO) pada tahun 2020, jumlah balita yang mengalami stunting di seluruh dunia mencapai 150,8 juta atau setara dengan 22,2% dari total populasi balita. WHO mengidentifikasi lima daerah dengan tingkat prevalensi stunting tertinggi, termasuk Indonesia yang tergolong dalam wilayah Asia Tenggara dengan prevalensi sebesar 36,4% (Rita Kirana, Aprianti, 2022). Meskipun angka prevalensi mengalami penurunan di tahun 2021 menjadi 24,4% dan tahun 2022 menjadi 21,6%, (Kesehatan, 2022) namun menurut WHO jika prosentase masih diatas 20% maka masalah kesehatan stunting itu masih dianggap kronis. (Car et al., 2023) Selain itu, persentase tersebut masih di bawah target nasional yang ditetapkan yaitu 14% pada tahun 2024, (Perpres, 2020) sehingga pemerintah mendorong seluruh masyarakat untuk memberikan prioritas pada isu kesehatan ini.

Stunting adalah kondisi di mana pertumbuhan dan perkembangan bayi (0-11 bulan) dan anak balita

(12-59 bulan) terhambat karena kekurangan gizi kronis, terutama terjadi dalam 1.000 hari pertama kehidupan. Sedangkan stunting menurut Rahmadhita (Sandi & Rahmadhita, 2020) merupakan suatu kondisi dimana anak mengalami gagal tumbuh, baik dari tubuh maupun otak. Salah satu ciri utamanya adalah tinggi badan yang tidak sesuai dengan usia. (Arnita, S., Rahmadhani, D. Y., & Sari, 2020)

Dampak stunting pada balita sangat beragam karena dapat memengaruhi pertumbuhan dan perkembangannya. Hal ini dapat mengganggu pertumbuhan balita terkait tinggi dan berat badannya, yang berdampak pada keterlambatan dalam kemampuan berjalan dan motorik yang kurang optimal. Selain itu, stunting juga dapat mempengaruhi kemampuan belajar balita karena dapat menyebabkan penurunan IQ dibandingkan dengan anak sebaya mereka. Oleh karena itu, stunting memiliki dampak yang signifikan terhadap pertumbuhan dan perkembangan anak, baik secara fisik maupun kognitif. (Setiawan, E., & Machmud, 2018; Alifariki, 2020) Stunting juga berdampak jangka panjang pada kehidupan dewasa, termasuk performa akademik buruk, rendahnya tingkat pendidikan, dan pendapatan minim. Anak-anak

stunting rentan terhadap penyakit dan risiko overweight serta obesitas, yang meningkatkan kemungkinan penyakit degeneratif serius. Stunting juga mencerminkan rendahnya kualitas sumber daya manusia suatu negara, mengurangi kemampuan kognitif, produktivitas, dan kesehatan masyarakat, sehingga berpotensi menyebabkan kerugian ekonomi jangka panjang bagi negara. (Trihono., Trihono and Atmarita., Atmarita and Tjandrarini., Dwi Hapsari and Irawati., Anies and Nurlinawati., Iin and Utami., Nur Handayani and Tejayanti., 2015; Nugroho et al., 2021)

Anak yang mengalami stunting dapat dikenali dari tinggi atau panjang tubuhnya yang berada di bawah -2 SD berdasarkan tabel Z-Score. (Damanik, S. M., Sitorus, E., & Mertajaya, 2021) Periode 1.000 hari pertama kehidupan dianggap sebagai fase kritis yang akan memiliki dampak jangka panjang dan berulang dalam siklus kehidupan seseorang. (Panigoro, 2020; Fauziah et al., 2023) menjelaskan bahwa gejala stunting pada balita mencakup berbagai kondisi, antara lain tinggi badan yang pendek atau sangat pendek, tidak mendapatkan ASI eksklusif, dan memiliki riwayat makan yang kurang. Balita stunting sering sakit, memiliki penyakit penyerta, dan berasal dari keluarga dengan status ekonomi rendah. Kenaikan berat badan yang tidak sesuai, riwayat kehamilan ibu dengan Kekurangan Energi Kronis (KEK) atau anemia, serta ibu dengan tinggi badan pendek juga menjadi faktor. Konsumsi suplemen kehamilan yang tidak teratur, lingkungan yang kumuh, imunisasi yang tidak lengkap, serta balita yang mengalami cacangan dan gangguan motorik serta kognitif, turut menjadi gejala stunting. Selain itu, menurut pratiwi dalam Sari (Sari, D., Ningsih, A. D., & Azzahra, 2023) anak-anak stunting sering mengalami kesulitan dalam berkonsentrasi pada pelajaran, menunjukkan ketidakstabilan emosi, dan menunjukkan berkurangnya tingkat keterlibatan dalam lingkungan sekolah.

Penelitian terdahulu dilakukan oleh Haosen, Hang Zhao dan Yi Jiang pada tahun 2023 dengan judul penelitian "Machine Learning Algorithms for Predicting Stunting among Under-Five Children in Papua Nugini." Yang mana hasil penelitiannya menunjukkan bahwa model kombinasi LASSO-XGBoost memiliki kinerja terbaik dalam memprediksi stunting di Papua Nugini, dengan akurasi, presisi, recall, dan skor F1 masing-masing mencapai 0,728, 0,715, 0,628, dan 0,669. Faktor-faktor penting dalam prediksi stunting meliputi wilayah tempat tinggal, usia anak, status keuangan keluarga, dan ukuran lahir anak. (Shen et al., 2023)

Selain itu, terdapat beberapa penelitian sebelumnya yang telah mengeksplorasi berbagai pendekatan menggunakan algoritma seperti K-means, (N, 2023) Support Vector Machine, dan Random Forest (Banurea et al., 2023) dalam mendeteksi stunting pada anak. Namun, belum ada penelitian yang menyelidiki secara komprehensif

algoritma mana yang paling efektif untuk memprediksi stunting. Oleh karena itu, penelitian ini bertujuan untuk mengisi kesenjangan tersebut dengan menyelidiki beberapa algoritma machine learning yang berbeda, termasuk Support Vector Machine, K-Nearest Neighbor, dan Gaussian Naive Bayes. Dengan demikian, diharapkan penelitian ini dapat memberikan wawasan yang berharga bagi masyarakat, terutama dalam bidang kesehatan, dengan memfasilitasi diagnosa yang lebih akurat dan cepat terhadap stunting berdasarkan gejala medis yang dialami oleh anak-anak.

## 2. Metode

Jenis penelitian ini adalah eksperimental dengan pendekatan kuantitatif. Penelitian ini akan menggunakan berbagai algoritma machine learning untuk memprediksi stunting berdasarkan gejala medis. Sifat dari penelitian ini adalah deskriptif karena tujuannya adalah menganalisis model algoritma machine learning yang paling efektif dan menentukan fitur yang paling penting dalam prediksi stunting. Pendekatan ini melibatkan pengumpulan data yang sistematis dan analisis statistik mendalam terhadap data tersebut menggunakan teknik-teknik machine learning.

### 2.1 Metode Pengumpulan Data

Dalam penelitian ini, metode pengumpulan data yang digunakan adalah metode pengumpulan data primer dan sekunder. Data primer merujuk pada data yang dikumpulkan langsung oleh peneliti melalui observasi, wawancara, dan survei. Berikut penjelasan dari data primer yang digunakan dalam penelitian.

1. Peneliti melakukan pengamatan langsung terhadap anak-anak yang didiagnosis stunting di fasilitas kesehatan, seperti rumah sakit atau puskesmas. Observasi ini dapat mencakup pengamatan fisik, perilaku, dan kondisi kesehatan anak
2. Peneliti melakukan wawancara dengan orang tua atau pengasuh anak untuk mengumpulkan informasi tentang riwayat kesehatan, pola makan, dan faktor-faktor sosial ekonomi yang mungkin berkontribusi terhadap stunting. Wawancara ini dapat dilakukan secara terstruktur atau semi-terstruktur untuk mendapatkan data yang lebih mendalam.
3. Peneliti menyebarkan kuesioner kepada orang tua atau pengasuh anak untuk mengumpulkan data kuantitatif mengenai berbagai aspek yang berkaitan dengan kesehatan anak, seperti frekuensi makan, asupan gizi, dan kebiasaan kesehatan. Kuesioner ini dirancang untuk mengumpulkan informasi yang spesifik dan relevan dengan tujuan penelitian.

Data sekunder merupakan data yang telah terkumpulkan sebelumnya oleh pihak lain dan digunakan untuk tujuan penelitian. Data sekunder ini

diambil dari sumber data public yaitu dari Satu Data Indonesia (SDI) dengan tanggal publikasi 13 oktober 2016.

**2.2 Metode Analisis Data**

Metode analisis data yang digunakan dalam penelitian ini adalah analisis statistik multivariat, yang melibatkan penggunaan berbagai teknik machine learning seperti, SVM, KNN, dan Gaussian Naive Bayes. Data gejala medis anak-anak yang telah terdiagnosis stunting akan dianalisis menggunakan algoritma-algoritma tersebut untuk mengidentifikasi pola-pola dan hubungan yang relevan dengan risiko stunting. Analisis statistik multivariat ini akan memungkinkan peneliti untuk mengevaluasi kinerja relatif dari setiap algoritma dan mengidentifikasi fitur-fitur yang paling berpengaruh dalam prediksi stunting.

**2.3 Alur Penelitian**

Langkah pertama dalam penelitian ini adalah identifikasi masalah. Langkah ini bertujuan untuk menentukan tujuan penelitian dan masalah yang akan dipecahkan terkait diagnosa stunting menggunakan algoritma machine learning. Identifikasi masalah yang jelas dan spesifik sangat penting untuk memberikan arah yang tepat dalam seluruh proses penelitian.

Setelah masalah diidentifikasi, langkah selanjutnya adalah pengumpulan data. Data medis yang relevan dengan gejala stunting dikumpulkan dari sumber terpercaya seperti rumah sakit, dinas terkait atau database kesehatan. Data ini mencakup berbagai informasi medis yang dibutuhkan untuk mendiagnosa stunting dengan akurat.

Tahap berikutnya adalah preprocessing data. Pada tahap ini, data yang telah dikumpulkan dibersihkan dari nilai yang hilang atau tidak relevan, dinormalisasi untuk memastikan skala yang konsisten, dan dibagi menjadi set pelatihan dan set pengujian. Proses ini penting untuk memastikan bahwa data siap untuk dianalisis dan digunakan dalam model machine learning.

Setelah data diproses, langkah selanjutnya adalah implementasi algoritma machine learning. Dalam penelitian ini, tiga algoritma utama yang digunakan adalah Naive Bayes, Support Vector Machine (SVM), dan K-Nearest Neighbor (K-NN). Algoritma-algoritma ini diterapkan pada data pelatihan untuk membangun model prediksi yang kemudian akan digunakan untuk mendiagnosa stunting.

Kemudian, model yang telah dibangun diuji menggunakan set pengujian dan kinerjanya dievaluasi. Evaluasi kinerja dilakukan dengan menggunakan metrik yang relevan seperti akurasi, presisi, recall, dan F1-score. Langkah ini bertujuan untuk mengukur seberapa baik setiap algoritma dalam memprediksi diagnosa stunting berdasarkan data yang diberikan.

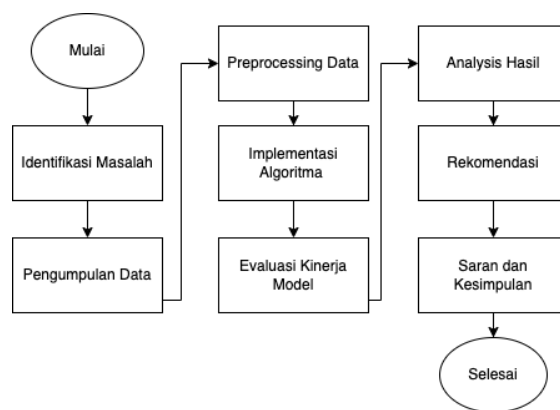
Selanjutnya, hasil evaluasi dianalisis untuk menentukan algoritma terbaik dan menilai signifikansi perbedaan kinerja di antara algoritma yang digunakan. Analisis ini memberikan wawasan mengenai kekuatan dan kelemahan masing-masing algoritma dalam konteks diagnosa stunting.

Berdasarkan analisis tersebut, kesimpulan dan rekomendasi disusun. Kesimpulan ini merangkum temuan utama dari penelitian dan memberikan rekomendasi mengenai penggunaan algoritma yang paling efektif dalam diagnosa stunting. Rekomendasi ini penting untuk memberikan panduan praktis bagi tenaga medis dan peneliti lain yang tertarik untuk menerapkan machine learning dalam diagnosa stunting.

Sebagai langkah opsional, validasi eksternal dapat dilakukan dengan menguji model yang dibangun pada data baru dari sumber yang berbeda. Validasi ini bertujuan untuk memastikan bahwa model yang dibangun memiliki generalisasi yang baik dan dapat digunakan pada data lain di luar set data yang digunakan untuk pelatihan dan pengujian awal.

Terakhir, hasil penelitian disusun dalam bentuk artikel ilmiah dan dipublikasikan di jurnal yang relevan. Penulisan dan publikasi ini bertujuan untuk menyebarkan temuan penelitian kepada komunitas ilmiah dan praktisi di bidang kesehatan, sehingga hasil penelitian dapat memberikan kontribusi nyata dalam mendiagnosa dan menangani stunting secara lebih efektif.

Dengan mengikuti alur penelitian ini, proses penelitian akan dilakukan secara sistematis dan komprehensif, memastikan bahwa hasil yang diperoleh valid dan dapat diandalkan untuk aplikasi praktis dalam mendiagnosa stunting berdasarkan gejala medis. Alur penelitian dapat dilihat pada Gambar 1.



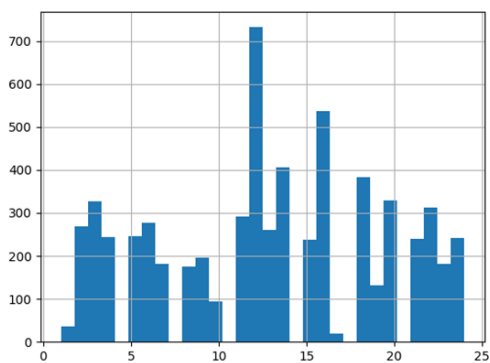
Gambar 1. Alur Penelitian

**3. Hasil dan Pembahasan**

Disini data yang digunakan adalah data laporan stunting, di dalamnya terdapat dua jenis status stunting yaitu “Balita Pendek” dan “Balita Sangat Pendek” yang berada pada kolom TB/U atau Tinggi

Badan/Umur. Total ada 6.342 jumlah data, dimana data tersebut berasal dari balita berusia mulai 2 bulan sampai dengan 24 bulan. Di dalam data tersebut terdapat 16 jenis gejala dimana gejala tersebut akan mempengaruhi status dari balita tersebut, 16 gejala yang ada menggunakan huruf "T" dan "F" dimana "T" mempunyai maksud "true" yang berarti balita tersebut mempunyai gejala pada bagian tersebut dan "F" yang berarti "false", memberitahukan bahwa balita tidak mempunyai gejala pada bagian tersebut. Data berupa nama balita kemudian jenis kelamin dari balita tersebut, setelah itu terdapat berat badan saat lahir dan tinggi badan saat balita tersebut lahir beserta provinsi dari balita tersebut berasal, data dari balita berasal dari provinsi Banten. Kemudian terdapat data berupa usia, berat badan dan tinggi pengukuran terakhir dan beberapa kondisi dari balita tersebut.

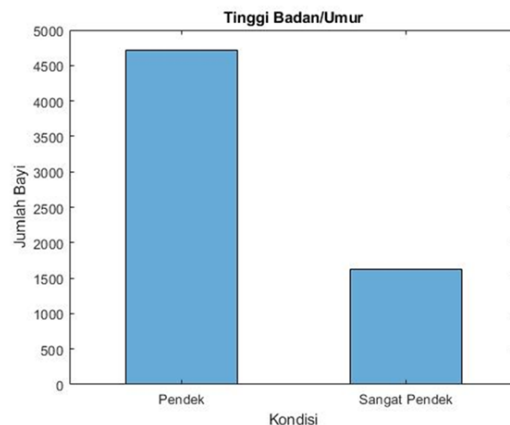
Setelah data tersedia, maka langkah selanjutnya dianalisa data mentah diatas untuk mengetahui isi data tersebut secara lebih terperinci, sehingga bisa ditentukan apakah data tersebut bisa digunakan untuk data dalam machine learning atau tidak. Hal pertama yang sudah diketahui sebelumnya bahwa ada 6.342 baris data balita, selanjutnya akan dianalisa umur yang di pakai oleh data mentah tersebut tersebut, gambar histogram untuk umur dapat dilihat pada Gambar 2.



Gambar 2. Histogram Umur

Pada gambar tersebut dapat diketahui bahwa rentang umur balita dari data yang akan digunakan adalah umur 1-24 bulan, dan dominasi umur balita adalah 12 bulan atau 1 tahun dengan jumlah data sekitar 730, dilanjutkan dengan umur 16 bulan yang memiliki data sekitar 540 dan untuk berikutnya adalah data umur dengan umur balita 14 bulan dengan banyak data 405 dan umur 18 bulan dengan data sebanyak 380.

Selanjutnya untuk data paling sedikit adalah data balita umur 17 bulan yaitu sekitar 25 data dan umur 1 bulan sebanyak 30 data, dari sini untuk umur masih bisa dikatakan bahwa variasi data umur masih tersebar merata dan data bisa digunakan untuk machine learning. Sebelumnya juga akan dicek status stunting untuk mengetahui apakah datanya juga bisa dipakai untuk machine learning. Gambar histogram untuk status stunting bisa dilihat pada Gambar 3.

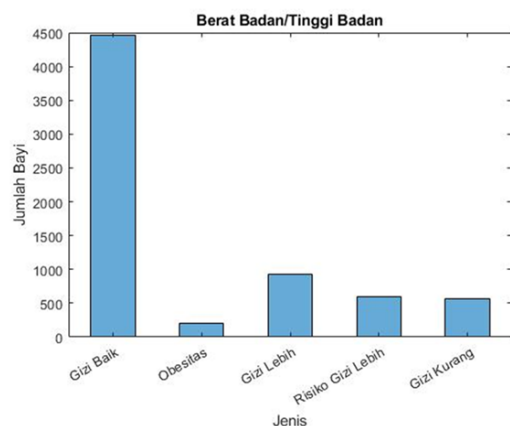


Gambar 3. Histogram Status Stunting

Dari gambar diatas, ada 2 kotak yang mempunyai panjang yang berbeda-beda, dimana secara berurutan adalah 2 status dari data stunting, yaitu status "Pendek" dan "Sangat Pendek", disini ada total 6.343 jumlah data dan data terbanyak adalah kondisi "Pendek" yang mempunyai 4716 data, kemudian "Sangat Pendek" yang memiliki 1625 data.

Dari gambar data sampel, diketahui bahwa balita ada di kategori sangat pendek ketika semua 16 gejala dimiliki oleh balita tersebut yang berarti semua gejala pada balita stunting bernilai T atau true, lebih lanjut diketahui bahwa data pendek sangat dominan karena memiliki 4716 data dari 6343 total data. Hal ini tentu bisa di simpulkan bahwa sebagian besar balita mengalami stunting kategori pendek yang berarti juga bisa diketahui selama beberapa tahun kedepan bisa diprediksi bahwa sebagian besar balita masih akan mengalami stunting dalam kategori pendek ataupun sangat pendek.

Selanjutnya akan dicari kondisi gizi balita dari data tersebut, seperti diketahui bahwa salah satu penyebab stunting adalah masalah gizi, dengan memperbaiki gizi yang diberikan ke balita maka resiko balita tersebut terkena stunting juga akan menurun, dibawah ini adalah data histogram data gizi yang bisa dilihat pada Gambar 4.

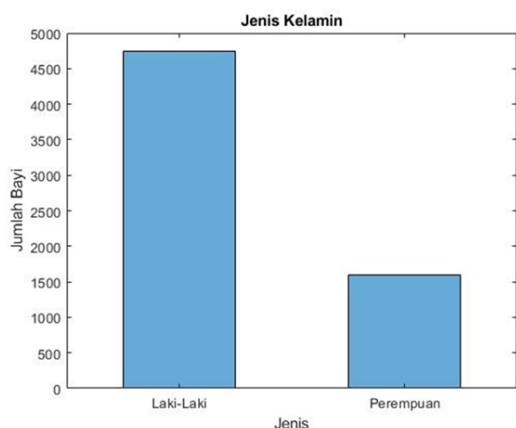


Gambar 4. Histogram Gizi Balita

Dari Gambar diatas, diketahui ada 5 kategori gizi untuk data balita yaitu gizi baik, obesitas, gizi

lebih, resiko gizi lebih dan gizi kurang. Dimana gizi baik mempunyai total 4467 data kemudian obesitas mempunyai 207 dan gizi lebih mempunyai 930 data dilanjutkan resiko gizi lebih 602 data dan terakhir gizi kurang yaitu 573 data. Seperti yang diketahui stunting terjadi karena tinggi badan yang tidak sesuai atau masa pertumbuhan seharusnya bertambah tinggi ke atas bukan ke samping, dari grafik 4.3 balita stunting mempunyai gizi yang baik hanya saja karena beberapa gizi lain tidak dilengkapi saat masa pertumbuhan sehingga membuat balita mengalami stunting.

Pada data diatas diketahui ada data tentang jenis kelamin, karena itu akan dianalisa terlebih dahulu banyaknya data laki-laki dan perempuan yang mengalami stunting dari data diatas, hasil histogram untuk mengathui jumlah pasti laki-laki dan perempuan yang mengalami stunting dari data diatas dapat dilihat pada Gambar 5.

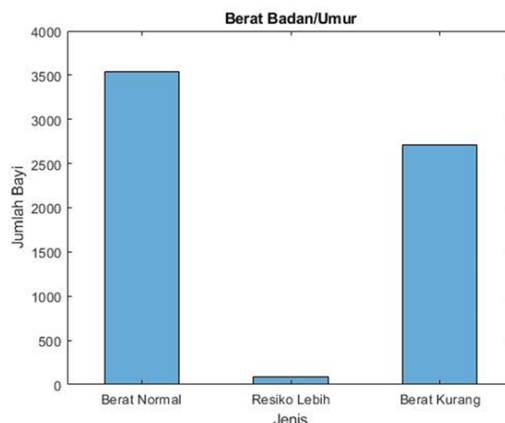


Gambar 5. Histogram Jenis Kelamin

Dari Gambar diatas ada 2 jenis data yaitu laki-laki dan perempuan, dimana laki-laki mempunyai 4748 data balita yang mengalami stunting dan perempuan memiliki 1593 data balita yang mengalami stunting, dari sini dapat diketahui bahwa laki-laki lebih dominan mengalami stunting daripada perempuan yang juga berarti balita berjenis kelamin laki-laki pada masa pertumbuhan 1000 hari pertama memiliki presentase terkena stunting lebih tinggi daripada perempuan jika gizi yang diberikan tidak lengkap pada masa pertumbuhan tersebut.

Hal terakhir yang akan dianalisa dari data diatas adalah berat badan dari balita tersebut, berat badan adalah salah satu faktor yang mempengaruhi stunting, ada masa dimana balita mempunyai berat badan normal tapi mengalami stunting karena kurang tinggi yang berarti pada masa pertumbuhan balita diberi gizi yang cukup sehingga mempunyai berat badan yang normal tetapi kurang lengkap sehingga pertumbuhannya semakin bertambah untuk berat badannya tapi tingginya pertumbuhannya terlalu lambat sehingga dikategorikan mengalami stunting. Ada juga masa dimana balita mempunyai berat badan kurang yang berarti pada waktu pertumbuhan gizinya kurang dan akhirnya menjadi stunting, hasil

histogram untuk berat badan dari data diatas dapat dilihat pada Gambar 6.



Gambar 6. Histogram Berat Badan

Dari gambar diatas dapat diketahui bahwa berat normal mempunyai 3542 data sedangkan resiko lebih berat badan adalah 90 data dan berat kurang adalah 2769 data, dari histogram tersebut menunjukkan bahwa balita yang mengalami stunting sebagian besar memiliki berat badan normal dan memiliki berat badan kurang atau bertubuh kurus, hal ini menunjukkan kurangnya gizi yang membuat balita bertubuh kurus, bisa saja terjadi karena beberapa faktor seperti balita susah makan atau karena faktor ekonomi keluarga yang membuat balita tidak tercukupi kebutuhan gizinya karena terkendala biaya.

Dari analisa data diatas setelah setelah dilakukan pengecekan isi data secara detail, maka langkah selanjutnya adalah menggunakan data diatas sebagai input untuk machine learning yang akan digunakan nantinya. Pada penelitian ini ada beberapa algoritma machine learning yang akan digunakan dimana semua inputnya akan menggunakan data yang sama, ada beberapa algoritma machine learning yang akan digunakan pada penelitian ini diantaranya adalah Gaussian Naïve Bayes, Support Vector Machine (SVM) dan KNN (K-Nearest Neighbor)

Semua algoritma akan dikerjakan menggunakan python dengan presentase data latih dan data uji yang sama, karena seperti diketahui machine learning bekerja dan melakukan kalkulasi pada tahap belajar membutuhkan data latih untuk melatih data, semakin banyak data latih semakin baik tapi semua itu tergantung kompleksitas data yang diberikan, semakin kompleks data tersebut semakin banyak data latih yang dibutuhkan machine learning untuk mencari pola dari data tersebut, begitu juga sebaliknya rendah kesulitan data tersebut algoritma machine learning hanya butuh sedikit data untuk belajar pola dari data tersebut

### 3.1 Gaussian Naïve Bayes

Naive Bayes lebih dikenal dengan Naive Bayes Classification yang berarti algoritma yang biasanya digunakan untuk melakukan pengelompokan suatu

data atau klasifikasi data atau juga digunakan untuk prediksi kelas atau kelompok data. Contoh penggunaan Naive Bayes adalah untuk melakukan prediksi foto, seperti foto kualitas buah, deteksi foto wajah orang dll.

Langkah pertama dalam pembuatan script untuk algoritma Naive Bayes adalah menampilkan histogram data untuk mengetahui detail data berdasarkan kategori. Algoritma Naive Bayes melakukan prediksi berdasarkan kategori, sehingga hasil akhirnya adalah prediksi kategori atau kelas, bukan nilai atau angka. Dataset terdiri dari 6343 data balita dengan dua kategori, yaitu "Pendek" dan "Sangat Pendek".

Setelah data diinput untuk Naive Bayes, diketahui bahwa terdapat 4716 balita dengan status stunting pendek dan 1625 balita dengan status sangat pendek. Ini menunjukkan bahwa data kategori sangat pendek memiliki jumlah data yang lebih sedikit, sedangkan data pendek mendominasi dengan jumlah tiga kali lebih banyak.

Langkah berikutnya adalah mengetahui histogram data training untuk Naive Bayes. Data training akan menggunakan 60% dari total data yang ada, atau sekitar 3805 data balita. Histogram menunjukkan bahwa terdapat 2830 balita dengan status stunting pendek dan 975 balita dengan status sangat pendek. Histogram ini mirip dengan histogram data input, hanya berbeda dalam jumlah data yang digunakan.

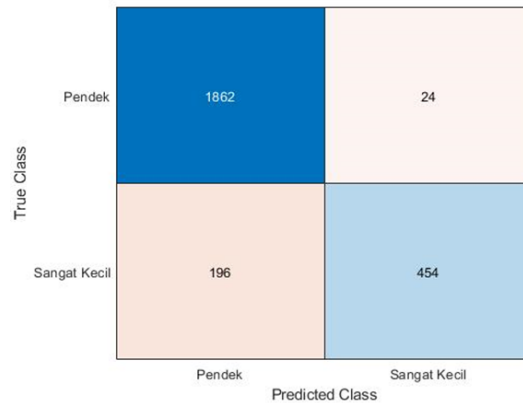
Selanjutnya adalah mengetahui isi data uji, yang merupakan 40% dari total data atau sekitar 2538 data balita. Histogram data uji menunjukkan bahwa terdapat 1886 balita dengan status stunting pendek dan 650 balita dengan status sangat pendek. Bentuk histogram data uji serupa dengan histogram data input dan data training.

Setelah mengetahui isi dari data uji, maka akan masuk langkah selanjutnya yaitu melakukan prediksi berdasar data yang telah disediakan dan sudah dimasukkan di data training dan data uji, Naive Bayes membutuhkan setidaknya 2 jenis data berbeda untuk algoritmanya agar bisa melakukan kalkulasi dengan baik, dengan catatan setiap kategori harus mempunyai variasi nilai, tapi jika salah satu kategori tidak mempunyai variasi nilai, maka kategori lain harus mempunyai variasi nilai untuk menutupi kategori yang tidak memiliki variasi nilai tersebut saat kalkulasi sedang berlangsung.

Hasil pertama dari output prediksi Naive Bayes adalah matrix dari kategori yang dipakai untuk kalkulasi, matrix disini digunakan untuk mengetahui apakah kategori yang dibuat sebelumnya yaitu pendek dan sangat pendek telah sesuai dan terbaca dengan baik tanpa perlu ada kelebihan atau kekurangan kategori. Hasil prediksi matrix Naive Bayes dapat dilihat pada Gambar 7.

Gambar tersebut adalah hasil matrix dari prediksi Naive Bayes, dapat diketahui terdapat 2 kategori yaitu pendek dan sangat kecil yang berarti

kategori sudah sesuai dengan data yang dibuat, disini terdapat total 4 kotak dimana terdapat 4 kotak yang berisi nilai yaitu kelas pendek sebanyak 1862 data balita, kemudian sangat kecil sebanyak 454 data balita dan antara kelas pendek dan sangat pendek sebanyak 24 data balita serta terakhir diantara sangat pendek dan pendek sebanyak 196 data balita.



Gambar 7. Matrix Prediksi Naive Bayes

Matrix diatas berfungsi untuk mengetahui bagaimana Naive Bayes melakukan kalkulasi untuk machine learning, dari gambar diatas dapat diketahui bahwa algoritma belajar dengan baik, diketahui dari kotak matrix berwarna biru tua dan biru muda, itu adalah tempat bagaimana Naive Bayes bisa melakukan klasifikasi kelas dengan jelas, sedangkan yang berwarna merah muda adalah dimana Naive Bayes memiliki kebingungan antara 2 jawaban yang ada atau dengan kata lain kemungkinan mempunyai jawaban lebih dari satu. Jika tidak ada nilai dikotak matrix lain menandakan bahwa algoritma bisa membedakan data setiap kategori dengan jelas. Semakin kompleks data dan banyaknya jenis data maka hasil matrix akan mempunyai sedikit atau banyak data yang berada diantara sisa kotak berwarna biru yang menandakan bahwa beberapa data dari kategori tersebut bisa saja punya 2 jawaban yang berbeda.

Setelah hasil matrix selanjutnya adalah prediksi kategori dari Naive Bayes, sebelumnya saat pembuatan script Naive Bayes untuk parameter hasil prediksi dimasukkan nilai 30, yang berarti nanti Naive Bayes akan melakukan kalkulasi dan melakukan prediksi dan menampilkan 30 data tabel prediksi. Setelah melakukan running script di python, kemudian akan dipanggil hasil prediksi dari Naive Bayes.

Hasil dari sebagian prediksi Naive Bayes dari total 30 hasil yang telah dibuat, disini dapat diketahui bahwa hasil output dari Naive Bayes adalah prediksi berupa nama dari kategori atau kelas tersebut, output dari Naive Bayes berbentuk tabel dari kategori yang dipakai berupa kata bukan angka. Diketahui hasil outputnya berupa kategori atau kelas yang telah dibuat sebelumnya untuk dilakukan prediksi, untuk membuatnya lebih mudah dibaca karena hasil

outputnya berupa kata, yaitu “pendek” dan “sangat pendek”

Hasil dari prediksi Naïve Bayes, untuk kategori pendek terdapat 24 balita yang berada pada status stunting pendek dan sangat pendek terdapat 6 balita. Dapat diketahui bahwa Naive Bayes melakukan prediksi berdasar kategori atau kelas dimana nantinya outputnya berupa kelas atau kategori tersebut, yang berarti algoritma Naive Bayes memiliki prediksi berupa range nilai karena kategori atau kelas harus mempunyai variasi nilai atau range nilai untuk Naive Bayes melakukan prediksi.

### 3.2 Support Vector Machine (SVM)

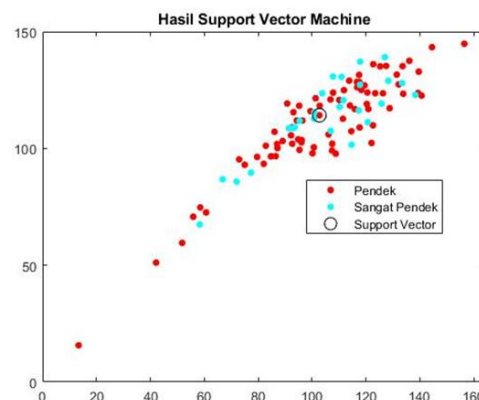
Support Vector Machine atau SVM adalah salah satu algoritma machine learning Dimana algoritma ini juga dikenal dengan nama Support Vector Machine Classification yang berarti algoritma ini bekerja dengan cara melakukan klasifikasi berdasar data latih dan data uji yang digunakan saat proses pelatihan berlangsung.

Sesuai dengan namanya Support Vector Machine yang berarti algoritma ini menggunakan vector sebagai caranya dalam melakukan kalkulasi, sebelumnya akan dicari tahu terlebih dahulu yang pertama adalah data yang akan dipakai oleh Support Vector Machine ini.

Dapat diketahui bahwa ada total 4716 balita dengan status stunting pendek, kemudian ada sekitar 1625 balita masuk dalam status sangat pendek. Dari sini dapat diketahui bahwa data kategori sangat pendek memiliki jumlah data yang sedikit sedangkan data pendek mendominasi dengan 3x lebih banyak jumlahnya dari data sangat pendek. langkah berikutnya adalah melihat histogram data training dengan rasio 60% dari total data. Data training terdiri dari 2830 balita stunting pendek dan 975 balita stunting sangat pendek.

Selanjutnya, data uji yang merupakan 40% dari total data dianalisis. Data uji terdiri dari 1886 balita stunting pendek dan 650 balita stunting sangat pendek. Secara keseluruhan, data training dan data uji memiliki distribusi kategori yang serupa dengan data input, hanya berbeda dalam jumlah balita yang digunakan.

Setelah mengetahui isi dari data uji, maka akan masuk langkah selanjutnya yaitu melakukan prediksi berdasar data yang telah disediakan dan sudah dimasukkan di data training dan data uji, Support Vector Machine membutuhkan setidaknya 2 jenis data berbeda untuk algoritmanya agar bisa melakukan kalkulasi dengan baik, dengan catatan setiap kategori harus mempunyai variasi nilai, tapi jika salah satu kategori tidak mempunyai variasi nilai, maka kategori lain harus mempunyai variasi nilai untuk menutupi kategori yang tidak memiliki variasi nilai tersebut saat kalkulasi sedang berlangsung. Hasil dari Support Vector Machine dapat dilihat pada Gambar 8.



Gambar 8. Hasil Support Vector Machine

Gambar diatas adalah hasil dari Support Vector Machine, karena algoritma bekerja dengan menggunakan vector maka hasil yang didapat adalah sebuah dot atau titik yang mempunyai berbagai macam warna, Dimana terdapat 2 warna berbeda yang mewakili 2 kelas yang digunakan oleh data tersebut yaitu pendek dan sangat pendek. Disini pendek berada pada titik atau dot berwarna merah sedangkan untuk kelas sangat pendek berada pada warna biru muda dan ada satu lingkaran berwarna hitam yang disebut support vector.

Lingkaran dari support vector adalah hasil tengah berdasar pelatihan dan uji yang dilakukan sebelumnya, hasil lingkaran support vector bisa lebih dari satu tergantung dari kompleksitas data yang diberikan dan hasil dari pelatihannya. Karena hasil dari Support Vector Machine sudah ditemukan maka selanjutnya adalah mencari nilai akurasi untuk algoritma ini.

### 3.3 KNN (K-Nearest Neighbor)

Pada bagian ini setelah menggunakan algoritma Support Vector Machine akan digunakan algoritma K-Nearest Neighbor, disini K-Nearest Neighbor adalah salah satu algoritma dari machine learning yang juga dikenal dengan algoritma KNN, algoritma ini bekerja dengan cara yang sama dengan algoritma sebelumnya, K-Nearest Neighbor mempunyai 2 fungsi yaitu fungsi klasifikasi dan fungsi prediksi sehingga algoritma ini bukan hanya digunakan untuk melakukan klasifikasi tapi juga melakukan prediksi.

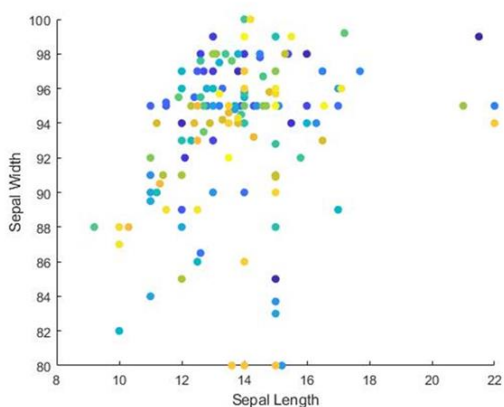
Pada bagian ini yang akan digunakan dari algoritma ini adalah fungsi klasifikasi atau classification yang akan dipakai untuk data yang telah ada, klasifikasi yang dipakai algoritma ini bekerja seperti random tree yaitu satu titik atau dot yang mewakili setiap kelas akan melakukan pencabangan dari data yang ada untuk mencari apakah ada hubungan dari dot atau titik dari kelas yang ada. Sebelum menjalankan script algoritma KNN, hal pertama yang akan dilakukan adalah mencari tahu input yang akan diberikan ke algoritma ini.

Hasil data input untuk K-Nearest Neighbor, dapat diketahui bahwa data kategori sangat pendek

memiliki jumlah data yang sedikit sedangkan data pendek mendominasi dengan 3x lebih banyak jumlahnya dari data sangat pendek. Kemudian, dilakukan ujicoba dengan data training akan mendapatkan 60% dari total data yang ada untuk dipakai sebagai data training. Hasil training memperlihatkan terdapat 2830 balita yang berada pada status stunting pendek, kemudian ada kategori sangat pendek sebanyak 975 balita yang berada pada status stunting sangat pendek. Dapat diketahui bahwa disini balita dengan status stunting pendek mempunyai jumlah data yang terbanyak sebanyak 2839 balita.

Selanjutnya setelah mengetahui isi data training adalah mengetahui isi data uji, disini karena sebelumnya data training mendapat 60% dari total jumlah data maka data uji mendapat sisa dari data tersebut yaitu 40%, disini data uji mendapat 40% total data dari 6343 data yang ada, untuk kategori pendek terdapat 1886 balita stunting yang masuk dalam kategori stunting pendek kemudian kategori sangat pendek dengan 650 balita yang berada dalam status stunting sangat pendek.

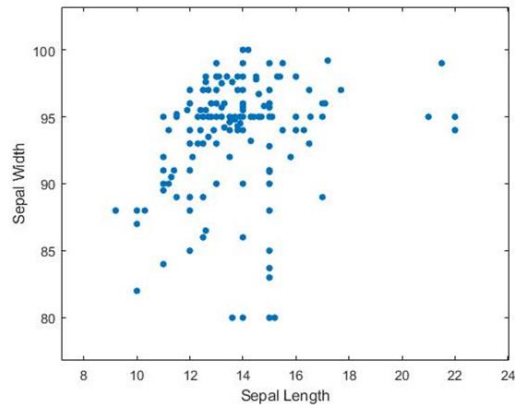
Setelah mengetahui isi dari data uji, maka akan masuk langkah selanjutnya yaitu melakukan prediksi berdasar data yang telah disediakan dan sudah dimasukkan di data training dan data uji, K-Nearest Neighbor membutuhkan setidaknya 2 jenis data berbeda untuk algoritmanya agar bisa melakukan kalkulasi dengan baik, dengan catatan setiap kategori harus mempunyai variasi nilai, tapi jika salah satu kategori tidak mempunyai variasi nilai, maka kategori lain harus mempunyai variasi nilai untuk menutupi kategori yang tidak memiliki variasi nilai tersebut saat kalkulasi sedang berlangsung. Hasil dari K-Nearest Neighbor dapat dilihat pada Gambar 9.



Gambar 9. Hasil Persebaran Data KNN

Gambar tersebut adalah salah satu hasil dari KNN yaitu berupa dot atau titik dengan berbagai macam warna, warna-warna ini merupakan data yang didapat dari analisa nilai X, beberapa warna berdekatan satu dengan lainnya beberapa warna berada sangat jauh yang menunjukkan banyak nilai yang berdekatan untuk setiap kelasnya dan beberapa sangat jauh perbedaan nilainya.

Selanjutnya terdapat satu hasil lagi yang didapat dari algoritma ini yaitu data persebaran kelas, jika digambar sebelumnya merupakan data persebaran nilai X, pada hasil selanjutnya adalah data persebaran pada kelas yang ada. Dan hasilnya dapat dilihat pada Gambar 10.



Gambar 10. Hasil Persebaran Kelas KNN

Gambar di atas adalah hasil persebaran data dari kelas yang ada, disini dapat diketahui hanya terdapat satu warna yaitu warna biru, dot atau titik diatas mewakili nilai kelas dari data yang ada. Beberapa diantaranya berdekatan dan beberapa diantara berada sangat jauh.

### 3.4 Akurasi Skor Semua Algoritma

Naive Bayes menghasilkan skor akurasi sebesar 90%, yang berarti algoritma ini dapat memprediksi dengan benar 90% dari keseluruhan data uji. Skor ini sangat tinggi, menandakan bahwa Naive Bayes mampu mempelajari pola data dengan baik dan membuat prediksi yang akurat berdasarkan data yang diberikan. Keberhasilan ini terutama disebabkan oleh kemampuan algoritma untuk memanfaatkan probabilitas dalam menghasilkan prediksi yang akurat, terutama ketika tersedia banyak data.

Sementara itu, Support Vector Machine (SVM) menghasilkan skor akurasi sebesar 86,56%. Ini juga merupakan nilai yang sangat tinggi, karena berada di atas 80%. SVM dikenal karena kemampuannya menangani data berdimensi tinggi dan mencari *hyperplane* yang memisahkan kelas-kelas dengan margin terbesar. Dengan akurasi yang tinggi ini, SVM menunjukkan bahwa ia mampu mengenali pola dalam data dan melakukan prediksi dengan tingkat kesalahan yang rendah, menjadikannya pilihan yang kuat untuk klasifikasi.

Di sisi lain, K-Nearest Neighbor (K-NN) menghasilkan skor akurasi sebesar 79,43%. Meskipun nilai ini sedikit lebih rendah dibandingkan dengan Naive Bayes dan SVM, tetap saja nilai ini cukup tinggi, terutama untuk dataset yang besar dengan lebih dari 6000 baris data. K-NN bekerja dengan mencari sejumlah tetangga terdekat dari data uji dan melakukan prediksi berdasarkan mayoritas



kelas dari tetangga tersebut. Akurasi yang mendekati 80% menunjukkan bahwa K-NN cukup efektif dalam memprediksi kelas berdasarkan kedekatan data, meskipun mungkin tidak sekuat Naive Bayes dan SVM dalam situasi tertentu.

Dari ketiga algoritma tersebut, Naive Bayes menunjukkan performa terbaik dengan skor akurasi 90%, menjadikannya pilihan yang sangat baik untuk dataset ini. Naive Bayes unggul dalam kecepatan komputasi dan efisiensi dengan data yang besar, serta bekerja dengan baik pada data dengan banyak fitur. Sementara SVM dan K-NN juga memiliki kelebihan masing-masing, seperti kemampuan SVM dalam menangani data berdimensi tinggi dan kemudahan implementasi K-NN, Naive Bayes tetap menjadi pilihan utama karena akurasinya yang tertinggi dan efisiensi yang ditawarkannya. Namun, penting untuk mempertimbangkan karakteristik spesifik data dan kebutuhan aplikasi sebelum memilih algoritma yang paling sesuai.

#### 4. Kesimpulan

Dari ketiga algoritma machine learning utama, yaitu Naive Bayes, Support Vector Machine (SVM), dan K-Nearest Neighbor (K-NN), telah diuji untuk menentukan kemampuan mereka dalam mendiagnosis stunting berdasarkan gejala medis. Hasil uji menunjukkan bahwa Naive Bayes menghasilkan skor akurasi tertinggi sebesar 90%, diikuti oleh SVM dengan akurasi 86,56%, dan K-NN dengan akurasi 79,43%. Naive Bayes, dengan akurasi yang sangat tinggi, menunjukkan keunggulan dalam mempelajari pola data dan memberikan prediksi yang akurat. Kecepatan komputasinya yang tinggi dan efisiensi dengan data besar menjadikannya pilihan yang sangat baik untuk diagnosa stunting. SVM, meskipun sedikit lebih rendah dalam akurasi, menunjukkan kemampuan yang kuat dalam menangani data berdimensi tinggi dan memberikan prediksi yang konsisten. K-NN, meskipun memiliki akurasi yang sedikit lebih rendah, tetap memberikan hasil yang memuaskan terutama untuk dataset besar.

Untuk penelitian selanjutnya, disarankan agar dilakukan validasi eksternal pada model yang telah dikembangkan dengan menggunakan dataset dari sumber yang berbeda, guna memastikan generalisasi model pada populasi yang lebih luas. Selain itu, eksplorasi algoritma machine learning lainnya, dapat memberikan wawasan lebih lanjut mengenai efektivitas berbagai pendekatan dalam mendiagnosa stunting. Penelitian mendatang juga sebaiknya melakukan analisis lebih mendalam terhadap fitur-fitur yang berkontribusi pada prediksi stunting, sehingga dapat mengidentifikasi faktor-faktor yang paling berpengaruh dan membantu dalam pengembangan intervensi yang lebih tepat sasaran. Melakukan studi longitudinal untuk memantau perkembangan anak-anak yang terdiagnosa stunting dari waktu ke waktu juga akan memberikan informasi berharga tentang faktor-faktor yang mempengaruhi

status gizi mereka. Selain itu, integrasi data dari berbagai sumber, seperti data lingkungan, sosial, dan ekonomi, dapat memberikan gambaran yang lebih komprehensif tentang faktor-faktor yang mempengaruhi stunting. Pengembangan aplikasi berbasis teknologi yang dapat digunakan oleh tenaga medis untuk mendiagnosa stunting secara real-time juga sangat dianjurkan, sehingga meningkatkan aksesibilitas dan efisiensi dalam diagnosis. Terakhir, penting untuk mengembangkan program pendidikan dan pelatihan bagi tenaga medis mengenai penggunaan algoritma machine learning dalam diagnosa stunting, untuk meningkatkan pemahaman dan penerapan teknologi ini di lapangan. Dengan mengikuti saran-saran ini, diharapkan penelitian selanjutnya dapat memberikan kontribusi yang lebih besar dalam upaya penanganan dan pencegahan stunting di masyarakat.

#### Daftar Pustaka:

- Alifariki, L. O. 2020. *Gizi Anak dan Stunting* (H. J. S. and M. K. Mariany (ed.)). LeutikaPrio.
- Arnita, S., Rahmadhani, D. Y., & Sari, M. T. 2020. Hubungan Pengetahuan dan Sikap Ibu dengan Upaya Pencegahan Stunting pada Balita di Wilayah Kerja Puskesmas Simpang Kawat Kota Jambi. *Jurnal Akademika Baiturrahim Jambi*, 9(1). <https://doi.org/10.36565/jab.v9i1.149>
- Banurea, M., Betaria Hutagaol, D., & Sihombing, O. 2023. Klasifikasi Penyakit Stunting Dengan Menggunakan Algoritma Support Vector Machine Dan Random Forest. *Jurnal TEKINKOM*, 6(2), 540–549. <https://doi.org/10.37600/tekinkom.v6i2.927>
- Car, A., Trisuchon, J., Ayaragarnchanakul, E., Creutzig, F., Javid, A., Puttanapong, N., Tirachini, A., Irawan, M. Z., Belgiawan, P. F., Tarigan, A. K. M., Wijanarko, F., Henao, A., Marshall, W. E., Chalermpong, S., Kato, H., Thaithakul, P., Ratanawaraha, A., Fillone, A., Hoang-Tung, N., ... Chalermpong, S. 2023. Collaborative Governance pada Penerapan Perwali No. 79 Tahun 2022 Tentang Percepatan Penurunan Stunting di Kota Surabaya. *International Journal of Technology*, 47(1), 100950. <https://doi.org/10.1016/j.tranpol.2019.01.002> <https://doi.org/10.1016/j.cstp.2023.100950> <https://doi.org/10.1016/j.geoforum.2021.04.007> <https://doi.org/10.1016/j.trd.2021.102816> <https://doi.org/10.1016/j.tra.2020.03.015> <https://doi.org/10.1016/j>
- Damanik, S. M., Sitorus, E., & Mertajaya, I. M. 2021. Sosialisasi Pencegahan Stunting pada Anak Balita di Kelurahan Cawang Jakarta Timur. *JURNAL Komunitas Servizio*, 3(1), 552–560. <https://doi.org/https://doi.org/10.33541/cs.v3i1.2909>

- Fauziah, J., Trisnawati, K. D., Rini, K. P. S., & Putri, S. U. 2023. Stunting: Penyebab, Gejala, dan Pencegahan. *Jurnal Parenting Dan Anak*, 1(2), 11. <https://doi.org/10.47134/jpa.v1i2.220>
- Kesehatan, K. 2022. *Buku Saku: Hasil Studi Satus Gizi Indonesia (SSGI)* (1st ed.). Badan Kebijakan Pembangunan Kesehatan KEMENTERIAN KESEHATAN RI. [https://repository.badankebijakan.kemkes.go.id/id/eprint/4855/3/Buku Saku SSGI 2022 rev 270123 OK.pdf](https://repository.badankebijakan.kemkes.go.id/id/eprint/4855/3/Buku_Saku_SSGI_2022_rev_270123_OK.pdf)
- N, S. 2023. Prediction of Stunting Among Under5 Children in Rwanda Using Machine Learning Techniques. *Journal of Preventive, Medicine & Public Health*, 56, 41–49.
- Nugroho, M. R., Sasongko, R. N., & Kristiawan, M. 2021. Faktor-faktor yang Mempengaruhi Kejadian Stunting pada Anak Usia Dini di Indonesia. *Jurnal Obsesi : Jurnal Pendidikan Anak Usia Dini*, 5(2), 2269–2276. <https://doi.org/10.31004/obsesi.v5i2.1169>
- Panigoro. 2020. Upaya Pencegahan Dan Penanggulangan Stunting Pada Balita Di Wilayah Kerja Puskesmas Tilongkabila. *Jurnal Ilmu Kesehatan Dan Gizi*, 1(1), 79–91. Perpres. 2020. *Peraturan Presiden No. 28. 1.*
- Rita Kirana, Aprianti, N. W. H. 2022. No TitlePengaruh Media Promosi Kesehatan Terhadap Perilaku Ibu Dalam Pencegahan Stunting di Masa Pandemi Covid-19 (Pada Anak Sekolah Tk Kuncup Harapan Banjarbaru). *Jurnal Inovasi Penelitian*, 02(9), 2899–2906.
- Sandi, & Rahmadhita, K. 2020. Permasalahan Stunting dan Pencegahannya Stunting Problems and Prevention. *Jurnal Ilmiah Kesehatan Husada*, 11(1), 225–229. <https://doi.org/10.35816/jiskh.v10i2.253>
- Sari, D., Ningsih, A. D., & Azzahra, A. 2023. Pencegahan Stunting Pada Anak Usia Dini Serta Dampaknya Pada Faktor Pendidikan Dan Ekonomi. *Jurnal Pengabdian Kepada Masyarakat Nusantara*, 4(3), 2679–2678.
- Setiawan, E., & Machmud, R. 2018. Faktor-Faktor yang Berhubungan dengan Kejadian Stunting pada Anak Usia 24-59 Bulan di Wilayah Kerja Puskesmas Andalas Kecamatan Padang Timur Kota Padang Tahun 2018. *Jurnal Kesehatan Andalas*, 7(2), 275–284.
- Shen, H., Zhao, H., & Jiang, Y. 2023. Machine Learning Algorithms for Predicting Stunting among Under-Five Children in Papua New Guinea. *Children*, 10(10). <https://doi.org/10.3390/children10101638>
- Trihono., Trihono and Atmarita., Atmarita and Tjandrarini., Dwi Hapsari and Irawati., Anies and Nurlinawati., Iin and Utami., Nur Handayani and Tejayanti., &Teti. 2015. *Pendek (Stunting) di Indonesia, Masalah dan Solusi* (M. Sudomo (ed.)). Lembaga Penerbit Balitbangkes.