

Penerapan Data Mining Klasifikasi Gizi Bayi Dengan Algoritma Decision Tree C4.5

Marini^{1*}, Heru Satria Tambunan², Zulia Almaida Siregar³, Agus Perdana Windarto⁴, Fitri Rizki⁵

^{1,2,3,4,5}STIKOM Tunas Bangsa, Pematangsiantar, Indonesia

Email: ¹marini@gmail.com, ²herusatriatambunan@amiktunasbangsa.ac.id, ³zuliaalmaidasiregar@gmail.com,

⁴agus.perdana@amiktunasbangsa.ac.id, ⁵fitririzki@amiktunasbangsa.ac.id

(* : marini@gmail.com)

Abstract

Malnutrition is a condition experienced by a person due to lack of nutritional intake or the amount of nutrients consumed is below the standard. Nutrition Nutritional problems in children under 5 years old (toddlers) can have serious short-term and long-term impacts. Toddlers who are malnourished and undernourished can have an impact on morbidity even in developing countries, malnutrition is one of the factors causing child mortality. Indonesia's human development is currently still constrained by nutritional problems, especially child nutrition. This can be seen from the basic health research data (Risksedas) in 2013 where the national prevalence of child malnutrition reached 5.7%, while malnutrition was 13.9%. For this reason, this study was carried out by classifying the nutritional status of infants and toddlers at Posyandu Gunung Maligas using the C4.5 algorithm as a reference to related parties to solve problems and find out the most influential factors on infant and toddler malnutrition.

Keywords: *Malnutrition, Infants, Posyandu, Data Mining, Decision Tree*

Abstrak

Gizi buruk merupakan kondisi yang dialami seseorang dikarenakan kekurangan asupan nutrisi atau jumlah nutrisi yang dikonsumsi berada dibawah standar. Nutrisi Masalah gizi pada balita usia dibawah 5 tahun (balita) dapat berdampak serius secara jangka pendek maupun jangka panjang. Balita yang mengalami gizi buruk dan kurang dapat berdampak terhadap penderita. Bahkan di negara-negara berkembang kekurangan gizi merupakan salah satu faktor penyebab kematian anak. Hal tersebut tampak dari data riset kesehatan dasar (Risksedas) tahun 2013 dimana prevalensi gizi buruk anak secara nasional mencapai angka 5,7%, sedangkan gizi kurang sebanyak 13,9%. Untuk itu dilakukan penelitian ini dengan mengklasifikasi status gizi pada bayi dan balita di Posyandu Gunung Maligas menggunakan algoritma C4.5 sebagai referensi kepada pihak terkait untuk menyelesaikan permasalahan dan mengetahui faktor yang paling berpengaruh pada gizi buruk bayi dan balita.

Kata Kunci: Gizi Buruk, Bayi, Posyandu, Data Mining, Decision Tree

1. PENDAHULUAN

Menurut UNICEF (United Nation of Children and Education Federation) Masalah gizi dalam pembangunan kependudukan masih merupakan persoalan yang dianggap menjadi masalah utama dalam tatanan kependudukan dunia. Oleh karena itu, persoalan ini menjadi salah satu butir penting yang menjadi kesepakatan global dalam MDGs (Milleneum Development Goals). Setiap negara secara bertahap harus mampu mengurangi jumlah balita bergizi buruk atau gizi kurang mencapai 15,5 persen setiap tahunnya. Sebagai salah satu negara dengan kompleksitas kependudukan yang sangat beraneka ragam, Indonesia dihadapi oleh dinamika persoalan gizi buruk. Walaupun proses pembangunan di Indonesia telah mampu mengatasi persoalan ini, tetapi dilihat dari kecenderungan data statistik, masih banyak persoalan yang perlu diselesaikan terutama yang menyangkut persoalan balita gizi kurang [1]. Menurut buku saku pemantauan gizi tahun 2017 oleh Kementerian kesehatan, Sebanyak 3,8% balita mempunyai status gizi buruk dan 14,0% balita mempunyai status gizi kurang. Persentase underweight/berat badan kurang/gizi kurang (gizi buruk + gizi kurang) pada kelompok balita (17,8%) lebih tinggi dibandingkan kelompok baduta (14,8%). Meningkatnya risiko kesakitan dan kematian anak akibat gizi buruk berkaitan dengan penyakit infeksi yang sering menyertai gizi buruk seperti infeksi pernapasan akut, diare, campak dan beberapa penyakit infeksi lainnya. World Health Organization (WHO) menyatakan bahwa 54% kematian balita pada tahun 2002 disebabkan oleh gizi buruk.

Gizi buruk pada anak secara langsung dipengaruhi oleh asupan nutrisi sedangkan secara tidak langsung disebakan oleh penyakit infeksi dimana kedua hal tersebut berhubungan dengan faktor maternal, sosial ekonomi, demografi dan perilaku. 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 balita. Agar memperoleh hasil yang tepat, diberikan suatu patokan sebagai pedoman. Z-Skor merupakan index antropometri yang digunakan secara internasional untuk penentuan status gizi dan pertumbuhan, yang diekspresikan sebagai satuan standar deviasi (SD) populasi. Z-Skor digunakan untuk menghitung status gizi secara antropometri pada berat badan terhadap umur (BB/U), tinggi badan terhadap umur (TB/U), berat badan terhadap tinggi badan (BB/TB). Tabel Untuk BB/U, TB/U, dan BB/TB [2].

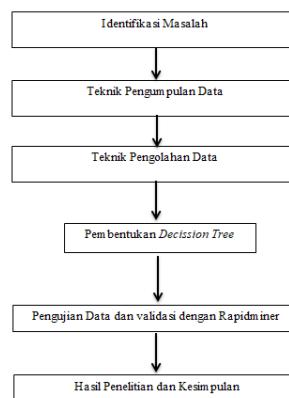
Angka bayi dan balita gizi buruk yang sangat tinggi, maka diperlukan pendekatan awal bagi balita untuk mengurangi kasus gizi buruk, salah satu langkah yang dapat dilakukan ialah dengan menguji status gizi balita, yang bertujuan untuk mendapatkan informasi status balita gizi secara cepat agar mempermudah instansi kesehatan melakukan penanganan pada balita. Untuk melakukan pengujian status gizi balita, dapat dilakukan dengan menggunakan sebuah teknik dalam data mining yaitu klasifikasi. Klasifikasi adalah proses penemuan model (atau fungsi) yang

menggambarkan dan membedakan kelas data atau konsep yang bertujuan agar bisa digunakan untuk memprediksi kelas dari objek yang label kelasnya tidak diketahui [3]. Klasifikasi adalah sebuah metode untuk mengelompokkan data secara sistem menurut aturan dan kaidah yang telah ditetapkan. Klasifikasi juga dapat diartikan pengelompokan data atau objek baru berdasarkan variabel yang diamati dengan tujuan untuk memprediksi suatu objek dari yang masih belum diketahui kelas atau kategorinya [4]. Untuk membantu pihak terkait dalam proses klasifikasi status gizi bayi pernah dilakukan oleh [5], dimana hasil penelitian menunjukkan bahwa data balita dengan jenis kelamin “laki-laki”, berat badan 7,4 kg dan tinggi badan 64,6 cm mendapatkan hasil normalisasi yang terdiri dari sangat kurus sebesar 0,07278948, kurus sebesar 0,56996389, normal sebesar 0,33702379 dan gemuk sebesar 0,02022283. Setiap hasil normalisasi dibandingkan berdasarkan nilai terbesar, maka didapatkan nilai 0,56996389 yang paling besar sehingga hasil klasifikasi status gizi balita adalah “Kurus”. Untuk itu penelitian ini dilakukan dengan menggunakan algoritma yang berbeda yang terdapat dalam datamining klasifikasi yaitu algoritma C4.5.

Algoritma C4.5 merupakan pengembangan dari algoritma ID3. Algoritma C4.5 [6] dan ID3 diciptakan oleh seorang peneliti dibidang kecerdasan buatan bernama J. Rose Quinlan pada akhir tahun 1970-an. Algoritma C4.5 membuat pohon keputusan dari atas ke bawah, dimana atribut paling atas merupakan akar, dan yang paling bawah dinamakan daun[7]. Hasil penelitian ini diharapkan dapat menjadi referensi kepada pihak terkait untuk mengetahui status gizi bayi di posyandu Gunung Maligas Simalungun. Dan menjadi tolak ukur dalam meminimalisir status bayi dan balita gizi buruk.

2. METODOLOGI PENELITIAN

Rancangan penelitian yang dilakukan dalam penelitian ini adalah pendekatan kuantitatif. dalam pendekatan penelitian kuantitatif yang dilakukan adalah (1) mengidentifikasi variabel-variabel masukan dan keluaran yang menjadi pusat perhatiannya; (2) mengeleminir atau mengontrol variabel-variabel; (3) memilih subyek dengan secara random; (4) melakukan treatment; dan (5) membandingkan pengaruh treatment dengan menggunakan batas kesesatan tertentu [8]. Rancangan penelitian digunakan untuk menguraikan dan menyelesaikan masalah dalam penelitian. Rancangan penelitian dapat dilihat pada *Flowchart* berikut ini :



Gambar 1. Rancangan Penelitian

Proses alur pada Gambar 1 adalah

- Identifikasi Masalah
Mengidentifikasi masalah apa yang terlihat pada Posyandu Gunung Maligas.
- Teknik Pengumpulan data
Pengumpulan data dalam penelitian ini dilakukan dengan wawancara kepada pihak posyandu dan memperoleh data mentah status gizi bayi dan balita.
- Teknik Pengolahan Data
Penelitian ini melakukan pengolahan data menggunakan Microsoft Excel 2010.
- Pembentukan *Decision Tree*
Penelitian ini melakukan pembentukan *Decision Tree* dengan menyatakan pengujian terhadap suatu atribut yang menghasilkan rule.
- Pengujian Data dan validasi Dengan *software* Rapidminer
Dalam pengujian data dilakukan dengan menggunakan *software* Rapidminer.
- Hasil Penelitian dan Kesimpulan
Setelah dilakukan pengujian *software* didapat hasil yang menjadi rekomendasi dalam proses pengambilan keputusan.

2.1. Analisis Data

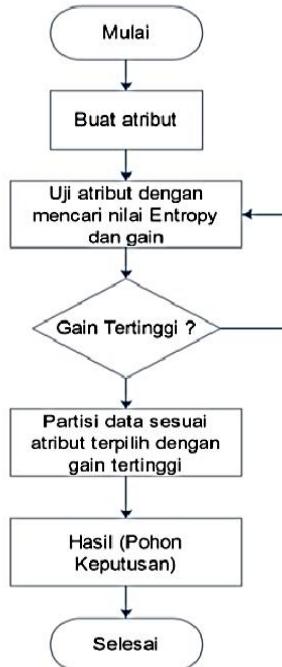
Untuk mengolah sebuah data menjadi sebuah informasi baru yang diperlukan analisis data agar karakteristik data mudah dimengerti dan dipahami sistem. Berikut adalah data gizi bayi dan balita yang diperoleh dari posyandu Gunung Maligas :

Tabel 1. Data Status Gizi Bayi Gunung Maligas

| No | Umur (Bulan) | JK | BB (kg) | TB (cm) | Gaji Orang Tua (Ribu) | Status Gizi | No | Umur (Bulan) | JK | BB (kg) | TB (cm) | Gaji Orang Tua (Ribu) | Status Gizi |
|----|--------------|----|---------|---------|-----------------------|-------------|-----|--------------|----|---------|---------|-----------------------|-------------|
| 1 | 11 | L | 12,2 | 90 | 2200000 | Baik | 61 | 35 | P | 11 | 110 | 2100000 | Buruk |
| 2 | 23 | L | 11,5 | 111 | 3000000 | Baik | 62 | 17 | L | 10,5 | 87 | 2100000 | Buruk |
| 3 | 35 | P | 11 | 110 | 2100000 | Buruk | 63 | 20 | L | 9 | 85 | 2500000 | Buruk |
| 4 | 17 | L | 10,5 | 87 | 2100000 | Buruk | 64 | 7 | L | 9,3 | 74 | 1000000 | Buruk |
| 5 | 20 | L | 9 | 85 | 2500000 | Baik | 65 | 4 | P | 6,8 | 67 | 1200000 | Buruk |
| 6 | 7 | L | 9,3 | 74 | 1000000 | Buruk | 66 | 1 | L | 4 | 55 | 1000000 | Buruk |
| 7 | 4 | P | 6,8 | 67 | 1200000 | Buruk | 67 | 42 | P | 12,6 | 112 | 3200000 | Baik |
| 8 | 1 | L | 4 | 55 | 1000000 | Buruk | 68 | 30 | P | 11,3 | 103 | 1500000 | Buruk |
| 9 | 42 | P | 12,6 | 112 | 3200000 | Baik | 69 | 11 | L | 12,2 | 90 | 2200000 | Baik |
| 10 | 30 | P | 11,3 | 103 | 1500000 | Buruk | 70 | 23 | L | 11,5 | 111 | 3000000 | Buruk |
| 11 | 27 | P | 10,2 | 89 | 2300000 | Buruk | 71 | 35 | P | 11 | 85 | 2100000 | Buruk |
| 12 | 14 | P | 25 | 120 | 2500000 | Buruk | 72 | 17 | L | 10,5 | 74 | 2100000 | Buruk |
| 13 | 45 | L | 15,4 | 115 | 3000000 | Baik | 73 | 20 | L | 9 | 67 | 2500000 | Buruk |
| 14 | 29 | L | 10 | 87 | 2500000 | Buruk | 74 | 7 | L | 9,3 | 55 | 1000000 | Buruk |
| 15 | 18 | L | 13,8 | 100 | 2200000 | Baik | 75 | 4 | P | 6,8 | 112 | 1200000 | Buruk |
| 16 | 14 | L | 13,6 | 99 | 2200000 | Baik | 76 | 1 | L | 4 | 103 | 1000000 | Buruk |
| 17 | 24 | L | 9,1 | 86 | 2500000 | Buruk | 77 | 42 | P | 12,6 | 112 | 3200000 | Baik |
| 18 | 7 | P | 7,8 | 75 | 1200000 | Buruk | 78 | 30 | P | 11,3 | 103 | 1500000 | Buruk |
| 19 | 11 | L | 12,2 | 90 | 2200000 | Baik | 79 | 11 | L | 12,2 | 90 | 2200000 | Baik |
| 20 | 23 | L | 11,5 | 111 | 3000000 | Baik | 80 | 23 | L | 11,5 | 111 | 3000000 | Baik |
| 21 | 35 | P | 11 | 110 | 2500000 | Buruk | 81 | 35 | P | 9,3 | 110 | 2100000 | Buruk |
| 22 | 17 | L | 10,5 | 87 | 1200000 | Buruk | 82 | 17 | L | 6,8 | 87 | 2100000 | Buruk |
| 23 | 20 | L | 9 | 85 | 2100000 | Buruk | 83 | 20 | L | 4 | 85 | 2500000 | Buruk |
| 24 | 7 | L | 9,3 | 74 | 2100000 | Buruk | 84 | 7 | L | 12,6 | 74 | 1000000 | Buruk |
| 25 | 4 | P | 6,8 | 67 | 2500000 | Buruk | 85 | 4 | P | 11,3 | 67 | 1200000 | Buruk |
| 26 | 1 | L | 4 | 55 | 1000000 | Buruk | 86 | 1 | L | 12,2 | 55 | 1000000 | Buruk |
| 27 | 42 | P | 12,6 | 112 | 1200000 | Buruk | 87 | 42 | P | 12,6 | 112 | 3200000 | Baik |
| 28 | 30 | P | 11,3 | 103 | 1500000 | Buruk | 88 | 30 | P | 11,3 | 103 | 1500000 | Buruk |
| 29 | 27 | P | 10,2 | 89 | 2300000 | Buruk | 89 | 27 | P | 10,2 | 89 | 2300000 | Buruk |
| 30 | 32 | P | 25 | 120 | 2500000 | Buruk | 90 | 60 | P | 25 | 120 | 2500000 | Buruk |
| 31 | 45 | L | 15,4 | 115 | 3000000 | Baik | 91 | 45 | L | 15,4 | 115 | 3000000 | Baik |
| 32 | 29 | L | 10 | 87 | 2500000 | Buruk | 92 | 29 | L | 10 | 87 | 2500000 | Buruk |
| 33 | 10 | P | 12,6 | 112 | 3200000 | Baik | 93 | 18 | L | 13,8 | 100 | 2200000 | Baik |
| 34 | 16 | P | 11,3 | 103 | 1500000 | Buruk | 94 | 14 | L | 13,6 | 99 | 2200000 | Baik |
| 35 | 27 | P | 10,2 | 89 | 2300000 | Buruk | 95 | 24 | L | 9,1 | 86 | 2500000 | Buruk |
| 36 | 20 | P | 25 | 120 | 2500000 | Buruk | 96 | 7 | P | 7,8 | 75 | 1200000 | Buruk |
| 37 | 15 | L | 15,4 | 115 | 3000000 | Baik | 97 | 7 | P | 7,8 | 75 | 1200000 | Buruk |
| 38 | 10 | L | 10 | 87 | 2500000 | Buruk | 98 | 14 | L | 13,6 | 99 | 2200000 | Baik |
| 39 | 12 | L | 13,8 | 100 | 2200000 | Baik | 99 | 57 | P | 14,8 | 114 | 3200000 | Buruk |
| 40 | 15 | L | 13,6 | 99 | 2200000 | Baik | 100 | 34 | P | 12,1 | 100 | 1200000 | Buruk |
| 41 | 16 | L | 9,1 | 86 | 2500000 | Buruk | | | | | | | |
| 42 | 7 | P | 7,8 | 75 | 1200000 | Buruk | | | | | | | |
| 43 | 35 | P | 11 | 110 | 2100000 | Buruk | | | | | | | |
| 44 | 17 | L | 10,5 | 87 | 2100000 | Buruk | | | | | | | |
| 45 | 20 | L | 9 | 85 | 2500000 | Buruk | | | | | | | |
| 46 | 7 | L | 9,3 | 74 | 1000000 | Buruk | | | | | | | |
| 47 | 4 | P | 6,8 | 67 | 1200000 | Buruk | | | | | | | |
| 48 | 1 | L | 4 | 55 | 1000000 | Buruk | | | | | | | |
| 49 | 42 | P | 12,6 | 112 | 3200000 | Baik | | | | | | | |
| 50 | 30 | P | 11,3 | 103 | 1500000 | Buruk | | | | | | | |
| 51 | 16 | L | 10,2 | 89 | 2500000 | Buruk | | | | | | | |
| 52 | 13 | L | 25 | 120 | 1200000 | Buruk | | | | | | | |
| 53 | 15 | L | 9,1 | 115 | 2100000 | Buruk | | | | | | | |
| 54 | 29 | L | 7,8 | 87 | 2100000 | Buruk | | | | | | | |
| 55 | 17 | P | 7,8 | 100 | 2500000 | Baik | | | | | | | |
| 56 | 14 | P | 13,6 | 99 | 1000000 | Baik | | | | | | | |
| 57 | 24 | L | 14,8 | 86 | 1200000 | Buruk | | | | | | | |
| 58 | 7 | P | 11,3 | 75 | 1000000 | Buruk | | | | | | | |
| 59 | 11 | L | 12,2 | 90 | 1200000 | Baik | | | | | | | |
| 60 | 23 | L | 11,5 | 111 | 3000000 | Baik | | | | | | | |

2.2. Pemodelan Metode

Metode yang digunakan dalam penelitian ini adalah metode datamining klasifikasi C4.5 yang bertujuan untuk mengklasifikasi metode dalam menyelesaikan permasalahan yang telah dipaparkan. Selanjutnya akan dilakukan pengujian data menggunakan aplikasi *Rapidminer*. Pemodelan metode C45 terlihat seperti pada Gambar 2:



Gambar 2. Flowchart C4.5

Proses pemodelan C4.5 yang tertera pada gambar dapat dijelaskan pemodelan metode dalam menyelesaikan masalah sebagai berikut:

- 1) Buat atribut yang akan dihitung;
- 2) Menghitung *Entropy* dan gain dari setiap atribut;
- 3) Melihat atribut yang memiliki gain tertinggi;
- 4) Menjadikan atribut dengan gain tertinggi sebagai root (Akar);
- 5) Bagi data sesuai atribut dengan gain tertinggi;
- 6) Membentuk pohon keputusan;
- 7) Selesai.

3. HASIL DAN PEMBAHASAN

3.1. Algoritma C4.5

Berikut adalah tahapan penyelesaian untuk membentuk sebuah pohon keputusan dari data kuisioner untuk mengetahui yang menjadi faktor utama status gizi bayi.

a) Pilih Atribut akar

Untuk memilih atribut sebagai akar, didasarkan pada nilai gain tertinggi dari atribut atribut yang ada. Untuk menghitung gain dan *Entropy* dapat dilihat pada rumus pada bab II. Dengan menggunakan persamaan diatas, maka akan didapatkan *Entropy* dan gain yang digunakan sebagai akar dalam membuat pohon keputusan.

1) *Entropy* Total

$$\begin{aligned} \text{Entropy Total} &= \left(-\frac{14}{100} * \log_2 \left(\frac{14}{100} \right) \right) + \left(-\frac{66}{100} * \log_2 \left(\frac{66}{100} \right) \right) + \left(-\frac{20}{100} * \log_2 \left(\frac{20}{100} \right) \right) \\ &= 1,2571408 \end{aligned}$$

Entropy total adalah menghitung hasil Lebih (14) Baik (66) dan Kurang (20) dimana 100 adalah keseluruhan jumlah kasus.

2) *Entropy* Atribut Umur

Untuk atribut Umur terdiri dari 2 nilai yaitu : umur ≤ 23 Bulan, dan umur ≥ 24 Bulan dimana nilai *Entropy* masing masing adalah sebagai berikut

a) umur ≤ 23 Bulan

(Jumlah kasus = 64, Hasil Lebih=14 , Hasil Baik = 39, Hasil Kurang =11

$$\begin{aligned} \text{Entropy} &= \left(-\frac{14}{64} * \log_2 \left(\frac{14}{64} \right) \right) + \left(-\frac{39}{64} * \log_2 \left(\frac{39}{64} \right) \right) + \left(-\frac{11}{64} * \log_2 \left(\frac{11}{64} \right) \right) \\ &= 1,3517593 \end{aligned}$$

b) umur ≥ 24 Bulan

(Jumlah Kasus = 36, Hasil Lebih=0 , Hasil Baik = 27, Hasil Kurang =9

$$\begin{aligned} Entropy &= \left(-\frac{0}{36} * \log_2 \left(\frac{0}{23} \right) \right) + \left(-\frac{27}{36} * \log_2 \left(\frac{27}{36} \right) \right) + \left(-\frac{9}{36} * \log_2 \left(\frac{9}{36} \right) \right) \\ &= 0 \end{aligned}$$

3) *Entropy* Jenis Kelamin

Untuk atribut jenis kelamin terdiri dari 2 nilai yaitu :jenis kelamin laki laki, dan jenis kelamin perempuan, dimana nilai *Entropy* masing masing adalah sebagai berikut :

a) Jenis kelamin Laki-Laki

(Jumlah Kasus = 57, Hasil Lebih = 12, Hasil Baik = 28, Hasil Kurang = 17)

$$\begin{aligned} Entropy &= \left(-\frac{12}{57} * \log_2 \left(\frac{12}{57} \right) \right) + \left(-\frac{28}{57} * \log_2 \left(\frac{28}{57} \right) \right) + \left(-\frac{17}{57} * \log_2 \left(\frac{17}{57} \right) \right) \\ &= 1,4975855 \end{aligned}$$

b) Jenis Kelamin Perempuan

(Jumlah Kasus = 43, Hasil Lebih = 2, Hasil Baik = 38, Hasil Kurang = 3

$$\begin{aligned} Entropy &= \left(-\frac{2}{43} * \log_2 \left(\frac{2}{43} \right) \right) + \left(-\frac{38}{43} * \log_2 \left(\frac{38}{43} \right) \right) + \left(-\frac{3}{43} * \log_2 \left(\frac{3}{43} \right) \right) \\ &= 0.631471 \end{aligned}$$

4) *Entropy* atribut Berat Badan

Untuk atribut berat badan terdiri dari 2 nilai yaitu: berat badan dengan berat kurang dari 10Kg, dan berat badan dengan berat lebih dari 10 Kg, dimana nilai *Entropy* masing masing adalah sebagai berikut :

a) Berat kurang dari 10Kg

(Jumlah Kasus = 37, Hasil Lebih =1 , Hasil Baik = 19, Hasil Kurang = 17

$$\begin{aligned} Entropy &= \left(-\frac{1}{37} * \log_2 \left(\frac{1}{37} \right) \right) + \left(-\frac{19}{37} * \log_2 \left(\frac{19}{37} \right) \right) + \left(-\frac{17}{37} * \log_2 \left(\frac{17}{37} \right) \right) \\ &= 1,1500617 \end{aligned}$$

b) Berat badan lebih dari 10Kg

(Jumlah Kasus = 63, Hasil Lebih =13 , Hasil Baik = 47, Hasil Kurang = 3

$$\begin{aligned} Entropy &= \left(-\frac{13}{63} * \log_2 \left(\frac{13}{63} \right) \right) + \left(-\frac{47}{63} * \log_2 \left(\frac{47}{63} \right) \right) + \left(\frac{3}{63} * \log_2 \left(\frac{3}{63} \right) \right) \\ &= 0.9943231 \end{aligned}$$

5) *Entropy* Tinggi Badan

Untuk atribut Tinggi badan terdiri dari 2 nilai yaitu : Tinggi badan dengan tinggi kurang dari 87 cm dan tinggi badan lebih dari 88 cm dimana nilai *Entropy* masing masing adalah sebagai berikut :

a) Tinggi badan kurang dari 87 cm

(Jumlah Kasus = 43, Hasil Lebih =0 , Hasil Baik = 25, Hasil Kurang = 18

$$\begin{aligned} Entropy &= \left(-\frac{0}{43} * \log_2 \left(\frac{0}{43} \right) \right) + \left(-\frac{25}{43} * \log_2 \left(\frac{25}{43} \right) \right) + \left(-\frac{18}{43} * \log_2 \left(\frac{18}{43} \right) \right) \\ &= 0 \end{aligned}$$

b) Tinggi badan lebih dari 88 cm

(Jumlah Kasus = 57, Hasil Lebih = 14, Hasil Baik = 41, Hasil Kurang = 2

$$\begin{aligned} Entropy &= \left(-\frac{14}{57} * \log_2 \left(\frac{14}{57} \right) \right) + \left(-\frac{41}{57} * \log_2 \left(\frac{41}{57} \right) \right) + \left(-\frac{2}{57} * \log_2 \left(\frac{2}{57} \right) \right) \\ &= 0.1,0089847 \end{aligned}$$

6) *Entropy* atribut Jumlah penghasilan orangtua

Untuk atribut penghasilan orangtua terdiri dari 2 nilai yaitu : gaji dengan nilai kurang dari Rp. 2000000, dan gaji dengan nilai lebih dari Rp. 2100000 dimana nilai *Entropy* masing masing adalah sebagai berikut :

a) Gaji kurang dari Rp. 2000000

(Jumlah Kasus = 35, Hasil Lebih = 2, Hasil Baik = 28, Hasil Kurang = 5

$$\begin{aligned} Entropy &= \left(-\frac{2}{35} * \log_2 \left(\frac{2}{35} \right) \right) + \left(-\frac{2}{35} * \log_2 \left(\frac{2}{35} \right) \right) + \left(-\frac{5}{35} * \log_2 \left(\frac{5}{35} \right) \right) \\ &= 0.8945522 \end{aligned}$$

b) Gaji lebih dari 2100000

(Jumlah Kasus = 65, Hasil Lebih = 12, Hasil Baik = 38, Hasil Kurang = 15

$$\begin{aligned} Entropy &= \left(-\frac{12}{65} * \log_2 \left(\frac{12}{65} \right) \right) + \left(-\frac{38}{65} * \log_2 \left(\frac{38}{65} \right) \right) + \left(-\frac{15}{65} * \log_2 \left(\frac{15}{65} \right) \right) \\ &= 1.3909193 \end{aligned}$$

7) Gain atribut umur

Gain untuk atribut umur adalah :

$$\begin{aligned} &= 1,2571408 - \left(\left(\frac{64}{100} * 1,3517593 \right) + \left(\frac{36}{100} * 0 \right) \right) \\ &= 0,392015 \end{aligned}$$

8) Gain atribut Jenis Kelamin

Gain untuk atribut jenis Kelamin adalah :

$$= 1,2571408 - \left(\left(\frac{57}{100} * 1,4975855 \right) + \left(\frac{43}{100} * 0,631471 \right) \right) \\ = 0,131985$$

9) Gain atribut Berat Badan

Gain untuk atribut berat badan adalah :

$$= 1,2571408 - \left(\left(\frac{37}{100} * 01,1500617 \right) + \left(\frac{63}{100} * 0,9943231 \right) \right) \\ = 0,205194$$

10) Gain atribut Tinggi Badan

Gain untuk atribut Tinggi badan adalah :

$$= 1,2571408 - \left(\left(\frac{43}{100} * 0 \right) + \left(\frac{57}{100} * 1,0089847 \right) \right) \\ = 0,682019$$

11) Gain atribut Gaji Orangtua

Gain untuk atribut Gaji Orangtua adalah :

$$= 1,2571408 - \left(\left(\frac{35}{100} * 0,8945522 \right) + \left(\frac{65}{100} * 1,3909193 \right) \right) \\ = 0,03995$$

Setelah seluruh nilai *Entropy* dan gain diperoleh, selanjutnya hasil dari perhitungan tersebut dimasukkan ke dalam Tabel 2.

Tabel 2. Tabel perhitungan Node 1

| Node | | Jumlah | Lebih | Baik | Entropy | Gain |
|-----------|-----------------|--------|-------|------|-----------|----------|
| 1 | | 100 | 29 | 71 | 0,8687212 | |
| Umur | ≤ 23 Bulan | 64 | 21 | 43 | 0,9129992 | 0,009288 |
| | ≥ 24 | 36 | 8 | 28 | 0,7642045 | |
| Jenis K | Laki-Laki | 57 | 21 | 36 | 0,949452 | 0,135065 |
| | Perempuan | 43 | 2 | 35 | 0,4476021 | |
| TB | < 87 | 37 | 2 | 35 | 0,3033748 | 0,135779 |
| | > 88 | 63 | 27 | 36 | 0,9852281 | |
| BB | < 10 | 43 | 0 | 42 | 0 | 0,230327 |
| | > 10 | 57 | 28 | 29 | 0,999778 | |
| Gaji ortu | < 2000000 | 35 | 2 | 33 | 0,3159971 | 0,121615 |
| | > 2100000 | 65 | 27 | 38 | 0,9792415 | |
| | | | | Max | 0,230327 | |

Pada perhitungan Tabel 2. dapat diketahui bahwa atribut dengan gain tertinggi adalah atribut Berat Badan dengan nilai 0,230427, sehingga atribut Berat Badan merupakan atribut yang menjadi node akar. Perhitungan yang sama dilakukan seperti perhitungan sebelumnya sampai ke node akhir dan memperoleh keputusan berikut adalah pohon keputusan yang terbentuk.

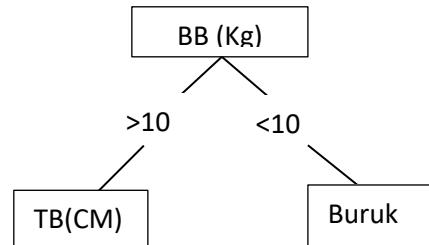
b) Cabang untuk setiap nilai

Pada atribut Tinggi Badan yang dijadikan akar, terdapat dua nilai yaitu "Berat Badan Kurang dari 10 Kg" dan "dan tinggi badan lebih dari 10 Kg". Dari nilai tersebut harus dilakukan perhitungan kembali sehingga dapat membentuk pohon keputusan. Selanjutnya adalah menghitung node 1.1 sebagai akar, sama dengan cara yang dilakukan pada perhitungan sebelumnya. Mengitung *Entropy* dari atribut yang tersisa . setelah dihitung *Entropy*, lalu kemudian dihitung gain untuk setiap atribut yang tersisa.

Tabel 3. Hasil Perhitungan Node 1.1

| Node | | Jumlah | Lebih | Baik | Entropy |
|---------|-----------------|--------|-------|------|-----------|
| 1.1 | | 57 | 14 | 41 | 0,0089847 |
| umur | ≤ 23 Bulan | 29 | 0 | 15 | 0,693034 |
| | ≥ 24 | 28 | 2 | 26 | 0,6431862 |
| Jenis K | Laki-Laki | 25 | 12 | 13 | 0,63365 |
| | Perempuan | 32 | 2 | 28 | 0,6685644 |
| TB | < 87 | 5 | 1 | 4 | 0,069206 |
| | > 88 | 52 | 13 | 37 | 1,0301418 |
| Gaji | < 2000000 | 14 | 2 | 11 | 0,9463729 |
| | > 2100000 | 43 | 12 | 30 | 1,0023981 |
| | | | | | 0,069206 |

Pada perhitungan Tabel 3. dapat diketahui bahwa atribut dengan gain tertinggi adalah atribut Tinggi badan dengan nilai 0,069206, sehingga atribut umur menjadi node berikutnya. Perhitungan yang sama dilakukan seperti perhitungan sebelumnya sampai ke node akhir dan memperoleh keputusan. Berikut adalah pohon keputusan yang terbentuk.



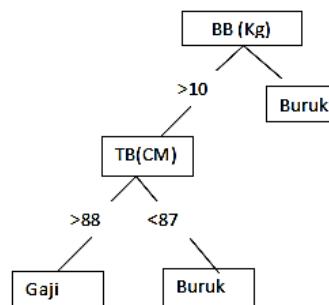
Gambar 3. Hasil Perhitungan Node 1.1

Berikut adalah perhitungan dari node 1.1.2 seperti yang ditunjukkan pada Tabel 4 berikut:

Tabel 4. Hasil Perhitungan Node 1.1.2

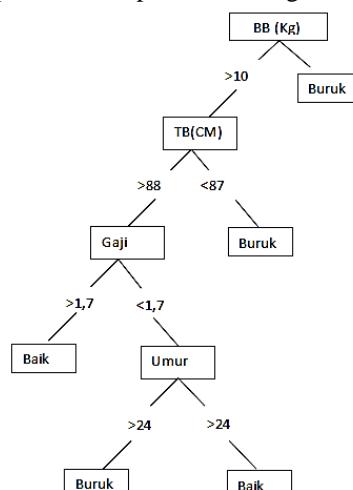
| | | | | | |
|------------|-----------|----|----|-----------|-----------|
| Node 1.1.2 | 28 | 2 | 26 | 0,6431862 | |
| Jenis K | Laki-Laki | 3 | 0 | 3 | 0 |
| | Perempuan | 25 | 4 | 21 | 0,7943096 |
| Umur | <23 | 1 | 0 | 1 | 0 |
| | >24 | 27 | 4 | 23 | 0,7533347 |
| Gaji Ortu | <2000000 | 8 | 2 | 6 | 1,0612781 |
| | >2100000 | 20 | 3 | 17 | 0,7475847 |
| | | | | Max | 0,19403 |

Pada perhitungan Tabel 4 dapat diketahui bahwa atribut dengan gain tertinggi adalah atribut Gaji Orangtua dengan nilai 0,19403, sehingga atribut umur menjadi node berikutnya. Dan dapat dibentuk pohon keputusan sebagai berikut:



Gambar 4. Hasil Perhitungan Node 1.1.2

Proses pembentukan pohon keputusan dari node ke node sudah terlihat pada proses diatas. Sehingga dari hasil perhitungan dan pembentukan pohon keputusan, didapat hasil sebagai berikut ini:

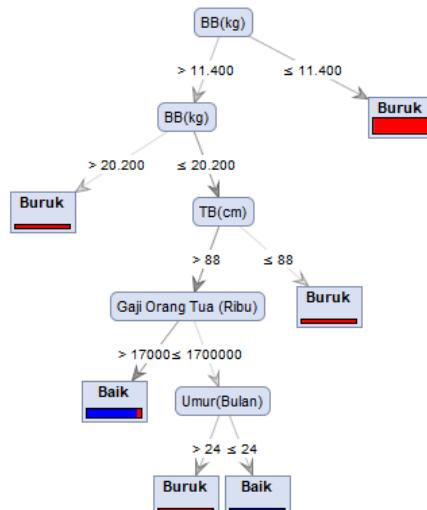


Gambar 5. Pohon Keputusan Akhir

Berdasarkan pohon keputusan yang telah dibuat, didapat rule rule sebagai berikut:

- Jika BB>20 Kg dan TB >88 Cm dan gaji orangtua >1,7Jt maka status Gizi BAIK;
- Jika BB>20 Kg dan TB >88 Cm dan gaji orangtua <1,7Jt dan umur >24 maka status gizi BURUK;
- Jika BB>20 Kg dan TB >88 Cm dan gaji orangtua <1,7Jt dan umur <24 maka status gizi BAIK;
- Jika BB>20 Kg dan TB >88 Cm Maka Status Gizi BURUK;
- Jika BB <<10 maka status gizi Buruk.

Berikut adalah hasil menggunakan *software* RapidMiner



Gambar 6. Pohon Keputusan yang terbentuk

Pada Gambar 6 dapat dilihat pohon keputusan yang terbentuk menggunakan *software* Rapidminer. Atribut Tinggi Badan menjadi node awal atau node akar.

Tree

```

BB(kg) > 11.400
|   BB(kg) > 20.200: Buruk {Baik=0, Buruk=5}
|   BB(kg) <= 20.200
|   |   TB(cm) > 88
|   |   |   Gaji Orang Tua (Ribu) > 1700000: Baik {Baik=25, Buruk=2}
|   |   |   Gaji Orang Tua (Ribu) <= 1700000
|   |   |   |   Umur(Bulan) > 24: Buruk {Baik=0, Buruk=2}
|   |   |   |   Umur(Bulan) <= 24: Baik {Baik=2, Buruk=0}
|   |   |   TB(cm) <= 88: Buruk {Baik=0, Buruk=3}
BB(kg) <= 11.400: Buruk {Baik=2, Buruk=59}
  
```

Gambar 7. Rule yang terbentuk oleh *Software* Rapidminer

Dari hasil perhitungan manual dan *software* dapat dilihat Atribut Tinggi badan menjadi node akar, yang berarti Atribut Tinggi badan merupakan faktor utama penentuan status gizi bayi dalam kasus ini.

4. KESIMPULAN

Berdasarkan penelitian yang telah dilakukan maka dapat disimpulkan bahwa Data Mining klasifikasi dengan algoritma C4.5 terbukti mampu mengetahui hasil analisis penentuan status gizi bayi di posyandu gunung maligas dengan hasil atribut Tinggi badan merupakan faktor utama penentuan status gizi bayi di posyandu Gunung Maligas. Hasil uji coba Software Rapidminer membuktikan uji coba data mining decision tree merupakan algoritma yang dapat membantu kasus penentuan status gizi bayi di posyandu gunung maligas.

REFERENCES

- [1] F. A. Aryanti, C. N. Istiqomah, D. Indrawan, and ..., "Door To Door Edukasi, Pemantauan Status Gizi Dan Program Mama Keren Untuk Pencegahan Gizi Kurang Pada Balita Di Masa Pandemi ...," ... *Masyarakat LPPM UMJ*, 2021.
- [2] M. Septikasari, *Status Gizi Anak dan Faktor yang Mempengaruhi*, vol. 1, no. 2. 2018.
- [3] I. dan A. Mutiara, "Penerapan K-Optimal Pada Algoritma Knn Untuk Prediksi Kelulusan Tepat Waktu Mahasiswa Program Studi Ilmu Komputer Fmipa Unlam Berdasarkan Ip Sampai Dengan Semester 4," *Klik - Kumpulan Jurnal Ilmu Komputer*, vol. 2, no. 2, pp. 159–173, 2015.
- [4] M. Sadli, F. Fajriana, W. Fuadi, E. Ermatita, and I. Pahendra, "Penerapan Model K-Nearest Neighbors Dalam Klasifikasi Kebutuhan Daya Listrik Untuk Masing-Masing Daerah Di Kota Lhokseumawe," *Jurnal ECOTIPE*, vol. 5, no. 2, pp. 11–18, 2018, doi: 10.33019/ecotipe.v5i2.646.
- [5] E. Darnila, M. Maryana, and M. Azmi, "Aplikasi Klasifikasi Status Gizi Balita Menggunakan Metode Naïve Bayes Berbasis Android," *METHOMIKA Jurnal Manajemen Informatika dan Komputerisasi Akuntansi*, vol. 5, no. 2, pp. 135–141, 2021, doi: 10.46880/jmika.vol5no2.pp135-141.
- [6] W. Katrina, H. J. Damanik, F. Parhusip, D. Hartama, A. P. Windarto, and A. Wanto, "C.45 Classification Rules Model for Determining Students Level of Understanding of the Subject," *Journal of Physics: Conference Series*, vol. 1255, no. 1, 2019, doi: 10.1088/1742-6596/1255/1/012005.
- [7] N. D. Prayoga, "Penerapan Algoritma C.45 Dalam Memprediksi Kelulusan Tepat Waktu Pada Perguruan Tinggi (Studi Kasus : Stmik Royal Kisaran)," 2018, doi: 10.31227/OSF.IO/UNQT4.
- [8] H. Ahyar *et al.*, *Buku Metode Penelitian Kualitatif & Kuantitatif*, no. March. 2020.