

PENERAPAN ALGORITMA DECISION TREE, SVM, NAÏVE BAYES DALAM DETEKSI STUNTING PADA BALITA

¹Kharis Hudaiby Hanif ✉, ²Novita Ranti Muntiar

¹Teknik Komputer, Universitas Borneo Tarakan, Tarakan, Indonesia

²Sistem Informasi, Politeknik Bisnis Kaltara, Tarakan, Indonesia

Email: hudaiby21@borneo.ac.id

DOI: <https://doi.org/10.46880/jmika.Vol8No1.pp105-109>

ABSTRACT

Stunting is a toddler's body condition that is short according to body length according to age (PB/U), ≤ 2 Standard Deviations (SD), with a z-score between -3 standard deviations (SD). Where checking the stunting status of toddlers takes quite a long time because it is done manually and is also prone to errors. Therefore, it is hoped that a system can classify toddler examination data quickly and accurately to predict children's stunting status. Building a system that uses an algorithm to classify the stunting status of toddlers using decision tree, naïve bayes, and SVM. With what level of accuracy is the best of the 3 algorithms? Results from testing with 30% testing data and 70% training data using an algorithm decision tree, naïve bayes, and SVM. Accuracy level test results decision tree by 99%, naïve bayes of 48%, and SVM of 95%. So, the algorithm with the highest level of accuracy is decision tree amounts to 99%. Wallet hire decision tree better for detecting stunting in toddlers.

Keyword: *Stunting, SVM, Decision Tree, Naïve Bayes.*

ABSTRAK

Stunting adalah kondisi tubuh balita yang pendek menurut panjang badan menurut umur (PB/U), ≤ 2 Standar Deviasi (SD), dengan z-score antara -3 standar deviasi (SD). Dimana pemeriksaan status stunting balita memerlukan waktu yang cukup lama karena dilakukan secara manual dan juga rentan terhadap kesalahan. Oleh karena itu, diharapkan sebuah sistem yang dapat mengklasifikasikan data pemeriksaan balita dengan cepat dan akurat untuk memprediksi status stunting anak. Membangun sistem yang menggunakan algoritma untuk mengklasifikasikan status stunting balita menggunakan decision tree, naïve bayes, dan SVM. Dengan tingkat akurasi mana yang paling baik dari 3 algoritma tersebut. Hasil dari pengujian dengan data testing 30 % dan data training 70 % menggunakan algoritma decision tree, naïve bayes, dan SVM. Hasil uji tingkat akurasi decision tree sebesar 99%, naïve bayes sebesar 48%, dan SVM sebesar 95%. Jadi, algoritma yang tingkat akurasi paling tinggi yaitu decision tree sebesar 99%. Maka algoritma decision tree lebih baik untuk deteksi stunting pada balita.

Kata Kunci: *Stunting, SVM, Decision Tree, Naïve Bayes.*

PENDAHULUAN

Stunting gangguan kesehatan yang disebabkan oleh kekurangan atau ketidakselimbangan zat gizi yang diperlukan untuk pertumbuhan, aktivitas berpikir, dan semula hal-hal penting lainnya dalam kehidupan (Tinendung & Zufria, 2023). Stunting adalah kondisi tubuh balita yang pendek menurut panjang badan menurut umur (PB/U), dengan z-score antara -3 standar deviasi (SD) dan ≤ 2 Standar Deviasi (SD) (Makmur, et al., 2023).

Beberapa faktor dapat menyebabkan stunting, seperti praktik pengasuhan yang buruk, pemahaman ibu kurang tentang kesehatan dan nutrisi, layanan kesehatan terbatasnya, kurangnya akses keluarga ke makanan bergizi, dan kurangnya akses ke air bersih dan sanitasi (TNP2K, 2017).

Dimana pemeriksaan status stunting balita memerlukan waktu yang cukup lama karena dilakukan secara manual dan juga rentan terhadap kesalahan. Diharapkan sebuah sistem yang dapat mengklasifikasikan data pemeriksaan balita dengan cepat dan akurat untuk mendeteksi anak balita termasuk katagori stunting atau tidak.

Machine learning juga dikenal sebagai penambangan data, proses penggalian data untuk menemukan masalah tertentu diselesaikan menggunakan pola dan hubungan (Fauzan et al., 2023). Metode pengkategorian yaitu klasifikasi untuk mengidentifikasi kelas data (Andiani et al., 2019). Tujuan utama klasifikasi adalah memprediksi label kelas. Proses klasifikasi terdiri dari dua tahap. Pembelajaran adalah algoritma klasifikasi digunakan

untuk menganalisa data, klasifikasi menggunakan uji data untuk memprediksi ketepatan klasifikasi. (Lonang & Normawati, 2022).

Algoritma dalam *Machine learning* antara lain *neural network*, *k-nearest*, dan sebagainya. Penelitian ini menggunakan SVM, *naïve bayes* dan *decision tree*. Algoritma *decision tree* adalah metode klasifikasi data berbasis pohon, algoritma seringkali terlalu sesuai dengan data dan dapat ditingkatkan kembali (Kristanto et al., 2023). *Naïve bayes* karena memiliki kemampuan untuk meningkatkan akurasi menggunakan jumlah data pelatihan yang minimal (Zeniarta et al., 2020). Algoritma Support Vector Machine (SVM) memungkinkan penyelesaian masalah keputusan, terutama dalam kasus sampel data yang kecil (Lonang & Normawati, 2022).

Peneliti terdahulu menurut Makmur Mulyono, dkk (Makmur et al., 2023) dengan judul “Klasifikasi Status Stunting Balita Menggunakan Metode *Naïve Bayes Gaussian* Berbasis Web “ digunakan untuk klasifikasi status stunting balita. Hasil pengujian menunjukkan bahwa tipe data numerik untuk atribut usia, berat badan, tinggi badan, dan nominal jenis kelamin, Inisiasi Menyusui Dini (IMD), memiliki akurasi tertinggi pada perbandingan data 80:20, atau 80,34%, dengan 1172 data total.

Kesimpulan maka penelitian akan membuat sistem mengklasifikasi stunting balita dengan algoritma *decision tree*, *naïve bayes*, dan *SVM*. Dengan tingkat akurasi mana yang paling baik dari 3 algoritma tersebut.

METODE PENELITIAN



Gambar 1. Alur Penelitian

Berdasarkan Gambar 1 alur penelitian terdiri dari pengumpulan data, pengolahan data, klasifikasi hingga evaluasi.

Pengumpulan Data

Data diperoleh dari Kaggle Stunting Toddler (Balita) oleh Eduardo d’anjour. File didapat ekstensi .csv dengan jumlah data 120.999 dengan data balita stunting 33.684 dan tidak stunting 87.315. Memiliki 4 atribut yaitu umur (bulan), jenis kelamin, tinggi badan (cm), dan status gizi. Atribut ditunjukkan pada Tabel 1.

Tabel 1 Dataset Atribut

Nama Atribut	Keterangan
Umur (bulan)	Umur balita (0-59bulan)
Jenis Kelamin	Laki-laki, Perempuan
Tinggi Badan (cm)	Tinggi badan balita (cm)
Status Gizi	<i>severely stunted</i> : 0, <i>stunted</i> : 1, normal: 2, tinggi: 3

Preprocessing

Preprocessing proses penting karena data yang dikumpulkan tidak lengkap atau tidak konsisten (Lonang & Normawati, 2022). Selanjutnya proses dapat dimulai karena data status gizi balita pemeriksaan digunakan konsisten.

Transformasi Data

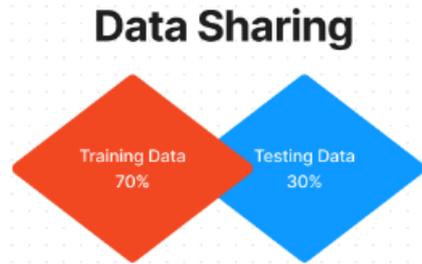
Tujuan dari proses transformasi adalah untuk mengubah data yang memiliki Untuk membuat implementasi mudah, nilai kategori diubah menjadi data bernilai angka. (Nikmatun & Waspada, 2019). Jenis kelamin dan status gizi adalah dua atribut yang akan di transformasi. Jenis kelamin 0 perempuan akan diberi nilai 1. Data tinggi badan dibandingkan dengan umur (bulan) terbagi menjadi empat kategori: *severely stunted*, *stunted*, normal, tinggi lebih. Nilai-nilai ini akan diubah menjadi 0,1,2,3. Transformasi Data ditunjukkan pada Tabel 2.

Tabel 2 Hasil Transformasi

Jenis Kelamin	Status Gizi
0	1
0	3
0	2
0	2

Klasifikasi

Dalam klasifikasi di *machine learning* dibagi menjadi dua bagian. Melatih model klasifikasi digunakan data latih, dan data uji digunakan untuk uji (Muntiari et al., 2021). 70 % data digunakan untuk klasifikasi, dan 30% data digunakan untuk pengujian ditunjukkan pada Gambar 2.



Gambar 2. Klasifikasi

Decision Tree

Decision tree dibentuk dengan menggunakan pohon keputusan, yang merupakan salah satu pendekatan yang paling umum. *Decision tree* diklasifikasikan sebagai pohon yang terdiri dari node yang membentuk akar dan tidak memiliki input. Node internal atau node pengujian adalah node lain yang bukan *root* tetapi memiliki salah satu inputnya. Node lainnya disebut daun. Target tepat dari setiap kelas diwakili oleh daun (Muntiari & Hanif, 2022).

Naïve Bayes

Naive bayes memanfaatkan teorema *bayes* kemungkinan masa depan berdasarkan data atau pengalaman sebelumnya. Menurut korelasi hipotesis, bukti klasifikasi, dan hubungan *naive bayes* dengan klasifikasi, teorema *bayes* menggambarkan label kelas menjadi sasaran pemetaan. Karakteristik yang digunakan dalam proses klasifikasi (Ginantra et al., 2021).

SVM

Metode klasifikasi vektorisasi instruksional diperkenalkan oleh Vapnik 1995 adalah *Support Vector Machine* (SVM). Pengklasifikasi diskriminan digunakan untuk menemukan *hyperplane* yang ideal berdasarkan maksimalisasi margin. Digunakan biasanya untuk data terpisah linier dan non-linier (Muntiari et al., 2023).

Confusion Matrix

Tabel 3. Confusion Matrix

	Prediction Class	
	+	-
+	TP	FN
-	FP	TN

Tabel 3 menunjukkan *confusion matrix* yang digunakan sebagai pengukuran kinerja parameter algoritma yaitu *f1-measure*, *recall*, akurasi, dan *presisi*. Rumus untuk menghitung *precision* ditunjukkan dalam Persamaan (1). Persamaan (2) rumus menghitung *recall*, di mana *recall* merupakan parameter untuk mengukur ketepatan suatu algoritma, dan Persamaan (3) rumus menghitung F. Rumus untuk mengukur ketepatan menggunakan persamaan (4). Nilai tertinggi 1. Salah satu perhitungan yang sering digunakan untuk mengevaluasi kinerja sebuah algoritma adalah akurasi, yang dihitung dengan membandingkan rasio jumlah data yang diprediksi secara akurat oleh algoritma dengan jumlah total data yang ada dalam dataset.

$$Precision = \frac{TP}{TP+FP} \tag{1}$$

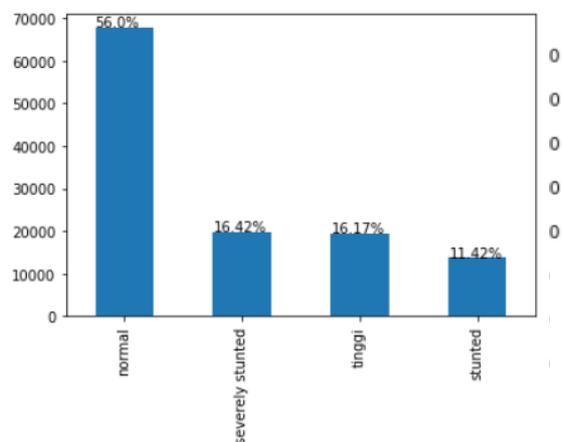
$$Recall = \frac{TP}{TP+FN} \tag{2}$$

$$F\text{-measure} = \frac{2 \times precision \times recall}{precision + recall} \tag{3}$$

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{4}$$

HASIL DAN PEMBAHASAN

Pengujian menggunakan data dengan jumlah data 120.999 dengan data balita stunting 13.815, stunting parah 19.869, tinggi 19.560, dan normal 67.755 dan tidak stunting 87.315. Persentase data ditunjukkan pada Gambar 3.



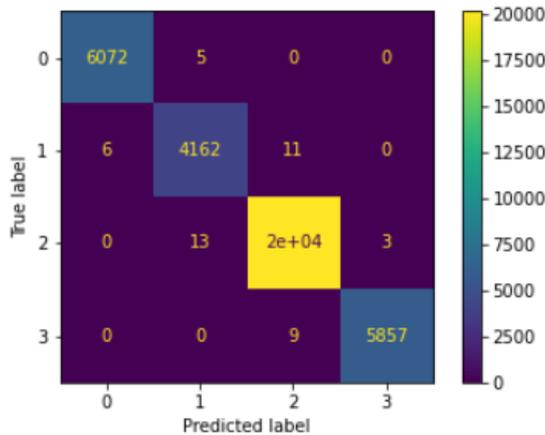
Gambar 3. Data Uji

Berdasarkan data uji diatas klasifikasi data, data testing 30 % dan training 70 %. Dilakukan pengujian menggunakan *decision tree*, *SVM*, dan *naïve bayes*.

Akurasi model Decision Tree: 0.9987052341597796				
confusion_matrix Decision Tree:				
	precision	recall	f1-score	support
0	1.00	1.00	1.00	6077
1	1.00	1.00	1.00	4179
2	1.00	1.00	1.00	20178
3	1.00	1.00	1.00	5866
accuracy			1.00	36300
macro avg	1.00	1.00	1.00	36300
weighted avg	1.00	1.00	1.00	36300

Gambar 4. Hasil Uji *Decision Tree*

Hasil uji yang ditunjukkan pada Gambar 4 algoritma *decision tree* hasil akurasi rata-rata sebesar 99%, *recall* 100%, *precison* 100%, dan *f1-score* 100%.



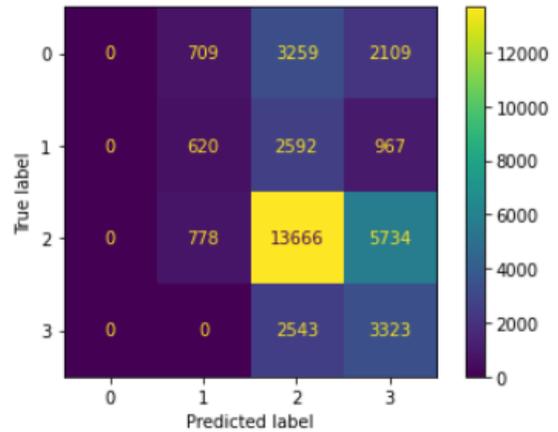
Gambar 5. *Confusion Matrix Decision Tree*

Gambar 5 menjelaskan *confusion matrix* pada algoritma *decision tree*.

Hasil uji yang ditunjukkan pada Gambar 6 algoritma *naïve bayes* hasil akurasi rata-rata akurasi 48%, *recall* sebesar 35%, *precison* 30%, dan *f1-score* 30%.

Akurasi model naive bayes: 0.48509641873278236				
	precision	recall	f1-score	support
0	0.00	0.00	0.00	6077
1	0.29	0.15	0.20	4179
2	0.62	0.68	0.65	20178
3	0.27	0.57	0.37	5866
accuracy			0.49	36300
macro avg	0.30	0.35	0.30	36300
weighted avg	0.42	0.49	0.44	36300

Gambar 6. Hasil Uji *Naïve Bayes*



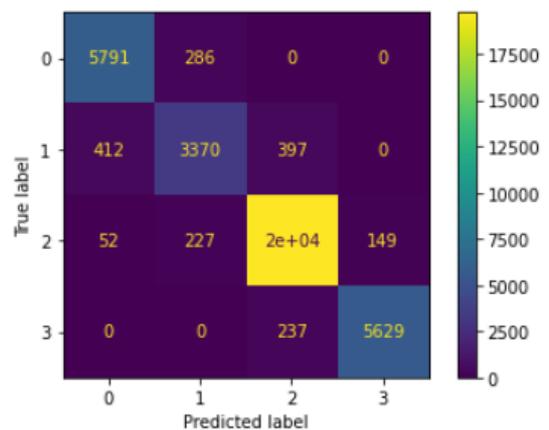
Gambar 7. *Confusion Matrix Naïve Bayes*

Gambar 7 menjelaskan *confusion matrix* pada algoritma *naïve bayes*.

Akurasi model SVM: 0.9515151515151515				
confusion_matrix SVM:				
	precision	recall	f1-score	support
0	0.93	0.95	0.94	6077
1	0.87	0.81	0.84	4179
2	0.97	0.98	0.97	20178
3	0.97	0.96	0.97	5866
accuracy			0.95	36300
macro avg	0.93	0.92	0.93	36300
weighted avg	0.95	0.95	0.95	36300

Gambar 6. Hasil Uji SVM

Hasil uji yang ditunjukkan pada Gambar 8 algoritma *SVM* hasil akurasi rata-rata sebesar 95%, *recall* 92%, *precison* 93%, dan *f1-score* 93%.



Gambar 1. *Confusion Matrix SVM*

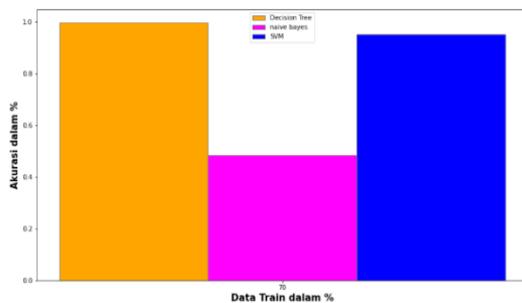
Gambar 9 menjelaskan *confusion matrix* pada algoritma *SVM*.

Berdasarkan hasil uji dari 3 algoritma *decision tree*, *naïve bayes*, dan *SVM*. Perbandingan hasil algoritma ditunjukkan pada Tabel 4.

Tabel 4. Perbandingan Algoritma

Algoritma	Tingkat Akurasi (%)
<i>Decision tree</i>	99 %
<i>Naïve bayes</i>	48 %
<i>SVM</i>	95 %

Hasil perbandingan pada Tabel 4 ditunjukkan Gambar 10.



Gambar 10. Grafik Perbandingan Algoritma

Seperti yang ditunjukkan dalam Tabel 4, ada hasil perbandingan akurasi dari algoritma *decision tree*, *naïve bayes*, dan SVM. Algoritma *decision tree* menunjukkan tingkat akurasi tertinggi, yaitu 99%, sehingga dapat disimpulkan bahwa menggunakan algoritma ini lebih baik untuk deteksi stunting pada balita.

KESIMPULAN

Hasil dari pengujian dengan data testing 30 % dan data training 70 % menggunakan algoritma *decision tree*, *naïve bayes*, dan SVM. Hasil uji tingkat akurasi *decision tree* 99%, SVM 95%, *naïve bayes* 48%, dan. Jadi, algoritma yang tingkat akurasi paling tinggi yaitu *decision tree* 99%. Maka algoritma *decision tree* lebih baik untuk deteksi stunting pada balita.

DAFTAR PUSTAKA

Andiani, L., Sukemi, & Palupi Rini, D. (2019). Analisis Penyakit Jantung Menggunakan Metode KNN Dan Random Forest. *Prosiding Annual Research Seminar*, 5(1), 1–5.

Fauzan A., Budianita, E., Nazir, A., & Syafria, F. (2023). Klasifikasi Status Stunting Balita Menggunakan Metode C4.5 Berbasis Web. *ZONAsi: Jurnal Sistem Informasi*, 5(3), 215–225. <https://doi.org/10.31849/zn.v5i3.15828>

Ginantra, N. L. W. S. R., Arifah, F. N., Wijaya, A. H., Septarini, R. S., Ahmad, N., Ardiana, D. P. Y., Effendy, F., Iskandar, A., Hazriani, H., Sari, I. Y., Gustiana, Z., Prianto, C., Gustian, D., & Negara, E. S. (2021). *Data Mining dan Penerapan Algoritma*.

Kristanto, F., Winarno, W. W., & Nasiri, A. (2023). Perbandingan Algoritma Naïve Bayes dan Decision Tree Pada Analisis Sentimen Data Komentar Siswa Pada Aplikasi Digital Teacher Assessment. *Jurnal Ilmiah Teknik Informatika Dan Sistem Informais*, 538–548.

Lonang, S., & Normawati, D. (2022). Klasifikasi Status Stunting Pada Balita Menggunakan K-Nearest Neighbor Dengan Feature Selection Backward Elimination. *Jurnal Media Informatika Budidarma*, 6(1), 49. <https://doi.org/10.30865/mib.v6i1.3312>

Makmur, M., Elvia B., Alwis N., F. S. (2023). Klasifikasi Status Stunting Balita Menggunakan Metode Naïve Bayes Gaussian Berbasis Web. *ZONAsi: Jurnal Sistem Informasi*, 5(3), 215–225. <https://doi.org/10.31849/zn.v5i3.15828>

Muntiari, N. R., & Hanif, K. H. (2022). Klasifikasi Penyakit Kanker Payudara Menggunakan Perbandingan Algoritma Machine Learning. *Jurnal Ilmu Komputer Dan Teknologi*, 3(1), 1–6. <https://doi.org/10.35960/ikomti.v3i1.766>

Muntiari, N. R., Hanif, K. H., & Nisa, I. C. (2021). *Perbandingan Algoritma Regresi Logistik , Support Vector Machine , dan Gradient Boosting Pada Analisis Sentimen Data Komentar Siswa*. 2(1), 1–7.

Muntiari, N. R., Nisa, K., Sandi, A. S., Ashari, I. A., Kharis Hudaiby Hanif, A., & Dwinanto, R. W. (2023). *Comparison of random forest algorithm, support vector machine, and k-nearest neighbor for diabetes disease classification*. May.

Nikmatun, I. A., & Waspada, I. (2019). Implementasi Data Mining Untuk Klasifikasi Masa Studi Mahasiswa Menggunakan Algoritma K-Nearest Neighbor. *Jurnal SIMETRIS*, 10(2), 421–432.

Tinendung, I. S., & Zufria, I. (2023). Pengelompokan Status Stunting Pada Anak Menggunakan Metode K-Means Clustering. *Jurnal Media ...*, 7, 2014–2023. <https://doi.org/10.30865/mib.v7i4.6908>

TNP2K. (2017). *100 Kabupaten/Kota Prioritas untuk Intervensi Anak Kerdil (Stunting): Tim Nasional Percepatan Penanggulangan Kemiskinan*. 2, 2017.

Zeniarja, J., Widia, K., & Sani, R. R. (2020). Penerapan Algoritma Naïve Bayes dan Forward Selection dalam Pengklasifikasian Status Gizi Stunting pada Puskesmas Pandanaran Semarang. *JOINS (Journal of Information System)*, 5(1), 1–9. <https://doi.org/10.33633/joins.v5i1.2745>