

SENASTIKA Universitas Malikussaleh

DETEKSI STUNTING PADA BALITA MENGGUNAKAN METODE K-NEAREST NEIGHBOR

Raisya Kamila¹, Mulkan², Cut Novita³, Yasmin Amirah Harahap⁴, Munirul Ula⁵

^{1,2,3,4} Teknik Informatika, Universitas Malikussaleh, Lhokseumawe, Aceh

Email: raisya.210170207@mhs.unimal.ac.id, mulkan.210170234@mhs.unimal.ac.id,

cut.210170213@mhs.unimal.ac.id, yasmin.210170214@mhs.unimal.ac.id, munirulula@unimal.ac.id

Abstrak

Stunting adalah kondisi kekurangan gizi pada anak balita sehingga pertumbuhannya tidak berkembang, yang berdampak negatif pada perkembangan fisik dan mental. Penelitian ini bertujuan untuk menerapkan metode K-Nearest Neighbor (K-NN) dalam mendeteksi *stunting* pada balita dengan menganalisis data dan menganalisis akurasi. K-Nearest Neighbor adalah algoritma machine learning yang memiliki fungsi untuk klasifikasi data berdasarkan data train dan data test yang mana data itu diambil dari K tetangga terdekatnya. Data yang dikumpulkan dari penelitian ini adalah data usia balita dari 0 bulan sampai dengan 5 tahun dengan klasifikasi umur, jenis kelamin, tinggi badan, dan status gizi yang berjumlah 121.000 kolom. Data pada penelitian ini dikonversi terlebih dahulu yaitu konversi data pada atribut jenis kelamin dan status gizi dimana status jenis kelamin laki-laki menjadi 0 dan perempuan menjadi 1. Kemudian untuk status gizi severely stunted 0, stunted 1, normal 2 dan tinggi 3. Kemudian, data ini dibagi menjadi dua yakni 80% data train dan 20% data test. Tingkat akurasi yang ada pada penelitian ini mencapai 99% yang mana angka tersebut nyaris sempurna. Ini menunjukkan bahwa metode KNN ini adalah metode yang cocok digunakan untuk penelitian ini.

Keywords: *stunting, K-Nearest Neighbor, machine learning.*

1. PENDAHULUAN

Stunting merupakan kondisi kekurangan gizi pada anak balita sehingga pertumbuhan badannya menyusut atau memendek. Menurut data UNICEF, sekitar 149 juta anak diseluruh dunia mengalami *stunting* sehingga hal ini berdampak negatif pada perkembangan fisik dan mental mereka. Hal ini dapat dilihat dari gejala umum seperti lambatnya pertumbuhan tinggi badan, penurunan kemampuan kognitif dan resiko penyakit menular. Deteksi dini *stunting* sangat penting dilakukan untuk mencegah dampak jangka panjang yang dapat mengganggu kualitas hidup anak dan produktivitasnya di masa depan. Di era digital saat ini, pemanfaatan teknologi dan metode analisis data seperti *K-Nearest Neighbor* (KNN) dapat membantu dalam mendeteksi *stunting* secara lebih efisien dan akurat.

Beberapa penelitian sebelumnya telah mengkaji berbagai metode dalam mendeteksi *stunting*, baik melalui pendekatan tradisional maupun berbasis teknologi. Misalnya, penelitian oleh Smith et al (2020) menggunakan model regresi logistik untuk mengidentifikasi faktor-faktor risiko *stunting*. Sedangkan, Umairah dan kolega (2024) menerapkan algoritma *machine learning* lainnya, seperti Naive Bayes dan Chi-Square untuk memprediksi risiko *stunting*. Namun, hanya beberapa penelitian yang menerapkan metode K-NN pada deteksi *stunting*. Meskipun metode ini memiliki keunggulan dalam kesederhanaan dan efisiensi dalam pengolahan data.

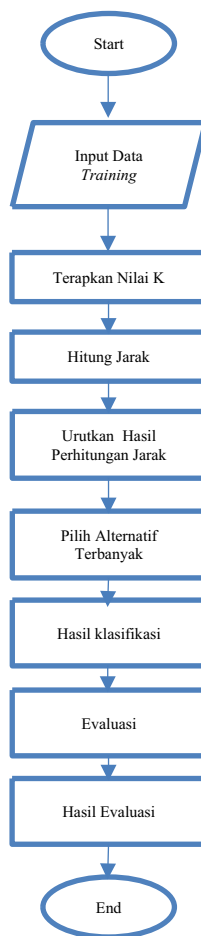
Tujuan penelitian ini untuk memberikan alternatif metode yang dapat digunakan dalam deteksi *stunting*. Selain itu, penelitian ini dibuat untuk mengembangkan model K-NN yang dapat menganalisis data dan menganalisis akurasi model K-NN.

Dengan menerapkan metode K-NN, diharapkan bisa memberikan solusi yang lebih cepat dan akurat dalam indentifikasi kasus *stunting*, sehingga pihak terkait seperti pemerintah dan lembaga kesehatan dapat mengambil langkah pencegahan yang tepat. Selain itu, penulis juga berharap agar penelitian ini dapat menjadi referensi atau acuan bagi peneliti lain yang tertarik dalam penggunaan teknologi dan data analisis dalam bidang kesehatan anak. Dengan demikian, hasil penelitian ini tidak cuma berkontribusi pada ilmu pengetahuan akan tetapi juga pada upaya global dalam mengurangi populasi *stunting* diseluruh dunia.

2. METODE PENELITIAN

2.1 Pengumpulan Data

Berikut ini adalah flowchart dari metode K-Nearest Neighbor (KNN) dalam penerapan pada penelitian ini. Ini merupakan gambaran umum penjelasan dari alur penelitian.



Gambar 1. Skema K-NN

Jarak ini kemudian digunakan untuk menentukan kedekatan data baru dengan data pelatihan yang sudah ada, yang menjadi dasar dalam proses klasifikasi. Setelah jarak *Euclidean* antara data yang akan diklasifikasikan dan setiap data pelatihan dihitung, untuk tahap selanjutnya yaitu mengurutkan hasil dari perhitungan jarak dari yang terkecil hingga terbesar. Pengurutan ini penting untuk menentukan tetangga terdekat dari data yang akan diklasifikasikan. Setelah itu, dari K tetangga terdekat yang telah diurutkan, dipilih kelas yang paling sering muncul sebagai hasil klasifikasi. Kelas dengan frekuensi terbanyak tersebut akan menjadi prediksi atau hasil klasifikasi untuk data yang baru. Setelah mendapatkan hasil klasifikasi, dilakukan evaluasi terhadap performa model untuk menilai seberapa akurat hasil prediksi yang dihasilkan oleh model KNN. Evaluasi ini biasanya dilakukan dengan membandingkan hasil klasifikasi dengan data sebenarnya dan menggunakan metrik tertentu seperti akurasi. Setelah evaluasi selesai, didapatkan hasil yang menunjukkan performa model, dan dengan ini, proses klasifikasi KNN dianggap selesai.

2.2 K-Nearest Neighbor

K-Nearest Neighbor merupakan algoritma *machine learning* yang memiliki fungsi untuk klasifikasi data berdasarkan data *train* dan data *test* yang mana data itu diambil dari K tetangga terdekatnya. Proses KNN dilanjutkan dengan menghitung jarak antara data yang akan diklasifikasikan dengan setiap data dalam set pelatihan. Perhitungan ini dilakukan menggunakan rumus jarak *Euclidean*, yang berfungsi untuk mengukur seberapa dekat dua titik data satu sama lain dalam ruang multidimensi. Rumus jarak *Euclidean* antara dua titik data $x = (x_1, x_2, \dots, x_n)$ dan $y = (y_1, y_2, \dots, y_n)$ adalah sebagai berikut:

$$d(x,y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_n - y_n)^2}$$

Keterangan : $d(x,y)$ = Jarak *Euclidean* antara dua titik

x = Titik pertama yang direpresentasikan dalam bentuk koordinat n yaitu $x = (x_1, x_2, \dots, x_n)$

Y = Titik kedua yang direpresentasikan dalam bentuk koordinat n yaitu $y = (y_1, y_2, \dots, y_n)$
 N = Jumlah dimensi ruang *Euclidean*.

2.3 Pengumpulan Data

Data yang dipakai pada penelitian ini bersifat penelitian kuantitatif karena melibatkan data numerik dan penerapan metode algoritma K-Nearest Neighbor (KNN) dalam melakukan deteksi *stunting*. Pengumpulan data penelitian ini adalah data usia balita dari 0 bulan sampai dengan 5 tahun dengan klasifikasi umur, jenis kelamin, tinggi badan, dan status gizi. Penulis dapat mengakses data balita ini melalui kaggle dengan judul *Stunting Toddler (Balita) Detection* (121K rows) dengan sumber data pada penelitian ini yang bisa di akses pada tautan ini <https://www.kaggle.com/datasets/rendiputra/stunting-balita-detection-121k-rows>, dataset ini mencakup hasil *stunting* dengan tingkat *stunted*, normal, tinggi, dan juga *severly stunted*, tabel berikut ini menampilkan dataset dari penelitian ini.

| | Umur (bulan) | Jenis Kelamin | Tinggi Badan (cm) | Status Gizi |
|-------------------------|--------------|---------------|-------------------|------------------|
| 0 | 0 | laki-laki | 44.591973 | stunted |
| 1 | 0 | laki-laki | 56.705203 | tinggi |
| 2 | 0 | laki-laki | 46.863358 | normal |
| 3 | 0 | laki-laki | 47.508026 | normal |
| 4 | 0 | laki-laki | 42.743494 | severely stunted |
| ... | ... | ... | ... | ... |
| 120994 | 60 | perempuan | 100.600000 | normal |
| 120995 | 60 | perempuan | 98.300000 | stunted |
| 120996 | 60 | perempuan | 121.300000 | normal |
| 120997 | 60 | perempuan | 112.200000 | normal |
| 120998 | 60 | perempuan | 109.800000 | normal |
| 120999 rows x 4 columns | | | | |

Gambar 2. Dataset *Stunting*

3. HASIL DAN PEMBAHASAN

3.1 Pre-processing Data

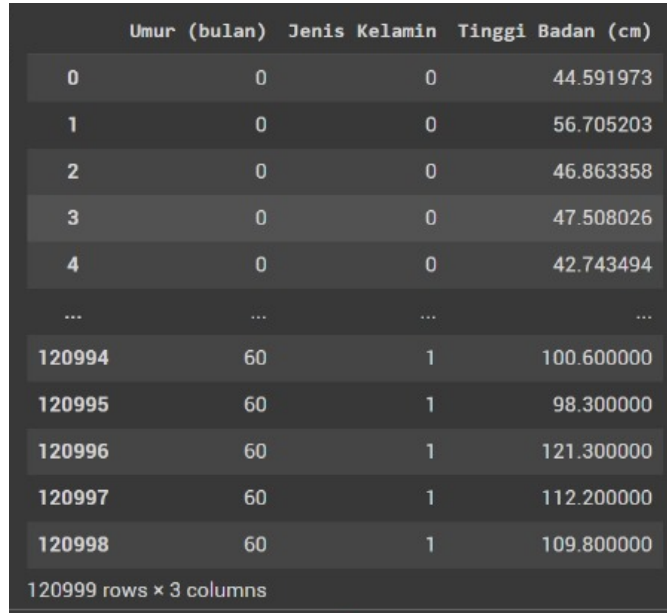
Adapun pre-processing pada penelitian ini hanya meliputi konversi data. Konversi data yang dimaksud adalah mengonversi data kategori ke data numerik. Hal ini perlu dilakukan karena K-NN menggunakan rumus euclidian distance. Dalam penelitian ini, data yang di konversi yaitu data pada atribut jenis kelamin dan status gizi dimana status jenis kelamin laki-laki menjadi 0 dan perempuan menjadi 1. Kemudian untuk status gizi *severly stunted* 0, *stunted* 1, normal 2 dan tinggi 3. berikut ini merupakan tampilan data yang sudah di pre-processing.

| | Umur (bulan) | Jenis Kelamin | Tinggi Badan (cm) | Status Gizi |
|--------|--------------|---------------|-------------------|-------------|
| 0 | 0 | 0 | 44.591973 | 1 |
| 1 | 0 | 0 | 56.705203 | 3 |
| 2 | 0 | 0 | 46.863358 | 2 |
| 3 | 0 | 0 | 47.508026 | 2 |
| 4 | 0 | 0 | 42.743494 | 0 |
| ... | ... | ... | ... | ... |
| 120994 | 60 | 1 | 100.600000 | 2 |
| 120995 | 60 | 1 | 98.300000 | 1 |
| 120996 | 60 | 1 | 121.300000 | 2 |

Gambar 3. Pre-processing data

3.2 Split Data

Fungsi *split* data adalah untuk membagi data menjadi data *training* dan data *testing*. Pada penelitian ini, 20% dari data digunakan sebagai data uji, sedangkan 80% sisanya digunakan sebagai data *training*. Di bawah ini merupakan hasil dari *split* data, sebagai berikut.



| | Umur (bulan) | Jenis Kelamin | Tinggi Badan (cm) |
|--|--------------|---------------|-------------------|
| | 0 | 0 | 44.591973 |
| | 1 | 0 | 56.705203 |
| | 2 | 0 | 46.863358 |
| | 3 | 0 | 47.508026 |
| | 4 | 0 | 42.743494 |
| | ... | ... | ... |
| | 120994 | 60 | 100.600000 |
| | 120995 | 60 | 98.300000 |
| | 120996 | 60 | 121.300000 |
| | 120997 | 60 | 112.200000 |
| | 120998 | 60 | 109.800000 |

120999 rows x 3 columns

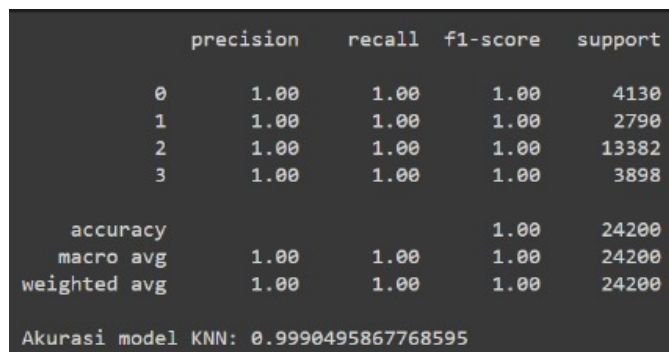
Gambar 4. Tampilan hasil *split* data

3.3 Accuracy

Tingkat akurasi yang dihasilkan menggunakan metode K-Nearest Neighbor adalah 99%, yang mana nilai akurasi tersebut mendekati nilai sempurna yaitu 100%. Hal ini membuktikan bahwa metode ini sangat akurat jika diimplementasikan ke dalam penelitian ini.

3.4 Evaluasi

Langkah terakhir dari penelitian ini adalah evaluasi. Pada proses sebelumnya yaitu *split* data, peneliti telah membagi data menjadi dua yakni data *train* dan data *test*. Pengujian akan dilakukan pada data *test* untuk mengukur dan mendapatkan hasil dari presisi, recall dan f1-score. Berikut ini merupakan hasil dari evaluasinya.



| | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 0 | 1.00 | 1.00 | 1.00 | 4130 |
| 1 | 1.00 | 1.00 | 1.00 | 2790 |
| 2 | 1.00 | 1.00 | 1.00 | 13382 |
| 3 | 1.00 | 1.00 | 1.00 | 3898 |
| accuracy | | | 1.00 | 24200 |
| macro avg | 1.00 | 1.00 | 1.00 | 24200 |
| weighted avg | 1.00 | 1.00 | 1.00 | 24200 |

Akurasi model KNN: 0.9990495867768595

Gambar 5. Tampilan akurasi data

4. DISKUSI

Dari penelitian Umairah dan kolega (2024) menyatakan bahwa metode Naïve Bayes dan Chi-Square dalam prediksi *stunting* dapat meningkatkan nilai akurasi, recall dan presisi tapi tidak terlalu signifikan sehingga tingkat akurasi penelitiannya mencapai 94.3%. Sedangkan, dalam penelitian ini dimana metode yang digunakan adalah K-Nearest Neighbor mendapat tingkat akurasi yang tinggi yaitu sebesar 99%. Maka, dari kedua metode ini dapat disimpulkan bahwa tingkat akurasi dalam memprediksi *stunting* yang paling mendekati nilai sempurna ialah menggunakan metode K-Nearest Neighbor. Hal ini didukung juga dengan banyaknya data yang tersedia.

5. KESIMPULAN

Dari data yang dikumpulkan dari penelitian ini, untuk prediksi status *stunting* balita berdasarkan umur, tinggi badan, jenis kelamin dan status gizi, dapat disimpulkan bahwa, dengan menggunakan metode K-Nearest Neighbor berhasil memprediksi status gizi balita dengan baik. Dilihat dari penelitian yang dilakukan oleh penulis, bahwa penerapan metode KNN untuk deteksi *stunting* pada balita dapat menghasilkan akurasi 99% yang mana nilai akurasi tersebut mendekati nilai sempurna. Ini menunjukkan bahwa metode KNN adalah salah satu metode yang tepat untuk penelitian ini.

UCAPAN TERIMAKASIH

Terima kasih kepada orangtua, dosen pengampu mata kuliah riset teknologi informasi yaitu Bapak Munirul Ula, S.T., M.Eng., Ph.D dan seluruh rekan yang meluangkan waktunya untuk turut membantu mengumpulkan data atas penelitian yang peneliti lakukan baik secara langsung maupun tidak langsung dan terima kasih juga kepada seluruh pihak yang membantu proses penelitian.

DAFTAR PUSTAKA

- [1] WHO, "Malnutrition." Accessed: Oct. 09, 2024. [Online]. Available: <https://www.who.int/news-room/fact-sheets/detail/malnutrition>
- [2] L. C. Smith and L. Haddad, "Reducing Child Undernutrition: Past Drivers and Priorities for the Post-MDG Era," *World Dev.*, vol. 68, no. 1, pp. 180–204, 2015, doi: 10.1016/j.worlddev.2014.11.014.
- [3] Kemenkes, "Penyakit Stunting." [Online]. Available: <https://ayosehat.kemkes.go.id/penyakit/stunting>
- [4] F. D. Wahyudi, D. Remawati, and P. Harsadi, "Sistem Pakar Deteksi Kerusakan Mesin Bubut Dengan Metode Knn," *J. Teknol. Inf. dan Komun.*, vol. 6, no. 2, pp. 7–13, 2019, doi: 10.30646/tikomsin.v6i2.370.
- [5] R. L. Hasanah, M. Hasan, W. E. Pangesti, F. F. Wati, and W. Gata, "Klasifikasi Penerima Dana Bantuan Desa Menggunakan Metode Knn (K-Nearest Neighbor)," *J. Techno Nusa Mandiri*, vol. 16, no. 1, pp. 1–6, 2019, doi: 10.33480/techno.v16i1.25.
- [6] A. R. Wulandari and D. Avianto, "Sistem Pakar Diagnosa Kelainan Stunting Balita Menggunakan Metode KNN Berbasis Web," *J. Indones. Manaj. Inform. dan Komun.*, vol. 5, no. 1, pp. 1064–1072, 2024, doi: 10.35870/jimik.v5i1.587.
- [7] N. A. Sholikhin and S. Atmojo, "Aplikasi Web Untuk Klasifikasi Stunting Pada Balita Dengan Menggunakan Metode K-Nearest Neighbours (Studi Kasus Posyandu Jawa Kidul)," *J. Syst. Eng. Technol. Innov.*, vol. 1, no. 02, pp. 44–47, 2022, doi: 10.38156/jisti.v1i02.23.
- [8] H. I. Purwanto and S. Wibisono, "AHP-CBR Untuk Deteksi Dini Stunting Pada Balita Menggunakan Algoritma Similaritas KNN," *INFORMATICS Educ. Prof. J. Informatics*, vol. 7, no. 1, p. 64, 2023, doi: 10.51211/itbi.v7i1.2121.
- [9] U. R. Gurning, S. F. Octavia, D. R. Andriyani, N. Nurainun, and I. Permana, "Prediksi Risiko Stunting pada Keluarga Menggunakan Naïve Bayes Classifier dan Chi-Square," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. 1, pp. 172–180, 2024, doi: 10.57152/malcom.v4i1.1074.
- [10] H. G. Pangestu, R. Y. Sinaga, F. Z. Ulya, U. Athiyah, A. W. Muhammad, and F. Alameka, "Analisis Efisiensi Metode K-Nearest Neighbor dan Forward Chaining Untuk Prediksi Stunting Pada Balita," *Inform. Mulawarman J. Ilm. Ilmu Komput.*, vol. 18, no. 2, p. 78, 2023, doi: 10.30872/jim.v18i2.10169.
- [11] P. Aisha, M. Fathurahman, and S. Prangga, "Implementasi Metode Neighbor Weighted K-Nearest Neighbor Pada Pengklasifikasian Status Gizi Balita Di Wilayah Kerja Puskesmas Wonorejo Kota Samarinda," *Var. J. Stat. Its Appl.*, vol. 6, no. 1, pp. 11–20, 2024, doi: 10.30598/variancevol6iss1page11-20.

- [12] R. Mutiara Sari and S. Sutikno, "Perbandingan Metode Knn Dan Mknn Untuk Deteksi Dini Diabetes Mellitus," *J. Mnemon.*, vol. 6, no. 2, pp. 96–101, 2023, doi: 10.36040/mnemonic.v6i2.6021.
- [13] Endang Etriyanti, "Perbandingan Tingkat Akurasi Metode Knn Dan Decision Tree Dalam Memprediksi Lama Studi Mahasiswa," *J. Ilm. Bin. STMIK Bina Nusant. Jaya Lubuklinggau*, vol. 3, no. 1, pp. 6–14, 2021, doi: 10.52303/jb.v3i1.40.
- [14] C. Suryanti and M. G. Rohman, "Klasifikasi Kualitas Buah Apel Berdasarkan Warna dan Bentuk Menggunakan Metode KNN," *Gener. J.*, vol. 8, no. 1, pp. 34–41, 2024, doi: 10.29407/gj.v8i1.21052.
- [15] S. L. Louis, A. N. Mirania, and E. Yuniarti, "The Relationship Between Exclusive Breastfeeding with Stunting on Toddles Children," *Matern. Neonatal Heal. J.*, vol. 3, no. 1, pp. 7–11, 2022, [Online]. Available: <https://journal.neoelectura.com/index.php/mnhj>
- [16] S. R. Raysyah, Veri Arinal, and Dadang Iskandar Mulyana, "Klasifikasi Tingkat Kematangan Buah Kopi Berdasarkan Deteksi Warna Menggunakan Metode Knn Dan Pca," *JSiI (Jurnal Sist. Informasi)*, vol. 8, no. 2, pp. 88–95, 2021, doi: 10.30656/jsii.v8i2.3638.