

PENERAPAN ALGORITMA C4.5 UNTUK PREDIKSI ANAK STUNTING DI KOTA PAGAR ALAM

Revaldo Xsanal Hakim, Ferry Putrawansyah, Riduan Syahri

Teknik Informatika, Institut Teknologi Pagar Alam

Jln. Masik Siagim No.75 Simpang Mbacang, Kota Pagar Alam, Sumatera Selatan, Indonesia

feypurawansyah@gmail.com

ABSTRAK

Di Pagar Alam, prediksi dan pengukuran tingkat Stunting masih mengandalkan analisis sekunder. Kader Posyandu melibatkan diri dalam mengukur kondisi balita, dan hasilnya diserahkan kepada ahli untuk menilai apakah balita tersebut mengalami Stunting atau tidak. Tujuan dari penelitian ini adalah mengaplikasikan Algoritma C4.5 untuk melakukan prediksi terkait kasus Stunting pada anak. Dari permasalahan yang ada diatas, maka metode yang dapat menyelesaikan permasalahan ini yaitu Algoritma C4.5 yang termasuk dalam Pohon Keputusan pada data mining. Proses data mining peneliti menggunakan salah satu metode CRISP-DM dan pengujian menggunakan *black box testing* memperoleh total 4.35 masuk ke kategori sangat layak. Hasil dalam penelitian ini berupa sebuah sistem. Pohon keputusan diwujudkan sebagai aturan dalam sistem. Tujuan sistem ini adalah membantu tenaga kesehatan dalam membuat keputusan terkait prediksi status gizi balita. Sistem prediksi ini bermanfaat untuk mengidentifikasi balita berisiko gizi buruk sehingga tindakan pencegahan dapat dilakukan dengan lebih efektif.

Kata kunci : Data Mining; C4.5 Algorithm; Prediction; Stunting; Crisp Dm

1. PENDAHULUAN

Perkembangan pesat dalam teknologi informasi membuka pintu lebar-lebar untuk kemudahan akses informasi melalui internet, memungkinkan akses di berbagai lokasi dan waktu. Internet menjadi sumber beragam informasi yang mencakup segala aspek kehidupan, termasuk Kesehatan, yang memegang peranan krusial dalam kehidupan [1]. saat ini komputer dapat mempercepat proses pengolahan data seperti halnya untuk memprediksi suatu kejadian hanya dengan memasukkan beberapa variabel yang telah ditentukan. Seorang stoner hanya tinggal memasukkan beberapa data, sehingga dari data tersebut sistem dapat langsung memprediksi suatu kejadian. Dengan memanfaatkan sistem komputer yang ada dan menggunakan jaringan komputer yang terintegrasi, maka akan didapatkan suatu sistem prediksi yang cepat, tepat dan akurat. Dalam proses kegiatannya salah satu permasalahan yang memerlukan sistem prediksi adalah untuk membantu memprediksi anak yang kekurangan gizi di kota Pagar Alam. Hingga saat ini, Indonesia masih dihadapkan pada berbagai masalah kesehatan, termasuk isu gizi balita sebagai salah satu contohnya [2].

Seperti yang diungkapkan oleh [3], gizi adalah proses di mana organisme menggunakan makanan dengan cara normal melalui proses pencernaan, penyerapan, transportasi, penyimpanan, dan pengeluaran zat-zat yang tidak diperlukan untuk mendukung kelangsungan hidup, pertumbuhan, fungsi normal organ tubuh, dan penghasilan energi.

Kelompok usia anak balita disebut sebagai "usia emas" karena pada periode ini, anak sedang membangun dasar untuk pertumbuhan tubuhnya di masa depan. Pada fase ini, anak sangat rentan terhadap masalah gizi karena membutuhkan asupan gizi yang memadai untuk mendukung proses tumbuh kembangnya [4].

Kekurangan gizi kronis dapat mengakibatkan kondisi Stunting.

Dalam pandangan [5], gizi buruk adalah kondisi yang terjadi ketika seseorang mengalami kekurangan asupan nutrisi atau jumlah nutrisi yang dikonsumsi berada di bawah standar. Stunting, yang umumnya terjadi pada balita, menjadi isu kesehatan masyarakat di Indonesia, berpotensi menurunkan produktivitas dan kualitas sumber daya manusia di masa depan. Stunting dapat meningkatkan risiko penyakit, kematian, gangguan perkembangan motorik otak, dan mengurangi tingkat produktivitas [6].

Meskipun prevalensi penyakit menular di Indonesia mencapai 24,4, hal ini dianggap sebagai masalah karena melebihi batas standar yang ditetapkan oleh Organisasi Kesehatan Dunia (WHO), di mana tingkat prevalensi penyakit menular dalam satu negara seharusnya tidak melebihi 20 [7].

Menurut [8], implementasi mengacu pada tindakan yang dilakukan oleh individu atau kelompok untuk mencapai tujuan yang telah ditetapkan dan disepakati dalam suatu keputusan. Sementara menurut [9], penerapan (implementasi) merupakan perluasan aktivitas yang menyesuaikan proses interaksi antara tujuan dan tindakan guna mencapainya, melibatkan jaringan pelaksana dan birokrasi yang efektif.

Menurut [10], CRISP-DM adalah standar proses penambangan data yang dikembangkan oleh tiga perusahaan pembuat perangkat lunak penambangan data, yaitu Daimler Chrysler (Daimler-Benz), SPSS (ISL), dan NCR. Standar ini kemudian mengalami pengembangan melalui serangkaian lokakarya antara tahun 1997 hingga 1999. Sebuah penelitian serupa berjudul "Prediksi Status Gizi Balita Menggunakan Algoritma C4.5" oleh [6] menggunakan Algoritma C4.5 untuk mengklasifikasikan dan memprediksi status gizi balita di kecamatan Bringin. Pengujian

dilakukan pada sistem yang dibangun oleh penulis, serta melibatkan pengujian homemade dan menggunakan RapidMiner sebagai pembanding. Penelitian "Simulasi Penerapan Metode Pohon Keputusan (C4.5) dalam Penentuan Status Gizi Balita" oleh (bertujuan untuk mengetahui status gizi balita. Dalam penelitian ini, melibatkan 123 data balita yang lahir antara tahun 2015 hingga 2019. Hasil penelitian menunjukkan bahwa Algoritma C4.5 dapat diaplikasikan untuk menentukan status gizi balita, dengan berat badan (C1) dan tinggi badan (C2) sebagai faktor penentu.

Dalam penelitian "Komparasi Algoritma C4.5 dengan Algoritma Naïve Bayes Classifier untuk Klasifikasi Status Gizi Balita" [11], terdapat lima atribut yang memengaruhi status gizi balita, termasuk jenis kelamin, umur, BB/U, TB/U, dan BB/TB. Penelitian ini menggunakan Algoritma C4.5 dan Algoritma Naïve Bayes Classifier untuk mengklasifikasikan status gizi balita.

Observasi dan wawancara di Dinas Kesehatan Kota Pagar Alam menunjukkan bahwa prediksi stunting masih menggunakan metode Analisis Data Sekunder. Deteksi stunting pada balita dapat dilakukan secara dini melalui pemantauan rutin kurva pertumbuhan anak di Posyandu terdekat, yang melibatkan pengukuran kondisi balita oleh kader Posyandu dan evaluasi hasil oleh ahli. Meskipun pendekatan ini memberikan wawasan tentang prevalensi stunting dan faktor risiko, analisis data sekunder memiliki kelemahan seperti keterbatasan data, ketidakonsistenan definisi dan metode, serta ketidakmungkinan memasukkan variabel tertentu yang dapat mengakibatkan prediksi tidak akurat dan tidak representatif. Oleh karena itu, solusi yang diusulkan adalah menggunakan teknik Data Mining. Menurut [3], Algoritma C4.5 merupakan salah satu algoritma Decision Tree yang dikembangkan oleh J.Ross Quinlan. Sementara menurut [4], Algoritma C4.5 juga dikenal sebagai Decision Tree. Pohon Keputusan adalah metode klasifikasi yang melibatkan konstruksi pohon keputusan dengan simpul keputusan terhubung dari simpul akar ke simpul daun akhir.

Data Mining, atau Penambangan Data, adalah proses ekstraksi informasi, pengidentifikasian pola, dan relasi dalam suatu himpunan data untuk menyelesaikan masalah tertentu. Tujuan utama Data Mining adalah mengenali relasi dan pola yang tidak terlihat langsung oleh analis manusia, memanfaatkan teknik komputasi dan statistik untuk menganalisis data dengan tujuan menghasilkan wawasan yang bermanfaat. Salah satu metode Data Mining untuk prediksi adalah Metode Pohon Keputusan.

Menurut [12], prediksi adalah peramalan atau perkiraan, sementara menurut [13], ramalan atau soothsaying adalah ilmu yang mempelajari prediksi kejadian di masa depan. Berdasarkan latar belakang ini, penelitian ini berjudul "Penerapan Algoritma C4.5 untuk Prediksi Stunting di Kota Pagar Alam." Harapannya, hasil penelitian ini dapat membantu

Dinas Kesehatan dan masyarakat Kota Pagar Alam dalam memprediksi pemenuhan gizi anak-anak, sehingga kekurangan gizi dapat dicegah dan diminimalisir. Dengan mengurangi stunting di Kota Pagar Alam, produktivitas dan kualitas masyarakat diharapkan meningkat, membuat sumber daya manusia di kota tersebut lebih kompetitif di masa depan. Tujuan penelitian ini adalah menerapkan Algoritma C4.5 untuk memprediksi stunting pada anak di Kota Pagar Alam.

2. TINJAUAN PUSTAKA

2.1. Penerapan/Implementasi

Menurut [7], penerapan merujuk pada langkah-langkah yang diambil baik oleh individu maupun kelompok, yang bertujuan mencapai sasaran yang telah ditetapkan dalam suatu keputusan. Sementara menurut [8],

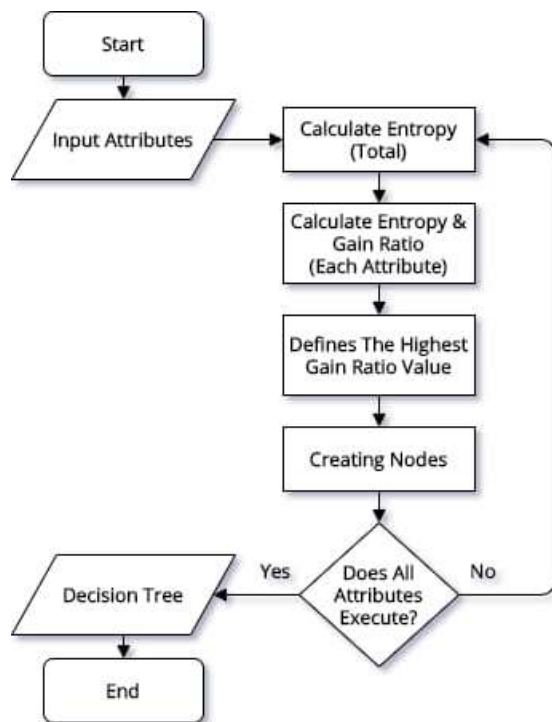
Implementasi, atau penerapan, melibatkan perluasan kegiatan yang secara bersinergi menyelaraskan interaksi antara tujuan dan tindakan guna mencapainya. Proses ini memerlukan kerjasama dari jaringan pelaksana kegiatan dan birokrasi yang sesuai. Berdasarkan penjelasan para ahli di atas, dapat disimpulkan bahwa penerapan adalah suatu metode yang digunakan dalam kegiatan untuk menerapkan pengetahuan dengan tepat dan akurat kepada masyarakat. Menurut [3]

2.2. Algoritma C4.5

Algoritma C4.5 adalah termasuk salah satu algoritma untuk membangun pohon keputusan yang dikembangkan oleh J.Ross Quinlan. *Algoritma C4.5* merupakan bagian dari algoritma *Decision Tree*. Sementara menurut [3], Algoritma C4.5 juga dikenal sebagai Pohon Keputusan atau decision tree. Berdasarkan opini ahli tersebut, dapat disimpulkan Algoritma C4.5 merupakan suatu algoritma yang efektif untuk melakukan prediksi menggunakan struktur pohon keputusan.

Menurut [11], proses pelatihan dalam Algoritma C4.5 dapat didefinisikan sebagai berikut:

- Mencari akar pohon menggunakan gain ratio, di mana atribut eksperimen untuk simpul dipilih berdasarkan gain ratio tertinggi. Proses perhitungan gain ratio melibatkan:
 - Menghitung Entropi pada setiap atribut dalam dataset.
 - Menghitung nilai gain untuk setiap atribut menggunakan metode information gain.
 - Menghitung split info untuk setiap atribut.
 - Setelah mendapatkan nilai-nilai tersebut, perhitungan gain ratio dilakukan untuk semua atribut.



Gambar 1. Flowchart Algoritma C4.5

- b. Menentukan Leaf Node berikutnya: a. Label yang mencakup nilai Positif (+) dan Negatif (-) menjadi leaf node berikutnya. b. Node yang dipilih adalah node dengan gain ratio tertinggi. c. Jika beberapa node memiliki gain ratio yang sama, keputusan diambil oleh ahli (expert).

Melakukan langkah-langkah serupa untuk leaf node berikutnya.

S	=	Himpunan kasus A
kasus A	=	Jumlah Kasus A
Atribut c	=	Jumlah Kasus c
Si	=	Jumlah kasus pada partisi ke-i
s	=	Jumlah kasus dalam s

Dengan demikian, dapat disimpulkan bahwa dalam Algoritma C4.5, pelatihan melibatkan perhitungan gain ratio untuk membentuk pohon keputusan yang digunakan untuk prediksi.

Setelah pohon keputusan terbentuk, langkah selanjutnya adalah merumuskan aturan berdasarkan struktur pohon keputusan tersebut. Aturan yang terbentuk dari pohon keputusan untuk membuat model.

Apabila terdapat hasil data yang bertentangan pada proses ini maka diperlukan seorang pakar untuk pengambilan keputusan.

Beberapa tahap dalam menggunakan Algoritma C4.5 yaitu :

- Menyusun data latihan. Data latihan umumnya diambil dari catatan sejarah peristiwa sebelumnya yang telah dikelompokkan ke dalam kategori-kategori tertentu.

- Menentukan simpul akar pohon. Dalam pembentukan pohon keputusan menggunakan metode algoritma Decision Tree (C4.5), simpul akar diambil dari atribut terpilih dengan menghitung nilai gain pada setiap atribut.
- Sebelum menghitung nilai gain dari atribut tersebut, perlu dilakukan perhitungan nilai entropy menggunakan rumus:

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

Keterangan :

S = Himpunan kasus

n = Jumlah partisi S

Pi = Proporsi dari Si terhadap S

Kemudian hitung nilai gain dengan metode *information gain* :

Gain(S, A) =

$$Entropy(S) - \sum_{i=1}^n \frac{|s_i|}{|s|} * Entropy(S_i) \quad (2)$$

Information gain memeriksa pola data dan menggunakan pola tersebut sebagai pedoman untuk memilih atribut yang akan digunakan dalam pembuatan pohon keputusan. Atribut yang dipilih akan menghasilkan pembagian data yang seragam dan berpotensi menghasilkan pohon keputusan yang sederhana dengan jumlah percabangan yang minimal. Untuk menilai gain ratio, perlu dilakukan perhitungan pemisahan informasi (split info). Metode penghitungan split info melibatkan atribut tersebut, perlu dilakukan perhitungan nilai entropy menggunakan rumus:

$$SplitInfo(S, A) = - \sum_{i=1}^c \frac{|s_i|}{|s|} \log_2 \frac{|s_i|}{|s|} \quad (3)$$

keterangan :

S = Himpunan

kasus A = Jumlah Kasus A

Atribut c = Jumlah Kasus c

Si = Jumlah kasus pada partisi ke-i

s = Jumlah kasus dalam s

dengan cara :

$$Gain(S, A) = SplitInfo(S, A) - Entropy(S, A) \quad (4)$$

Keterangan :

S = Himpunan kasus

A = Atribut

- Setelah memperoleh nilai gain ratio, langkah berikutnya melibatkan penentuan variabel berikutnya: a) Variabel yang mencakup nilai Positif (+) dan Negatif (-) dianggap sebagai variabel berikutnya. b) Variabel yang dipilih

adalah variabel yang memiliki nilai gain ratio tertinggi.

- Proses serupa diulang untuk variabel berikutnya. Setelah terbentuk Pohon Keputusan, langkah selanjutnya adalah merumuskan aturan berdasarkan struktur pohon keputusan yang telah dibuat.

2.3. Prediksi

Menurut [12], prediksi merujuk pada tindakan meramalkan atau memperkirakan suatu peristiwa. Sementara menurut [14], Ramalan atau peramalan adalah disiplin ilmu yang mengeksplorasi upaya untuk meramalkan peristiwa yang akan terjadi di masa mendatang. Dengan merujuk pada definisi para ahli tersebut, dapat disimpulkan bahwa prediksi adalah suatu proses perkiraan terhadap kejadian di masa depan berdasarkan data yang telah dikumpulkan dan dipelajari.

3. METODE PENELITIAN

Penelitian menggunakan metode CRISP DM:

3.1. Metode Pengembangan Sistem

Pendekatan Pengembangan Sistem Cross Industry Standard Process for Data Mining (CRISP-DM) adalah suatu proses standar untuk Data Mining yang dirancang oleh tiga inisiator utama dalam industri Data Mining, yaitu Daimler Chrysler (Daimler-Benz), SPSS (ISL), dan NCR. Proses standar ini kemudian mengalami perkembangan melalui serangkaian lokakarya yang berlangsung antara tahun 1997 dan 1999 [10]. Gambaran tahap proses pada pengembangan sistem ini dapat dilihat pada ilustrasi [14].

Penelitian mengadopsi metode Cross Industry Standard Process untuk Data Mining (CRISP aDM) menurut [15]. Tahap-tahap dalam metode ini melibatkan:



Gambar 2. Metode CRIPS_DM

- Pengertian Bisnis:** Fase ini secara rinci mendefinisikan tujuan dan persyaratan proyek dalam kerangka keseluruhan unit penelitian [16]. Tujuan dan batasan diterjemahkan menjadi rumus untuk merinci masalah penambahan data.
- Pemahaman Data:** Pada tahap ini, analisis data digunakan untuk mengumpulkan, mengidentifikasi, dan mencari pengetahuan awal dalam data [17]

- Persiapan Data:** Fase ini melibatkan aktivitas untuk menyusun data akhir (data yang telah diproses dalam fase pemodelan) dari data mentah [18].
- Pemodelan:** Pada tahap ini, berbagai teknik pemodelan dieksplorasi dan dievaluasi dengan penyesuaian parameter untuk mencapai hasil optimal [19].
- Evaluasi:** Model yang dihasilkan dievaluasi untuk memastikan kualitas dan efektivitasnya sebelum diimplementasikan secara lebih luas [20].
- Penyebaran:** Pada tahap ini, pengetahuan dan data yang diperoleh disusun untuk penggunaan praktis. Proses penyebaran bisa berupa penyusunan laporan atau implementasi siklus penggalian data secara berulang di organisasi [21].

3.2. Metode Pengumpulan Data :

Penelitian menggunakan teknik pengumpulan data yang melibatkan studi pustaka dan studi lapangan yaitu observasi dan wawancara [22]:

- Metode observasi:** Langsung melakukan pengamatan ke lapangan, seperti Dinas Kesehatan Kota Pagar Alam [23].
- Wawancara:** Menjadi alat verifikasi untuk informasi yang diperoleh dari narasumber [24], khususnya dengan Seksi Kesehatan Keluarga dan Gizi untuk mendapatkan informasi mengenai sistem prediksi anak stunting, variabel penentu, dan ketersediaan data.
- Studi pustaka:** Melibatkan penelitian terhadap jurnal, artikel, skripsi, atau tesis untuk menambah referensi yang relevan dengan desain program sistem prediksi stunting [25].

4. HASIL DAN PEMBAHASAN

Hasil dari penelitian oleh penulis menggunakan pendekatan teknik CRISP-DM untuk merancang Sistem Prediksi Anak Stunting di Kota Pagar Alam berbasis website.

- Sistem Perencanaan:** Memahami sistem yang berjalan saat ini di Dinas Kesehatan Kota Pagar Alam dalam memprediksi stunting.
- Pengumpulan Data:** Mempelajari struktur data yang ada dan menentukan data yang akan digunakan pada proses data mining. Data yang diambil mencakup atribut seperti Nama, Jenis Kelamin, Usia Saat Ukur, Tanggal Pengukuran, Berat, dan Tinggi.
- Persiapan Data:** Melibatkan pemilihan, pemrosesan, dan konversi data, termasuk pemilihan atribut dan pemetaan data menggunakan Algoritma C4.5.
- Proses Pemodelan C4.5:** Menyertakan langkah-langkah seperti menyiapkan atribut, menghitung Entropy, menentukan akar pohon keputusan, dan menggambarkan proses untuk mengatasi permasalahan spesifik. Proses pemodelan C4.5 menjelaskan pemodelan metode dalam menyelesaikan masalah sebagai berikut:

1. Buat atribut yang akan dihitung;
2. Menghitung *Entropy* dan gain dari setiap atribut;

a. Pilih Atribut akar

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

1) *Entropy* Total

$$\begin{aligned} \text{Entropy Total} &= \left(-\frac{51}{100} * \log_2 \left(\frac{51}{100} \right) \right) + \left(-\frac{41}{100} * \right. \\ &\quad \left. \left(\frac{41}{100} \right) \right) + \left(-\frac{8}{100} * \log_2 \left(\frac{8}{100} \right) \right) \\ &= 1,314323 \end{aligned}$$

2) *Entropy* Atribut Umur

Untuk atribut Umur terdiri dari fase nilai yaitu umur ≤ 23 Bulan, umur ≥ 24 Bulan, nilai *Entropy* nya adalah

a) umur ≤ 23 Bulan

(Jumlah kasus = 0, Hasil Lebih=0 , Hasil Baik = 0, Hasil Kurang =0)

$$\begin{aligned} \text{Entropy Atribut umur} &= \left(-\frac{0}{0} * \log_2 \left(\frac{0}{0} \right) \right) + \left(-\frac{0}{0} * \right. \\ &\quad \left. \left(\frac{0}{0} \right) \right) + \left(-\frac{0}{0} * \log_2 \left(\frac{0}{0} \right) \right) \\ &= 0 \end{aligned}$$

b) umur ≥ 24 Bulan

(Jumlah Kasus = 100, Hasil Lebih=51 , Hasil Baik = 41, Hasil Kurang = 8)

$$\begin{aligned} \text{Entropy Atribut umur} &= \left(-\frac{51}{100} * \log_2 \left(\frac{51}{100} \right) \right) + \\ &\quad \frac{1}{90} * \log_2 \left(\frac{41}{100} \right) + \\ &\quad = \frac{3}{90} * \log_2 \left(\frac{8}{100} \right) \\ &= 1,314323 \end{aligned}$$

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 :

a) Jenis kelamin Laki-Laki

(Jumlah Kasus = 47, Hasil Lebih = 26, Hasil Baik = 18, Hasil Kurang = 3)

b) Jenis Kelamin Perempuan

(Jumlah Kasus = 53, Hasil Lebih = 25, Hasil Baik = 23, Hasil Kurang = 3)

$$\begin{aligned} \text{Kelamin Perempuan} &= \left(-\frac{25}{53} * \log_2 \left(\frac{25}{53} \right) \right) + \left(-\right. \\ &\quad \left. \left(\frac{23}{53} \right) \right) + \left(-\frac{3}{53} * \log_2 \left(\frac{3}{53} \right) \right) \\ &= 1,355317 \end{aligned}$$

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 Ringan

(Jumlah Kasus = 8, Hasil Lebih =8 , Hasil Baik = 0, Hasil Kurang = 0)

$$\begin{aligned} \text{Entropy Berat Badan} &= \left(-\frac{8}{8} * \log_2 \left(\frac{8}{8} \right) \right) + \left(-\frac{0}{8} * \right. \\ &\quad \left. \left(\frac{0}{8} \right) \right) + \left(-\frac{0}{8} * \log_2 \left(\frac{0}{8} \right) \right) \\ &= -0,301029 \end{aligned}$$

b) Berat Sedang

(Jumlah Kasus = 41, Hasil Lebih =0 , Hasil Baik = 41, Hasil Kurang = 0)

$$\begin{aligned} \text{Entropy Berat Badan} &= \left(-\frac{0}{41} * \log_2 \left(\frac{0}{41} \right) \right) + \\ &\quad \frac{1}{1} * \log_2 \left(\frac{41}{41} \right) + \left(-\frac{0}{41} * \right. \\ &\quad \left. \left(\frac{0}{41} \right) \right) \\ &= -0,301029 \end{aligned}$$

c) Massa Tubuh yang Berlebihan

(Jumlah Kejadian = 51, Output Lebih = 0, Output Baik = 0, Output Kurang = 51)

$$\begin{aligned} \text{Entropy Berat Badan} &= \left(-\frac{0}{51} * \log_2 \left(\frac{0}{51} \right) \right) + \\ &\quad \frac{1}{1} * \log_2 \left(\frac{0}{51} \right) + \left(-\frac{51}{51} * \right. \\ &\quad \left. \left(\frac{51}{51} \right) \right) \\ &= -0,301029 \end{aligned}$$

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 Rendah

(Jumlah Kasus = 7, Hasil Lebih =6 , Hasil Baik = 1, Hasil Kurang = 0)

$$\begin{aligned} \text{Entropy Tinggi Rendah} &= \left(-\frac{6}{7} * \log_2 \left(\frac{6}{7} \right) \right) + \\ &\quad * \log_2 \left(\frac{1}{7} \right) + \left(-\frac{0}{7} * \right. \\ &\quad \left. \left(\frac{0}{7} \right) \right) \\ &= -0,2277924173 \end{aligned}$$

$$\begin{aligned} \text{Kelamin Laki-Laki} &= \left(-\frac{26}{47} * \log_2 \left(\frac{26}{47} \right) \right) + \\ &\quad \frac{3}{7} * \log_2 \left(\frac{18}{47} \right) + \left(-\frac{3}{47} * \right. \\ &\quad \left. \left(\frac{3}{47} \right) \right) \\ &= -1,256185 \end{aligned}$$

b) Postur Tubuh Normal

(Jumlah Kejadian = 36, Output Lebih = 2,
Output Baik = 24, Output Kurang = 10)

$$\begin{aligned} \text{Tinggi} &= \left(-\frac{2}{36} * \log_2 \left(\frac{2}{36} \right) \right) + \\ \text{Sedang} &\quad \frac{4}{5} * \log_2 \left(\frac{24}{36} \right) + \left(-\frac{10}{36} * \right. \\ &= \left. \left(\frac{10}{36} \right) \right) \\ &= -0,1579478372 \end{aligned}$$

c) Elevasi (ketinggian)

(Jumlah Kejadian = 57, Output Lebih = 0,
Output Baik = 16, Output Kurang = 41)

$$\begin{aligned} \text{Entropy Tinggi (Tinggi)} &= \left(-\frac{0}{57} * \log_2 \left(\frac{0}{57} \right) \right) + \left(-\frac{16}{57} * \right. \\ &\quad \left(\frac{16}{57} \right) + \left(-\frac{41}{57} * \log_2 \left(\frac{41}{57} \right) \right) \\ &= -0,1794690987 \end{aligned}$$

6). Gain untuk atribut umur adalah :

$$\begin{aligned} \text{Gain Umur} &= 1,314323 - \left(\left(\frac{0}{100} * 0 \right) + \right. \\ &\quad \left. \left(\frac{100}{100} (-1,314323) \right) \right) \end{aligned}$$

= -0

7). Hitung Gain atribut

Gain untuk atribut jenis Kelamin adalah :

$$\begin{aligned} \text{Gain JK} &= 1,314323 - \left(\left(\frac{47}{100} * (-1,256185) \right) + \right. \\ &\quad \left. \left(\frac{53}{100} * (-1,355317) \right) \right) \\ &= 0,00559804 \end{aligned}$$

8. Keuntungan atribut Berat Badan
Keuntungan yang terkait dengan atribut
berat badan adalah:

$$\begin{aligned} \text{Gain BB} &= 1,314323 \\ &- \left(\left(\frac{8}{100} * -0,301029 \right) + \left(\frac{41}{100} * -0,301029 \right) \right) \\ &\quad + \left(\frac{51}{100} * -0,301029 \right) \\ &= 1,151466311 \end{aligned}$$

8) Gain atribut Tinggi Badan

Gain untuk atribut Tinggi badan berikut ini :

$$\begin{aligned} \text{Gain TB} &= \\ &1,314323 \left(\left(\frac{7}{100} * -0,22779 \right) + \right. \\ &\quad \left. \left(\frac{36}{100} * -0,1579 \right) + \left(\frac{57}{100} * -0,1794 \right) \right) \\ &= 1,1392189231 \end{aligned}$$

3. Melihat atribut yang memiliki gain tertinggi

Gain tertinggi yaitu pada Atribut Berat Badan 1,151466311, gain tertinggi yaitu pada Atribut tinggi badan 1,1392189231, Gain tertinggi ketiga

yaitu pada Atribut Jenis Kelamin 0,00559804, dan Gain yang paling terendah pada Atribut Umur -0

4. Atribut yang memiliki gain tertinggi dijadikan sebagai root (akar)

Atribut Berat Badan dipilih sebagai node akar karena memiliki Gain tertinggi.

5. Membagi data berdasarkan atribut dengan gain tertinggi. Dengan BB sebagai Akar, dilanjutkan dengan Atribut TB, lalu Atribut Jenis Kelamin dan yang terakhir yaitu Atribut Umur

6. Membentuk pohon keputusan;



Gambar 3. Pohon Keputusan yang terbentuk

7. Pengujian Confusion Matrix

accuracy: 84.63%

	true Normal	true Elevasi	true Tinggi	class precision
pred Normal	16	4	0	80.00%
pred Elevasi	2	22	0	91.67%
pred Tinggi	2	0	4	66.67%
class recall	80.00%	84.62%	100.00%	

Gambar 4. Pengujian dengan Confusion Matrix

8. Selesai.

Pada tahap Evaluasi model yang dihasilkan sudah membentuk pohon keputusan dan sudah memiliki cabang berdasarkan entropy dan gain yang sudah dihitung. Cabang yang berada paling atas merupakan atribut yang memiliki nilai Gain tertinggi. Dari proses data mining Algoritma C4.5 yang telah dihitung menghasilkan beberapa aturan atau Rules yang nantinya akan di Implementasikan ke dalam sebuah sistem prediksi yang akan di bangun.

```

Rule 0 = 1.500 Elevasi (Normal=0, Elevasi=2, Tinggi=0)
Rule 1 = 1.500
1. Berat = 14.500
2. Jika Berat Badan > 54.500 Normal (Normal=0, Elevasi=0, Tinggi=0)
3. Jika Berat Badan < 54.500
4. Jika Berat Badan < 54.500
5. Berat > 0.500 Elevasi (Normal=0, Elevasi=2, Tinggi=0)
6. Berat < 0.500 Normal (Normal=1, Elevasi=0, Tinggi=0)
7. Jika Berat Badan < 54.500 Normal (Normal=0, Elevasi=0, Tinggi=0)
8. Berat < 14.500 Tinggi (Normal=1, Elevasi=0, Tinggi=0)
  
```

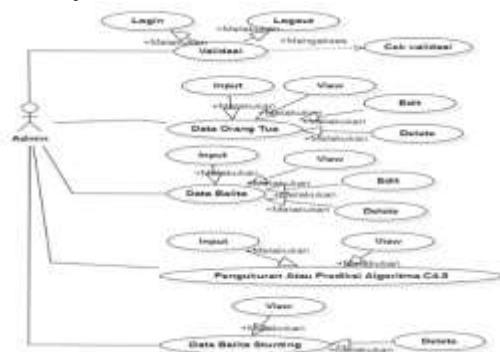
Gambar 5. Aturan yang terbentuk

Pada tahapan *deployment* merupakan tahapan akhir dari proses *crisp-dm*. Peneliti membangun suatu sistem untuk mengukur nilai *support* dan *confidence* pada data transaksi penjualan sehingga diharapkan dapat meningkatkan omset penjualan.

a. Use Case Diagram

Menerangkan adanya satu entitas yang bertanggung jawab dalam sistem, yang dikenal sebagai Admin. Admin memiliki peran sentral dalam administrasi sistem, dengan tanggung jawab termasuk kemampuan untuk login ke dalam sistem

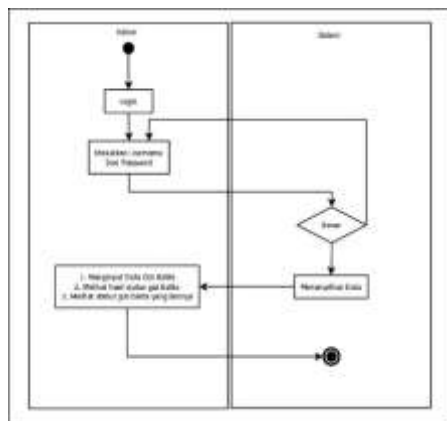
menggunakan kombinasi username dan password yang valid. Tugas-tugas admin mencakup manajemen data, peninjauan hasil perhitungan Stunting, dan pengelolaan informasi pengguna lainnya.



Gambar 6. Usecase Diagram

b. *Activity Diagram*

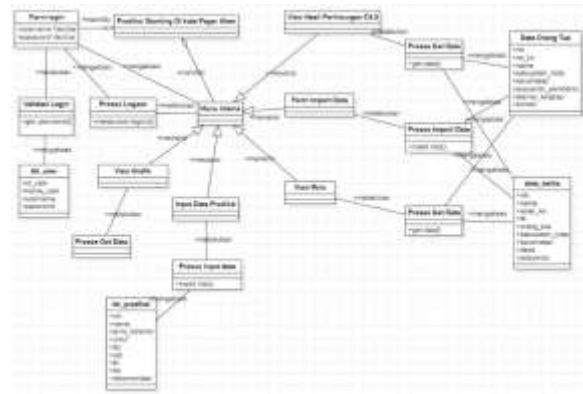
Ketika Admin pertama kali mengakses situs web, halaman utama akan ditampilkan bersama dengan formulir Login. supaya bisa menggunakan Sistem, Admin harus meng input Username dan password yang benar. Jika informasi yang dimasukkan oleh Admin tidak akurat, Sistem akan mengembalikan formulir Login. Tetapi, jika Login berhasil, menu utama akan terlihat, memberikan akses ke opsi seperti Data Balita, Informasi Orang Tua, proses Prediksi menggunakan Algoritma C4.5, dan Hasil Prediksi. Setelah Admin selesai menggunakan situs web, mereka dapat keluar dengan melakukan logout.



Gambar 7. Activity Diagram

Dari ilustrasi yang disajikan, dapat dijelaskan bahwa setelah berhasil membuka situs web, Admin akan diarahkan langsung ke halaman Login, dimana mereka diminta memasukkan Username dan password. Jika terjadi kegagalan saat Login karena kesalahan pada Username atau password, Admin akan tetap berada di halaman Login. Namun, jika informasi yang dimasukkan benar, proses Login akan berhasil, dan Admin dapat mengakses menu untuk Mengelola Data Kue, mengelola data transaksi, melakukan proses FP-

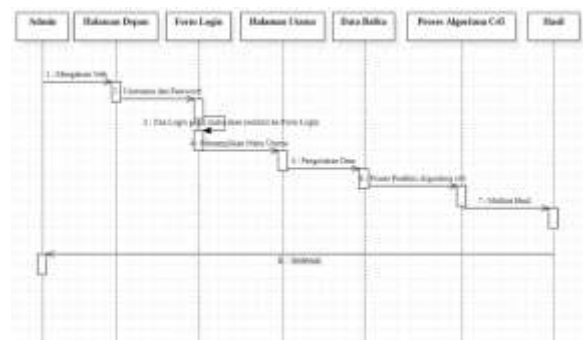
Growth, melihat hasil analisis, dan terakhir melakukan logout.



Gambar 8. Sequence Diagram

c. *Class diagram*

Digunakan untuk memvisualisasikan struktur kelas yang ada dalam sistem. Diagram kelas memberikan gambaran statis dan realistis tentang hubungan antara kelas-kelas tersebut. Dalam diagram kelas ini, terdapat beberapa kelas, seperti kelas User, kelas Orang Tua Balita, kelas Data Balita, kelas Data Balita Stunting, kelas Prediksi, dan kelas Hasil.



Gambar 9. Class Diagram

d. *Halaman Menu Utama* akan muncul sebagai tampilan awal saat Admin ingin melakukan Login ke dalam Sistem.



Gambar 10. Tampilan Depan



Gambar 11. Menu Dashborad

- e. *Halaman Menu Pengukuran* merupakan menu yang digunakan untuk melakukan pengukuran *Stunting*. *Admin* dapat memasukkan data dan melihat hasil dari pengukuran.



Gambar 12. Halaman Pengukuran/Prediksi

- f. *Halaman Menu Hasil Prediksi* merupakan menu yang berisikan semua data Pengukuran yang ada di dalam *Sistem*. *Admin* dapat melakukan edit, melakukan hapus dan melihat data.



Gambar 13. Halaman Hasil Prediksi

- g. *Pengujian Blackbox Testing*:

Pada sistem ini, pengujian dilakukan dengan menggunakan metode pengujian *black box* untuk menemukan dan memastikan bahwa sistem yang dihasilkan bebas dari kesalahan atau kegagalan dari keseluruhan sistem.

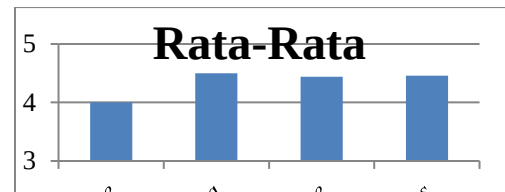
Berdasarkan hasil kuisisioner, pengujian *database* menghasilkan skor rata-rata 4.0, pengujian algoritma mendapatkan Skor rata-rata 4.5, kemudian pengujian *Interface* skor rata-rata 4.4 dan pengujian *fungsionalitas* skor rata-rata 4.5. Berikut ini tabel ringkasan kuisisioner melalui uji kotak hitam.

Tabel 1. Tabel Hasil Pengujian

Pengujian	Rata-Rata
-----------	-----------

Database	4.00
Algoritma	4.50
Interface	4.44
Fungsionalitas	4.46

Di bawah ini adalah bagan ringkasan kuisisioner melalui uji kotak hitam.



Gambar 14. Grafik Rata - Rata Koesioner

Berdasarkan hasil grafik diatas menunjukan bahwa Penerapan Metode asosiasi algoritma apriori memiliki nilai skor rata-rata 4.00 dan untu total skor mendapatkan total 4.35 sangat layak untuk diimplementasikan.

5. KESIMPULAN DAN SARAN

Berdasarkan penilaian terhadap desain dan uji coba yang dilakukan oleh penulis, penulis dapat menyimpulkan hasil dari "Implementasi Algoritma C4.5". Untuk Prediksi Anak Stunting Di Kota Pagar Alam" sebagai berikut: Penggunaan algoritma sangat dianjurkan untuk memprediksi status gizi Anak dengan menggunakan metode algoritma C4.5 di kota Pagar Alam. Karena mampu menghasilkan pohon keputusan dan efisien dalam menangani atribut diskrit serta numerik, algoritma ini menghasilkan akurasi yang dapat diterima.. sistem Prediksi status gizi Anak yang dirancang penulis layak karena dapat mengkategorikan status gizi balita. balita secara otomatis berdasarkan zscore yang ditetapkan dan hanya terdapat selisih 6% dari pengujian prediksi dataset yang sama menggunakan *Rapid Miner*. Sistem yang penulis rancang dapat lebih cepat dan efektif dalam mengklasifikasikan status gizi Anak. Berdasarkan Berdasarkan data hasil uji sistem, dapat disimpulkan bahwa akurasi Algoritma C.4.5 untuk memprediksi anak stunting yaitu 84,00% tergolong baik. Diharapkan bagi penelitian selanjutnya untuk menggunakan metode lain dalam menerapkan klasifikasi kasus status gizi Anak, selain itu juga diharapkan kedepannya dapat dikembangkan ke dalam versi aplikasi android/ ios.

DAFTAR PUSTAKA

- [1] R. A. Anggraini and A. J. Mulyana, "Implementasi Metode Prototyping Pada Rancang Sistem Informasi Kesehatan Gizi Berbasis Mobile di Kota Tasikmalaya," vol. 9, no. 1, pp. 9–18, 2022.
- [2] A. Prasetio, "Simulasi Penerapan Metode Decision Tree (C4.5) Pada Penentuan Status Gizi Balita," *J. Nas. Komputasi dan Teknol. Inf.*, vol. 4, no. 3, pp. 209–214, 2021, doi:

- 10.32672/jnkti.v4i3.2983.
- [3] M. Mahpuz, A. Muliawan Nur, and L. M. Samsu, "Penerapan Algoritma C4.5 Dalam Mengklasifikasi Status Gizi Balita Pada Posyandu Desa Dames Damai Kabupaten Lombok Timur," *Infotek J. Inform. dan Teknol.*, vol. 5, no. 1, pp. 72–81, 2022, doi: 10.29408/jit.v5i1.4414.
 - [4] J. Ilmiah and W. Pendidikan, "3 1,2,3," vol. 8, no. July, pp. 116–125, 2022.
 - [5] H. S. Tambunan, Z. A. Siregar, A. P. Windarto, and F. Rizki, "Penerapan Data Mining Klasifikasi Gizi Bayi Dengan Algoritma," vol. 1, no. 2, pp. 88–96, 2022.
 - [6] S. Mujiyono and N. Syaiful, "Prediksi Status Gizi Balita Menggunakan C4 . 5 Algoritma," vol. 8984, no. 1, pp. 41–52, 2023.
 - [7] F. Adzim *et al.*, "KLASIFIKASI STATUS STUNTING BALITA MENGGUNAKAN METODE C4 . 5 BERBASIS WEB," vol. 5, no. 3, pp. 515–525, 2023.
 - [8] D. Memberikan, P. Prima, P. Wisatawan, D. I. Desa, and W. Pentingsari, "Jurnal Administrasi Bisnis Terapan (JABT)," vol. 2, no. 2, 2020, doi: 10.7454/jabt.v2i2.1071.
 - [9] F. Revita, "Penerapan Penilaian Kinerja Karyawan Pada Politeknik Kutaraja," *J. Kreasi Rakyat*, vol. 1, no. 1, pp. 25–035, 2023, [Online]. Available: <https://jurnal.kreyatcenter.com/index.php/jkr/index>
 - [10] N. S. Mianti, S. Hartati, and M. Arafat, "Membuat Website Uptd Puskesmas Batumarta Ii Menggunakan Php & Mysql," *J. Inform. dan Komput.*, vol. 14, no. 1, pp. 82–88, 2023, [Online]. Available: <https://garuda.ristekbrin.go.id/documents/detail/1064467>
 - [11] A. Parkhan and M. Sugarindra, "... Kualitas Mekanis Kain Tenun Menggunakan Metode Vikor Optimal Design of Woven Fabric Mechanical Quality Using Vikor ...," *J. Disprotek*, vol. 13, no. 2, pp. 137–145, 2022, doi: 10.34001/jdpt.v12i2.
 - [12] C. Algoritma, U. Prediksi, and B. Siswa, "BERDASARKAN TIPOLOGI HIPPOCRATES-GALENUS," vol. 11, no. 1, pp. 1–8, 2023.
 - [13] A. Wahyudi and S. Mujilawati, "Implementasi Metode Time Series untuk Prediksi dan Monitoring Pendapatan Masla Delivery Berbasis Website," vol. 8, no. 2, pp. 1–6, 2023.
 - [14] R. Wirth and J. Hipp, "CRISP-DM: towards a standard process model for data mining. Proceedings of the Fourth International Conference on the Practical Application of Knowledge Discovery and Data Mining, 29-39," *Proc. Fourth Int. Conf. Pract. Appl. Knowl. Discov. Data Min.*, no. 24959, pp. 29–39, 2000, [Online]. Available: https://www.researchgate.net/publication/239585378_CRISP-DM_Towards_a_standard_process_model_for_data_mining
 - [15] H. Asyraf and M. E. Prasetya, "Implementasi Metode CRISP DM dan Algoritma Decision Tree Untuk Strategi Produksi Kerajinan Tangan pada UMKM A," vol. 8, pp. 94–105, 2024, doi: 10.30865/mib.v8i1.7050.
 - [16] W. Komala, R. Kurniawan, and Y. A. Wijaya, "ANALISIS DATA PENJUALAN PRODUK PAKAIAN MENGGUNAKAN K-MEANS DENGAN CLUSTER DISTANCE PERFORMANCE 2023). Pendekatan yang umum dilakukan dalam analisis cluster adalah dengan menggunakan algoritma K-Means (No , Trisnawati , and Purnamasari 2023). Pada penelitian ini menggunakan metode Cross Industry Standard Process for Data Mining (CRISP-DM). CRISP- DM merupakan standar dalam proses data mining yang digunakan untuk melakukan tahapan analisis industri sebagai dasar strategi pemecahan masalah bisnis atau suatu penelitian (Triyandana , Putri , and," vol. 8, no. 1, pp. 43–50, 2024.
 - [17] V. Issue, Y. Nababan, and I. Nugraha, "JUTIN : Jurnal Teknik Industri Terintegrasi Penerapan Data Mining Produksi Padi di Pulau Sumatera Menggunakan Analisis Regresi Linear," vol. 7, no. 1, 2024.
 - [18] A. Wulandari, "Analisis Dan Perancangan Sistem Informasi Pada Kantor Desa Bukit Rawi Berbasis Web," *Braz Dent J.*, vol. 33, no. 1, pp. 1–12, 2022.
 - [19] V. Issue, "JUTIN: Jurnal Teknik Industri Terintegrasi Penerapan Association Rule-Market Basket Analysis (AR-MBA) Dalam Menentukan Strategi Product Bundling: Studi Kasus Pada Minimarket AKPRIND," vol. 7, no. 1, pp. 379–386, 2024, doi: 10.31004/jutin.v7i1.24873.
 - [20] I. Swarsih, S. Sofian, J. T. Handoko, and M. S. Hasibuan, "Analisis Sentimen Tingkat Kepuasan Pelanggan Terhadap Layanan Kurir J & T Express Menggunakan Algoritma Support Vector Machine (SVM) Berdasarkan Ulasan Pengguna Di Google Playstore," pp. 92–98, 2023.
 - [21] F. Prasetyo and H. Hasugian, "Analisis Pola Pembelian Produk Makanan Menggunakan Algoritma FP-Growth Untuk Strategi Penjualan," *Idealis Indones. J. Inf. Syst.*, vol. 7, no. 1, pp. 11–20, 2023, [Online]. Available: <http://jom.fti.budiluhur.ac.id/index.php/IDEALIS/indexFachrulPrasetyo%7Chttp://jom.fti.budiluhur.ac.id/index.php/IDEALIS/index%7C>
 - [22] M. A. Rafi, R. D. Dana, and A. R. Dikananda, "PENGELOMPOKAN SKEMA MBKM BERBASIS ANALISIS CLUSTER KOMPETENSI MAHASISWA DI BIDANG

- PEMROGRAMAN MENGGUNAKAN METODE K-MEANS,” vol. 7, no. 1, pp. 464–469, 2023.
- [23] A. S. Puspaningrum, E. R. Susanto, and A. Sucipto, “Penerapan Metode Forward Chaining untuk Mendiagnosa Penyakit Tanaman Sawi,” *INFORMAL Informatics J.*, vol. 5, no. 3, p. 113, 2020, doi: 10.19184/isj.v5i3.20237.
- [24] M. I. S. A. Pontinak, “IMPLEMENTASI METODE DISKUSI DALAM PEMBELAJARAN DI KELAS X OLEH GURU PAI SMA KAPUAS PONTIANAK TAHUN PELAJARAN 2022 / 2023,” vol. 2, pp. 336–347, 2024.
- [25] S. Safitri and N. Nurhayati, “Studi Pustaka: Pengaruh Perhatian Orang Tua Terhadap Prestasi Belajar Siswa Di Sekolah,” *J. Educ. Rev. Res.*, vol. 1, no. 2, p. 64, 2018, doi: 10.26737/jerr.v1i2.1624.