

**PERBANDINGAN ALGORITMA C4.5 DAN ALGORITMA C5.0  
DALAM KLASIFIKASI STATUS GIZI BALITA *STUNTING* (Studi  
Kasus: Balita *Stunting* di Puskesmas IV Koto Mudiak)**

**SKRIPSI**

*Diajukan sebagai salah satu persyaratan memperoleh gelar*

*Sarjana Statistika*



**Oleh:**

**DHEA AFRILA HARELVI**

**NIM.19337044**

**PROGRAM STUDI SARJANA STATISTIKA  
DEPARTEMEN STATISTIKA  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS NEGERI PADANG  
2024**

## PENGESAHAN LULUS UJIAN SKRIPSI

Nama : Dhea Afrila Harelvi  
NIM : 19337044  
Program Studi : S1 Statistika  
Departemen : Statistika  
Fakultas : Matematika dan Ilmu Pengetahuan Alam

### PERBANDINGAN ALGORITMA C4.5 DAN ALGORITMA C5.0 DALAM KLASIFIKASI STATUS GIZI BALITA *STUNTING* (Studi Kasus: Balita *Stunting* di Puskesmas IV Koto Mudiak)

Dinyatakan lulus setelah dipertahankan di depan Tim Penguji Skripsi  
Departemen Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam  
Universitas Negeri Padang

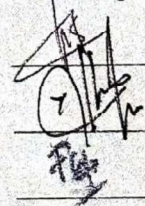
---

Padang, 21 Agustus 2024

#### Tim Penguji

	Nama
Ketua	: Admi Salma, S.Pd., M.Si
Anggota	: Dr. Yenni Kurniawati, S.Si., M.Si
Anggota	: Fadhilah Fitri, S.Si., M.Stat

#### Tanda Tangan





## PERSETUJUAN SKRIPSI

### PERBANDINGAN ALGORITMA C4.5 DAN ALGORITMA C5.0 DALAM KLASIFIKASI STATUS GIZI BALITA *STUNTING* (Studi Kasus: Balita *Stunting* di Puskesmas IV Koto Mudiak)

Nama : Dhea Afrila Harelvi  
NIM : 19337044  
Program Studi : S1 Statistika  
Departemen : Statistika  
Fakultas : Matematika dan Ilmu Pengetahuan Alam

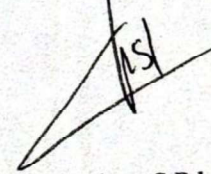
Padang, 21 Agustus 2024

Mengetahui:  
Ketua Departemen Statistika



Dr. Yenni Kurniawati, S.Si., M.Si  
NIP. 198402232010122005

Disetujui Oleh:  
Pembimbing



Admi Salma, S.Pd., M.Si  
NIDN.0025129003

## SURAT PERNYATAAN TIDAK PLAGIAT

Saya yang bertandatangan di bawah ini:

Nama : Dhea Afrila Harelvi  
NIM : 19337044  
Program Studi : S1 Statistika  
Departemen : Statistika  
Fakultas : Matematika dan Ilmu Pengetahuan Alam

Dengan ini menyatakan bahwa, skripsi saya dengan judul **“Perbandingan Algoritma C4.5 Dan Algoritma C5.0 Dalam Klasifikasi Status Gizi Balita Stunting (Studi Kasus: Balita Stunting di Puskesmas IV Koto Mudiak)”** adalah benar merupakan hasil karya saya dan bukan merupakan plagiat dari karya orang lain atau pengutipan dengan cara-cara yang tidak sesuai dengan etika yang berlaku dalam tradisi keilmuan.

Apabila suatu saat terbukti saya melakukan plagiat maka saya bersedia diproses dan menerima sanksi akademis maupun hukum sesuai dengan hukum dan ketentuan yang berlaku, baik di institusi UNP maupun di masyarakat dan negara.

Demikian pernyataan ini saya buat dengan penuh kesadaran rasa tanggung jawab sebagai anggota masyarakat ilmiah.

Diketahui oleh,  
Ketua Departemen Statistika,



Dr. Yenni Kurniawati, S.Si., M.Si  
NIP. 198402232010122005

Saya yang menyatakan,



Dhea Afrila Harelvi  
NIM. 19337044

**PERBANDINGAN ALGORITMA C4.5 DAN ALGORITMA C5.0  
DALAM KLASIFIKASI STATUS GIZI BALITA *STUNTING*  
(Studi Kasus : Balita *Stunting* di Puskesmas IV Koto Mudiak)**

**Dhea Afrila Harelvi**

**ABSTRAK**

Klasifikasi memiliki beberapa metode salah satunya yaitu metode *decision tree*. Salah satu algoritma yang dapat digunakan untuk pembentukan *decision tree* adalah C4.5 dan C5.0. Algoritma C4.5 memiliki kelebihan yaitu dapat menghasilkan pohon keputusan yang mudah untuk diinterpretasikan, memberikan nilai akurasi yang tinggi, dan efisiensi dalam menangani atribut bertipe kategorik numerik. Algoritma C5.0 merupakan pengembangan dari algoritma C4.5. Algoritma C5.0 hampir sama dengan algoritma C4.5, namun algoritma C5.0 memiliki langkah kerja yang lebih sedikit dari algoritma C4.5. Meskipun demikian, perbandingan antara kedua algoritma ini sering kali menghasilkan hasil yang tidak konsisten, maka pada penelitian ini akan dilakukan perbandingan algoritma C4.5 dengan algoritma C5.0 untuk melihat metode mana yang baik dan akurat dalam menangani kasus status gizi balita *stunting*.

Penelitian ini merupakan penelitian terapan dengan menerapkan metode Perbandingan Algoritma C4.5 dan Algoritma C5.0 dalam klasifikasi status gizi balita *stunting*. Data yang digunakan dalam penelitian ini adalah data sekunder. Data yang digunakan yaitu data status gizi balita di Puskesmas IV Koto Mudiak Kabupaten Pesisir Selatan Tahun 2022, dengan jumlah sebanyak 429 data.

Berdasarkan hasil penelitian yang dilakukan dari klasifikasi status gizi balita *stunting* menggunakan algoritma C4.5 dan C5.0 dengan metode evaluasi *k-fold cross-validation*, didapatkan bahwa algoritma C5.0 menghasilkan akurasi klasifikasi yang lebih tinggi dibandingkan algoritma C4.5. Algoritma C5.0 memberikan akurasi rata-rata sebesar 90%, sedangkan algoritma C4.5 memberikan akurasi rata-rata sebesar 83%. Dapat disimpulkan bahwa algoritma C5.0 lebih efektif dalam mengklasifikasikan status gizi balita *stunting*.

**Kata kunci :** Algoritma C4.5, Algoritma C5.0, Klasifikasi Status Gizi Balita *Stunting*.

**COMPARISON OF C4.5 AND C5.0 ALGORITHMS IN CLASSIFYING  
THE NUTRITIONAL STATUS OF STUNTED TODDLERS  
(A Case Study of : Stunted Toddlers at Puskesmas IV Koto Mudiak)**

**Dhea Afrila Harelvi**

**ABSTRACT**

Classification has several methods, one of which is the decision tree method. One of the algorithms that can be used for decision tree formation is C4.5 and C5.0. The C4.5 algorithm has the advantage of being able to produce decision trees that are easy to interpret, provide high accuracy values, and efficiency in handling numeric categorical type attributes. The C5.0 algorithm is a development of the C4.5 algorithm. The C5.0 algorithm is similar to the C4.5 algorithm, but the C5.0 algorithm has fewer work steps than the C4.5 algorithm. However, comparisons between these two algorithms often produce inconsistent results, so this study will compare the C4.5 algorithm with the C5.0 algorithm to see which method is good and accurate in handling cases of stunting toddler nutritional status.

This research is applied research by applying the C4.5 Algorithm Comparison method and the C5.0 Algorithm in the classification of nutritional status of stunting toddlers. The data used in this research is secondary data. The data used is toddler nutritional status data at Puskesmas IV Koto Mudik, South Coastal Regency in 2022, with a total of 429 data.

Based on the results of research conducted from the classification of nutritional status of stunting toddlers using the C4.5 and C5.0 algorithms with the k-fold cross-validation evaluation method, it is found that the C5.0 algorithm produces higher classification accuracy than the C4.5 algorithm. The C5.0 algorithm provides an average accuracy of 90%, while the C4.5 algorithm provides an average accuracy of 83%. It can be concluded that the C5.0 algorithm is more effective in classifying the nutritional status of stunted toddlers.

**Keywords:** Algorithm C4.5, Algorithm C5.0, Classification, Toddler Nutrition Status

## KATA PENGANTAR

Alhamdulillah, rasa syukur penulis ucapkan kehadiran Allah SWT atas berkat rahmat dan karunia-Nya, sehingga penulis dapat menyelesaikan Tugas Akhir dengan judul **“Perbandingan Algoritma C4.5 dan Algoritma C5.0 Dalam Klasifikasi Status Gizi Balita *Stunting* (Studi Kasus : Balita *Stunting* di Puskesmas IV Koto Mudiak”**. Penulisan Skripsi ini disusun untuk memenuhi salah satu syarat dalam menyelesaikan pendidikan pada Program Studi Statistika (S1) Departemen Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Padang. Dalam menyelesaikan Skripsi ini, tidak terlepas dari bimbingan dan bantuan dari berbagai pihak. Penulis menyadari bahwa skripsi ini memiliki keterbatasan sesuai dengan ilmu yang dimiliki, oleh sebab itu penulis menerima saran dan kritikan dari pembaca demi kesempurnaan isi skripsi ini penulis mengucapkan terima kasih kepada.

1. Ibu Admi Salma, S.Pd., M.Si., Dosen Pembimbing Skripsi dan Penasehat Akademik yang telah membimbing dan memberikan arahan selama masa perkuliahan dan penyusunan skripsi ini.
2. Ibu Dr. Yenni Kurniawati, S.Si., M.Si., Kepala Departemen Statistika, sekaligus Koordinator Program Studi S1 Statistika dan Dosen Penguji Skripsi yang telah memberikan masukan serta saran pada skripsi ini.
3. Ibu Fadhilah Fitri, S.Si., M.Stat., Dosen Penguji Skripsi yang telah memberikan masukan serta saran pada skripsi ini.
4. Bapak dan Ibu Staf Pengajar dan Tenaga Kependidikan Departemen Statistika Fakultas Matematika Ilmu Pengetahuan Alam Universitas Negeri Padang

yang telah membantu penulis selama menimba ilmu di Departemen Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Padang.

5. Teristimewa kepada kedua orang tua yang telah memberikan kasih sayang, semangat, nasehat, dukungan dan doa selama masa perkuliahan dan dalam menyelesaikan skripsi ini.
6. Semua teman-teman yang selalu memberikan semangat serta dukungan kepada penulis selama menempuh perkuliahan di Departemen Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Padang.

Semoga semua kebaikan dan ketulusan dibalas oleh Allah SWT sebagai amal ibadah. Penulis menyadari bahwa skripsi ini masih jauh dari kata sempurna dan masih terdapat kekurangan didalamnya. Oleh karena itu, penulis mengharapkan masukan dan saran yang membangun dari pembaca. Semoga penulisan skripsi ini dapat bermanfaat bagi seluruh pihak yang memerlukannya.

Padang, Juni 2024

Penulis

Dhea Afrila Harelvi



## DAFTAR ISI

ABSTRAK.....	i
KATA PENGANTAR .....	iii
DAFTAR ISI .....	v
DAFTAR TABEL .....	vii
DAFTAR GAMBAR .....	viii
DAFTAR LAMPIRAN .....	ix
BAB I PENDAHULUAN .....	1
A. Latar Belakang Masalah.....	1
B. Batasan Masalah.....	7
C. Rumusan Masalah .....	8
D. Tujuan Penelitian .....	8
E. Manfaat Penelitian .....	9
BAB II KAJIAN TEORI .....	10
A. <i>Data Mining</i> .....	10
B. Klasifikasi .....	11
C. <i>Decision Tree</i> (Pohon Keputusan) .....	12
D. Algoritma C4.5.....	12
E. Algoritma C5.0.....	16
F. <i>K-Fold Cross Validation</i> .....	18
G. Status Gizi Balita.....	21
BAB III METODOLOGI PENELITIAN .....	26
A. Jenis Penelitian.....	26
B. Jenis dan Sumber Data .....	26
C. Variabel Penelitian .....	26
D. Struktur Data .....	27
E. Teknik Analisis Data.....	27
BAB IV HASIL DAN PEMBAHASAN .....	30
A. Deskripsi Data.....	30
B. Hasil .....	34

C. Pembahasan.....	39
BAB V PENUTUP .....	42
A. Kesimpulan .....	42
B. Saran.....	43
DAFTAR PUSTAKA .....	44
LAMPIRAN.....	48

## DAFTAR TABEL

Tabel	Halaman
1. Variabel Penelitian .....	26
2. Struktur Data .....	27
3. Hasil Akurasi Algoritma C4.5 .....	34
4. Nilai <i>Gain Ratio</i> Algoritma C4.5 .....	36
5. Hasil Akurasi Algoritma C5.0 .....	37
6. Nilai <i>Gain Ratio</i> Algoritma C5.0.....	38

## DAFTAR GAMBAR

Gambar	Halaman
1. Grafik Prevelensi Masalah Status Gizi Balita di Sumatera Barat tahun 2022 .....	5
2. Ilustrasi Metode <i>K-Fold Cross Validation</i> .....	19
3. Bagian Alir Tahapan Analisis .....	29
4. Kasus Balita <i>Stunting</i> .....	30
5. Data Balita Stunting Berdasarkan Jenis Kelamin .....	31
6. Data Balita Stunting Berdasarkan Umur .....	32
7. Data Balita Stunting Berdasarkan Berat Badan .....	33
8. Data Balita Stunting Berdasarkan Tinggi Badan .....	34



## DAFTAR LAMPIRAN

Lampiran	Halaman
1. Data Status Gizi di Puskesmas IV Koto Mudiak.....	48
2. Sintaks Algoritma C4.5.....	49
3. Sintaks Algoritma C5.0 .....	50
4. Hasil Pohon Klasifikasi Algoritma C4.5 .....	51
5. Hasil Pohon Klasifikasi Algoritma C5.0 .....	52

# BAB I

## PENDAHULUAN

### A. Latar Belakang

*Data mining* merupakan proses pengumpulan informasi baru yang diambil dari kumpulan data besar yang membantu dalam pengambilan keputusan. *Data mining* adalah suatu proses analisis data yang menggunakan teknik matematika dan statistika, untuk mengidentifikasi pola atau pengetahuan yang bermanfaat dari kumpulan data besar. Tujuan utama dari *data mining* adalah untuk menggali wawasan yang tidak terlihat secara langsung atau informasi yang signifikan dari data yang kompleks. Munculnya *data mining* didasarkan pada kenyataan bahwa jumlah data yang tersimpan dalam basis data semakin besar (Prasetyo, 2012).

*Data mining* memiliki beberapa teknik analisis yaitu klasifikasi, asosiasi, klusterisasi, prediksi, dll. Klasifikasi adalah proses mengelompokkan objek atau data ke dalam kategori atau kelas tertentu berdasarkan ciri-ciri dari atribut yang dimiliki oleh data tersebut. Tujuan utama klasifikasi adalah membuat suatu model yang dapat memprediksi kelas atau kategori objek yang baru berdasarkan informasi yang telah dipelajari dari objek yang sudah diketahui sebelumnya (Kohavi, 1995).

Klasifikasi memiliki beberapa metode salah satunya yaitu metode *decision tree*. *Decision tree* adalah suatu model yang ditampilkan ke dalam bentuk pohon yang digunakan untuk membuat keputusan atau prediksi. Model ini memetakan berbagai kondisi atau pertanyaan (atribut) ke keputusan atau hasil akhir. Setiap "cabang" dalam pohon mewakili pilihan atau keputusan berdasarkan kondisi tertentu, dan "daun" di ujung cabang memberikan hasil atau prediksi. Banyak

algoritma yang dapat digunakan untuk pembentukan *decision tree* seperti ID3 (*Iterative Dichotomiser 3*), CART (*Classification and Regression Trees*), CHAID (*Chi-square Automatic Interaction Detection*), C4.5 dan C5.0 (North, 2012).

Algoritma C4.5 merupakan pengembangan dari algoritma ID3 yang diciptakan oleh J.Ross Quinlan pada tahun 1996. Perbedaan yang membedakan algoritma C4.5 dengan ID3 adalah C4.5 dapat menangani fitur dengan tipe numerik dan tipe kategorik, sedangkan ID3 hanya bisa menangani fitur dengan tipe kategorik. Algoritma C4.5 memiliki kelebihan yaitu dapat menghasilkan pohon keputusan yang mudah untuk diinterpretasikan, memberikan nilai akurasi yang tinggi, dan efisiensi dalam menangani atribut bertipe kategorik numerik (Prasetyo, 2014:65).

Algoritma C5.0 merupakan salah satu algoritma *data mining* yang khususnya diterapkan pada *decision tree*. Algoritma C5.0 merupakan pengembangan dari metode sebelumnya yaitu ID3 dan C4.5. Algoritma ini nantinya akan memilih atribut berdasarkan nilai *gain ratio* tertinggi. Atribut yang memiliki *gain ratio* tertinggi akan dipilih sebagai akar pada simpul selanjutnya (Ernawati, 2008). Algoritma C5.0 hampir sama dengan algoritma C4.5, namun algoritma C5.0 memiliki langkah kerja yang lebih sedikit dari algoritma C4.5. Menghasilkan pohon keputusan yang ringkas dan juga memberikan nilai akurasi tinggi dari prediksi (Pandya, 2015).

Penelitian oleh Putri dkk (2016) yang menyatakan bahwa algoritma C4.5 menghasilkan akurasi yang lebih tinggi, namun penelitian yang dilakukan oleh Dewi dkk (2019) menyatakan bahwa algoritma C5.0 memberikan nilai akurasi yang lebih tinggi. Karena hasil dari penelitian sebelumnya yang tidak konsisten, maka

pada penelitian ini akan dilakukan perbandingan algoritma C4.5 dengan algoritma C5.0 menggunakan metode *k-fold cross validation* untuk melihat metode mana yang baik dan akurat dalam menangani kasus status gizi balita *stunting*.

*Cross validation* adalah suatu metode evaluasi kinerja model statistik yang umum digunakan untuk mengukur seberapa baik model dapat memprediksi data yang belum pernah dilihat sebelumnya. *Cross validation* berfungsi membagi data menjadi dua bagian yaitu data *training* dan data *testing* secara berulang, sehingga setiap data memiliki kesempatan yang sama untuk dibagi kembali menjadi data *training* dan *testing*. Ada beberapa jenis dari *cross validation* diantaranya yaitu *K-fold cross validation*, *Leave-one-out-cross validation* (LOOCV), dan *Stratified k-fold cross validation*.

Jenis *cross validation* yang sering digunakan yaitu *k-fold cross validation*. *K-fold cross validation* adalah salah satu dari jenis pengujian *cross validation* yang berfungsi untuk menilai kinerja proses sebuah metode algoritma dengan membagi sampel data secara acak dan mengelompokkan data tersebut sebanyak nilai *K*. Pengujian menggunakan *k-fold Cross Validation* memiliki kelebihan dapat melihat model yang memiliki akurasi terbaik karena data secara random dibagi menjadi *K*-partisi sehingga dapat diketahui komposisi model terbaik (Rohani, 2018 ; Arisandi, 2022).

*K-fold cross validation* dapat digunakan untuk menilai kinerja proses algoritma klasifikasi. Klasifikasi banyak dimanfaatkan dalam bidang kesehatan untuk mengelompokkan informasi tentang penyakit. Salah satu permasalahan kesehatan yang dapat menggunakan klasifikasi yaitu *status gizi balita stunting*.



Kecukupan gizi merupakan salah satu faktor terpenting dalam mengembangkan kualitas sumber daya manusia, sebagai indikator keberhasilan pembangunan suatu bangsa. Dalam hal ini gizi ternyata sangat berpengaruh terhadap kecerdasan dan produktivitas kerja manusia (Almatsier, 2001)

Salah satu persoalan utama dalam pembangunan manusia adalah masalah gizi. Hal ini sejalan dengan pendapat Saputra (2013) yang menyatakan bahwa Indonesia sebagai salah satu negara dengan kompleksitas kependudukan yang sangat beraneka ragam, Indonesia dihadapkan oleh dinamika persoalan gizi.

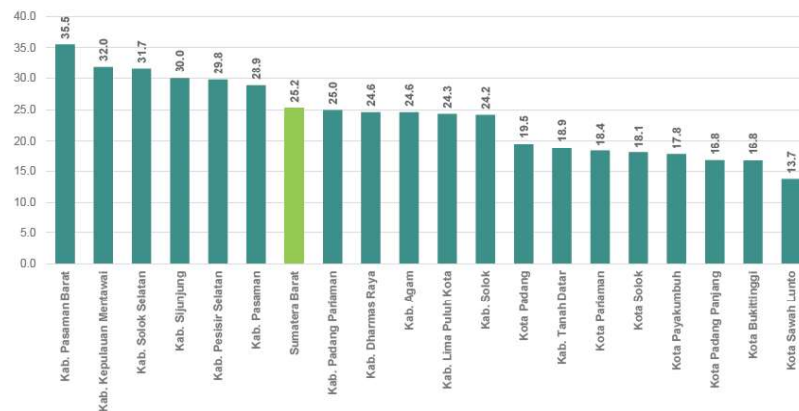
Status gizi merupakan keadaan yang diakibatkan oleh keseimbangan antara asupan zat gizi dari makanan dengan kebutuhan zat gizi yang diperlukan untuk metabolisme tubuh. Kelompok usia yang sering terkena masalah gizi adalah usia balita. Anak yang mengalami masalah gizi pada usia dini akan mengalami gangguan tumbuh kembang, mengalami kesakitan bahkan kematian (Proverawati, 2009).

Status gizi balita dapat dilihat dari berbagai kondisi kesehatan yang mencerminkan aspek-aspek nutrisi dan pertumbuhan anak-anak. Salah satu jenis status gizi balita yang umumnya dinilai adalah *Stunting*, menggambarkan kondisi di mana tinggi badan anak lebih pendek dari yang dianggap normal atau sehat untuk usianya. Hal ini sering kali menjadi indikator kekurangan gizi kronis pada masa pertumbuhan (Kemenkes, 2017).

Berdasarkan hasil studi status gizi indonesia (SSGI) tahun 2022 memberikan gambaran masalah status gizi balita yang paling tinggi yaitu *stunting*. Jumlah kasus *stunting* di Indonesia pada tahun 2022 mencapai angka 21,6%. Hasil SSGI di

Sumatera barat pada tahun 2022 kasus *stunting* mencapai angka 25,20%, dimana hasil ini mengalami kenaikan dari tahun 2021 yang sebelumnya hanya 23,30%.

Di Sumatera Barat masalah status gizi yang paling tinggi adalah *stunting* dibandingkan dengan masalah status gizi yang lain. Salah satu daerah di Sumatera Barat yang memiliki angka *stunting* tertinggi yaitu Kabupaten Pesisir Selatan. Grafik angka masalah status gizi balita *stunting* di Kabupaten Pesisir Selatan dapat dilihat pada Gambar 1.



Sumber : Kemenkes RI, 2022

**Gambar 1.** Grafik Presentase Masalah Status Gizi Balita di Sumatera Barat Tahun 2022

Berdasarkan Gambar 1 dapat diketahui bahwa Kabupaten Pesisir menjadi salah satu daerah yang mempunyai kasus *stunting* yang paling tinggi di Sumatera Barat. Angka presentase *stunting* di Kabupaten Pesisir Selatan yaitu 29,80%.

Kabupaten Pesisir Selatan merupakan salah satu Kabupaten dengan angka penduduk miskin paling tinggi di Sumatera Barat. Angka penduduk miskin yang tinggi merupakan salah satu faktor yang dapat menyebabkan tingginya angka *stunting*. Karena angka *stunting* yang tinggi, pemerintah Kabupaten Pesisir Selatan memiliki beberapa program kerja untuk menurunkan angka *stunting* salah satunya adalah Program Bapak/Bunda Asuh Stunting (BAAS). Program ini merupakan

salah satu program yang diusahakan oleh Pemda Pesisir Selatan untuk mengatasi *stunting*.

Dalam upaya mendukung program tersebut, klasifikasi status gizi balita *stunting* dengan metode *decision tree* menggunakan algoritma C4.5 dan algoritma C5.0 perlu dilakukan untuk memperoleh informasi balita *stunting* di Kabupaten Pesisir Selatan. Metode *decision tree* dipilih karena kemampuannya untuk menghasilkan model yang mudah dipahami dan diinterpretasikan, yang dapat membantu dalam mengidentifikasi variabel-variabel penting yang mempengaruhi status gizi balita. Menurut Adzim (2023) variabel yang dapat digunakan untuk mengukur status gizi balita *stunting* yaitu usia balita, jenis kelamin, berat badan, tinggi badan, dan status gizi balita menurut(TB/U).

Beberapa penelitian terdahulu yang dilakukan oleh Rahmadani, dkk (2022) membandingkan algoritma membandingkan algoritma *decision tree* C4.5 dan CART untuk mendapatkan hasil klasifikasi yang terbaik. Hasil terbaik didapatkan oleh algoritma C4.5 dengan akurasi sebesar 95,76%, sedangkan akurasi algoritma CART hanya 95,11%. Penelitian Rahmayuni (2014) tentang Perbandingan Performansi Algoritma C4.5 dan CART Dalam Klasifikasi Data Nilai Mahasiswa Prodi Teknik Komputer Politeknik Negeri Padang hasilnya algoritma C4.5 memberikan akurasi paling baik (85,61%), sedangkan algoritma CART memberikan hasil sedikit dibawahnya (84,95%).

Penelitian Yogi (2007) tentang Perbandingan Performansi Algoritma *Decision Tree* C5.0, CART, dan CHAID: Kasus Prediksi Status Resiko Kredit Di Bank X. Hasilnya adalah algoritma C5.0 memberikan rata-rata tingkat keakuratan sebesar

87,72%, CART 87,27%, dan CHAID 87,15%. Penelitian Hartianto (2022) tentang Optimasi Prediksi Kelulusan Mahasiswa dengan Menggunakan C5.0 dan Regresi Linear. Hasil penelitian menunjukkan bahwa penggunaan algoritma C5.0 menghasilkan performa yang lebih baik dalam memprediksi kelulusan mahasiswa dengan tingkat akurasi sebesar 86%, sedangkan regresi linear menghasilkan tingkat akurasi sebesar 75%.

Dari hasil penelitian terdahulu yang telah dijelaskan dapat dilihat dari perbandingan algoritma C4.5 dan C5.0 dengan algoritma lain, hasil yang diperoleh yaitu algoritma C4.5 dan algoritma C5.0 menghasilkan tingkat akurasi yang tinggi dibandingkan algoritma yang lain. Dalam hal ini akan dilakukan perbandingan algoritma C4.5 dengan algoritma C5.0 melihat algoritma manakah yang kinerja modelnya lebih baik dan memberikan hasil akurasi yang tinggi dengan menggunakan metode *k-fold cross validation* sebagai metode evaluasi kinerja model dalam menangani status gizi balita *stunting* sehingga didapatkan klasifikasi yang paling tepat dan akurat. Berdasarkan uraian diatas, maka dilakukanlah penelitian yang berjudul **“Perbandingan Algoritma C4.5 Dan Algoritma C5.0 dalam Klasifikasi Status Gizi Balita *Stunting* (Studi Kasus Balita *Stunting* di Puskesmas IV Koto Mudiak)”**.

## **B. Batasan Masalah**

Berdasarkan latar belakang diatas terdapat beberapa batasan masalah dalam penelitian ini yaitu :

1. Algoritma *decision tree* yang digunakan yaitu algoritma C4.5 dan algoritma C5.0



2. Variabel yang digunakan ada 4 atribut penjelas dan 1 atribut kelas target, yaitu Jenis Kelamin Balita, Usia Balita (Bulan), Berat Badan Balita (Kg), Tinggi Badan Balita (cm), dan Status Gizi Balita (TB/U).
3. Data set yang digunakan dalam penelitian ini adalah data balita yang ada di Puskesmas IV Koto Mudiak Kabupaten Pesisir Selatan.

### **C. Rumusan Masalah**

Berdasarkan latar belakang di atas, maka rumusan masalah pada penelitian ini adalah sebagai berikut :

1. Bagaimana hasil klasifikasi dari algoritma C4.5 pada kasus status gizi balita *stunting*?
2. Bagaimana hasil klasifikasi dari algoritma C5.0 pada kasus status gizi balita *stunting*?
3. Bagaimana perbandingan hasil akurasi algoritma C4.5 dan C5.0 dalam klasifikasi status gizi balita *stunting* dengan menggunakan *k-fold cross validation*?

### **D. Tujuan Penelitian.**

Berdasarkan rumusan masalah diatas maka tujuan penelitian ini adalah sebagai berikut :

1. Mengetahui hasil klasifikasi dari algoritma C4.5 pada kasus status gizi balita *stunting*.
2. Mengetahui hasil klasifikasi dari algoritma C5.0 pada kasus status gizi balita *stunting*.

3. Mengetahui hasil perbandingan algoritma C4.5 dan algoritma C5.0 pada klasifikasi status gizi balita *stunting* dengan menggunakan *k-fold cross validation*

#### **E. Manfaat Penelitian**

Adapun manfaat dari penelitian ini adalah :

1. Bagi peneliti

Hasil dari penelitian ini berguna untuk menambah wawasan pengetahuan tentang algoritma klasifikasi C4.5 dan algoritma C5.0.

2. Bagi pembaca

Penelitian ini diharapkan dapat menjadi pedoman serta referensi mengenai kondisi status gizi balita *stunting* dengan mengklasifikasi menggunakan algoritma C4.5 dan C5.0.

3. Bagi Puskesmas IV Koto Mudik

Hasil penelitian ini berupa klasifikasi status gizi balita, semoga bisa membantu dan dijadikan bahan pendukung pihak Puskesmas IV Koto Mudik dalam mengklasifikasikan status gizi balita *stunting*.