JURNAL INSY PRO Information System and Processing
Volume 10, Nomor 1, Mei 2025 P–ISSN 2597-419X, E-ISSN 2579-468X
Ojs : http://journal.uinalauddin.ac.id/index.php/insypro
Email : insypro@uin-alauddin.ac.id

PERBANDINGAN ALGORITMA NAIVE BAYES DAN K-MEANS UNTUK MENENTUKAN STATUS GIZI STANTING

COMPARISON OF NAIVE BAYES ALGORITHM AND K-MEANS TO DETERMINE STUNTING NUTRITION STATUS

Ariastuti Rahman[1)], Akhmad Qashlim[2)] , Nur Marifah[3)]

[1,2,3] Program Studi Sistem Informasi, Fakultas Imu Komputer, Universitas Al Asyariah Mandar,
[1,2,3]Jl. Budi Utomo No. 2 Kecamatan Polewali, Polewali Mandar, Sulawesi Barat 91311

E-mail: ariastuti.rahman@mail.unasman.ac.id[1)], qashlim@mail.unasman.ac.id[2)], marifah@gmail.com[3)]

Abstrak – Stunting sebagai suatu keadaan dimana anak di bawah usia lima tahun menderita kekurangan asupan gizi dalam jangka panjang dan menyebabkan gangguan tumbuh kembang, meningkatkan morbiditas dan mortalitas. Penanganan status gizi stunting di Indonesia khsusnya Kabupaten Polewali Mandar merupakan hal yang penting mengingat kasus stunting pada balita menjadi masalah yang serius. Untuk membantu petugas kesehatan dalam penentuan status gizi stunting maka dibutuhkan sebuah metode yang tepat dan efektif. Penelitian ini akan memberikan alternatif cara penentuan status gizi stunting menggunakan teknik Data mining yaitu metode Naive Bayes dan K-Means dengan indikator antropometri. Terdapat 173 data balita yang diproses. Hasil penelitian menunjukkan bahwa penggabungan dari kedua metode tersebut untuk nilai accuracy algoritma Naive Bayes mempunyai nilai 89% sedangkan K-Mens mempunyai nilai 54%. Dengan menggunakan parameter precission algoritma Naive Bayes mempunyai nilai 84% sedangkan K-Mens mempunyai nilai 58%, Dengan menggunakan parameter recall algoritma Naive Bayes mempunyai nilai 97% sedangkan K-Mens mempunyai nilai 60%. Berdasarkan nilai yang diperoleh maka dapat disimpulkan bahwa algoritma Naive Bayes memiliki akurasi, precission dan recall yang lebih tinggi bila dibandingkan algoritma K-Means

Kata Kunci: Stunting, K-Means, Naïve Bayes, Perbandingan Algoritma

Abstract *– Stunting is a condition where children under five years of age suffer from long-term nutritional deficiencies and cause growth and development disorders, increasing morbidity and mortality. Handling the nutritional status of stunting in Indonesia, especially in Polewali Mandar Regency, is important considering that stunting cases in toddlers are a serious problem. To assist health workers in determining the nutritional status of stunting, an appropriate and effective method is needed. This study will provide an alternative way to determine the nutritional status of stunting using Data mining techniques, namely the Naive Bayes and K-Means methods with anthropometric indicators. There are 173 toddler data processed. The results of the study show that the combination of the two methods for the accuracy value of the Naive Bayes algorithm has a value of 89% while K-Mens has a value of 54%. By using the precision parameter, the Naive Bayes algorithm has a value of 84% while K-Mens has a value of 58%. By using the recall parameter, the Naive Bayes algorithm has a value of 97% while K-Mens has a value of 60%. Based on the values obtained, it can be concluded that the Naive Bayes algorithm has higher accuracy, precision and recall when compared to the K-Means algorithm..*

Keywords: *Stunting, K-Means, Naïve Bayes, Algorithm Comparison*

PENDAHULUAN

Salah satu masalah gizi kronis yang sedang dialami dunia kesehatan adalah stunting (T. Prasetiya, dkk., 2020) yang merupakan suatu kondisi dimana anak usia lima tahun ke bawah mengalami kekurangan asupan gizi dalam jangka waktu lama dan menyebabkan gangguan tumbuh kembang. Stunting dapat berdampak dalam jangka pendek, pada peningkatan morbiditas dan mortalitas, perkembangan mental, keterampilan motorik, pengetahuan, dan perkembangan bicara anak yang tidak berkembang secara optimal, sementara dampak dalam jangka panjang dari stunting dapat berupa pengerdilan pada bentuk tubuh anak dan meningkatkan risiko obesitas R. Puspitarahayu and E. Nasution (2020).

Penentuan status gizi stunting oleh tenaga kesehatan dilakukan secara fisik dengan berbagai indikator salah satunya adalah indikator antropometri atau pengukuran tubuh manusia (T. Prasetiya, dkk., 2020; R. Setiawan and A. Triayudi , 2022) seperti umur, berat badan, tinggi badan dan lingkar kepala, lingkar lengan dan lingkar perut, serta tinggi lutut data mining untuk mendapatkan dengan melihat perkembangan pada buku KIA (O. Saeful Bachri and R. M. Herdian Bhakti, 2021) walaupun terdapat

indikator lain seperti pengaruh lingkungan dan pola asuh orang tua (T. Prasetiya, dkk., 2020). Indikator antropometri dapat diolah kedalam metode klastering dan klasifikasi keputusan status gizi balita menjadi lebih cepat dengan mempertimbangkan pola data sebelumnya (R. Setiawan and A. Triayudi , 2022), sosial ekonomi dan faktor demografi dipilih sebagai variabel penjelas berdasarkan literatur (S. M. J. Rahman *et al, 2021*). Beberapa metode klasifkasi yang dapat digunakan adalah Naïve Bayes dan K-Nearest Neigbor (T. Prasetiya, dkk., 2020; R. Setiawan and A. Triayudi , 2022; K. A. K-nn and R. G. Whendasmoro¸2022; N. Syafina, 2022) sementara metode klastering yang banyak digunakan untuk data kesehatan adalah K-Means (S. Priyadarshini, M. Abdul, and B. Ssali, 2022)

Penerapan machine learning di bidang kesehatan masyarakat semakin meningkat dari hari ke hari, masalah gizi anak adalah topik hangat di bidang kesehatan masyarakat serta epidemiologi secara global. Banyak penelitian tentang gizi anak dan malnutrisi yang hanya berfokus pada identifikasi faktor risiko malnutrisi menggunakan model klasik seperti regresi logistik. Oleh karena itu, perlu untuk mengusulkan model dan algoritma lainnya dan mesin learning (ML) memiliki daya tarik untuk digunakan pada berbagai jenis data medis/biomedis. (S. M. J. Rahman *et al.*, 2021)

Menggunakan metode data mining yaitu Algoritma K-Means termasuk dalam clustering yang mana prosesnya mempartisi data yang digunakan menjadi satu cluster atau lebih dari satu, proses pemecahan masalah diselesaikan dengan meminimalkan kesalahan berulang, serta pembelajaran sederhana (W. I. Rahayu, 2021) sementara. Algoritma Naive Bayes termasuk dalam salah satu metode klasifikasi. Keakuratan dan kecepatan metode Naive Bayes Classifier sangat tinggi bila digunakan pada aplikasi database dengan banyak data. Algoritma Naive Bayes Classifier mengurangi tingkat kesalahan jika dibandingkan dengan algoritma klasifikasi lainnya (M. Y. Titimeidara and W. Hadikurniawati, 2021). Sebuah penelitian yang dilakukan dengan kombinasi tiga indeks antropometri menggunakan Naive Bayes Classifier dengan akurasi 100% yang mengolah 225 data balita (T. E. Putri, 2020).

Algoritma Naïve Bayes

Algoritma klasifikasi Naive Bayesian banyak digunakan dalam analisis big data dan berbagai bidang lainnya karena struktur algoritmanya yang sederhana dan cepat (H. Chen, S. Hu, R. Hua, and X. Zhao, 2021) Naive Bayes Classifier dan entropy classifier digunakan untuk tujuan klasifikasi (S. S and H. Wang,, 2021) berbagai kasus dapat diselesaikan dengan menggunakan algoritma Naïve bayes diantaranya adalah, strategi diagnosis COVID-19 menggunakan Fitur Correlated Naïve Bayes (FCNB) yang terdiri dari empat fase, yaitu; Seleksi Fitur Phase (FSP), Feature Clustering Phase (FCP), Master Feature Weighting Phase (MFWP), dan Feature Correlated Naïve Fase Bayes (FCNBP), membuktikan keefektifan FCNB strategi karena mencapai akurasi 99% (N. A. Mansour, 2022). Naïve bayes juga digunakan untuk melakukan klasifikasi daerah rawan pangan dengan menggunakan 517 data, naïve bayes memberikan hasil akurasi 68%, lebih rendah dari akurasi menggunakan algoritma C4.5 yaitu 84% untuk kasus yang sama (S. M. J. Rahman *et al.*, 2021). Berbagai upaya pun dilakukan untuk meningkatkan dan memperbaiki hasil akurasi pada algoritma ini seperti pembobotan fitur dan Kalibrasi Laplace. Melalui simulasi numerik, diketahui bahwa ketika memiliki jumlah sampel yang besar, akurasi algoritma klasifikasi naive Bayes dapat meningkat lebih dari 99%, dan sangat stabil ketika atribut sampel kurang dari 400 dan apabila jumlah kategori kurang dari 24, maka keakuratan naive Bayes hanya dapat maeningkat lebih dari 95%. Selain itu terdapat penelitian yang menggunakan metode cerdas hibrid untuk mengintegrasikan Naïve Bayes classifier dan sistem fuzzy paralel untuk klasifikasi diabetes tipe 2. Metode yang diusulkan menunjukkan lebih baik akurasi klasifikasi sebesar 90,26% saat diuji menggunakan dataset diabetes Pima (T. T. Ramanathan, 2022). Penelitian empiris menunjukkan bahwa ketika algoritma klasifikasi naive Bayes ditingkatkan akurasinya maka meningkat pula tingkat kebenaran analisis dan akurasi yang lebih tinggi. (H. Chen, et., al., 2021) Perbedaan nilai akurasi antara algoritma naive bayes dengan algoritma klasifikasi lainnya dapat dipengaruhi oleh beberapa faktor seperti banyaknya label/kelas, banyaknya data yang digunakan untuk data latih dan data uji, serta perubahan pola atau karakteristik data (H. Rasmita and S. Hendra, 2022). Model machine learning naïve bayes dirumuskan sebagai berikut:

$$P(H \mid X) = \mathrm{P(X \mid H).P(H) / P(X)} \qquad (1)$$

(1)

Keterangan:
$P(H \mid X)$: Probabilitas A terjadi, bukti bahwa B telah terjadi
P(X | H) : Probabilitas B terjadi, bukti bahwa A telah terjadi
P(H) : Peluang terjadinya H
P(X) : Peluang terjadinya X
X : Data dengan class yang belum diketahui

H : Hipotesis data merupakan suatu class spesifik

Algoritma K-Means

Standar algoritma K-Means menetapkan pengelompokan centroids secara acak untuk cluster. Pengelompokan titik data yang serupa ke dalam kluster akan dilakukan berdasarkan pada centroid awal (T. S. Priyadarshini, 2022)

Secara sederhana algoritma K-Means dimulai dari tahap berikut :

a. Menentukan nilai K sebagai jumlah kluster yang ingin dibentuk
b. Membangkitkan nilai random untuk pusat cluster awal (centroid) sebanyak k.
c. Menghitung jarak setiap data input terhadap masing–masing centroid menggunakan rumus jarak Euclidean (Euclidean Distance) hingga ditemukan jarak yang paling dekat dari setiap data dengan centroid. Berikut adalah persamaan Euclidian Distance:

$$d(xi, \mu j) = \sqrt{\sum(xi - \mu j)^2} \quad (2)$$

d. Mengklasifikasikan setiap data berdasarkan kedekatannya dengan centroid (jarak terkecil).
e. Memperbaharui nilai Nilai centroid baru di peroleh dari rata-rata cluster yang bersangkutan dengan menggunakan rumus:

$$\mu j(t+1) = \frac{1}{Nsj} \sum_{j \in sj} xj \quad (3)$$

Melakukan perulangan dari langkah 3 hingga 5, sampai anggota tiap cluster tidak ada yang berubah

Berdasarkan penjelasan di atas maka perlu dilakukan pengelompokan data yaitu membagi data menjadi beberapa kelompok sesuai dengan kesamaan karakteristik masing-masing data berdasarkan kelompok yang ada, agar data yang diperoleh lebih akurat. Oleh karena itu, peneliti menggunakan dua metode algoritma, yaitu Algoritma Naive Bayes dan K-Means untuk menentukan status gizi stanting, kedua metode yang digunakan masing-masing memiliki tingkat akurasi berbeda, sehingga dilakukan pengujian untuk mengurangi terjadinya kesalahan dan untuk memastikan bahwa output yang diperoleh memenuhi persyaratan. Hasil yang diperoleh dari algoritma Naive Bayes dan K-Means diharapkan dapat digunakan sebagai nilai referensi, dan hasilnya dipilih sesuai dengan referensi yang ditentukan.

Pengukuran Kinerja

ROC (Receiver Operating Characteristics) dan area under the curve (AUC) adalah alat ukur performance untuk masalah classification yang mampu menentukan threshold dari suatu model dan merupakan alat yang ampuh untuk menilai kemampuan prediksi fitur (T. Gneiting and E. M. Walz, 2022). Standar Kurva ROC menunjukkan visualisasi antara true positive rate (TPR) dan false positive rate (FPR) sementara skor AUC memiliki kemampuan model untuk membedakan antar kelas. Semakin tinggi AUC, semakin baik model dalam membedakan antara baliata stunting dan tidak stunting (T. S. Priyadarshini, 2022). Walaupun kedua alat ukur ini terlalu umum karena mengevaluasi semua ambang keputusan termasuk yang tidak realistis.Sehingga pada kasus tertentu sebaiknya diukur pada ambang tunggal yang optimal yang spesifik misalnya dapat melalui normalisasi (A. M. Carrington *et al. 2022*)

Semakin mendekat ke sudut kiri atas sebuah Classifier maka semakin menunjukkan kondisi *perfect classifier* atau kinerja yang semakin baik. Agar memiliki dasar pada evaluasi kinerja model maka dibuat classifier acak *(random classifier)* berupa garis putus-putus yang terletak di sepanjang diagonal (FPR = TPR). Kondisi akan menunjukkan classifier semakin tidak akurat apabila kurva classifier semakin mendekat ke diagonal 45 derajat dari ruang ROC.

Selain ROC dan AUC, Evaluasi performance algoritma Machine Learning (ML) khususnya supervised learning, juga dapat menggunakan acuan Confusion Matrix yang merepresentasikan hasil prediksi dan kondisi sebenarnya(aktual) dari sekumpulan data yang diolah menggunakan algoritma ML. dari model Confusion Matrix, kita bisa menentukan Accuracy, Precission, Recall dan Specificity.

Tabel 1.1 Confusion Matrix

n=	Aktual Positif	Aktual Negatif
Prediksi Positif	TP	FP
Prediksi Negatif	FN	TN

Keterangan
True Positive (TP) : Interpretasi: Anda memprediksi positif dan itu benar.
True Negative (TN): Interpretasi: Anda memprediksi negatif dan itu benar.
False Positive (FP): (Kesalahan Tipe 1): Interpretasi: Anda memprediksi positif dan itu salah.
False Negative (FN): (Kesalahan Tipe 2, kesalahan tipe 2 ini sangat berbahaya) Interpretasi: Anda memprediksi negatif dan itu salah.

Adapun akurasi diukur dengan penjelasan sebagai berikut:

a. Accuracy yaitu Merupakan rasio prediksi Benar (positif dan negatif) dengan keseluruhan data. Adapun formula yang digunakan sebagai berikut

Accuracy = (TP + TN) / (TP+FP+FN+TN) (4)

b. Precission yaitu Merupakan rasio prediksi benar positif dibandingkan dengan keseluruhan hasil yang diprediksi positf. Precission menjawab pertanyaan "Berapa persen balita yang benar *Stunting* dari keseluruhan balita yang diprediksi Stunting?"

Precission = (TP) / (TP+FP) (5)

c. Recall Merupakan rasio prediksi benar positif dibandingkan dengan keseluruhan data yang benar positif. Recall menjawab pertanyaan "Berapa persen balita yang diprediksi Stunting dibandingkan keseluruhan balita yang sebenarnya Stunting".

Recall = (TP) / (TP + FN) (6)

Stunting

Stunting berdampak negatif terhadap perkembangan negara karena hubungannya dengan morbiditas anak dan risiko kematian (WHO (World Health Organization, 2018) Pengukuran antropometri, status gizi ideal berdasarkan pada ambang batas nilai z-score (M. K. RI,, 2020) sebagai berikut:

Tabel 2.1 Batas ambang nilai z-score dalam status gizi anak (M. K. RI,, 2020)

Indeks	Kategori status gizi anak	Ambang batas (*z-score*)
Berat Badan Menurut Umur(BB/U) Anak umur 0-60 bulan	Berat badan sangat kurang	< -3 SD
	Berat badan kurang	-3 sampai dengan < -2 SD
	Berat badan normal	-2 sampai dengan +1SD
	Resiko berat badan lebih	> +1 SD
Panjang Badan Menurut Umur (PB/U) atau Tinggi Badan Menurut Umur (TB/U) Anak umur 0-60 bulan	Sangat pendek	< -3 SD
	Pendek	-3 sampai dengan < -2 SD
	Normal	-2 sampai dengan +3 SD
	Tinggi	> +3 SD

METODOLOGI PENELITIAN

Tahapan Penelitian

Prosedur penelitian sebagai pedoman dalam menyelesaikan penelitian diuraikan pada Gambar 3.1. yang menunjukkan prosedur penelitian dan menjelaskan secara sistematis aktifitas dan output yang dihasilkan pada setiap tahapan yang dikerjakan.

Data Penelitian

Sebanyak 173 data balita digunakan dan status gizi dilihat berdasarkan indikator antropmentri menggunakan parameter umur, berat badan dan tinggi badan. Data akan dikelompokkan kedalam dua cluster, cluster 1 (Tidak beresiko) dan cluster 2 (Beresiko stunting) menggunakan metode K-Means klastering sementara metode Naive Bayes melakukan dua klasifikasi yaitu Presence (Beresiko stunting), Absence (Tidak beresiko).

Status gizi stunting terjadi jika pertumbuhannya baita -2 sampai < -3(Berat badan: kurang, tinggi badan; pendek), dan dibawah < -3 (Berat badan: sangat kurang, tinggi badan: sangat pendek) selain dari kondisi tersebut maka termasuk normal atau tidak beresiko stunting (R. Puspitarahayu and E. Nasution,2020). Pengukuran antropometri, status gizi ideal berdasarkan pada ambang batas nilai z-score pada tabel 2.1 (M. K. RI,, 2020)

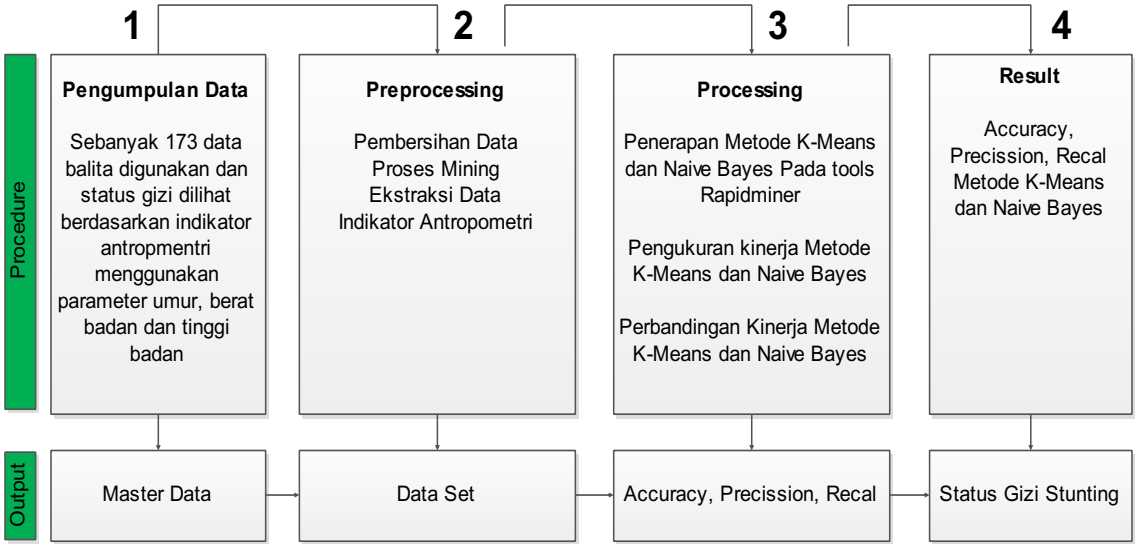
Flowchart System

Adapun sistem yang diusulkan dalam menentukan status gizi stunting yaitu dengan membandingkan dua algoritma yaitu Naive Bayes dan K-Means untuk mendapatkan hasil dengan akurasi yang lebih baik. Aktivitas dimulai dengan pengolahan data posyandu untuk diekstrak sesuai dengan data yang dibutuhkan setelah itu datanya di ubah ke dalam CSV untuk memudahkan perhitungan pada aplikasi RapidMiner.

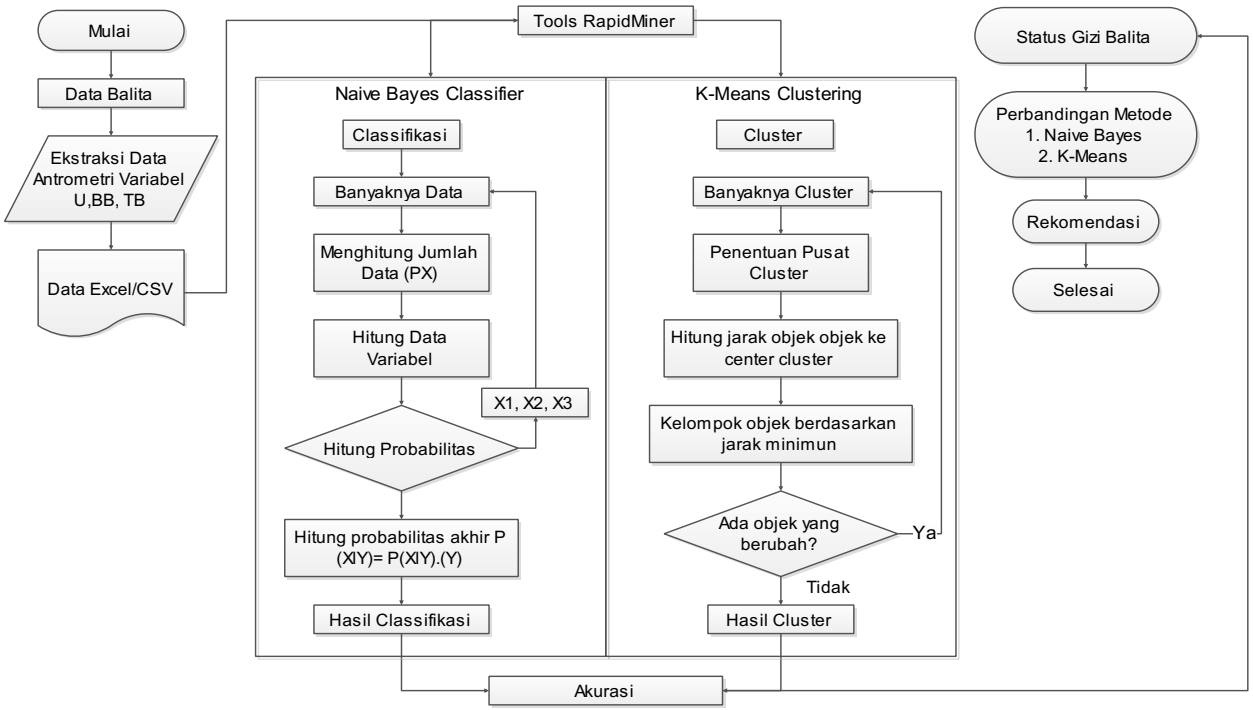
Pengolahan dalam aplikasi RapidMiner terbagi menjadi dua yaitu Naive Bayes pada klasterisasi sedangkan K-Means di klastering, dalam proses algoritma Naive Bayes menentukan jumlah data yang akan diolah, kemudian perhitungan variabel yang akan digunakan, setelah itu menentukan probabilitas setiap variabel sehingga menghasilkan akurasi dari Naive Bayes. Sedangakan pada K-Means yang pertama dilakukan yaitu binning data proses mengubah data menjadi nilai binner kemudian menentukan jumlah claster, penentuan pusat claster, menghitung jarak objek ke center claster,

sehingga kelompok objek berdasarkan jarak minimum apabila ada objek yang berubah, maka akan melakukan literasi atau perhitungan ulang untuk menentukan claster baru sehingga menghasilkan akurasi K-Means, dari hasil akurasi tersebut digunakan untuk rekomendari algoritma dalam menentukan status gizi stunting di tempat penelitian. Flowchart sistem disajikanpada gambar 3.2.

Untuk menguraikan hasil penelitian maka pertama-pertama disajkan hasil represntasi data. Berdasarkan hasil dari pengumpulan data yang sudah dilakukan berupa umur, berat badan dan tinggi badan, maka diketahui jumlah seluruh data balita terdiri dari 173 balita dengan status gizi yang terdiri dari 98 balita



Gambar 3.1. Tahapan Sistematis Penelitian



Gambar 3.2. Flowchrt System

HASIL DAN PEMBAHASAN

Hasil penelitian yang disajikan berupa nilai akurasi pada masing-masing metode, perbandingan kinerja dan rekomendasi metode terbaik untuk kasus data stunting. absence (tidak beresiko), dan 75 balita presence (beresiko).

Clasifikasi Data Menggunakan Naïve Bayes

Tahap pertama penggunaan algoritma Naive Bayes yaitu dengan menkategorikan data sebagaimana disajikan pada tabe. 4.1 sebagai berikut:

Tabel 4.1. Kategori Atribut Data

Atribut Data	Kategori
Umur	1. 0-23 Bulan = baduta (anak usia dibawah dua tahun) 2. 23-59 Bulan = balita(anak usian dibawah lima tahun)
Berat Badan	1. Berat badan sangat kurang 2. Berat badan kurang 3. Berat badan normal 4. Resiko berat badan lebih
Tinggi Badan	1. Sangat pendek 2. Pendek 3. Normal 4. Tinggi

Status gizi ideal anak berdasarkan atribut yang digunakan selalu merujuk pada tabel 2.1 Batas ambang nilai z-score dalam status gizi anak [20]. Untuk klasifikasi status gizi disajikan pada tabel 4.2

Tabel 4.5 Atribut Status Gizi

Kategori Status	Keterangan
Absence	Tidak Beresiko
Presence	Beresiko

Adapun decission system yang telah diproses tersebut yaitu sebagai berikut

Tabel 4.6. Hasil Decision System Naïve Bayes

NAMA	UMUR	BB	TB	STATUS
Balita 1	Baduta	Normal	Pendek	Presence
Balita 2	Baduta	Normal	Pendek	Presence
Balita 3	Baduta	Normal	Normal	Absence
Balita 4	Baduta	Normal	Normal	Absence
Balita 5	Baduta	Normal	Normal	Absence
Balita 6	Baduta	Normal	Normal	Absence
Balita 7	Baduta	Normal	Pendek	Presence
Balita 8	Baduta	Normal	Normal	Absence
Balita 9	Baduta	Normal	Normal	Absence
Balita 10	Baduta	Normal	Normal	Absence
Balita 11	Baduta	Normal	Normal	Absence
Balita 12	Baduta	Normal	Sangat Pendek	Presence
Balita 13	Baduta	Normal	Normal	Absence
Balita 14	Baduta	Normal	Normal	Absence
Balita 15	Baduta	Kurang	Sangat Pendek	Presence
Balita 16	Baduta	Normal	Normal	Absence
Balita 17	Baduta	Kurang	Sangat Pendek	Presence

Sebanyak 173 Data yang telah diproses oleh naïve bayes dan menghasilkan keputusan pada kolom status tabel 4.6 hanya menyajikan 17 sebagai perwakilan data dari 173 total data. Proses selanjutnya adalah menghitung nilai Probabilitas dari masing-masing atribut mulai dari umur, berat badan, tinggi badan dan atribut keputusan yaitu status gizi stunting. Adapun tabel probabilitas tersebut sebagai berikut:

Tabel 4.7 Nilai Probabilitas Umur

P (Umur)	Absence	Presence
Baduta	38%	40%
Balita	62%	60%
Total	100%	100%

Tabel 4.8 Nilai Probabilitas Berat Badan

P (Berat Badan)	Absence	Presence
Berat Badan Sangat Kurang	0%	3%
Berat Badan Kurang	2%	25%
Berat Badan Normal	97%	69%
Resiko Berat Badan Lebih	1%	3%
Total	100%	100%

Tabel 4.9 Nilai Probabilitas Tinggi Badan

P (Berat Badan)	Absence	Presence
Sangat Pendek	0%	21%
Pendek	0%	53%
Normal	99%	25%
Tinggi	1%	0%
Total	100%	100%

Setelah mendapatkan nilai probabilitas diatas, maka akan dilakukan uji test berikut:

1. Apabila diketahui hasil status gizi dengan umur kategori baduta, berat badan normal, dan tinggi badan pendek, Maka dapat dihitung :

   Absence = 38% × 97% × 0% = 0 × 0.566 = 0
   Presence = 40% × 69% × 53% = 0.146 × 0.433 = 0.063

Jadi, keputusan status gizi balita 1 adalah Presence

2. Apabila diketahui hasil status gizi dengan umur kategori baduta, berat badan normal, dan tinggi badan normal, Maka dapat dihitung :

   Absence = 38% × 97% × 99% = 0.364 × 0.566 = 0.206
   Presence = 40% × 69% × 25%= 0.065 × 0.433 = 0.028

Jadi, keputusan status gizi balita 1 adalah *Absence.*

Tabel 4.10 Hasil Uji Data Balita

Nama	Absence	Presence	Class Prediction
Balita 1	0.00%	6.40%	Presence
Balita 2	0.00%	6.40%	Presence
Balita 3	20.50%	3.00%	Absence
Balita 4	20.50%	3.00%	Absence
Balita 5	20.50%	3.00%	Absence
Balita 6	20.50%	3.00%	Absence
Balita 7	0.00%	6.40%	Presence
Balita 8	20.50%	3.00%	Absence
Balita 9	20.50%	3.00%	Absence
Balita 10	20.50%	3.00%	Absence
Balita 11	20.50%	3.00%	Absence
Balita 12	0.00%	2.60%	Presence
Balita 13	20.50%	3.00%	Absence
Balita 14	20.50%	3.00%	Absence
Balita 15	0.00%	0.90%	Presence
Balita 16	20.50%	3.00%	Absence
Balita 17	0.00%	0.90%	Presence

Metode Clasifikasi Data Metode K-Means

Penetapan nilai Initial *Cluster* Center awal yang telah di pilih secara random pada Tabel 4.11 dibawah.

1. Inisiasi nilai pusat cluster

Tabel 4.11 *Centroid* Awal

Cluster	Umur	BB	TB	Keterangan
C1	19	10.8	79.9	Absence
C2	20	9.6	77.2	Presence

2. Pengelompokan Data Cluster

Tabel 4.12 Hasil *Cluster* dengan Jarak Terdekat

Nama	C1	C2	Jarak terdekat	Kelompok data
Balita 1	2.85	1.42	1.42	C2
Balita 2	3.12	0.00	0.00	C2
Balita 3	0.00	3.12	0.00	C1
Balita 4	1.99	3.16	1.99	C1
Balita 5	1.57	2.87	1.57	C1
Balita 6	1.57	2.72	1.57	C1
Balita 7	11.57	9.60	9.60	C2
Balita 8	8.47	7.15	7.15	C2
Balita 9	5.31	6.11	5.31	C1
Balita 10	6.71	7.01	6.71	C1
Balita 11	9.12	8.07	8.07	C2
Balita 12	16.24	14.21	14.21	C2
Balita 13	10.88	10.03	10.03	C2
Balita 14	13.73	12.47	12.47	C2
Balita 15	17.51	15.70	15.70	C2
Balita 16	7.13	7.07	7.07	C2
Balita 17	11.77	9.54	9.54	C2

3. Penemtuam Nilai Centroid Baru

Tabel 4.17 *Centroid* Baru

Cluster	Umur	BB	TB	Keterangan
C1	42.3	13.1	91.2	Absence
C2	13.6	8.6	71.9	Presence

Hasil Pengujian

Hasil pengujian kedua algoritma menggunakan data yang sama menunjukkan hasil klasifikasi dan cluster yang berbeda pula. Penelitian ini menggunakan parameter Accuracy, Precission, Recall untuk mengukur performance algoritma naïve bayes dan K-Means

Confusion Matriks

1. Algoritma *Naive Bayes*
Tabel 4.11 merupakan tabel *Confusion Matrix* dengan mengunakan algoritma *Naive Bayes.*

Tabel 4.11 *Confusion Matrix Naive Bayes*

Predicted	*Class*	
	Absence	*Presence*
Absence	37	0
Presence	2	31

Keterangan :
TP = *True Positive* sebanyak 37
TN = *True Negative* sebanyak 31
FP = *False Positive* sebanyak 0
FN = *False Negative* sebanyak 2

2. Algoritma *K-Means*
Tabel 4.12 merupakan tabel *Confusio Matrik* dengan mengunakan algoritma *K-Means*.

Tabel 4.12. Tabel *Confusion Matrix K-Means*

Predicted	*Class*	
	Absence	*Presence*
Absence	57	41
Presence	38	37

Keterangan :
TP = *True Positive* sebanyak 57
TN = *True Negative* sebanyak 37
FP = *False Positive* sebanyak 41
FN = *False Negative* sebanyak 38

4.3.2.Accuracy

1. Algoritma *Naive Bayes*
Accuracy = ((37+31)/(37+0+2+31)) x 100%
= (68/70) x 100%
= 0.97 x 100%
= 97 %

2. Algortima *K-Mens*
Accuracy = ((57+37)/(57+41+38+37)) x 100%
= (94/173) x 100%
= 0.54 x 100%
= 54%

Dengan menggunakan rumus *confusion matrix* diperoleh, nilai *accuracy* dari kedua algoritma. Algoritma *Naive Bayes* mempunyai nilai *accuracy* 97% sedangkan algoritma *K-Mens* mempunyai nilai *accuracy* 54%.

4.3.3. Precission

1. Algoritma *Naive Bayes*
Precission = (37)/(37+0)) x 100%
= (37/37) x 100%
= 1 x 100%
= 100%

2. Algortima *K-Mens*
Precission = (57/(57+41)) x 100%
= (57/98) x 100%
= 0. 58 x 100%
= 58 %

Dengan menggunakan rumus *confusion matrix* diperoleh, nilai precission dari kedua algoritma. Algoritma *Naive Bayes* mempunyai nilai *precission* 100% sedangkan algoritma *K-Mens* mempunyai nilai *precission* 58%.

4.3.4. Recall

1. Algoritma *Naive Bayes*
Recall = (37/(37+2)) x 100%
= (37/39) x 100%
= 0.94 x 100%
= 94 %

2. Alortima *K-Means*
Recall = (57/(57+38)) x 100%
= (57/95) x 100%
= 0.6 x 100%
= 60 %

Dengan menggunakan rumus *confusion matrix* diperoleh, nilai recall dari kedua algoritma. Algoritma *Naive Bayes* mempunyai nilai *recall* 97% sedangkan algoritma *K-Mens* mempunyai nilai *recall* 60%.

Comparison Analysis:

Hasil perbandingan akurasi diagnosis Balita Sunting menggunakan k-means clustering dan Naïve Bayes classifier disajikan pada tabel 4.13

Tabel 4.13. Perbandingan Nilai Akurasi

	K-Means	Naïve Bayes
Accuracy	54%	97%
Precission	98%	100%
Recall	60%	94%

Tabel 4.13 menunjukkan nilai accuracy pada naïve bayes lebih tinggi dari K-Means ini berarti bahwa Akurasi sangat bagus karena dataset memiliki jumlah data False Negatif dan False Positif yang sangat mendekati Symmetric. Dengan demikin jumlah data latih dan model data akan mempengaruhi nilai akurasi pada kedua algortima tersebut. Sementara nilai precission juga menunjukkan naïve bayes lebih tinggi dari k-menas karena hal ini karena kita lebih menginginkan terjadinya True Positif dan sangat tidak menginginkan terjadinya False Positif contohnya kita lebih memilih yang sebenarnya tidak stunting tetapi prediksi masuk dalam stunting (Sehingga tetap akan mendapat penanganan) daripada balita yang sebenarnya stunting tetapi masuk dalam tidak stunting (sehingga tidak mendapat penanganan). Pada bagian Reacall, Naïve bayes juga lebih tinggi dari K-Means hal ini menunjuukan bahwa naïve bayes lebih memilih False Positif lebih baik terjadi daripada False Negatif misalnya Algoritma naïve bayes memprediksi balita positif Stunting tetapi sebenarnya tidak stunting daripada salah memprediksi bahwa balita diprediksi tidak stunting padahal sebenarnya balita stunting.

KESIMPULAN

Semulah data balita ada 173 data yang terdiri dari 98 data absence dan 75 data presence. Pada proses klasifikasi dikelompokkan menjadi dua klasifikasi dengan data training sebanyak 103 data dan pada data testing sebanyak 70 data pada hasil klasifikasi pada status Absence 37 data sedangkan pada data Presence 33 data. Untuk hasil proses Clustering, cluster pertama yaitu absence terdiri dari 95 data. Cluster kedua yaitu presence terdiri dari 78 data. Maka, hasil perhitungan implementasi metode yang didapatkan pada kedua metode tersebut untuk nilai accuracy algoritma Naive Bayes mempunyai nilai 97% sedangkan K-Mens mempunyai nilai 54%. Dengan menggunakan parameter precission algoritma Naive Bayes mempunyai nilai 100% sedangkan K-Mens mempunyai nilai 58%, Dengan menggunakan parameter recall algoritma Naive Bayes mempunyai nilai 94% sedangkan K-Mens mempunyai nilai 60%.

Hasil dari perhitungan yang dilakukan dapat ditarik kesimpulan bahwa algoritma Naive Bayes memiliki accuracy, precission dan recall yang tinggi. Hal ini dipengaruhi oleh model data dan jumlah data training yang digunakan utamanya pada algoritma K-Means yang perlu menggunakan aloritma normalisasi, pengaturan data dan model clustering seperti *Partition Entroy* (PE), Elbow criterion atau G-Means hal ini dilakukan untuk melihat apakah suatu cluster sudah dalam keadaan terdistribusi secara normal atau tidak

Untuk meningkatkan kinerja sistem maka sebaiknya menambahkan indikator barupa parameter selain fisik untuk memperjelas keadaan balita apakah berpengaruh terhadap resiko stunting, penambahan data sampel agar hasil analisis sistem lebih akurat, selain itu juga dapat melakukan komparasi dengan menggunakan metode yang berbeda atau melakukan normalisasi data menggunakan metode yang tepat.

UCAPAN TERIMA KASIH

Terima kasih banyak yang sebesar besarnya kami ucapkan kepada pengelola laboratorium riset Fakultas Ilmu Komputer Universitas Al Asyariah Mandar yang telah menfasilitasi pengolahan dan analisis data penelitian, juga kepada staf puskesmas yang membantu dalam menyediakan data penelitan, tidak lupa untuk segenap pihak yang turut berpartisipasi pada penyelesaian penelitian ini.

DAFTAR PUSTAKA

A. M. Carrington *et al.*, “Deep ROC Analysis and AUC as Balanced Average Accuracy, for Improved Classifier Selection, Audit and Explanation,” *IEEE Trans. Pattern Anal. Mach. Intell.*, no. August, 2022, doi: 10.1109/TPAMI.2022.3145392

H. Chen, S. Hu, R. Hua, and X. Zhao, “Improved naive Bayes classification algorithm for traffic risk management,” *EURASIP J. Adv. Signal Process.*, vol. 2021, no. 1, 2021, doi: 10.1186/s13634-021-00742-6.

H. Rasmita and S. Hendra, “Comparative Analysis of C4 . 5 And Naïve Bayes Algorithms for Classification of Food Vulnerable Areas,” vol. 3, no. 1, pp. 35–41, 2022.

K. A. K-nn and R. G. Whendasmoro, “Analisis Penerapan Normalisasi Data Dengan Menggunakan Z-Score Pada,” vol. 9, no. 4, pp. 872–876, 2022, doi: 10.30865/jurikom.v9i4.4526.

M. K. RI, *Peraturan Menteri Kesehatan Republik Indonesia Nomor 2 Tahun 2020 Tentang Standar Antropometri Anak*, vol. 21, no. 1. 2020, pp. 1–9. [Online]. Available: http://mpoc.org.my/malaysian-palm-oil-industry/

M. R. Yuliansyah, M. B, and A. Franz, “Perbandingan Metode K-Nearest Neighbors dan Naïve Bayes Classifier Pada Klasifikasi Status Gizi Balita di Puskesmas Muara Jawa Kota Samarinda,” *ATASI Adopsi Teknol. dan Sist. Inf.*, vol. 1, no. 1, pp. 08–21, 2022

M. Y. Titimeidara and W. Hadikurniawati, “Implementasi Metode Naïve Bayes Classifier Untuk Klasifikasi Status Gizi Stunting Pada Balita,” *J. Ilm. Inform.*, vol. 9, no. 01, pp. 54–59, 2021, doi: 10.33884/jif.v9i01.3741.

N. A. Mansour, A. I. Saleh, M. Badawy, and H. A. Ali, *Accurate detection of Covid-19 patients based on Feature Correlated Naïve Bayes (FCNB) classification strategy*, vol. 13, no. 1. Springer Berlin Heidelberg, 2022. doi: 10.1007/s12652-020-02883-2.

N. Syafina, M. Jainalabidin, A. Fawwaz, M. Amidon, and N. Ismail, “The k-Nearest Neighbor modelling by varying Mahalanobis and Correlation in distance metric for agarwood oil quality classification,” vol. 11, no. 3, pp. 242–252, 2022, doi: 10.11591/ijaas.v11.i3.pp242-252.

O. Saeful Bachri and R. M. Herdian Bhakti, “Penentuan Status Stunting pada Anak dengan Menggunakan Algoritma KNN,” *J. Ilm. Intech Inf. Technol. J. UMUS*, vol. 3, no. 02, pp. 130–137, 2021, doi: 10.46772/intech.v3i02.533

R. Puspitarahayu and E. Nasution, “An Overview of the Growth of Stunting Toddlers in the Supplementary Feeding Program (PMT) in Padang Tualang Village, Langkat Regency,” *Repos. Inst. Univ. Sumatra Utara*, p. 2020, 2020, [Online]. Available: https://repositori.usu.ac.id/handle/123456789/29893?show=full

R. Setiawan and A. Triayudi, “JURNAL MEDIA INFORMATIKA BUDIDARMA Klasifikasi Status Gizi Balita Menggunakan Naïve Bayes dan K-Nearest Neighbor Berbasis Web,” *J. Media Inform. Budidarma*, vol. 6, no. 2, pp. 777–785, 2022, doi: 10.30865/mib.v6i2.3566.

S. S and H. Wang, “Naive Bayes and Entropy based Analysis and Classification of Humans and Chat Bots,” *J. ISMAC*, vol. 3, no. 1, pp. 40–49, 2021, doi: 10.36548/jismac.2021.1.004.

S. M. J. Rahman *et al.*, “Investigate the risk factors of stunting, wasting, and underweight among under-five Bangladeshi children and its prediction based on machine learning approach,” *PLoS One*, vol. 16, no. 6 June 2021, pp. 1–11, 2021, doi: 10.1371/journal.pone.0253172.

T. Prasetiya, I. Ali, C. L. Rohmat, and O. Nurdiawan, “Klasifikasi Status Stunting Balita Di Desa

Slangit Menggunakan Metode K-Nearest Neighbor," *INFORMATICS Educ. Prof. J. Informatics*, vol. 5, no. 1, p. 93, 2020, doi: 10.51211/itbi.v5i1.1431

T. S. Priyadarshini, M. Abdul, and B. Ssali, "Deep Learning Prediction Model for predicting heart stroke using the combination Sequential Row Method integrated with Artificial Neural Network," *J. Posit. Sch. Psychol.*, vol. 6, no. 4, pp. 623–629, 2022.

T. Gneiting and E. M. Walz, "Receiver operating characteristic (ROC) movies, universal ROC (UROC) curves, and coefficient of predictive ability (CPA)," *Mach. Learn.*, vol. 111, no. 8, pp. 2769–2797, 2022, doi: 10.1007/s10994-021-06114-3.

T. E. Putri, R. T. Subagio, Kusnadi, and P. Sobiki, "Classification System of Toddler Nutrition Status using Naïve Bayes Classifier Based on Z-Score Value and Anthropometry Index," *J. Phys. Conf. Ser.*, vol. 1641, no. 1, 2020, doi: 10.1088/1742-6596/1641/1/012005.

T. T. Ramanathan, J. Hossen, and S. Sayeed, "Naïve Bayes Based Multiple Parallel Fuzzy Reasoning Method for Medical Diagnosis," *J. Eng. Sci. Technol.*, vol. 17, no. 1, pp. 472–490, 2022.

W. I. Rahayu, C. Prianto, and E. A. Novia, "Perbandingan Algoritma K-Means Dan Naïve Bayes Untuk Memprediksi Prioritas Pembayaran Tagihan Rumah Sakit Berdasarkan Tingkat Kepentingan Pada Pt. Pertamina (Persero)," *J. Tek. Inform.*, vol. 13, no. 2, pp. 1–8, 2021, [Online]. Available: https://ejurnal.poltekpos.ac.id/index.php/informatika/article/view/1383

WHO (World Health Organization), *World health statistics 2018: monitoring health for the SDGs, sustainable development goals*, vol. 66. 2018. [Online]. Available: http://apps.who.int/iris