

KOMPARASI ALGORITMA C4.5 DENGAN ALGORITMA NAÏVE BAYES CLASSIFIER UNTUK KLASIFIKASI STATUS GIZI BALITA

COMPARISON OF C4.5 ALGORITHM WITH THE NAVE BAYES CLASSIFIER ALGORITHM FOR CLASSIFICATION OF NUTRITIONAL STATUS OF TODDLER

Sofia Ulfah^{1*}

¹Politeknik Balekambang Jepara
Email : ¹*shofia.mahefi@gmail.com

*Penulis Korespondensi

Abstrak - Selama ini, Indeks Massa Tubuh (IMT) digunakan sebagai alat ukur untuk menilai status gizi seseorang. Apabila ada dua yang memiliki berat badan dan tinggi badan yang sama bisa jadi memiliki status gizi yang berbeda. Alat ukur antropometri menjadi sangat berperan untuk penentuan status gizi tersebut. Adapun pedoman antropometri bagi penentuan keadaan gizi merupakan parameter yang dipilih dan dianjurkan yaitu meliputi penilaian terhadap usia, berat badan, panjang badan atau tinggi badan. Disisi lain perkembangan teknologi saat ini semakin meningkat dan memberi pengaruh yang besar hampir disetiap sektor kehidupan dan kenegaraan. Proses globalisasi yang terjadi disetiap negara di dunia saat ini juga mendukung perkembangan dan penggunaan teknologi. Kebutuhan informasi akan semakin meningkat bersamaan dengan perkembangan teknologi. Beberapa metode yang digunakan untuk mengolah data yang sifatnya besar untuk menemukan pola yang terdapat didalamnya diantaranya adalah *Algoritma C4.5* dan *Algoritma Naïve Bayes Classification* yang merupakan algoritma dari metode *decision tree* dan *teorema bayes*. Algoritma C4.5 dan NBC tersebut akan digunakan untuk klasifikasi status gizi balita yang akan diklasifikasikan menjadi lima kelas yaitu normal, gemuk, sangat gemuk, kurus, atau sangat kurus.

Kata kunci: Status gizi balita, Antropometri, IMT, *Decision Tree*, Algoritma C4.5, Algoritma Naïve Bayes Classification

Abstract - So far, Body Mass Index (BMI) is used as a measuring tool to assess a person's nutritional status. If there are two who have the same weight and height, they may have different nutritional status. Anthropometric measuring tools play a very important role in determining the nutritional status. The anthropometric guidelines for determining nutritional status are the parameters chosen and recommended, which include an assessment of age, weight, body length or height. On the other hand, technological developments are currently increasing and giving a great influence in almost every sector of life and the state. The process of globalization that occurs in every country in the world today also supports the development and use of technology. The need for information will increase along with the development of technology. Several methods are used to process large data to find the patterns contained in them, including the C4.5 Algorithm and the Naïve Bayes Classification Algorithm which are algorithms from the decision tree method and Bayes theorem. The C4.5 and NBC algorithms will be used for the classification of the nutritional status of children under five, which will be classified into five classes, namely normal, fat, very fat, thin, or very thin.

Keywords: Nutritional status of toddlers, Anthropometry, BMI, *Decision Tree*, C4.5 Algorithm, Naïve Bayes Classification Algorithm

1. PENDAHULUAN

Perkembangan teknologi saat ini semakin meningkat dan memberi pengaruh yang besar hampir disetiap sektor kehidupan dan kenegaraan. Proses globalisasi yang terjadi disetiap negara di dunia saat ini juga mendukung perkembangan dan penggunaan teknologi. Kebutuhan informasi akan semakin meningkat bersamaan dengan perkembangan teknologi. Semakin banyak informasi yang dibutuhkan maka data yang dibutuhkan juga semakin banyak dan jumlahnya akan semakin besar. Kebutuhan akan jumlah data yang besar dapat kita temukan dalam dunia kesehatan. Hal ini dikarenakan, dunia kesehatan memiliki banyak cabang ilmu yang masing – masing memiliki banyak data dan mungkin akan terus meningkat.

Jumlah data yang terus meningkat ini memerlukan beberapa metode untuk mengolah dan mengambil kesimpulan dan informasi dari data tersebut. Beberapa metode yang digunakan untuk mengolah data yang sifatnya besar untuk

menemukan pola yang terdapat didalamnya diantaranya adalah: teknik klustering, analisis diskriminan, *teorema bayes*, *decision tree*, *artificial neural networks*, *support vector machine*, regresi linear, *support vector regres*i. Setiap metode tersebut memiliki algoritma-algoritma yang digunakan untuk memproses data yang ada. Namun pada kesempatan kali ini penulis mengangkat mengenai penggunaan *Algoritma C4.5* dan *Algoritma Naïve Bayes Classification* yang merupakan algoritma dari metode *decision tree* dan *teorema bayes* untuk pengklasifikasian status gizi balita. Algoritma C4.5 merupakan algoritma yang memiliki proses *learning* dan klasifikasi yang sederhana dan cepat. Secara umum, model algoritma C4.5 mempunyai tingkat akurasi yang tinggi [1]. Hal ini terbukti berdasarkan penelitian dari Ade Yuliana dan Duwi Bayu Pratomo dengan judul “Algoritma decision tree (C4.5) untuk memprediksi kepuasan mahasiswa terhadap kinerja dosen Politeknik TEDC Bandung” dengan kesimpulan bahwa tingkat akurasi algoritma C4.5 sebesar 94,62% [2]. Kemudian penelitian yang dilakukan oleh Uminingsih, Ign.Suraya dan Isnanto Nugroho dengan judul “Implementasi algoritma decision tree C4.5 Untuk Menentukan Kelas Calon Siswa di Lembaga Kursus Bahasa Inggris Berbasis Web” dengan kesimpulan algoritma C4.5 mempunyai akurasi sebesar 91,11% dan laju error besar 8,89% dengan data prediksi tepat [3]. Sedangkan *Naïve Bayes Classifier (NBC)* merupakan algoritma klasifikasi yang sangat efektif (mendapatkan hasil yang tepat) dan efisien (proses penalaran dilakukan memanfaatkan *input* yang ada dengan cara yang relatif cepat). Algoritma NBC bertujuan untuk melakukan klasifikasi data pada kelas tertentu. Unjuk kerja pengklasifikasi diukur dengan nilai *predictive accuracy* [4]. Penelitian Muhammad Husni Rifqo dan Ardi Wijaya dengan judul “Implementasi algoritma naïve bayes dalam penentuan pemberian kredit” dengan kesimpulan bahwa algoritma naïve bayes mempunyai akurasi sebesar 96,69% [5], kemudian penelitian Mawadatul Maulidah, Windu Gata dan Rizki Aulanita dengan judul “Algoritma klasifikasi decision tree untuk rekomendasi buku berdasarkan kategori buku” dengan kesimpulan bahwa algoritma naïve bayes mempunyai akurasi sebesar 99,95% [6]. Dari jurnal diatas menunjukkan bahwa algoritma C4.5 dan Naïve Bayes merupakan algoritma klasifikasi yang memiliki akurasi yang tinggi untuk klasifikasi sehingga penulis memilih algoritma C4.5 dan algoritma NBC untuk di komparasikan untuk mengklasifikasi status gizi.

Setiap orang memiliki status gizi yang berbeda sesuai dengan usianya. Status gizi dapat ditentukan melalui pemeriksaan laboratorium maupun secara antropometri. Antropometri merupakan cara penentuan status gizi yang paling mudah dan murah. Pengukuran antropometri adalah pengukuran yang digunakan untuk menentukan keadaan gizi seseorang. Agar memperoleh hasil yang tepat, diberikan suatu patokan sebagai pedoman. Adapun pedoman antropometri bagi penentuan keadaan gizi merupakan parameter yang dipilih dan dianjurkan, yang meliputi penilaian terhadap usia dan berat badan, panjang badan, atau tinggi badan, dan lingkaran lengan atas [7]. Namun, banyaknya jumlah balita yang ada membuat data – data gizi balita menjadi sulit untuk diklasifikasikan mana yang status gizinya normal, gemuk, obesitas, kurus, atau sangat kurus. Penelitian dengan tema klasifikasi status gizi balita pernah dilakukan oleh Hamsir Saleh, Muh. Faisal, dan Rachmat Irawan dengan judul “Klasifikasi status gizi balita menggunakan metode K-nearest Neighbor” dengan kesimpulan bahwa metode K-NN adalah metode klasifikasi yang bias digunakan untuk klasifikasi status gizi balita dengan hasil nilai $V(G) = CC$ yaitu =4 yang berarti kebenaran sebuah logika [8]. Penelitian lain tentang status gizi balita juga pernah dilakukan oleh Wiwid Wahyudi dengan judul “Optimasi klasifikasi status gizi balita berdasarkan indeks antropometri menggunakan algoritma C4.5 adaboost classification” dengan kesimpulan algoritma c4.5 adaboost classification memperoleh akurasi 90,23% [9].

Berdasarkan masalah yang telah diuraikan, maka penulis mengangkat judul “*KOMPARASI ALGORITMA C4.5 DENGAN NAÏVE BAYES CLASSIFIER (NBC) UNTUK KLASIFIKASI STATUS GIZI BALITA*” agar memudahkan dalam pengklasifikasian gizi balita.

2. METODE PENELITIAN

2.1 Metode Pengumpulan Data

Penulis menggunakan metode pengumpulan data sebagai berikut :

- 1) Studi Literatur
Pada metode ini penulis akan melakukan pencarian dan pembelajaran dari berbagai macam literatur dan dokumen yang menunjang penelitian ini khususnya yang berkaitan dengan klasifikasi status gizi balita.
- 2) Observasi
Merupakan teknik pengumpulan data dengan mengadakan pengamatan atau penelitian secara langsung dari objek penelitian, serta melakukan wawancara dengan pihak – pihak terkait yaitu pihak dari instansi dimana penelitian dilakukan.
- 3) *Browsing*
Melakukan pengamatan ke berbagai macam *website* di *internet* yang menyediakan informasi yang sesuai dengan permasalahan dalam penelitian ini.

2.2 Metode Penelitian

Metode penelitian merupakan suatu cara yang dilakukan untuk mendapatkan informasi yang dibutuhkan agar data yang diperoleh lebih akurat. Metode yang digunakan dalam penelitian ini adalah metode *Cross industry Standard Process for Data mining (CRISP-DM)*, yang memiliki beberapa tahapan proses di bawah ini [10]:

2.2.1 Pemahaman Bisnis (*Business Understanding*)

Sesuai dengan panduan metode CRISP-DM, proses *data mining* dalam penelitian ini dimulai dengan pemahaman terhadap proses bisnis dan lingkungan dimana asal data dimanfaatkan. Proses pemahaman bisnis yang dilakukan di Puskesmas pada saat ini adalah :

a. Penentuan tujuan proyek

Tujuan dari proyek penelitian ini adalah untuk menggali informasi yang terkandung dalam tumpukan data status gizi di Puskesmas.

b. Menyiapkan strategi awal untuk mencapai tujuan.

Strategi awal yang dilakukan untuk mencapai tujuan adalah dengan menentukan tujuan penelitian yang jelas serta pencarian data-data yang valid dan penganalisisan data yang sudah didapatkan, yang kemudian diolah menggunakan rumus-rumus Algoritma C4.5 dan Algoritma *Naïve Bayes Classifier (NBC)*.

2.2.2 Pemahaman Data (*Data Understanding*)

Pada tahapan pemahaman data penulis melakukan beberapa langkah untuk benar-benar dapat memahami alur dari data tersebut sehingga dapat mencapai keberhasilan dalam pengolahan data dan keakuratan dalam menyajikan hasil pengolahannya:

a. Mengumpulkan Data

Data-data yang diambil penulis adalah data status gizi balita yang sudah ada pada Puskesmas. Atribut-atribut yang diambil oleh penulis untuk pengolahan data adalah sebagai berikut:

1. Jenis Kelamin
2. Umur
3. Berat Badan terhadap Umur (BB/U)
4. Tinggi Badan terhadap Umur (TB/U)
5. Berat Badan terhadap Tinggi Badan (BB/U)
6. Indeks Masa Tubuh terhadap Umur (IMT/U)

b. Mengevaluasi Kualitas Data

Karena tidak semua data-data yang didapatkan dalam status gizi tersebut dapat digunakan maka penulis melakukan pembuangan data-data yang tidak diperlukan, misalkan data status gizi yang tidak lengkap, sehingga data yang tidak berguna akan dihapus atau tidak digunakan.

2.2.3 Pengolahan Data (*Data Preparation*)

Pada tahap pengolahan data penulis perlu melakukan beberapa hal diantaranya:

a. Mempersiapkan dari data awal

Penulis melakukan persiapan data awal dengan cara mencari data-data status gizi, kemudian melakukan pemisahan data-data yang dipakai dan yang tidak dipakai.

b. Memilih variabel

Tidak semua variabel yang ada dalam data status gizi itu dapat digunakan, sehingga penulis melakukan pembuangan variabel-variabel yang memang sekiranya tidak digunakan dan penulis hanya mengambil variabel-variabel yang dapat diolah.

c. Melakukan perubahan variabel

Karena variabel-variabel yang digunakan di dalam status gizi kadang berbeda sehingga perlu melakukan perubahan variabel untuk penyamaan variabel, apabila variabel tersebut tidak digunakan maka penulis melakukan pembuangan variabel tersebut, misalkan variabel nama balita, nama orang tua, tanggal lahir, dan alamat.

d. Mempersiapkan Data Awal untuk Mempersiapkan Pemodelan

Dalam tahapan ini penulis mengumpulkan semua berkas-berkas yang diperlukan guna penelitian yaitu data-data status gizi yang terdapat di Puskesmas, dari data-data tersebut penulis melakukan analisis dan pengolahan data sehingga siap untuk diterapkan dalam tahap pemodelan.

e. Pengolahan data

Pengolahan data dilakukan berdasarkan kebutuhan bentuk *input* yang diterima oleh masing-masing algoritma dengan bantuan beberapa program Matlab untuk memproses data teks yang tersedia.

2.2.4 Pemodelan (*Modelling*)

Berdasarkan kebutuhan bisnis dan karakteristik data yang telah dijabarkan sebelumnya, pada bagian ini akan dijelaskan suatu rumusan mengenai teknik-teknik yang digunakan dalam menjelajahi data dan juga perangkat lunak yang digunakan untuk kebutuhan tersebut. Sesuai dengan fokus penelitian yang disebutkan di awal, *Decision Tree C4.5* dan *Naïve Bayes Classifier* dipilih menjadi metode *data mining* kali ini. Hasil dari classification untuk membantu menafsirkan mengenai struktur kelompok yang tersembunyi dalam data.

2.2.5 Evaluasi (*Evaluation*)

Pola informasi yang dihasilkan dari proses *data mining* perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. Dengan membuat tabel yang dapat merepresentasikan hasil analisis pengolahan

data. Tahap ini merupakan bagian dari proses KDD yang disebut *interpretation* (pemahaman). Tahap ini mencakup pemeriksaan apakah pola informasi yang ditemukan bertentangan dengan fakta atau anggapan yang ada sebelumnya.

2.2.6 Deployment

Pembuatan dari model bukanlah akhir dari proyek *data mining*. Meskipun tujuan dari pemodelan adalah untuk meningkatkan pengetahuan dari data, pengetahuan data tersebut perlu dibangun dengan terorganisasi dan dibuat pada satu bentuk yang dapat digunakan oleh pengguna. Dalam tahap ini akan dilakukan penyusunan laporan atau presentasi dari pengetahuan yang didapat pada proses evaluasi sebelumnya.

2.3 Algoritma C4.5

Algoritma C4.5 yaitu algoritma yang pakai untuk menghasilkan sebuah pohon keputusan yang ditemukan oleh Ross Quinlan [11]. Algoritma ini digunakan untuk membangun sebuah pohon keputusan yang mudah dimengerti, fleksibel, dan menarik karena dapat divisualisasikan dalam bentuk gambar [12]. konsep *information gain* atau *entropy reduction* pada algoritma C4.5 digunakan untuk memilih pembagian akar yang optimal. Langkah-langkah dalam pembuatan sebuah *decision tree* menggunakan algoritma C4.5 [13] yaitu:

1. Menyiapkan data ujicoba atau training.
2. Menghitung Total *Entropy* sebelum dicari masing-masing *Entropy class*, rumus entropy sebagai berikut:

$$Entropy(S) = - \sum_{i=1}^n p_i \log_2 p_i$$

Keterangan:

S = Total Kasus

n = Jumlah Partisi S

Pi = Proporsi Si Terhadap S

3. Hitung nilai gain setiap atribut, dimana nilai gain tertinggi akan menjadi akar pertama. Dengan rumus sebagai berikut

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{S} * Entropy(S_i)$$

4. Ulangi langkah ke-2 dan ke-3 hingga semua *record* terpartisi.
5. Proses partisi pohon keputusan akan berhenti disaat:
 - a. Semua *record* dalam node N mendapatkan kelas yang sama
 - b. Tidak ada atribut didalam record yang dipartisi lagi
 - c. Tidak ada record didalam cabang yang kosong

2.4 Algoritma Naïve Bayes Classifier

Algoritma Naïve Bayes merupakan algoritma klasifikasi yang bisa dipakai untuk memprediksi probabilitas keanggotaan suatu class. NBC terbukti mempunyai akurasi dan kecepatan yang tinggi pada waktu diimplementasikan ke database yang besar [14]. Teorema Bayes memiliki bentuk umum sebagai berikut:

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

X = Data dengan *class* yang belum diketahui

H = Hipotesis data X merupakan suatu *class* spesifik

P(H|X) = Probabilitas hipotesis H berdasarkan kondisi x (posteriori prob.)

P(H) = Probabilitas hipotesis H (prior prob.)

P(X|H) = Probabilitas X berdasarkan kondisi tersebut

P(X) = Probabilitas dari X

3. HASIL DAN PEMBAHASAN

3.1 ANALISIS DAN PENGOLAHAN DATA

Analisis data Status Gizi Balita pada puskesmas bertujuan untuk menggali informasi yang terkandung dalam tumpukan data status gizi, penelitian ini bertujuan untuk dijadikan pengetahuan bagi Posyandu, Puskesmas, Rumah Sakit, Dinas Kesehatan serta Masyarakat.

3.1.1 Sumber Data

Data Status Gizi yang penulis dapatkan merupakan kumpulan data dari pengukuran status gizi pada bulan Juli sampai September tahun 2020.

Pengumpulan data dilakukan pada saat penerimaan pasien di masing-masing Puskesmas, selanjutnya dilakukanlah pengukuran status gizi oleh petugas Puskesmas yang kemudian data pengukuran status gizi tersebut didokumentasikan oleh petugas Puskesmas.

3.2 Pengumpulan Data

Dalam tahap pengumpulan data ini penulis mengambil data dari Pemantauan Status Gizi (PSG) Balita berupa file Excel. Data yang penulis dapatkan sebanyak 530 data.

Tabel 1. Data Status Gizi Balita

No	Jenis Kelamin	Umur (bulan)	Berat Badan (kg)	Tinggi Badan (cm)	IMT
1	L	17	8.8	73	16.51
2	L	20	10.1	77	17.03
3	P	19	10.5	76	18.18
4	P	26	13	88	16.79
5	L	38	13.5	92	15.95
X	P	30	10.7	88	13.82
X	L	18	10	73	18.77
X	L	3	5	57.5	15.12
529	L	44	12.8	100	12.80
530	P	25	11.5	86.5	15.37

3.3 Penerapan Algoritma C4.5

3.3.1 Tahapan Algoritma C4.4

1. Menghitung jumlah kasus

Jumlah kasus untuk keputusan IMT/U serta *Entropy* dari semua kasus yang dibagi berdasarkan atribut Jenis Kelamin, Umur, BB/U, TB/U, dan BB/TB.

Entropy(Total)

$$\begin{aligned}
 &= \left(-\frac{\Sigma \text{ sangat kurus}}{\Sigma \text{ kasus}} \times \log_2 \left(\frac{\Sigma \text{ sangat kurus}}{\Sigma \text{ kasus}} \right) \right) + \left(-\frac{\Sigma \text{ kurus}}{\Sigma \text{ kasus}} \times \log_2 \left(\frac{\Sigma \text{ kurus}}{\Sigma \text{ kasus}} \right) \right) \\
 &\quad + \left(-\frac{\Sigma \text{ normal}}{\Sigma \text{ kasus}} \times \log_2 \left(\frac{\Sigma \text{ normal}}{\Sigma \text{ kasus}} \right) \right) + \left(-\frac{\Sigma \text{ gemuk}}{\Sigma \text{ kasus}} \times \log_2 \left(\frac{\Sigma \text{ gemuk}}{\Sigma \text{ kasus}} \right) \right) \\
 &\quad + \left(-\frac{\Sigma \text{ sangat gemuk}}{\Sigma \text{ kasus}} \times \log_2 \left(\frac{\Sigma \text{ sangat gemuk}}{\Sigma \text{ kasus}} \right) \right) \\
 &= \left(-\frac{10}{430} \times \log_2 \left(\frac{10}{430} \right) \right) + \left(-\frac{23}{430} \times \log_2 \left(\frac{23}{430} \right) \right) + \left(-\frac{372}{430} \times \log_2 \left(\frac{372}{430} \right) \right) + \\
 &\quad \left(-\frac{16}{430} \times \log_2 \left(\frac{16}{430} \right) \right) + \left(-\frac{9}{430} \times \log_2 \left(\frac{9}{430} \right) \right) \\
 &= 0.126192204 + 0.225968629 + 0.18083875 + 0.176676943 + 0.116754443
 \end{aligned}$$

$$\mathbf{Entropy(Total) = 0.826430969}$$

2. Setelah itu melakukan penghitungan ***Gain*** untuk tiap-tiap atribut.

Contoh perhitungan *Gain* pada atribut Jenis Kelamin.

$$\mathbf{Gain(Jenis Kelamin)}$$

$$= \mathbf{Entropy(Total)}$$

$$\begin{aligned}
 &- \left(\left(\frac{\Sigma \text{ kasus JK_L}}{\Sigma \text{ kasus total}} \times \mathbf{Entropy(Jenis Kelamin_L)} \right) \right. \\
 &\quad \left. + \left(\frac{\Sigma \text{ kasus JK_P}}{\Sigma \text{ kasus total}} \times \mathbf{Entropy(Jenis Kelamin_P)} \right) \right) \\
 &= 0.826430969 - \\
 &\quad \left(\left(\frac{217}{430} \times 0.773479661 \right) + \left(\frac{213}{430} \times 0.86714726 \right) \right) \\
 &= 0.826430969 - 0.819877797
 \end{aligned}$$

$$\mathbf{Gain(Jenis Kelamin) = 0.006553172}$$

Tabel 2. Data Status Gizi Balita

No	Jenis Kelamin	Umur (bulan)	Berat Badan (kg)	Tinggi Badan (cm)	IMT
1	L	17	8.8	73	16.51
2	L	20	10.1	77	17.03
3	P	19	10.5	76	18.18
4	P	26	13	88	16.79
5	L	38	13.5	92	15.95
6	P	30	10.7	88	13.82
7	L	18	10	73	18.77
8	L	3	5	57.5	15.12
9	L	44	12.8	100	12.80
10	P	25	11.5	86.5	15.37

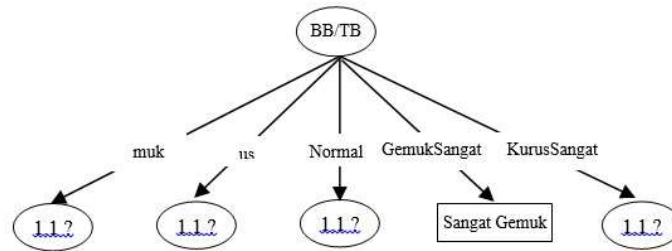
Setelah menghitung jumlah kasus pada tiap *atribut* maka selanjutnya menghitung *entropy* dan *gain*. Perhitungannya dapat dilihat pada tabel di bawah :

Tabel 3. Perhitungan Entropy dan Gain Node 1

NODE			Entropy	Gain
1	Total		0.826430969	
	JK			0.006553172
		L	0.773479661	
		P	0.86714726	
	Umur			0.021752455
		0-24	0.83917557	
		25-60	0.773828723	
	BB/U			0.092886016
		Sangat kurang	1.57662122	
		Kurang	0.987639382	
		Normal	0.673466675	
		Lebih	0	
	TB/U			0.009728235
		Sangat pendek	1.089884755	
		Pendek	0.829454733	
		Normal	0.781677111	
		Tinggi	0	
	BB/TB			0.29862317
		Sangat kurus	0.918295834	
		Kurus	1.31056738	
		Normal	0.452474524	
		Gemuk	1.370950594	
		Sangat gemuk	0	

Dari hasil pada Tabel 3 dapat diketahui bahwa atribut dengan Gain tertinggi adalah BB/TB yaitu sebesar 0.29862317. Dengan demikian BB/TB dapat menjadi node akar (*root*). Ada 5 nilai atribut dari BB/TB yaitu Sangat Kurus, Kurus, Normal, Gemuk, dan Sangat Gemuk. Dari kelima nilai atribut tersebut, hanya nilai atribut Sangat Gemuk

yang sudah terklasifikasi menjadi 1 kasus yaitu Sangat Gemuk dan sudah tidak perlu dilakukan perhitungan lebih lanjut lagi. Sedangkan nilai atribut Sangat Kurus, Kurus, Normal, dan masih perlu dilakukan perhitungan lagi. Dari hasil tersebut dapat digambarkan pohon keputusan sementara seperti berikut



Gambar 1. Pohon keputusan hasil perhitungan node 1

3.4 Penghitungan Algoritma *Naïve Bayes Classifier* (NBC)

Jika seorang balita X dengan data sebagai berikut :

Tabel 4. Contoh Data Testing Algoritma NBC

Jenis Kelamin	Umur	BB/U	TB/U	BB/TB	IMT/U
Laki-Laki	0-24	Normal	Normal	Normal	???

Jawab :

1. Menghitung Jumlah *Class*

$$P(\text{IMT/U} = \text{"Sangat Kurus"}) = \frac{\Sigma \text{kasus sangat kurus}}{\Sigma \text{kasus total}}$$

$$= 10/74 = 0.135135135$$

$$P(\text{IMT/U} = \text{"Kurus"}) = \frac{\Sigma \text{kasus kurus}}{\Sigma \text{kasus total}}$$

$$= 17/74 = 0.22972973$$

$$P(\text{IMT/U} = \text{"Normal"}) = \frac{\Sigma \text{kasus normal}}{\Sigma \text{kasus total}}$$

$$= 31/74 = 0.418918919$$

$$P(\text{IMT/U} = \text{"Gemuk"}) = \frac{\Sigma \text{kasus gemuk}}{\Sigma \text{kasus total}}$$

$$= 9/74 = 0.121621622$$

$$P(\text{IMT/U} = \text{"Sangat Gemuk"}) = \frac{\Sigma \text{kasus sangat gemuk}}{\Sigma \text{kasus total}}$$

$$= 7/74 = 0.094594595$$

2. Menghitung jumlah kasus yang sama dengan *class* yang sama

$$P(\text{Jenis Kelamin} = \text{"L"} \mid \text{IMT/U} = \text{"Sangat Kurus"})$$

$$= \frac{\Sigma \text{kasus JK "L" dengan IMT/U "Sangat Kurus"}}{\Sigma \text{kasus IMT/U "Sangat Kurus"}}$$

$$= 4/10 = 0.4$$

3. Kalikan semua hasil *variable* Sangat Kurus, Kurus, Normal, Gemuk, dan Sangat Gemuk

$$P(X \mid \text{IMT/U} = \text{"Sangat Kurus"})$$

$$= P(\text{Jenis Kelamin} = \text{"L"} \mid \text{IMT/U} = \text{"Sangat Kurus"}) \cdot P(\text{Umur} = \text{"0-24"} \mid \text{IMT/U} = \text{"Sangat Kurus"}) \cdot P(\text{BB/U} = \text{"Normal"} \mid \text{IMT/U} = \text{"Sangat Kurus"}) \cdot P(\text{TB/U} = \text{"Normal"} \mid \text{IMT/U} = \text{"Sangat Kurus"}) \cdot P(\text{BB/TB} = \text{"Normal"} \mid \text{IMT/U} = \text{"Sangat Kurus"})$$

$$= 0.4 \times 0.5 \times 0.5 \times 0.6 \times 0.4 = 0.024$$

4. Kalikan semua hasil *variable* Sangat Kurus, Kurus, Normal, Gemuk, dan Sangat Gemuk dengan jumlah *class* Sangat Kurus, Kurus, Normal, Gemuk, dan Sangat Gemuk

$$P(X \mid \text{IMT/U} = \text{"Sangat Kurus"}) \cdot P(\text{IMT/U} = \text{"Sangat Kurus"})$$

$$= 0.024 \times 0.135135135 = 0.003243243$$

$$P(X \mid \text{IMT/U} = \text{"Kurus"}) \cdot P(\text{IMT/U} = \text{"Kurus"})$$

$$= 0.01217024 \times 0.22972973 = 0.002795866$$

$$P(X \mid \text{IMT/U} = \text{"Normal"}) \cdot P(\text{IMT/U} = \text{"Normal"})$$

$$= 0.021125321 \times 0.418918919 = 0.008849797$$

$$\begin{aligned}
 &P(X|IMT/U = \text{"Gemuk"}) \cdot P(IMT/U = \text{"Gemuk"}) \\
 &= 0.121932632 \times 0.121621622 = 0.014829644 \\
 &P(X|IMT/U = \text{"Sangat Gemuk"}) \cdot P(IMT/U = \text{"Sangat Gemuk"}) \\
 &= 0.085678586 \times 0.094594595 = 0.008104731
 \end{aligned}$$

5. Bandingkan hasil *class* Sangat Kurus, Kurus, Normal, Gemuk, dan Sangat Gemuk. Karena hasil $(P|\text{Gemuk})$ lebih besar dari $(P|\text{Sangat Kurus})$, $(P|\text{Kurus})$, $(P|\text{Normal})$, dan $(P|\text{Sangat Gemuk})$, maka keputusannya balita tersebut masuk ke *class* $IMT/U = \text{"Gemuk"}$

Tabel 5. Hasil Data Testing Algoritma NBC

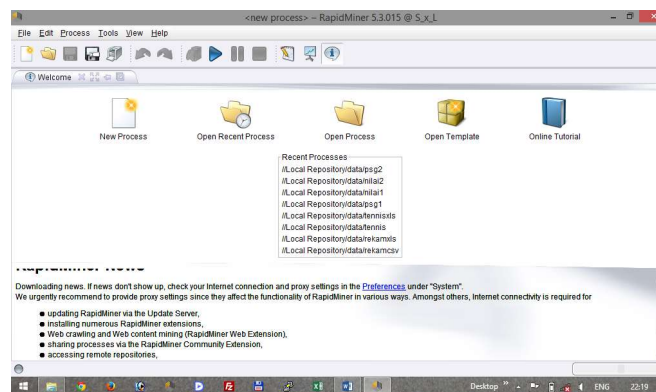
Jenis Kelamin	Umur	BB/U	TB/U	BB/TB	IMT/U
Laki-Laki	0-24	Normal	Normal	Normal	Gemuk

3.2 Pengolahan data dengan tools rapidminer

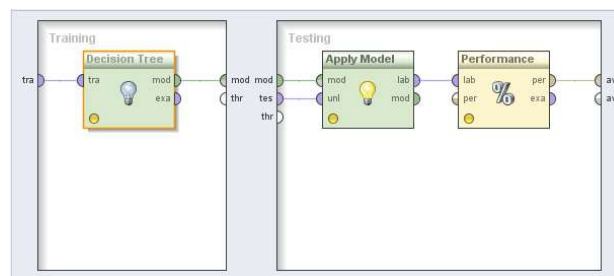
Setelah data-data melalui proses transformasi data dan melakukan perhitungan secara manual, maka akan dilanjutkan ke proses pengolahan data menggunakan Aplikasi RapidMiner, dikarenakan aplikasi rapidminer bisa membantu dan mempermudah menghasilkan probabilitas untuk dijadikan prediksi [15]

3.2.1 Algoritma C4.5

Dalam proses penghitungan algoritma C4.5 di aplikasi rapidminer, tahapan rapidminer melalui banyak tahapan yaitu tahapan memasukkan dataset ke main process, training dan testing decision tree c4.5 kemudian akan tampil akurasi, berikut tampilannya



Gambar 2. Tampilan Menu Utama Aplikasi RapidMiner



Gambar 3. Tampilan Training dan Testing Algoritma C4.5

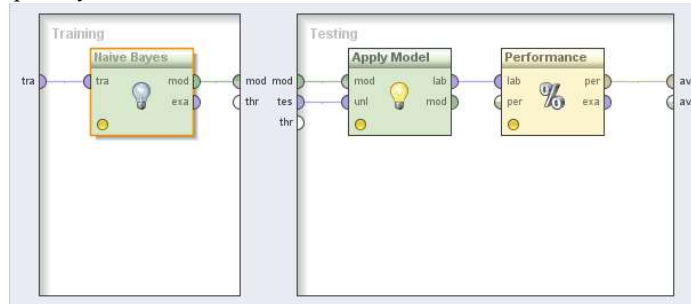
accuracy: 89.53% +/- 2.99% (mikro: 89.53%)						
	true Normal	true Kurus	true Gemuk	true Sangat gemuk	true Sangat kurus	class precision
pred. Normal	366	10	12	4	5	92.19%
pred. Kurus	3	10	0	0	3	62.50%
pred. Gemuk	0	0	4	2	0	66.67%
pred. Sangat gemuk	0	0	0	3	0	100.00%
pred. Sangat kurus	3	3	0	0	2	25.00%
class recall	98.39%	43.48%	25.00%	33.33%	20.00%	

Gambar 4. Tampilan Hasil Akurasi Algoritma C4.5 Untuk Klasifikasi Status Gizi Balita

Pada gambar 4 dapat dilihat bahwa klasifikasi status gizi balita menggunakan Algoritma C4.5 memiliki akurasi sebesar 89.53%

3.2.2 Algoritma Naïve Bayes Classifier

Dalam proses penghitungan algoritma Naïve Bayes di aplikasi rapidminer melalui banyak tahapan yaitu tahapan memasukkan dataset ke main process, training dan testing algoritma naïve bayes classifier kemudian akan tampil akurasi, berikut tampilannya



Gambar 5. Tampilan training dan testing algoritma naïve bayes classifier

accuracy: 88.60% +/- 4.70% (mikro: 88.60%)						
	true Normal	true Kurus	true Gemuk	true Sangat gemuk	true Sangat kurus	class precision
pred. Normal	361	8	11	4	5	92.80%
pred. Kurus	5	13	0	1	4	56.52%
pred. Gemuk	2	0	4	2	0	50.00%
pred. Sangat gemuk	0	0	1	2	0	66.67%
pred. Sangat kurus	4	2	0	0	1	14.29%
class recall	97.04%	56.52%	25.00%	22.22%	10.00%	

Gambar 6. Tampilan hasil akurasi algoritma naïve bayes classifier

Pada gambar 1.6 dapat dilihat bahwa klasifikasi status gizi balita menggunakan Algoritma Naïve Bayes Classifier (NBC) memiliki akurasi sebesar 88.60%.

4. Kesimpulan

Terdapat 5 atribut yang mempengaruhi status gizi balita yaitu (1) jenis kelamin, (2) umur, (3) BB/U, (4) TB/U, (5) BB/TB. Penelitian ini menggunakan Algoritma C4.5 dan Algoritma Naïve Bayes Classification yang merupakan algoritma dari metode decision tree dan teorema bayes untuk pengklasifikasian status gizi balita. Dari 530 data kotor setelah dilakukan pembersihan data tidak lengkap dan melalui proses seleksi sesuai yang dibutuhkan dalam analisa status gizi balita maka didapat data 430 data bersih.

Dari hasil pengujian keseluruhan data dengan memakai kedua algoritma di atas didapatkan akurasi untuk Algoritma C4.5 sebesar 89.53% dan Algoritma Naïve Bayes Classifier (NBC) sebesar 88.60%, kedua algoritma tersebut belum cukup akurat karena nilai kesalahannya masing – masing masih mencapai 10.47% dan 11.40%. Namun bila dibandingkan, Algoritma C4.5 lebih akurat dari pada Algoritma Naïve Bayes Classifier (NBC) dalam klasifikasi status gizi balita.

DAFTAR PUSTAKA

- [1] J. dan M. K. Han, *Data Mining Cocepts and Techniques Second Edition*, Second Edi. San Francisco, 2006.
- [2] A. Yuliana and D. B. Pratomo, "Memprediksi Kepuasan Mahasiswa Terhadap Kinerja Dosen Politeknik TEDC Bandung," *Semnasinotek 2017*, pp. 377–384, 2017.
- [3] I. Nugroho, P. Studi, and S. Komputer, "Kelas Calon Siswa Di Lembaga Kursus Bahasa Inggris Berbasis," 2019.
- [4] J. Zhang, H dan SU, *Naive bayesian classifiers for ranking*. 2007.
- [5] M. H. Rifqo and A. Wijaya, "Implementasi Algoritma Naive Bayes Dalam Penentuan Pemberian Kredit," *Pseudocode*, vol. 4, no. 2, pp. 120–128, 2017, doi: 10.33369/pseudocode.4.2.120-128.
- [6] M. Maulidah, W. Gata, R. Aulianita, and C. I. Agustyaningrum, "Algoritma Klasifikasi Decision Tree Untuk Rekomendasi Buku Berdasarkan Kategori Buku," *J. Ilm. Ekon. Dan Bisnis*, vol. 13, no. 2, pp. 89–96, 2020.
- [7] Proverawati, *Gizi untuk kebidanan*. Yogyakarta: Nuha Medika, 2009.
- [8] H. Saleh, M. Faisal, and R. I. Musa, "Klasifikasi Status Gizi Balita Menggunakan Metode K-Nearest Neighbor," *Simtek J. Sist. Inf. dan Tek. Komput.*, vol. 4, no. 2, pp. 120–126, 2019, doi:

- 10.51876/simtek.v4i2.60.
- [9] W. Wahyudi, "Optimasi Klasifikasi Status Gizi Balita Berdasarkan Indeks Antropometri Menggunakan Algoritma C4.5 Adaboost Classification," *Komputerisasi Akunt.*, vol. 12, no. 2, p. 45, 2019, [Online]. Available: <http://jurnal.stekom.ac.id/index.php/kompak>.
- [10] D. Larose, *Discovering Knowledge in Data: An Introduction to Data Mining*. USA: Jhon Wiley & Son, Inc, 2006.
- [11] I. Wahyudi, S. Bahri, and P. Handayani, "Klasifikasi Channel Youtube Indonesia Menggunakan Algoritma C4.5," vol. V, no. 1, pp. 135–138, 2019, doi: 10.31294/jtk.v4i2.
- [12] F. Gorunescu, *Data Mining: Concept and Technique*, Second. Morgan Kaufman, 2011.
- [13] D. . Larose, *Discovering Knowledge In Data An Introduction to Data Mining*. 2005.
- [14] K. and E. Luthfi, *Algoritma Data Mining*. Yogyakarta: Andi, 2019.
- [15] R. Nofitri and N. Irawati, "Analisis Data Hasil Keuntungan Menggunakan Software Rapidminer," *JURTEKSI (Jurnal Teknol. dan Sist. Informasi)*, vol. 5, no. 2, pp. 199–204, 2019, doi: 10.33330/jurteks.v5i2.365.