

Perbandingan Metode Naïve Bayes dan C4.5 Klasifikasi Status Gizi Bayi Balita

Alter Lasarudin¹, Hilmansyah Gani², Misran Tomayahu³

Program Studi Sistem Informasi,
Universitas Muhammadiyah Gorontalo^{1,2,3}
alterlasarudin@umgo.ac.id¹
hilmansyahgani@umgo.ac.id²
misrantomayahu@gmail.com³

Article Info

Article history:

Submitted December 2022

Revised December 2022

Accepted December 2022

Published December 2022

Keyword:

Data Mining
Classification
Naïve Bayes
C4.5

ABSTRACT

In supporting good services, an orderly, neat and meticulous work procedure are needed. So that information will be produced quickly, precisely and accurately. In health institution, there is a lot of data which increase every year. One of them is data on toddler growth and development. However, the excessive data have consequences more difficult to learn the data were only used as archives. The use of data mining techniques is expected to help in overcoming the imperferent development of toddlers. In this study, researcher compared the classification techniques of performance methods (C4.5) and Naïve Bayes. The attributes used in this research consist of Gender, Age, Weight, Body Length, Region and Growth with 500 data. The results of the research conducted, based on the accuracy and recall values of Naive Bayes were higher than the accuracy value of C4.5 with an accuracy value 94.20% for Naive Bayes, and C4.5 with an accuracy value of 85.80%. Although, in this study the Precision level is higher Naive Bayes which is 94.20% compared to C4.5 which is 85.80%. The final result of this research is that the Naive Bayes method is better used than the decision tree method.

Kata Kunci:

Data Mining
Klasifikasi
Naïve Bayes
C4.5

ABSTRAK

Sebagai upaya instansi kesehatan dalam mendukung pelayanan yang baik diperlukan tata kerja yang tertib, rapi dan teliti sehingga akan menghasilkan informasi yang cepat, tepat dan akurat. Dalam instansi kesehatan banyak data yang dari setiap tahunnya bertambah. Salah satunya yaitu data tumbuh kembang balita. Akan tetapi dengan sekian banyaknya data tumbuh kembang balita semakin sulit juga data tersebut dipelajari lebih lanjut, dan umumnya data tersebut hanya digunakan sebagai arsip saja. Pemanfaatan teknik data mining diharapkan dapat membantu dalam mengatasi tumbuh kembang balita yang kurang baik saat ini. Pada penelitian ini peneliti membandingkan teknik klasifikasi dari kinerja metode (C4.5) dan Naive Bayes. Atribut yang digunakan terdiri dari Gender, Umur, Berat badan, Panjang Badan Waktu, Wilayah dan Tumbuh kembang. Dengan menggunakan data sebanyak 500 data. Hasil dari penelitian yang dilakukan, berdasarkan dari nilai *accuracy* maupun *recallnya naive bayes* lebih tinggi dibandingkan dengan nilai *akurasi C4.5* dengan nilai akurasi 94,20% untuk *naive bayes*, dan *C4.5 nilai akurasi 85,80%*. Meskipun dalam penelitian ini tingkat *Precision-nya* lebih tinggi *Naive Bayes*

yaitu 94,20% dibandingkan C4.5 yaitu 85,80 %. Hasil akhir dari penelitian ini adalah metode *Naive Bayes* lebih baik digunakan dari pada metode *decision tree*.

1. PENDAHULUAN

Kehidupan Manusia dimulai dari sejak janin dalam rahim ibu. Sejak saat itu manusia kecil sudah memulai perjuangan untuk bertahan hidup, salah satunya dari kemungkinan kurangnya gizi yang diterima dari ibu yang mengandungnya (Purwati et al., 2017)

Setelah lahir, usia Bayi dan balita adalah usia yang paling rentan mengalami berbagai masalah kesehatan dan gizi. Kondisi kesehatan dan status gizi bayi dan balita merupakan salah satu tolok ukur cerminan keadaan gizi masyarakat secara luas. Kasus yang selalu terjadi tidak hanya menjadi beban keluarga tetapi juga menjadi beban Negara. Menurut (Purwati et al., 2017) bahwa status gizi bayi dan balita sangat dipengaruhi oleh asupan gizi pada menu makanan yang diberikan.

Namun saat ini masih banyak orang tua yang memiliki keterbatasan pengetahuan tentang kebutuhan gizi bayi dan balita, sehingga menyebabkan bayi dan balita mengalami masalah gizi. Faktor lain adalah keterlambatan dalam penanganan penyakit gizi dan kurangnya tenaga ahli atau pakar.

Determinan biologis seperti jenis kelamin, kondisi dalam Rahim, berat bayi saat lahir jumlah kelahiran orang tua serta ukuran tubuh orang mempengaruhi status gizi bayi. Menurut Devi (2010) dalam (Herawati, 2020) bahwa status gizi dapat ditentukan dengan melakukan pemeriksaan laboratorium maupun secara antropometri. Antropometri adalah cara dalam menentukan status gizi yang paling mudah. Indikator TB/U, BB/U, dan BB/TB direkomendasikan yang baik dalam menentukan status gizi balita.

Dalam penelitian ini penulis melakukan perbandingan metode *Naive Bayes* dan metode *C.45* dan untuk mengklasifikasikan status gizi balita. Berdasarkan indeks antropometri bayi dan balita harus memperoleh asupan makanan masa tumbuh kembang bayi dan balita adalah masa yang sangat beresiko bagi setiap kehidupan, maka sangat memperhatikan semua aspek yang mendukung dan yang mempengaruhi status gizi dua peristiwa yang berbeda namun saling berkaitan dan saling mempengaruhi pertumbuhan (*growth*) itu sendiri mempunyai pengertian yaitu berkaitan dengan masalah perubahan perubahan ukuran besar jumlah atau dimensi pada tingkat sel, organ maupun individu. Pertumbuhan bersifat kuantitatif sehingga dapat diukur satuan Umur (Bulan), Berat Berat (Gram, Kilogram) satuan Panjang Badan (Cm, M) dan keseimbangan *metabolic* (proses penguraian nutrisi dari makanan menjadi energi yang dibutuhkan oleh tubuh) (Murdianingsih et al., 2016).

Undang-undang No 20 Tahun 2003 menyebut anak bayi dan balita adalah anak dalam rentang usia 0-59 Bulan. Kelompok bayi dan balita berada dalam proses pertumbuhan dan perkembangan yang bersifat unik. artinya memiliki pola pertumbuhan dan perkembangan yang harus diperhatikan orang tua. Pada tahun 2008, sesuai pendataan Puskesmas dilihat sesuai dengan angka status gizi bahwa masih ada bayi dan balita yang asupan gizinya belum baik.

Status gizi bayi dan balita dapat kita klasifikasikan sesuai dengan pengelompokan umur, berat badan, dan panjang badan, untuk dapat bisa memudahkan kita dalam menentukan apakah bayi dan balita ini sehat atau tidak. Berdasarkan klasifikasi tersebut maka status gizi dengan mudah kita ketahui. Dengan demikian maka status gizi sangatlah penting untuk diklasifikasikan.

Terdapat banyak metode yang dapat dalam klasifikasi diantaranya, adalah *Artificial Neural Network (ANN)*, *Support Vector Machine (SVM)*, *Bayesian*, *Decision Tree (C4.5)* dan lain sebagainya. Metode klasifikasi *Naive Bayes* dan *C4.5* merupakan metode yang banyak digunakan untuk mengklasifikasi. Hal ini dibuktikan dengan banyaknya penelitian yang telah dilakukan sebelumnya.

(Nuha, 2020) melakukan penelitian dengan judul klasifikasi status gizi balita menggunakan metode *Naive Bayes Classifier* berbasis *Website*. Hasil akurasi yang didapatkan pada penelitian ini sebesar 90% dan galat sebesar 10% dengan menggunakan pengujian *confusion matrix*.

Penelitian lainnya yang dilakukan oleh (Farhurrokiem Azizi, 2021) tentang Klasifikasi Status Gizi Menggunakan Perhitungan *Body Mass Index* Dengan Metode Algoritma *C4.5* Pada

Masyarakat Di Kota Yogyakarta. Hasil klasifikasi menggunakan Metode *C4.5* pada penelitian ini mendapatkan akurasi sebesar 85%. Dengan demikian dapat disimpulkan bahwa metode *C4.5* mampu mengklasifikasikan status gizi balita.

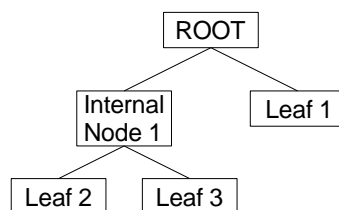
Berdasarkan permasalahan dan beberapa penelitian di atas, maka peneliti akan melakukan perbandingan akurasi metode *Naïve bayes* dan metode *C4.5* untuk mengklasifikasi status gizi pada bayi dan balita.

2. METODE PENELITIAN

2.1. Pohon Keputusan

Pohon keputusan adalah suatu metode klasifikasi dan prediksi yang sangat kuat dan terkenal. Metode pohon keputusan dapat mengubah fakta yang sangat besar menjadi pohon keputusan yang merepresentasikan aturan-aturan yang dengan mudah dipahami dengan bahasa alami. Dan berguna mengeksplorasi data, menemukan hubungan tersembunyi antara sejumlah calon variabel *input* dengan sebuah variabel target.

Pohon keputusan (*decision tree*) merupakan sebuah struktur yang dapat digunakan untuk membagi kumpulan data yang besar menjadi himpunan-himpunan *record* yang lebih kecil dengan menerapkan serangkaian aturan keputusan, dengan masing-masing rangkaian pembagian, anggota himpunan hasil menjadi mirip satu dengan yang lain. Metode *decision tree* digunakan untuk memperkirakan nilai *diskrit* dari fungsi target. Menurut (Yuan dan Shaw, 1995) yang dikutip dari (Rintiasti & Fauzi, 2018) bahwa pohon keputusan merupakan himpunan aturan IF, THEN. Dimana setiap path dalam tree terhubung dengan aturan, dimana premis terdiri atas sekumpulan node-node yang ditemui, dan kesimpulan dari aturan terdiri atas kelas yang terhubung dengan leaf dari path. Data dalam pohon keputusan biasanya dinyatakan dalam bentuk tabel dengan atribut dan *record*. Atribut menyatakan suatu parameter yang dibuat sebagai kriteria dalam pembentukan pohon. Misalkan untuk menentukan main *golf*, kriteria yang diperhatikan adalah cuaca, angin, dan temperatur. Salah satu atribut merupakan atribut yang menyatakan data solusi per *item* data yang disebut target atribut. Atribut memiliki nilai-nilai yang dinamakan dengan *instance*. Misalkan atribut cuaca mempunyai *instance* berupa cerah, awan, berawan, dan hujan. Proses pada pohon keputusan adalah mengubah bentuk data (tabel) menjadi model pohon, mengubah model pohon menjadi *rule*, dan menyederhanakan *rule* (Basuki dan Syarif, 2003) dalam (Ratniasih, 2019). Pohon keputusan dapat dilihat seperti contoh gambar 1 berikut ini



Gambar 1. Konsep Dasar Pohon Keputusan (Rintiasti & Fauzi, 2018)

2.2. Naïve Bayes (NB)

Naive Bayes merupakan teknik prediksi berbasis probabilitas sederhana yang berdasar pada penerapan aturan *bayes* dengan asumsi ketidaktergantungan yang kuat. Selain itu *Naive Bayes* juga dapat menganalisa variabel-variabel yang paling mempengaruhinya dalam bentuk peluang (Prasetyo, 2012) dalam (Subarkah et al., 2020).

$$P(H|X) = \frac{P(X|H) \cdot P(H)}{P(X)} \quad (2.1)$$

Berikut merupakan persamaan dari teorema *Bayes*.

Keterangan:

- X : data dengan *class* yang belum diketahui
 H : hipotesis data X merupakan suatu *class* spesifik
 $P(H|X)$: probabilitas akhir bersyarat suatu hipotesis H terjadi jika diberikan bukti X terjadi
 $P(H)$: Probabilitas awal (*priori*) hipotesis H terjadi tanpa memandang bukti apapun
 $P(X|H)$: probabilitas sebuah bukti X terjadi akan mempengaruhi hipotesis lain.

Naïve Bayes merupakan salah satu algoritma yang banyak digunakan dalam pengklasifikasian baik dokumen maupun teks. Hubungan *Naïve Bayes* dengan klasifikasi adalah bahwa dalam teorema Bayes terdapat label kelas yang menjadi target dalam klasifikasi. Formulasi *Naïve Bayes* digunakan untuk mencari nilai probabilitas dari setiap kata yang muncul pada kategori atau *class*. Setelah mendapatkan hasil probabilitas dari setiap kata pada kategori selanjutnya dilakukan penjumlahan probabilitas tersebut sesuai dengan kata yang di *testing*.

Klasifikasi dengan *Naïve Bayes* bekerja berdasarkan teori probabilitas yang memandang semua fitur dari kata sebagai bukti dalam probabilitas. Hal ini menjadi karakteristik *Naïve Bayes* sebagai berikut:

- 1) Metode *Naïve Bayes* teguh (*robust*) terhadap data-data yang terisolasi yang biasanya merupakan data dengan karakteristik berbeda (*outlier*). *Naïve Bayes* juga bisa menangani nilai atribut yang salah dengan mengabaikan data latih selama proses pembangunan model dan prediksi.
- 2) Tangguh menghadapi atribut yang tidak relevan
- 3) Atribut yang tidak mempunyai korelasi bisa mendegradasi kinerja klasifikasi *Naïve Bayes* karena asumsi independensi tersebut sudah tidak ada.

2.3. Tahapan Metode *Naïve Bayes*

Tahapan dari proses metode *Naïve Bayes* adalah:

1. Menghitung jumlah kelas/label
2. Menghitung jumlah kasus perkelas
3. Kalikan semua Variabel kelas
4. Bandingkan hasil perkelas

2.4. Metode C4.5

Algoritma data mining *C4.5* adalah salah satu algoritma yang sering digunakan untuk melakukan klasifikasi yang bersifat prediktif. Algoritma *C4.5* merupakan pengembangan dari algoritma ID3. Menurut Widiyastuti (2019) yang dikutip dari (Hafizan & Putri, 2020) mengungkapkan bahwa klasifikasi Algoritma *C4.5* dapat digunakan dengan membangun pohon keputusan (*Decision Tree*). Secara umum, metode *C4.5* untuk membangun pohon keputusan dimulai dari pemilihan atribut sebagai akar, membuat cabang untuk tiap-tiap nilai, membagi kasus dalam cabang dan mengulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama.

Menurut (Elisa, 2007) yang dikutip dari (Setio et al., 2020) bahwa terdapat empat hal yang membedakan algoritma *C4.5* dengan ID3 antara lain: 1) tahan (*robust*) terhadap data *noise*; 2) mampu menangani variabel dengan tipe diskrit maupun kontinu; 3) mampu menangani variabel yang memiliki *missing value*; dan 4) dapat memangkas cabang dari pohon keputusan.

2.5. Tahapan Metode C4.5

Secara umum metode *C4.5* untuk membangun pohon keputusan adalah sebagai berikut:

1. Mengumpulkan data *training*
2. Menyeleksi data
3. Menentukan nilai *gain*
4. Menentukan akar pohon
5. Menganalisa data

2.6. Kelebihan dan Kekurangan Metode C4.5

Metode *C4.5* memiliki kelebihan dan kelemahan dalam sistem analisisnya, Kelebihan metode ini adalah :

1. Mudah dipahami dan diinterpretasikan dalam bentuk pohon.

2. Biaya komputasi untuk membangunnya tidak mahal. Metode ini membutuhkan sedikit ruang penyimpanan untuk data latih dibandingkan metode klasifikasi yang lain.
3. Bisa digunakan untuk tipe data kategorikal dan numerik.
4. Merepresentasikan model seperti 'white-box', dimana proses logika keputusannya dapat diikuti dengan mudah mengikuti arah dalam *Decision Tree*-nya
5. Metode ini menggunakan teknik statistik yang klasik untuk membuat validasi model menjadi mungkin dilakukan.
6. Metodenya handal, cepat, dan memproses dengan baik pada latih yang banyak
7. Akurasiya dapat dibandingkan dengan metode klasifikasi yang lain bagi banyak data set sederhana.

Sedangkan kekurangan dari metode *C4.5* adalah sebagai berikut :

1. Bisa terjadi *overfitting* ketika ada data anomali didalamnya .
2. Bisa terjadi *under-fitting* ketika ada proses *pre-pruning* yang tidak diinginkan

3. Hasil dan Pembahasan

Penelitian ini bertujuan untuk membandingkan performa akurasi klasifikasi menggunakan metode *Naïve Bayes* dan *C4.5* untuk proses klasifikasinya dan algoritma *Information Gain* untuk seleksi atribut. Jumlah data set yang digunakan dalam penelitian ini sebanyak 500 data dimana dari jumlah data set tersebut dibagi menjadi 90% data *training* dan 10% data *testing*. Data set diambil dari Puskesmas Mootilango pada tahun 2018. Ada tiga atribut yang digunakan, yaitu Umur, Berat Badan, Panjang Badan. Pada penelitian ini menggunakan bantuan *tool* yang ada pada Rapidminer.

Tabel 1. Data Hasil Normalisasi

No.	UMUR	BB	PB	STATUS GIZI
1	BALITA	KURUS	PENDEK	KURANG
2	BALITA	NORMAL	PENDEK	BAIK
3	BALITA	NORMAL	TINGGI	BAIK
4	BALITA	NORMAL	TINGGI	BAIK
5	BALITA	NORMAL	TINGGI	BAIK
6	BALITA	NORMAL	TINGGI	BAIK
7	BALITA	NORMAL	TINGGI	BAIK
8	BALITA	NORMAL	TINGGI	BAIK
9	BALITA	NORMAL	TINGGI	BAIK
10	BALITA	NORMAL	TINGGI	BAIK
11	BALITA	NORMAL	TINGGI	BAIK
12	BALITA	NORMAL	PENDEK	KURANG
13	BALITA	NORMAL	PENDEK	BAIK
14	BALITA	KURUS	PENDEK	KURANG
15	BALITA	NORMAL	PENDEK	BAIK
16	BALITA	NORMAL	PENDEK	BAIK
17	BALITA	NORMAL	TINGGI	BAIK
18	BALITA	NORMAL	PENDEK	BAIK
.....
500	BALITA	NORMAL	PENDEK	BAIK

Tabel di atas merupakan hasil dari data awal yang sudah di normalisasi, yang artinya atribut yang dianggap tidak berpengaruh pada pengolahan data telah dihilangkan.

Dengan demikian dapat dijelaskan bahwa berat badan balita dikatakan kurus yaitu dengan menggunakan rumus berat badan dibagi dengan tinggi badan, panjang badan balita dikatakan pendek dengan menggunakan rumus tinggi badan dibagi dengan umur, dan status gizi balita dikatakan kurang, baik, buruk menggunakan rumus berat badan dibagi dengan umur. Pada masing-masing rumus dapat dilihat dalam kategori di bawah ini:

1. Status gizi kurus
BB/TB = 9/75
2. Status gizi pendek
TB/U = 89/35
3. Status gizi baik
BB/U = 11/10
4. Status gizi kurang
BB/U = 7/15
5. Status gizi buruk
BB/U = 6/19

3.1. Contoh Perhitungan Algoritma Naïve Bayes

Dengan menggunakan algoritma *Naïve Bayes* seperti pada rumus 2.1, peneliti melakukan klasifikasi dengan menggunakan sampel 15 data *training* dan 1 data *testing*, terlihat pada tabel 2.

Tabel 2. Data Training dan Testing

NO	UMUR	BB	PB	STATUS GIZI	Ket
1	BALITA	NORMAL	PENDEK	BAIK	Training
2	BALITA	NORMAL	TINGGI	BAIK	Training
3	BALITA	NORMAL	TINGGI	BAIK	Training
4	BALITA	NORMAL	TINGGI	BAIK	Training
5	BALITA	NORMAL	TINGGI	BAIK	Training
6	BALITA	KURUS	PENDEK	KURANG	Training
7	BALITA	NORMAL	PENDEK	KURANG	Training
8	BALITA	KURUS	PENDEK	KURANG	Training
9	BALITA	KURUS SEKALI	PENDEK	KURANG	Training
10	BALITA	NORMAL	PENDEK	KURANG	Training
11	BALITA	NORMAL	SANGAT PENDEK	BURUK	Training
12	BALITA	KURUS SEKALI	PENDEK	BURUK	Training
13	BAYI	KURUS SEKALI	SANGAT PENDEK	BURUK	Training
14	BAYI	KURUS SEKALI	SANGAT PENDEK	BURUK	Training
15	BALITA	KURUS	PENDEK	BURUK	Training
16	BALITA	NORMAL	PENDEK	?	Testing

Berdasarkan tabel 2 jumlah *training* yang digunakan pada percobaan dengan menggunakan metode *Naïve Bayes* terdapat 15 data *training* sampel dan 1 data *testing*. Dengan menggunakan rumus 2.1 maka nilai probabilitas pada masing-masing kategori sebagai berikut :

1. Menghitung *Class/label*

$$P(Y=\text{Baik}) = \frac{5}{15} \text{ "Jumlah data baik pada pengujian ini dibagi dengan jumlah data seluruhnya" } = \frac{5}{15}$$

$$P(Y = \text{Kurang}) = \frac{5}{15} \text{ "Jumlah data kurang pada pengujian ini dibagi dengan jumlah data seluruhnya" } = \frac{5}{15}$$

$$P(Y = \text{Buruk}) = \frac{5}{15} \text{ "Jumlah data buruk pada pengujian ini dibagi dengan jumlah data seluruhnya" } = \frac{5}{15}$$

2. Menghitung jumlah kasus perkelas.

$$P(Y=\text{Baik}) = \frac{5}{15} = 0,3$$

$$P(Y = \text{Kurang}) = \frac{5}{15} = 0,3$$

$$P(Y = \text{Buruk}) = \frac{5}{15} = 0,3$$

$$\begin{aligned}
P(\text{umur}=\text{Balita}|\text{Y}=\text{Baik}) &= \frac{5}{5} = 1 \\
P(\text{umur}=\text{Balita}|\text{Y}=\text{Kurang}) &= \frac{5}{5} = 1 \\
P(\text{umur}=\text{Balita}|\text{Y}=\text{Buruk}) &= \frac{3}{5} = 0,6 \\
P(\text{umur}=\text{Bayi}|\text{Y}=\text{Baik}) &= \frac{0}{5} = 0 \\
P(\text{umur}=\text{Bayi}|\text{Y}=\text{Kurang}) &= \frac{0}{5} = 0 \\
P(\text{umur}=\text{Bayi}|\text{Y}=\text{Buruk}) &= \frac{2}{5} = 0,4 \\
P(\text{BB}=\text{Normal}|\text{Y}=\text{Baik}) &= \frac{5}{5} = 1 \\
P(\text{BB}=\text{Normal}|\text{Y}=\text{Kurang}) &= \frac{2}{5} = 0,4 \\
P(\text{BB}=\text{Normal}|\text{Y}=\text{Buruk}) &= \frac{1}{5} = 0,2 \\
P(\text{BB}=\text{Kurus}|\text{Y}=\text{Baik}) &= \frac{0}{5} = 0 \\
P(\text{BB}=\text{Kurus}|\text{Y}=\text{Kurang}) &= \frac{2}{5} = 0,4 \\
P(\text{BB}=\text{Kurus}|\text{Y}=\text{Buruk}) &= \frac{1}{5} = 0,2 \\
P(\text{BB}=\text{KurusSekali}|\text{Y}=\text{Baik}) &= \frac{0}{5} = 0 \\
P(\text{BB}=\text{KurusSekali}|\text{Y}=\text{Kurang}) &= \frac{1}{5} = 0,2 \\
P(\text{BB}=\text{KurusSekali}|\text{Y}=\text{Buruk}) &= \frac{3}{5} = 0,6 \\
P(\text{PB}=\text{Tinggi}|\text{Y}=\text{Baik}) &= \frac{4}{5} = 0,8 \\
P(\text{PB}=\text{Tinggi}|\text{Y}=\text{Kurang}) &= \frac{0}{5} = 0 \\
P(\text{PB}=\text{Tinggi}|\text{Y}=\text{Buruk}) &= \frac{0}{5} = 0 \\
P(\text{PB}=\text{Pendek}|\text{Y}=\text{Baik}) &= \frac{1}{5} = 0,2 \\
P(\text{PB}=\text{Pendek}|\text{Y}=\text{Kurang}) &= \frac{5}{5} = 1 \\
P(\text{PB}=\text{Pendek}|\text{Y}=\text{Buruk}) &= \frac{2}{5} = 0,4 \\
P(\text{PB}=\text{SangatPendek}|\text{Y}=\text{Baik}) &= \frac{0}{5} = 0 \\
P(\text{PB}=\text{SangatPendek}|\text{Y}=\text{Kurang}) &= \frac{0}{5} = 0 \\
P(\text{PB}=\text{SangatPendek}|\text{Y}=\text{Buruk}) &= \frac{3}{5} = 0,6
\end{aligned}$$

3. Setelah nilai probabilitas ditemukan maka hitunglah nilai probabilitas disetiap kategori.

$$\begin{aligned}
&P(\text{Y}=\text{Baik}) * P(\text{umur}=\text{Balita}|\text{Y}=\text{Baik}) * P(\text{umur}=\text{Bayi}|\text{Y}=\text{Baik}) * P(\text{BB}=\text{Normal}|\text{Y}=\text{Baik}) * P(\text{BB}=\text{Kurus}|\text{Y}=\text{Baik}) * P(\text{BB}=\text{KurusSekali}|\text{Y}=\text{Baik}) * P(\text{PB}=\text{Tinggi}|\text{Y}=\text{Baik}) * P(\text{PB}=\text{Pendek}|\text{Y}=\text{Baik}) * P(\text{PB}=\text{SangatPendek}|\text{Y}=\text{Baik})
\end{aligned}$$

$$\begin{aligned}
&= \frac{5}{15} \times \frac{5}{5} \times \frac{0}{5} \times \frac{5}{5} \times \frac{0}{5} \times \frac{0}{5} \times \frac{4}{5} \times \frac{1}{5} \times \frac{0}{5} \\
&= 0,3 \times 1 \times 0 \times 1 \times 0 \times 0 \times 0,8 \times 0,2 \times 0 \\
&= 0
\end{aligned}$$

$$\begin{aligned}
&P(\text{Y}=\text{Kurang}) * P(\text{umur}=\text{Balita}|\text{Y}=\text{Kurang}) * P(\text{umur}=\text{Bayi}|\text{Y}=\text{Kurang}) * P(\text{BB}=\text{Normal}|\text{Y}=\text{Kurang}) * P(\text{BB}=\text{Kurus}|\text{Y}=\text{Kurang}) * P(\text{BB}=\text{KurusSekali}|\text{Y}=\text{Kurang}) * P(\text{PB}=\text{Tinggi}|\text{Y}=\text{Kurang}) * P(\text{PB}=\text{Pendek}|\text{Y}=\text{Kurang}) * P(\text{PB}=\text{SangatPendek}|\text{Y}=\text{Kurang})
\end{aligned}$$

$$\begin{aligned}
&= \frac{5}{15} \times \frac{5}{5} \times \frac{0}{5} \times \frac{2}{5} \times \frac{2}{5} \times \frac{1}{5} \times \frac{0}{5} \times \frac{5}{5} \times \frac{0}{5} \\
&= 0,3 \times 1 \times 0 \times 0,4 \times 0,4 \times 0,2 \times 0 \times 1 \times 0 \\
&= 0
\end{aligned}$$

$$\begin{aligned}
& P(Y=\text{Buruk}) * P(\text{umur}=\text{Balita}|Y=\text{Buruk}) * P(\text{umur}=\text{Bayi}|Y=\text{Buruk}) * \\
& P(\text{BB}=\text{Normal}|Y=\text{Buruk}) * P(\text{BB}=\text{Kurus}|Y=\text{Buruk}) * P(\text{BB}=\text{KurusSekali}|Y= \\
& \text{Buruk}) * P(\text{PB}=\text{Tinggi}|Y=\text{Buruk}) * P(\text{PB}=\text{Pendek}|Y=\text{Buruk}) * P(\text{PB}=\text{SangatPendek}|Y=\text{Buruk}) \\
& = \frac{5}{15} \times \frac{3}{5} \times \frac{2}{5} \times \frac{1}{5} \times \frac{1}{5} \times \frac{3}{5} \times \frac{0}{5} \times \frac{2}{5} \times \frac{3}{5} \\
& = 0,3 \times 0,5 \times 0,4 \times 0,2 \times 0,2 \times 0,6 \times 0 \times 0,4 \times 0,6 \\
& = 0
\end{aligned}$$

4. Bandingkan hasil kelas baik kurang dan buruk
 Dengan melihat hasil perhitungan di atas maka kesimpulannya yaitu nilai probabilitas termasuk pada kategori **Buruk** atau $P = \text{Buruk}$).

3.2. Pengujian Data Metode Naïve Bayes Menggunakan Tool Rapidminer

Data set dibagi menjadi dua bagian yaitu 90% data *training* dan 10% data *testing* dari 500 data yang digunakan.

Tabel 3. Pengujian Data Naïve Bayes

	<i>true</i> KURANG	<i>true</i> BAIK	<i>true</i> BURUK	<i>class precision</i>
<i>pred.</i> KURANG	38	1	2	92.68%
<i>pred.</i> BAIK	24	430	0	94.71%
<i>pred.</i> BURUK	2	0	3	60.00%
<i>class recall</i>	59.38%	99.77%	60.00%	

Tabel hasil pengujian metode *Naïve Bayes* dengan Rapidminer jumlah **True Kurang** (TK) adalah 38 *record* diklasifikasikan gizi kurang dan **True Baik** (TB) sebanyak 1 *record* diklasifikasikan sebagai gizi baik, berikutnya 2 *record* untuk **True Buruk** (TB) diklasifikasikan sebagai gizi bermasalah. Berdasarkan tabel tersebut menunjukkan bahwa tingkat akurasi dengan menggunakan metode *Naïve Bayes* sebesar 94,20%

3.3. Perhitungan Metode C4.5

Dengan menggunakan metode *C4.5* peneliti melakukan klasifikasi dengan menggunakan data set yang sama seperti yang di olah menggunakan algoritma *Naïve Bayes* yaitu 90% data *training* dan 10% data *testing* dari total 500 data set yang digunakan.

Secara umum metode *C4.5* untuk membangun pohon keputusan adalah sebagai berikut:

1. Mengumpulkan data *training*

Untuk mengumpulkan data *training* metode *C4.5* sama dengan mengumpulkan data dengan menggunakan metode *Naïve Bayes*.

2. Menyeleksi data menyeleksi data sama dengan metode *Naïve Bayes*.
3. Menentukan akar pohon.

Dalam menentukan akar pohon dapat dilakukan dengan persamaan rumus seperti di bawah ini :

$$Entropi (S) = \sum_i^n -i \cdot p_i \cdot \log_2 p_i$$

Keterangan :

S = Himpunan data set kasus

n = Banyaknya partisi S

p_i = Perhitungan yang didapat dari jumlah (Atribut), jumlah (baik), jumlah (kurang), jumlah (buruk) dibagi dengan total kasus.

4. Menentukan nilai *gain*

Setelah nilai *entropy* sudah diketahui, maka langkah selanjutnya adalah mencari nilai *gain* pada atribut jurusan, maka gunakan rumus

Untuk menghitung nilai *gain* dari tiap atribut dengan rumus sebagai berikut :

$$Gain (S,A) = entropy (s) - \sum_i^n 1 \frac{[si]}{s} Entropi (s)$$

Maka

$$Gain = Entropy (total) - \left(\left(\frac{Sum(bayi)}{Jumlah kasus} \right) \right) \times Entropy(Bayi) + \left(\frac{sum (balita)}{jumlah kasus} \right) \times Entropi Balita$$

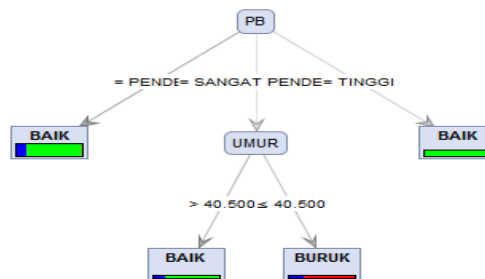
Jadi

$$Gain = 0.630734 - \left(\left(\frac{83}{500} \right) \right) \times 0.6549921 + \left(\left(\frac{417}{500} \right) \right) \times 0.6226944 = 1.04113324$$

Tabel 4. Hasil Perhitungan Entropy dan Gain

Atribut	JUMLAH KASUS	KURANG	BAIK	BURUK	ENTROPY	GAIN	
TOTAL	500	64	431	5	0.630734	1.0413324	
UMU R	Bayi	83	9	72	2	0.6549921	0.4375348
	Balita	417	55	359	3	0.6226944	-0.4159598
BB	Kurus Sekali	4	1	0	3	0	0
	Kurus	34	33	0	1	0	0
	Normal	462	30	431	1	0.3688005	0.1263439
	Sangat Pendek	9	2	4	3	1.5304931	1.4753953
PB	Pendek	361	62	297	2	0.9412933	-0.4179342
	Tinggi	130	0	130	0	0	0

Sesuai tabel di atas terlihat jumlah *entropy* dan *gain* pada atribut Panjang Badan (PB) adalah tertinggi dari beberapa atribut yang ada.



Gambar 2. Tampilan Hasil Rapid Miner Metode C4.5

Tabel 5 Hasil Confusion Matrix

	true KURANG	true BAIK	true BURUK	class precision
pred. KURANG	0	2	1	0.00%
pred. BAIK	62	427	2	86.97%
pred. BURUK	2	2	2	33.33%
class recall	0.00%	99.07%	40.00%	

Tabel hasil pengujian metode *C4.5* dengan Rapidminer jumlah *True Kurang* (TK) adalah 0 *record* diklasifikasikan sebagai baik dan *True Baik* 2 *record* diklasifikasikan sebagai sangat baik, dan 1 *record* untuk *True Buruk* (TB) diklasifikasikan sebagai gizi bermasalah. Berdasarkan tabel

tersebut menunjukkan bahwa tingkat akurasi dengan menggunakan metode *C4.5* sebesar 85,80% dan dapat dihitung untuk mencari nilai *accuracy sensitivity*.

Tabel 6. Perbandingan Metode

Metode	Akurasi
<i>Naïve Bayes</i>	94.20%
C 4.5	85,80%

Berdasarkan tabel 6 dapat dilihat bahwa hasil analisa dalam eksperimen yang telah dilakukan dengan penentuan model dari kedua metode yaitu *Naïve Bayes* dan *C4.5* yang digunakan untuk melakukan klasifikasi status gizi dengan akurasi *Naïve Bayes* sebesar 94.20% dan *C4.5* sebesar 85,80%. Dengan demikian hasil akurasi metode *Naïve Bayes* lebih baik dibandingkan metode *C4.5*. oleh karena itu metode *Naïve Bayes* dianggap lebih baik dalam mengklasifikasikan status gizi bayi dan balita.

4. PENUTUP

4.1. Simpulan

Dari hasil pengolahan data untuk mengklasifikasi status gizi pada bayi dan balita dengan menggunakan metode *Naïve Bayes* dan metode *C4.5* dengan menggunakan 90% data *training* dan 10% data *testing* yang akan di olah dengan menggunakan Rapid Miner untuk mencari nilai akurasi. Dari hasil percobaan yang dilakukan diperoleh nilai akurasi metode *Naïve Bayes* sebesar 94.20% dan metode *C4.5* sebesar 85,80%. Dengan demikian dapat disimpulkan bahwa metode *Naïve Bayes* lebih baik nilai akurasinya dalam mengklasifikasi status gizi bayi dan balita dibandingkan metode *C4.5*. Dan penelitian ini mendapatkan hasil akurasi yang lebih baik dari penelitian sebelumnya. Hal ini di buktikan dengan hasil penelitian yang dilakukan oleh (Nuha, 2020) dengan judul klasifikasi status gizi balita menggunakan metode *Naïve Bayes Classifier* berbasis *Website*. Dimana Hasil akurasi yang didapatkan pada penelitian tersebut sebesar 90% dan galat sebesar 10% sedangkan dalam penelitian ini hasil akurasi yang didapatkan dengan metode *Naïve Bayes* sebesar 94,20%.

Sedangkan untuk algoritma *C4.5* seperti penelitian yang dilakukan oleh (Farhurokhiem Azizi, 2021) tentang Klasifikasi Status Gizi Menggunakan Perhitungan *Body Mass Index* Dengan Metode Algoritma *C4.5* Pada Masyarakat Di Kota Yogyakarta. Dimana hasil penelitiannya mendapatkan akurasi sebesar 85%, sedangkan dalam penelitian yang dilakukan oleh peneliti mendapatkan hasil akurasi sebesar 85,80%.

4.2. Saran

Dalam mengklasifikasi status gizi pada bayi dan balita hendaknya kita menggunakan metode *Naïve Bayes* karena nilai akurasi pada probabilitas metode *Naïve Bayes* sangat tinggi dibandingkan dengan metode *C4.5*. Dan metode *Naïve bayes* bisa coba dengan data yang lain untuk memastikan bahwa metode ini lebih baik.

REFERENSI

- [1] Farhurokhiem Azizi, Y. (2021). *KLASIFIKASI STATUS GIZI MENGGUNAKAN PERHITUNGAN BODY MASS INDEX DENGAN METODE ALGORITMA C4.5 PADA MASYARAKAT DI KOTA YOGYAKARTA*.
- [2] Hafizan, H., & Putri, A. N. (2020). Penerapan Metode Klasifikasi Decision Tree Pada Status Gizi Balita Di Kabupaten Simalungun. *KESATRIA: Jurnal Penerapan Sistem Informasi (Komputer & Manajemen)*, 1(2), 68–72. <https://doi.org/10.30645/kesatria.v1i2.23>.
- [3] Herawati, D. (2020). Penerapan Metode K-Nearest Neighbor Untuk Menentukan Status Gizi Balita. *Universitas Teknologi Yogyakarta*.

-
- [4] Murdianingsih, Y., Syukur, A., & Soeleman, M. A. (2016). Penentuan Tingkat Kesejahteraan Anak Menggunakan Algoritma C4.5. *Jurnal Teknologi Informasi*, 12(April), 61–73.
- [5] Nuha, M. (2020). *Klasifikasi Status Gizi Balita Menggunakan Metode Naive Bayes Classifier Berbasis Website*.
- [6] Purwati, N., Agustina, C., & S, G. B. (2017). Komparasi Algoritma C . 45 Dan Backpropagation Untuk Klasifikasi Status Gizi Balita Berdasarkan Indeks Antropometri Bb / U Dan BB / PB. *Journal Speed*, 9(3), 26–33.
- [7] Ratniasih, N. L. (2019). Optimasi Data Mining Menggunakan Algoritma Naive Bayes Dan C4.5 Untuk Klasifikasi Kelulusan Mahasiswa. *Jurnal Teknologi Informasi Dan Komputer*, 5(1), 28–34. <https://doi.org/10.36002/jutik.v5i1.634>.
- [8] Rintiasti, A., & Fauzi, R. (2018). Data mining untuk menentukan parameter kunci syarat tumbuh optimal beberapa tanaman industri dengan menggunakan metode c4.5. *Jurnal Industri Hasil Perkebunan*, 13, 95–105.
- [9] Setio, P. B. N., Saputro, D. R. S., & Bowo Winarno. (2020). Klasifikasi Dengan Pohon Keputusan Berbasis Algoritme C4.5. *PRISMA, Prosiding Seminar Nasional Matematika*, 3, 64–71.
- [10] Subarkah, P., Pambudi, E. P., & Hidayah, S. O. N. (2020). Perbandingan Metode Klasifikasi Data Mining untuk Nasabah Bank Telemarketing. *MATRIK : Jurnal Manajemen, Teknik Informatika Dan Rekayasa Komputer*, 20(1), 139–148. <https://doi.org/10.30812/matrik.v20i1.826>.