

**KLASIFIKASI KELUARGA BERISIKO STUNTING MENGGUNAKAN
ALGORITMA C4.5**

SKRIPSI

Oleh:
ABDURRAHMAN NAUFAL
NIM. 17650074



**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2024**

**KLASIFIKASI KELUARGA BERISIKO STUNTING MENGGUNAKAN
ALGORITMA C4.5**

SKRIPSI

Diajukan kepada:
Universitas Islam Negeri Maulana Malik Ibrahim Malang
Untuk memenuhi Salah Satu Persyaratan dalam
Memperoleh Gelar Sarjana Komputer (S.Kom)

Oleh:
ABDURRAHMAN NAUFAL
NIM. 17650074

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2024**

HALAMAN PERSETUJUAN


**KLASIFIKASI KELUARGA BERISIKO STUNTING MENGGUNAKAN
ALGORITMA C4.5**

SKRIPSI

**Oleh:
ABDURRAHMAN NAUFAL
NIM. 17650074**

Telah Diperiksa dan Disetujui untuk Diuji:
Tanggal: 21 Juni 2024

Pembimbing I,



Okta Oamaruddin Aziz, M.Kom
NIP. 19911019 201903 1 013

Pembimbing II,


Dr. Totok Cholidy M.Kom
NIP. 19691222 200604 1 001

Mengetahui,
Ketua Program Studi Teknik Informatika
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang




Dr. Hajar Kurniawan, M.MT, IPM
19771020 200912 1 001

HALAMAN PENGESAHAN

**KLASIFIKASI KELUARGA BERISIKO STUNTING MENGGUNAKAN
ALGORITMA C4.5**

SKRIPSI

Oleh:
ABDURRAHMAN NAUFAL
NIM. 17650074

Telah Dipertahankan di Depan Dewan Penguji Skripsi
dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan
Untuk Memperoleh Gelar Sarjana Komputer (S.Kom)
Tanggal: 21 Juni 2024

Susunan Dewan Penguji

Ketua Penguji : Prof. Dr. Suhartono, M.Kom
NIP. 19680519 200312 1 001

Anggota Penguji I : Johan Ericka Wahyu P. M. Kom
NIP. 19831213 201903 1 004


Anggota Penguji II : Okta Qomaruddin Aziz, M.Kom
NIP. 19911019 201903 1 013

Anggota Penguji III : Dr. Totok Chamidy, M.Kom
NIP. 19691222 200604 1 001

()
()
()
()

Mengetahui dan Mengesahkan,
Ketua Program Studi Teknik Informatika
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang




Kurniawan, M.MT, IPM
NIP. 19771020 200912 1 001

PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan di bawah ini:

Nama : Abdurrahman Naufal
NIM : 17650074
Fakultas / Program Studi : Sains dan Teknologi / Teknik Informatika
Judul Skripsi : Klasifikasi Keluarga Berisiko *Stunting*
Menggunakan Algoritma C4.5

Menyatakan dengan sebenarnya bahwa skripsi yang saya tulis ini benar-benar merupakan hasil karya saya sendiri, bukan merupakan pengambil alihan data, tulisan, atau pikiran orang lain yang saya akui sebagai hasil tulisan atau pikiran saya sendiri, kecuali dengan mencantumkan sumber cuplikan pada daftar pustaka.

Apabila dikemudian hari terbukti atau dapat dibuktikan skripsi ini merupakan hasil jiplakan, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Malang, 28 Juni 2024
Yang membuat pernyataan,

A 10,000 Rupiah Indonesian postage stamp is shown, featuring the Garuda Pancasila emblem and the text 'REPUBLIK INDONESIA' and 'MELUKAI TEMPEL'. The stamp number '10000' is visible on the left. A handwritten signature in black ink is written over the stamp.

Abdurrahman Naufal
NIM. 17650074

MOTTO

“Focus on what you can do, not on what you cannot change. Your energy is best spent on the actions within your control”

“Belajarlah untuk melepaskan kekhawatiran tentang hal-hal yang tidak dapat diubah, maka kamu akan menemukan kedamaian batin.”

HALAMAN PERSEMBAHAN

Saya persembahkan karya ini kepada:

Ayah saya,

Alm. M. Ali Murtadha

Yang telah mendukung dan menyemangati saya dalam mengerjakan skripsi ini

Ibu saya,

Titik Yuliati

Yang telah mendukung dan menyemangati saya dalam mengerjakan skripsi ini

Teman-teman seperjuangan,

Teknik Informatika Angkatan 2017

Semoga kita semua selalu diberi kemudahan oleh Allah SWT

KATA PENGANTAR

Assalamualaikum Warahmatullahi Wabarakatuh.

Segala puji hanya milik Allah Subhanahu Wa Ta'ala atas segala nikmat dan kasih sayang-Nya yang telah memudahkan penulis untuk menyelesaikan skripsi yang berjudul “Klasifikasi Keluarga Berisiko Stunting Menggunakan Algoritma C4.5”. Semoga shalawat dan salam senantiasa terlimpah kepada Nabi Muhammad SAW. Dan semoga kita semua mendapat syafaatnya di hari kiamat nanti, Aamiin.

Penulis mengucapkan rasa syukur dan terima kasih yang tak terhingga kepada semua pihak- yang selalu memberikan bantuan dan motivasi kepada penulis untuk dapat menyelesaikan skripsi ini. Ucapan ini penulis sampaikan kepada:

1. Prof. Dr. H. M. Zainuddin, M.A., selaku rektor Universitas Islam Negeri Maulana Malik Ibrahim Malang.
2. Prof. Dr. Hj. Sri Hariani, M.Si., selaku dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang.
3. Dr. Fachrul Kurniawan, M.MT., IPM, selaku Ketua Program Studi Teknik Informatika Universitas Islam Negeri Maulana Malik Ibrahim Malang.
4. Okta Qomaruddin Aziz, M.Kom selaku dosen pembimbing I dan Dr. Totok Chamidy M.Kom selaku dosen pembimbing II yang telah memberikan bantuan dan arahan kepada penulis, sehingga bisa menuntaskan skripsi ini.
5. Prof. Dr. Suhartono, S.Si., M.Kom selaku dosen penguji I dan Johan Ericka Wahyu P, M. Kom selaku dosen penguji II yang telah menguji serta

memberikan masukan sehingga penulis dapat menuntaskan skripsi dengan baik.

6. Segenap dosen, admin, staf laboran dan teman-teman seperjuangan mahasiswa Teknik Informatika angkatan 2017 yang telah memberikan banyak dukungan selama pengerjaan skripsi ini.
7. Dr. Abdul Manab, S.Pt., MP. selaku paman saya yang bersedia meluangkan waktunya untuk mendampingi penulis dalam proses penyusunan skripsi ini meskipun berada di luar lingkungan akademik universitas
8. Ayah, Ibu serta saudara saya yang selalu memberikan dukungan dan motivasi untuk terus berusaha, dan doa yang tak putus-putusnya selalu disampaikan agar dapat menuntaskan skripsi ini dengan lancar dan baik.

Akhir kata, penulis mengakui bahwa penulisan pada skripsi ini masih banyak kekurangan. Peneliti berharap semoga skripsi ini dapat bermanfaat bagi pembaca dan dapat dijadikan referensi demi pengembangan ke arah yang lebih baik. Kebenaran datangnya dari Allah dan kesalahan datangnya dari diri penulis. Semoga Allah SWT senantiasa melimpahkan Rahmat dan Ridho-Nya kepada kita semua.
Wassalamualaikum Warahmatullahi Wabarakatuh.

Malang, 28 Juni 2024

Penulis

DAFTAR ISI

HALAMAN PENGAJUAN.....	ii
HALAMAN PERSETUJUAN	iii
HALAMAN PENGESAHAN.....	iv
PERNYATAAN KEASLIAN TULISAN.....	v
MOTTO.....	vi
HALAMAN PERSEMBAHAN	vii
KATA PENGANTAR.....	viii
DAFTAR ISI	x
DAFTAR TABEL	xii
DAFTAR GAMBAR.....	xiii
ABSTRAK.....	xiv
ABSTRACT	xv
مستخلص البحث	xvi
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Batasan Masalah	4
1.3 Tujuan Penelitian	4
1.4 Manfaat Penelitian.....	4
BAB II KAJIAN PUSTAKA.....	6
2.1 Penelitian Terkait	6
2.2 Keluarga Berisiko Stunting.....	8
2.3 Algoritma C4.5.....	9
BAB III METODE PENELITIAN	11
3.1 Desain Penelitian.....	11
3.2 Tempat dan Waktu Penelitian	12
3.3 Pengumpulan Sumber Data.....	12
3.4 <i>Pre-processing Data</i>	15
3.5 <i>Split Data</i>	15
3.6 Algoritma Klasifikasi <i>C4.5</i>	16
3.6.1 Perhitungan entropy	17
3.6.2 Perhitungan gain	17
3.6.3 Pembentukan Pohon Keputusan.....	18
3.7 Evaluasi	22
3.8 Skenario Pengujian.....	25
BAB IV UJI COBA DAN PEMBAHASAN.....	28
4.1 Pengujian	28
4.2 Implementasi.....	28
4.2.1 Input Data	29
4.2.2 Normalisasi Dataset.....	29
4.2.3 Pembagian Data	30
4.2.4 Modeling Decision Tree C4.5.....	30
4.3 Hasil Uji Coba.....	31
4.3.1 Hasil Uji Coba Skenario 1	31
4.3.2 Hasil Uji Coba Skenario 2	33
4.3.3 Hasil Uji Coba Skenario 3	34

4.3.4 Hasil Uji Coba Skenario 4	35
4.3.5 Hasil Uji Coba Skenario 5	37
4.4 Pembahasan	38
4.5 Integrasi AI -Quran dan Hadits	42
BAB V KESIMPULAN DAN SARAN	43
5.1 Kesimpulan	43
5.2 Saran.....	44
DAFTAR PUSTAKA

DAFTAR TABEL

Tabel 2.1 Penelitian Terkait	7
Tabel 3.1 Atribut <i>Dataset</i>	13
Tabel 3.2 Contoh Data Rekapitulasi Keluarga Berisiko <i>Stunting</i>	14
Tabel 3.3 Hasil Perhitungan Nilai Entropy Dan Gain Dari Tabel 3.2	19
Tabel 3.4 Hasil Perhitungan Nilai Entropy Dan Gain Dari Node Atribut Air Minum dengan Keterangan “Ya”	22
Tabel 3.5 Confusion Matrix	24
Tabel 3.6 Confusion Matrix Data Latih	25
Tabel 3.7 Skenario Pembagian <i>Dataset</i>	26
Tabel 3.8 Confusion Matrix Skenario 1	27
Tabel 3.9 Hasil Performa Scenario Uji Coba 1	28
Tabel 4.1 Skenario Uji Coba Pembagian Data Latih dan Data Uji	29
Tabel 4.2 Perbandingan Hasil Data Latih dengan Model C4.5 Skenario 1	32
Tabel 4.3 Confusion Matrix Skenario 1	33
Tabel 4.4 Perbandingan Hasil Data Latih dengan Model C4.5 Skenario 2	32
Tabel 4.5 Confusion Matrix Skenario 2.....	34
Tabel 4.6 Perbandingan Hasil Data Latih dengan Model C4.5 Skenario 3	34
Tabel 4.7 Confusion Matrix Skenario 3.....	36
Tabel 4.8 Perbandingan Hasil Data Latih dengan Model C4.5 Skenario 4	37
Tabel 4.9 Confusion Matrix Skenario 4.....	37
Tabel 4.10 Perbandingan Hasil Data Latih dengan Model C4.5 Skenario 5	38
Tabel 4.11 Confusion Matrix Skenario 5.....	39
Tabel 4.12 Hasil Pengujian Skenario 1-5.....	39
Tabel 4.13 Akurasi Hasil Pengujian Skenario 1–5 dengan Pengurangan Atribut.41	

DAFTAR GAMBAR

Gambar 3.1 Desain Sistem	11
Gambar 3.2 Diagram alir algoritma C4.5.....	11
Gambar 3.3 Root node (akar percabangan) decision tree	11
Gambar 3.3 Pembentukan decision tree yang telah selesai dari tabel 3.2	11
Gambar 4.1 Input data menggunakan library pandas	53
Gambar 4.2 Convert tipe data menjadi integer.....	54
Gambar 4.3 Pembagian data train test split	54
Gambar 4.4 Modeling decision tree menggunakan DecisionTreeClassifier	55
Gambar 4.5 Hasil pembentukan decision tree C4.5 pada skenario 1	53
Gambar 4.6 Hasil pembentukan decision tree C4.5 pada skenario 2	54
Gambar 4.7 Hasil pembentukan decision tree C4.5 pada skenario 3	54
Gambar 4.8 Hasil pembentukan decision tree C4.5 pada skenario 4	55
Gambar 4.9 Hasil pembentukan decision tree C4.5 pada skenario 5	53

ABSTRAK

Naufal, Abdurrahman. 2024. **Klasifikasi Keluarga Berisiko Stunting Menggunakan Algoritma C4.5**. Skripsi. Program Studi Teknik Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang. Pembimbing: (I) Okta Qomaruddin Aziz, M. Kom, (II) Dr. Totok Chamidy, M.Kom.

Kata kunci: *Stunting, C4.5, Data Mining, Classification*

Stunting adalah kondisi di mana anak mengalami pertumbuhan yang terhambat, terutama dalam hal tinggi badan, dibandingkan dengan usianya. Kondisi ini biasanya disebabkan oleh malnutrisi kronis dan infeksi yang terjadi dalam jangka waktu yang Panjang. Penanganan stunting memerlukan pendekatan yang holistik, termasuk perbaikan gizi, peningkatan akses terhadap layanan kesehatan, edukasi kepada orang tua, serta perbaikan sanitasi dan kondisi lingkungan. Penelitian ini bertujuan untuk mengklasifikasikan keluarga yang berisiko stunting atau tidak. Untuk mengatasi permasalahan tersebut, pada penelitian ini akan dirancang sistem klasifikasi keluarga berisiko stunting berdasarkan pemutakhiran data keluarga berisiko stunting menggunakan metode C4.5, dengan tujuan untuk mengetahui performa metode C4.5 dalam mengklasifikasikan keluarga yang berisiko stunting atau tidak. Data merupakan faktor yang sangat penting dalam pencegahan balita stunting, dengan mengklasifikasi data yang telah ada akan dapat dengan mudah diketahui keluarga mana yang berpotensi memiliki balita stunting. Data yang di uji coba sebanyak 5912 data dengan 10 atribut yang diperoleh dari Dinas Kesehatan Kecamatan Poncokusumo, Kabupaten Malang. Uji coba dilakukan dengan pembagian data uji dan data latih sehingga nantinya akan terdapat 5 skenario pengujian dengan data uji dan data latih yang berbeda. Hasil uji coba menunjukkan bahwa dari 5 skenario yang telah diuji, hasilnya adalah accuracy rata – rata sebesar 98,404%. precision rata – rata sebesar 98,408%, dan recall rata – rata sebesar 99,382%, dengan atribut ”PUS terlalu tua” berpengaruh paling signifikan dalam faktor penyebab keluarga berisiko stunting dengan nilai atribut sebesar 13% .

ABSTRACT

Naufal, Abdurrahman. 2024. **Classification of Families at Risk of Stunting Using the C4.5 Algorithm**. Thesis. Department of Informatics Engineering, Faculty of Science and Technology, Maulana Malik Ibrahim State Islamic University Malang. Supervisors: (I) Okta Qomaruddin Aziz, M. Kom, (II) Dr. Totok Chamidy, M.Kom. M.Kom.

Keywords: *Stunting, C4.5, Data Mining, Classification*

Stunting is a condition where a child experiences impaired growth, particularly in height, compared to their age. This condition is usually caused by chronic malnutrition and infections that occur over a long period. Addressing stunting requires a holistic approach, including nutritional improvements, increased access to healthcare services, education for parents, and improvements in sanitation and environmental conditions. This research aims to classify families at risk of stunting or not. To address this issue, this study will design a classification system for families at risk of stunting based on updated data using the C4.5 method, with the goal of assessing the performance of the C4.5 method in classifying families at risk of stunting. Data is a crucial factor in preventing stunting in toddlers, and by classifying existing data, it will be easier to identify which families have the potential to have stunted toddlers. The data tested consists of 5,912 entries with 10 attributes obtained from the Poncokusumo Health Office in Malang Regency. The trials were conducted by dividing the test data and training data, resulting in 5 testing scenarios with different test and training data. The trial results show that out of the 5 tested scenarios, the average accuracy is 98.404%, the average precision is 98.408%, and the average recall is 99.382%, with the attribute "couples of childbearing age are too old" being the most significant factor influencing the risk of stunting, with an attribute value of 13%.

مستخلص البحث

فياندي، أليف راديفان. ٢٠٢٤. تصنيف الأسر المعرضة لخطر التقزم باستخدام خوارزمية C4.5. رسالة بكالوريوس. برنامج دراسات هندسة الحاسوب، كلية العلوم والتكنولوجيا، جامعة مولانا مالك إبراهيم الإسلامية، مالانج. المشرفون: (١) أوكتا قمر الدين عزيز، ماجستير في علوم الكمبيوتر، (٢) الدكتور توتوك تشاميدي، ماجستير في علوم الكمبيوتر. ، تنقيب البيانات، التصنيف C4.5 الكلمات الرئيسية: التقزم،

التقزم هو حالة يعاني فيها الطفل من نمو متعثر، وخاصة في الطول، مقارنة بعمره. عادةً ما يكون سبب هذه الحالة هو سوء التغذية المزمن والالتهابات التي تحدث على مدار فترة زمنية طويلة. يتطلب التعامل مع التقزم نهجًا شاملاً، بما في ذلك تحسين التغذية، وزيادة الوصول إلى الخدمات الصحية، وثقيف الوالدين، وكذلك تحسين الصرف الصحي وظروف البيئة. تهدف هذه الدراسة إلى تصنيف العائلات المعرضة لخطر التقزم أو غير المعرضة له. للتعامل مع هذه المشكلة، ستقوم هذه الدراسة بتصميم نظام لتصنيف العائلات المعرضة لخطر التقزم بناءً على تحديث بيانات العائلات المعرضة لخطر التقزم باستخدام طريقة C4.5، بهدف معرفة أداء طريقة C4.5 في تصنيف العائلات المعرضة لخطر التقزم أو غير المعرضة له. تعتبر البيانات عاملاً بالغ الأهمية في الوقاية من التقزم عند الأطفال، ومن خلال تصنيف البيانات الموجودة، يمكن بسهولة معرفة أي العائلات لديها احتمال أن يكون لديها أطفال متقزمون. تم اختبار البيانات التي يبلغ عددها ٥٩١٢ بيانات مع ١٠ سمات تم الحصول عليها من دائرة الصحة في منطقة بونكوكوسومو، محافظة مالانج. تم إجراء الاختبارات من خلال تقسيم البيانات الاختبارية وبيانات التدريب، بحيث ستكون هناك ٥ سيناريوهات اختبار بيانات اختبار وبيانات تدريب مختلفة. أظهرت نتائج الاختبار أنه من بين ٥ سيناريوهات التي تم اختبارها، كانت النتائج هي متوسط دقة بنسبة ٩٨,٤٠٤٪، ومتوسط دقة بنسبة ٩٨,٤٠٨٪، ومتوسط استرجاع بنسبة ٩٩,٣٨٢٪، مع كون السمة "النساء المسنات في سن الإنجاب" الأكثر تأثيرًا بشكل كبير كعامل مؤثر في خطر العائلة بالتقزم بنسبة ١٣٪.

BAB I

PENDAHULUAN

1.1 Latar Belakang

Stunting merupakan salah satu masalah kesehatan serius yang dihadapi oleh banyak negara berkembang, termasuk Indonesia. Menurut (WHO, 2019), stunting adalah kondisi di mana pertumbuhan dan perkembangan anak terganggu akibat kekurangan gizi kronis dan infeksi berulang. Stunting ditandai oleh tinggi badan yang lebih rendah dibandingkan dengan standar tinggi badan rata-rata untuk usia tertentu, yaitu lebih dari dua standar deviasi di bawah median pada kurva pertumbuhan WHO. Berdasarkan data dari Badan Kesehatan Dunia (WHO), sekitar 21,3% anak-anak di bawah usia lima tahun di seluruh dunia mengalami stunting pada tahun 2019.

Keluarga berperan penting dalam pencegahan stunting, terutama melalui pemberian asupan gizi yang cukup dan pola asuh yang tepat. Namun, keluarga dengan risiko tinggi stunting, seperti keluarga dengan kondisi sosial ekonomi rendah, pengetahuan gizi yang kurang, dan akses terbatas terhadap layanan kesehatan, membutuhkan perhatian khusus. Berbagai upaya telah dilakukan untuk mengatasi masalah ini, mulai dari intervensi gizi, edukasi, hingga perbaikan akses layanan kesehatan. (Chowdhury dkk. 2016).

Orang tua mempunyai peran yang paling penting dalam pencegahan keluarga berisiko stunting dengan menjaga kesehatan anak dengan memberikan ASI hingga 2 tahun sebagai ibu dan menanggung nafkah untuk kebutuhan pokok keluarga sebagai ayah sebagaimana yang tertera dalam Q.S. Al-Baqarah ayat: 233.

وَالْوَالِدَاتُ يُرْضِعْنَ أَوْلَادَهُنَّ حَوْلَيْنِ كَامِلَيْنِ لِمَنْ أَرَادَ أَنْ يُنَمِّمَ الرِّضَاعَةَ وَعَلَى الْمَوْلُودِ لَهُ رِزْقُهُنَّ وَكِسْوَتُهُنَّ بِالْمَعْرُوفِ
 لَا تُكَلِّفُ نَفْسٌ إِلَّا وُسْعَهَا لَا تُضَارَّ وَالِدَةٌ بِوَلَدِهَا وَلَا مَوْلُودٌ لَهُ بِوَالِدِهِ وَعَلَى الْوَارِثِ مِثْلُ ذَلِكَ فَإِنْ أَرَادَا فِصَالًا
 عَنْ تَرَاضٍ مِنْهُمَا وَتَشَاوُرٍ فَلَا جُنَاحَ عَلَيْهِمَا وَإِنْ أَرَدْتُمْ أَنْ تَسْتَرْضِعُوا أَوْلَادَكُمْ فَلَا جُنَاحَ عَلَيْكُمْ إِذَا سَلَّمْتُمْ مَا آتَيْتُمْ
 بِالْمَعْرُوفِ وَاتَّقُوا اللَّهَ وَاعْلَمُوا أَنَّ اللَّهَ بِمَا تَعْمَلُونَ بَصِيرٌ ﴿٢٣٣﴾

“Dan ibu-ibu hendaklah menyusui anak-anaknya selama dua tahun penuh, bagi yang ingin menyusui secara sempurna. Dan kewajiban ayah menanggung nafkah dan pakaian mereka dengan cara yang patut. Seseorang tidak dibebani lebih dari kesanggupannya. Janganlah seorang ibu menderita karena anaknya dan jangan pula seorang ayah (menderita) karena anaknya. Ahli waris pun (berkewajiban) seperti itu pula. Apabila keduanya ingin menyapih dengan persetujuan dan permusyawaratan antara keduanya, maka tidak ada dosa atas keduanya. Dan jika kamu ingin menyusukan anakmu kepada orang lain, maka tidak ada dosa bagimu memberikan bayaran dengan cara yang patut. Bertakwalah kepada Allah dan ketahuilah bahwa Allah Maha Melihat apa yang kamu kerjakan.” (QS. Baqarah: 233).

Ayat ini menekankan pentingnya kolaborasi antara kedua orang tua dalam memberikan nutrisi terbaik bagi pertumbuhan anak. Sebagai ibu, disarankan untuk memahami aspek keseimbangan gizi dengan memperhatikan jenis makanan yang diberikan kepada anak, termasuk yang dianjurkan dan tidak, untuk memastikan anak mendapat nutrisi yang cukup. Islam mengajarkan pentingnya pemberian ASI eksklusif selama enam bulan pertama kehidupan anak, diikuti dengan pemberian makanan pendamping ASI yang memadai hingga dua tahun. Langkah ini sangat penting dalam mencegah stunting, sesuai dengan Al-Qur'an Surah Al-Baqarah ayat 233 yang merekomendasikan masa menyusui selama dua tahun untuk mendukung periode ini secara optimal.

Islam mengakui perlunya nutrisi yang baik pada tahap awal kehidupan untuk mendukung perkembangan fisik dan kognitif anak-anak. Dengan demikian, ayat

Al-Qur'an tersebut tidak secara langsung membahas pencegahan stunting, tetapi prinsip-prinsip yang terdapat di dalamnya tentang pemberian ASI yang cukup dan tepat waktu secara tidak langsung mendukung upaya pencegahan stunting pada anak-anak.

Data merupakan faktor yang sangat penting dalam pencegahan balita stunting, dengan mengklasifikasi data yang telah ada akan dapat dengan mudah diketahui keluarga mana yang berpotensi memiliki balita stunting. Dalam mengukur keberhasilan program stunting, saat ini terdapat 2 sumber data yaitu data survei status gizi indonesia (SSGI) dan data e-PPGBM yang merupakan data rutin hasil pengukuran antropometri Balita yang dikumpulkan melalui kegiatan posyandu (Achmad Sopian dan I Lina Widyastuti 2022).

Berdasarkan permasalahan yang disebutkan di atas, penting untuk diketahui bahwa mengabaikan faktor risiko keluarga stunting dapat mengakibatkan peningkatan jumlah keluarga yang berisiko melahirkan bayi dengan kondisi stunting. Dengan demikian, sangat penting menggunakan pendekatan yang dapat mengidentifikasi status keluarga apakah mereka berisiko memiliki stunting atau tidak.

Salah satu metode atau pendekatan yang dapat diterapkan yaitu metode klasifikasi *C4.5*. Klasifikasi adalah salah satu algoritma pada data mining yang mengelompokkan suatu data kedalam kategori tertentu dengan diantaranya adalah *C4.5*, Naive Bayes (NB), K-Nearest Neighbor (KNN), Learning Vector Quantization (LVQ), dll. Pada penelitian sebelumnya, algoritma *C4.5* diketahui melebihi algoritma Learning Vector Quantization (LVQ) dalam akurasi rata-rata

dan memiliki waktu proses yang lebih cepat daripada metode K-Nearest Neighbors (K-NN) dalam mengklasifikasikan kemampuan siswa (Fakhrurriqfi dan Wardoyo 2013). Studi lain yang menggunakan algoritma C4.5 juga menunjukkan kinerja yang baik untuk mengklasifikasikan nilai kelayakan calon debitur baru (Hermanto dan SN 2017) dan melebihi Naïve Bayes dalam memprediksi perilaku masa depan pelanggan (Sadikin dan Alfiandi 2018).

Dalam penelitian ini, metode C4.5 digunakan untuk mengklasifikasikan keluarga berisiko stunting berdasarkan data rekapitulasi keluarga. Akurasi pada hasil klasifikasi kemudian dievaluasi menggunakan tabel *confusion matrix* untuk menilai seberapa baik hasil klasifikasi selaras dengan data keluarga yang sebenarnya. Diharapkan hasil yang diperoleh bermanfaat bagi dunia kesehatan dan mampu untuk menurunkan prevalensi stunting pada balita.

1.2 Batasan Masalah

Batasan masalah dalam penelitian ini adalah sebagai berikut:

1. Data diperoleh dari Puskesmas Kec. Poncokusumo, Kab. Malang
2. Dataset yang dipakai adalah data rekapitulasi keluarga stunting pada tahun 2022

1.3 Tujuan Penelitian

Mengetahui performa algoritma C4.5 dalam mengklasifikasikan keluarga berisiko stunting berdasarkan data rekapitulasi keluarga.

1.4 Manfaat Penelitian

Manfaat dari penelitian ini adalah sebagai berikut:

1. Melakukan upaya promotive, preventif dan skrinning untuk menurunkan prevelansi stunting pada balita di Puskesmas Kecamatan Ponsokusumo
2. Penelitian ini diharapkan dapat menjadi sumber informasi mengenai faktor-faktor risiko stunting dalam keluarga bagi calon ibu atau ibu dengan balita

BAB II

KAJIAN PUSTAKA

2.1 Penelitian Terkait

(Nazir dkk. 2022) melakukan penelitian tentang Klasifikasi Status Gizi Balita Menggunakan C4.5 dan Particle Swarm Optimization berdasarkan indikator tinggi badan menurut umur (H/A) dan pengembangan aplikasi yang dapat memprediksi status gizi H/A balita dan memberikan informasi tentang stunting. Data yang digunakan adalah data Monitoring Status Nutrisi 2017 di Provinsi Riau yang berisi 3.961 data dengan 213 parameter. Ketepatan algoritma C4.5 dan PSO dalam mengklasifikasikan status gizi H/A balita adalah 94,49% dengan rasio 90:10 dan 4 parameter. Dari 19 parameter yang dipilih, terdapat 4 parameter yang dapat mencapai akurasi 94,49%, yaitu jenis kelamin, umur, tinggi badan, dan cara mengukur tinggi badan. Parameter-parameter tersebut merupakan yang paling penting dan berpengaruh dalam menentukan status gizi H/A balita, karena PSO berfungsi sebagai seleksi fitur untuk menghindari parameter-parameter yang tidak berhubungan dengan kinerja pengklasifikasi.

(Wijaya 2018) melakukan penelitian tentang Pemberian Asi Eksklusif Menggunakan Algoritma C4.5 & PSO". Alat yang digunakan adalah aplikasi RapidMiner dengan sampel data yang terdiri dari 250 responden. Pengujian dilakukan dengan membandingkan value dari 2 hingga 10. Kesimpulannya adalah C4.5 menunjukkan akurasi terbaik pada nilai validasi 6 dengan akurasi sebesar 61,61%. Sementara itu, algoritma C4.5 yang dikombinasikan dengan PSO

menghasilkan akurasi sebesar 69,20% pada nilai validasi 10. Dengan demikian, ada peningkatan akurasi sebesar 7,59% ketika menggunakan algoritma C4.5 + PSO.

(Prasetya dkk. 2020) melakukan penelitian tentang “Klasifikasi Stunting Balita Menggunakan KNN”, menyimpulkan bahwa metode K-Nearest Neighbor (KNN) efektif dalam mengklasifikasikan status stunting pada balita. Data yang digunakan adalah data antropometri dari balita di Desa Slangit dan mencapai tingkat akurasi yang tinggi. Hasil analisis menunjukkan bahwa metode KNN dapat menjadi alat yang andal dalam mendeteksi stunting, membantu dalam intervensi dini dan perencanaan program kesehatan yang lebih efektif.

Tabel 2.1 Penelitian Terkait

No.	Judul	Variabel	Metode	Hasil
1	Klasifikasi Status Gizi Balita	umur, tinggi badan, cara mengukur tinggi badan dan jenis kelamin	C4.5 & PSO	Akurasi algoritma sebesar 94,49% dengan rasio
2	Klasifikasi Tingkat Pengetahuan Ibu dalam Pemberian Asi Eksklusif	Umur, pendidikan, pekerjaan, tempat bersalin, dll	C4.5 & PSO	Akurasi algoritma sebesar 69.20%
2	Klasifikasi Status Stunting Balita Di Desa Slangit	Umur, berat badan, tinggi badan	KNN	akurasi algoritma sebesar 98,89% dengan status normal dan kurang.
3	Klasifikasi Keluarga Berisiko Stunting	PUS hamil atau tidak hamil, pendidikan, kondisi tempat tinggal, PUS 4T, dll	C4.5	

Pada tabel di atas dapat diketahui bahwa dalam penelitian 1, peneliti menggunakan metode C4.5 dan PSO untuk mengembangkan aplikasi yang dapat memprediksi status gizi H/A balita dan memberikan informasi tentang stunting.

Dengan menggunakan 4 variabel, akurasi yang didapatkan sebesar 94.59%. Dalam penelitian 2, peneliti melakukan perbandingan akurasi metode C4.5 dan C4.5 + PSO untuk klasifikasi tingkat pengetahuan Ibu dalam pemberian ASI eksklusif. Tingkat keakurasian yang didapatkan dengan metode C4.5 adalah 61.61% dan metode C4.5 + PSO adalah 69.20%. Dalam penelitian 3 peneliti melakukan penelitian tentang klasifikasi status stunting balita dengan aplikasi rapid miner, menghasilkan sampel data 300 balita dengan menggunakan rumus Indeks Massa tubuh (IMT). Kemudian dilakukan perhitungan dengan KNN Classification yang menghasilkan akurasi sebesar 98,89%. Untuk penelitian ini, peneliti melakukan penelitian tentang klasifikasi keluarga berisiko stunting dengan menggunakan metode C4.5

2.2 Keluarga Berisiko Stunting

Keluarga berisiko stunting adalah keluarga sasaran yang memiliki risiko untuk melahirkan anak stunting. Stunting merupakan salah satu masalah gizi kronis yang serius, terutama di negara-negara berkembang. Kondisi ini ditandai dengan tinggi badan yang lebih pendek dari standar usia anak dan dapat menyebabkan dampak jangka panjang terhadap kesehatan, perkembangan kognitif, dan produktivitas anak di masa depan (UNICEF, 2020).

Berbagai faktor risiko dapat meningkatkan kemungkinan stunting pada anak-anak. Faktor-faktor tersebut mencakup status ekonomi keluarga, pendidikan orang tua, akses terhadap layanan kesehatan, serta kondisi sanitasi dan lingkungan. Menurut penelitian oleh (Danaei dkk. 2016), status ekonomi rendah berhubungan langsung dengan tingginya prevalensi stunting. Keluarga yang hidup dalam

kemiskinan seringkali tidak mampu menyediakan makanan bergizi yang cukup untuk anak-anak mereka.

2.3 Algoritma C4.5

Algoritma C4.5 telah menjadi salah satu pendekatan yang populer dalam pembangunan model keputusan dan digunakan secara luas dalam berbagai aplikasi, termasuk di bidang kecerdasan buatan, pengenalan pola, dan analisis data. Algoritma C4.5 adalah sebuah algoritma pengklasifikasi yang dikembangkan oleh Ross Quinlan pada tahun 1993 sebagai peningkatan dari algoritma sebelumnya, ID3 (Iterative Dichotomiser 3). Algoritma ini digunakan untuk membangun pohon keputusan dalam proses pengklasifikasi data. (Quinlan, J. R. 2014)

Algoritma C4.5 menggunakan metode Gain Ratio untuk memilih atribut yang paling informatif pada setiap langkah dalam membangun pohon keputusan. Gain Ratio mengukur seberapa baik sebuah atribut dapat memisahkan dataset ke dalam kelas-kelas yang berbeda dengan mempertimbangkan besarnya informasi yang dimiliki oleh atribut tersebut. Setelah atribut terpilih, algoritma memisahkan dataset ke dalam subset berdasarkan nilai atribut tersebut. Proses ini diulang secara rekursif untuk membangun struktur pohon keputusan yang akhirnya digunakan untuk mengklasifikasikan data baru. Untuk menghindari overfitting, C4.5 melakukan pemangkasan pada pohon keputusan setelah proses pembangunan selesai. Berikut adalah rumus-rumus yang terkait dengan algoritma C4.5:

$$Entropy(S) = \sum_{i=1}^n - p_i * \log_2 p_i \quad (2.1)$$

Keterangan:

n = jumlah kelas

p_i = proporsi dari setiap kelas dalam himpunan data S

Entropi adalah ukuran ketidakpastian atau keacakan dalam dataset

$$Gain(S, A) = Entropy(S) - \sum_v \frac{|S_v|}{|S|} * Entropy(S_v) \quad (2.2)$$

Keterangan:

A = variable

v = nilai yang mungkin untuk variabel A

$|S_v|$ = jumlah sampel untuk nilai v

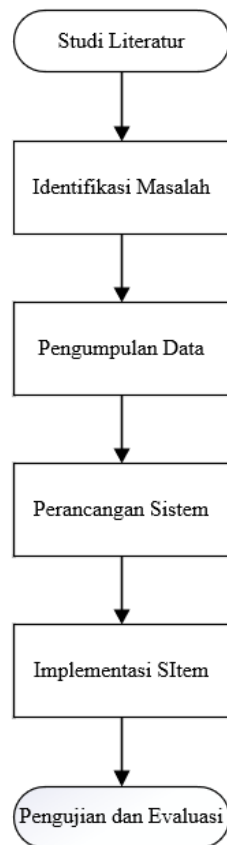
Information Gain digunakan untuk memilih atribut terbaik untuk dijadikan simpul pada pohon keputusan

BAB III

METODE PENELITIAN

3.1 Desain Penelitian

Desain penelitian adalah struktur yang digunakan untuk mengorganisir dan melaksanakan sebuah penelitian yang mencakup metode, prosedur, dan teknik yang digunakan dalam proses pengumpulan dan analisis data. Berikut adalah desain penelitian yang digunakan dalam penelitian ini



Gambar 3.1 Desain penelitian

Langkah pertama dalam melakukan penelitian adalah melakukan kajian literatur untuk memahami konteks relevansi masalah tersebut dan mengkaji

penelitian sebelumnya. Selanjutnya, mengidentifikasi masalah yang ingin dijawab dan melakukan kajian literatur untuk memahami konteks dan relevansi masalah tersebut. Kemudian, memastikan data yang dikumpulkan akurat dan relevan dan sistem mulai dirancang serta diimplementasikan. Setelah sistem berjalan, dilakukan pengujian dan evaluasi terhadap sistem tersebut, termasuk membahas pengujian sistem dan akurasi hasil prediksi. Setelah sistem berfungsi dan telah diuji, penelitian dapat dilanjutkan ke tahap berikutnya.

3.2 Tempat dan Waktu Penelitian

Penelitian dilakukan di Teknik Informatika, diharapkan bahwa penelitian ini dapat berjalan dengan baik berkat kerjasama dan fasilitas yang disediakan oleh UIN Malang. Penelitian ini diharapkan mampu berjalan secara baik karena kondisi universitas yang menunjang proses penelitian. Untuk waktu penelitian dimulai pada bulan November sampai selesai.

3.3 Pengumpulan Sumber Data

Pengumpulan sumber data dalam penelitian adalah proses di mana peneliti mengumpulkan semua sumber informasi yang akan digunakan untuk menjawab pertanyaan penelitian. Sumber data bisa berupa berbagai jenis, baik primer maupun sekunder, dan dikumpulkan melalui berbagai metode sesuai dengan desain penelitian. Data yang digunakan dalam penelitian ini adalah data rekapitulasi stunting di Kecamatan Poncokusumo. Proses pengumpulan sumber data yang efektif memerlukan perencanaan yang matang dan pemilihan metode yang sesuai dengan tujuan penelitian. Data yang dikumpulkan harus relevan, akurat, dan dapat

diandalkan untuk memastikan hasil penelitian yang valid dan dapat diinterpretasikan dengan baik.

Tabel 3.2 Atribut *Dataset*

No.	Nama Atribut	Deskripsi Atribut
1	Baduta	Keluarga sasaran memiliki bayi umur 0 – 23 bulan atau tidak
2	Balita	Keluarga sasaran memiliki bayi umur 24 – 59 bulan atau tidak
3	PUS	Keluarga sasaran tergolong Pasangan Usia Subur atau tidak
4	PUS Hamil	Keluarga sasaran tergolong Pasangan Usia Subur yang sedang hamil atau tidak
5	Air Minum	Keluarga sasaran memiliki sumber air minum yang layak atau tidak
6	Jamban	Keluarga sasaran memiliki jamban yang layak atau tidak
7	PUS Terlalu Muda	Pasangan Usia Subur yang berusia kurang dari 21 tahun
8	PUS Terlalu Tua	Pasangan Usia Subur yang berusia lebih dari 35 tahun
9	PUS terlalu Dekat	Pasangan Usia Subur yang masa kehamilan pertama dengan berikutnya kurang dari 2 tahun
10	PUS Terlalu Banyak	Pasangan Usia Subur yang memiliki anak lebih dari 2

Untuk contoh *dataset* yang digunakan bisa dilihat pada Tabel 3.2. berikut.

Tabel 3.2 Contoh data rekapitulasi keluarga berisiko stunting

	Baduta	Balita	PUS	PUS Hamil	Air Minum	Jamban	PUS Terlalu Muda	PUS Terlalu Tua	PUS Terlalu Dekat	PUS Terlalu Banyak	Berisiko Stunting
1	Ya	Ya	Ya	Tidak	Ya	Ya	Tidak	Tidak	Tidak	Tidak	Tidak
2	Tidak	Tidak	Tidak	Ya	Tidak	Tidak	Tidak	Tidak	Ya	Ya	Ya
3	Tidak	Tidak	Tidak	Ya	Tidak	Tidak	Ya	Tidak	Ya	Tidak	Ya
4	Tidak	Ya	Ya	Tidak	Ya	Tidak	Ya	Tidak	Tidak	Ya	Tidak
5	Ya	Ya	Tidak	Tidak	Ya	Tidak	Ya	Ya	Tidak	Tidak	Tidak
6	Tidak	Ya	Tidak	Tidak	Ya	Tidak	Ya	Tidak	Ya	Ya	Tidak
7	Tidak	Ya	Ya	Ya	Ya	Ya	Tidak	Tidak	Tidak	Ya	Ya
8	Ya	Tidak	Tidak	Tidak	Ya	Ya	Tidak	Tidak	Ya	Tidak	Ya
9	Tidak	Tidak	Tidak	Ya	Ya	Tidak	Tidak	Tidak	Tidak	Ya	Ya
10	Ya	Ya	Tidak	Ya	Ya	Tidak	Tidak	Tidak	Ya	Tidak	Tidak
11	Tidak	Tidak	Tidak	Ya	Ya	Ya	Ya	Ya	Tidak	Ya	Tidak
12	Ya	Ya	Tidak	Tidak	Tidak	Tidak	Tidak	Ya	Ya	Tidak	Ya
13	Tidak	Tidak	Ya	Tidak	Ya	Tidak	Ya	Tidak	Tidak	Ya	Tidak
14	Tidak	Tidak	Tidak	Tidak	Ya	Ya	Tidak	Tidak	Tidak	Tidak	Tidak
15	Ya	Ya	Ya	Ya	Ya	Tidak	Ya	Ya	Ya	Ya	Tidak
16	Ya	Ya	Ya	Tidak	Tidak	Tidak	Tidak	Ya	Tidak	Tidak	Ya
17	Tidak	Ya	Tidak	Ya	Tidak	Ya	Ya	Ya	Tidak	Ya	Ya
18	Tidak	Ya	Tidak	Tidak	Tidak	Tidak	Ya	Tidak	Ya	Tidak	Ya
19	Ya	Ya	Tidak	Tidak	Ya	Tidak	Ya	Tidak	Ya	Ya	Tidak
20	Tidak	Tidak	Tidak	Tidak	Tidak	Ya	Tidak	Ya	Tidak	Ya	Tidak

3.4 Pre-processing Data

Pre-processing - adalah langkah untuk mengatasi data kosong (missing value), duplikasi data dan data yang tidak konsisten. Pertama dilakukan penghapusan record yang mengandung data yang kosong sehingga data menjadi konsisten. Kemudian menghapus data yang duplikat dengan memeriksa kesamaan nilai pada parameter ID. Setelah melakukan kedua pengecekan tersebut, maka data akan dapat diproses ke dalam langkah selanjutnya seperti yang sudah ditunjukkan dalam diagram alir

3.5 Split Data

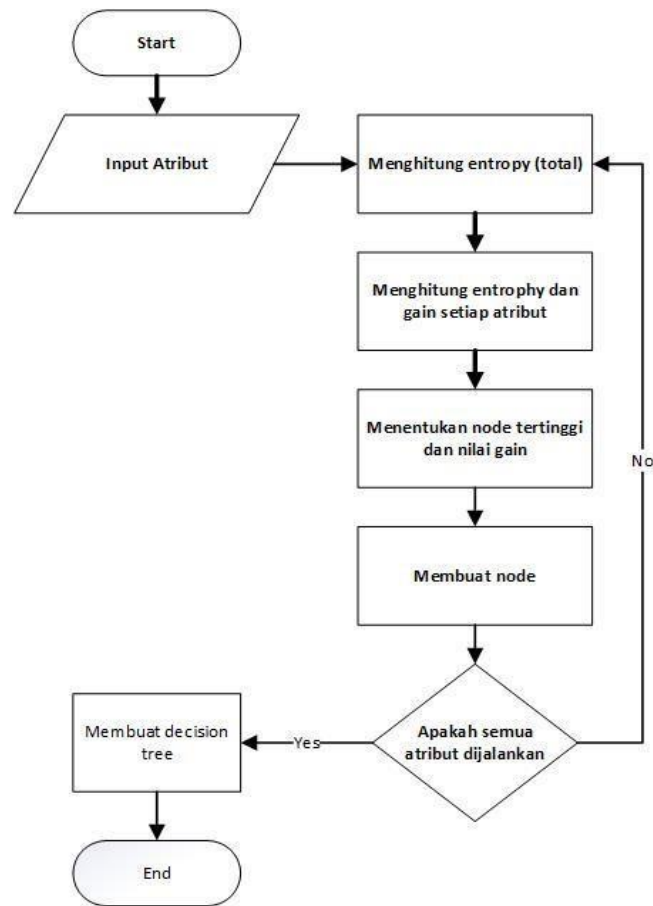
Salah satu langkah penting dalam proses pembuatan model machine learning adalah memisahkan data. Langkah ini melibatkan pembagian dataset menjadi dua bagian: data latih dan data uji. Data latih digunakan untuk melatih model machine learning, yaitu proses yang mengajarkan model mengenali pola atau hubungan antara fitur dan target dalam data. Data uji berfungsi untuk menguji kinerja model machine learning, yaitu proses yang mengukur seberapa baik model dapat memprediksi atau mengklasifikasi data baru yang belum pernah dilihat sebelumnya.

Split data perlu dilakukan agar model machine learning bisa beradaptasi dengan data yang beragam dan tidak terbatas. Jika model hanya dilatih dan diuji dengan data yang sama, maka model tersebut akan mengalami overfitting, yaitu kondisi di mana model terlalu spesifik pada data latih dan tidak dapat menggeneralisasi pada data baru. Overfitting dapat menyebabkan kinerja model yang buruk pada data uji, karena model tidak bisa menangkap variasi atau

ketidakpastian yang ada pada data uji. Oleh karena itu, pemisahan data membantu peneliti mencegah overfitting dan meningkatkan kinerja model.

3.6 Algoritma Klasifikasi C4.5

Algoritma C4.5 adalah salah satu algoritma yang paling terkenal untuk membangun pohon keputusan dari dataset yang diberikan. Berikut adalah diagram alir tahapan-tahapan utama dalam klasifikasi dengan algoritma C4.5:



Gambar 3.2 Diagram Alir Algoritma C4.5

Pada setiap node dalam pohon keputusan, algoritma C4.5 memilih atribut dari dataset yang paling efektif untuk memisahkan data menjadi kelas-kelas yang berbeda. Efektivitas ini diukur dengan menggunakan konsep information gain yang

berbasis pada entropi. Atribut yang memiliki information gain tertinggi dipilih sebagai atribut pembagi pada node tersebut (Zhang dkk. 2022).

3.6.1 Perhitungan entrophy

Entropy dalam algoritma C4.5 diimplementasikan sebagai alat untuk menilai seberapa tidak pastinya atau tingkat keacakan dalam suatu himpunan data. Fungsinya adalah untuk mengevaluasi seberapa seragam atau bervariasinya himpunan data dalam klasifikasi. Berikut adalah contoh perhitungan nilai entrophy atribut dataset menggunakan data dari table 3.2 di atas sebagai contoh perhitungannya.

$$S = \left(-\frac{b}{n} * \log_2 \left(\frac{b}{n} \right) \right) + \left(-\frac{a}{n} * \log_2 \left(\frac{a}{n} \right) \right) \quad 3.1$$

Keterangan:

- S : Entrophy
- a : Jumlah data dengan keterangan “ya”
- n : Jumlah keseluruhan
- b : Jumlah data dengan keterangan “tidak”
- d

$$\text{Entropy total} = \left(-\frac{11}{20} * \log_2 \left(\frac{11}{20} \right) \right) + \left(-\frac{9}{20} * \log_2 \left(\frac{9}{20} \right) \right) = 0,992774454$$

Terlihat dai perhitungan di atas hasil dari jumlah entrophy total dari table 3.2 adalah 0,992. Untuk langkah selanjutnya dilakukan perhitungan setiap entrophy dari semua atribut untuk nantinya digunakan untuk menghitung nilai gainnya.

3.6.2 Perhitungan gain

Gain merupakan parameter yang digunakan untuk mengevaluasi seberapa baik sebuah atribut dapat memisahkan data ke dalam kelas-kelas yang berbeda

dalam pembentukan pohon keputusan. Dalam algoritma C4.5, Gain Informasi dihitung dengan membandingkan entropy sebelum dan sesudah memisahkan dataset berdasarkan nilai atribut tertentu. Atribut yang memberikan Gain Informasi tertinggi dipilih sebagai atribut yang paling informatif untuk dijadikan simpul pada pohon keputusan. Berikut adalah contoh perhitungan nilai gain setelah didapatkan hasil dari perhitungan dari seluruh atribut dalam dataset table 3.2.

$$Gain = S - \left(\left(\frac{8}{n} * Sb \right) + \left(\frac{12}{n} * Sa \right) \right) \quad (3.2)$$

Keterangan:

- S : Entrophy total
- n : Jumlah keseluruhan data
- Sa : Entrophy data keterangan “ya”
- a : Jumlah data keterangan “ya”
- Sb : Entrophy data keterangan “tidak”
- b : Jumlah data keterangan “tidak”

Gain baduta :

$$0,992774454 - \left(\left(\frac{8}{20} * 1 \right) + \left(\frac{12}{20} * 0,954434003 \right) \right) = 0,011000853$$

Terlihat dai perhitungan di atas hasil dari perhitungan yang dilakukan, nilai gain dari atribut baduta adalah 0,110. Untuk langkah selanjutnya dilakukan perhitungan gain dari semua atribut untuk nantinya digunakan untuk menghitung membangun pohon keputusannya.

3.6.3 Pembentukan Pohon Keputusan

Setelah selesai melakukan perhitungan entropy dan gain dari dataset yang digunakan, langkah selanjutnya adalah membuat decision tree untuk menentukan sebuah keluarga apakah berisiko stunting atau tidak. Tahap pertama yang akan

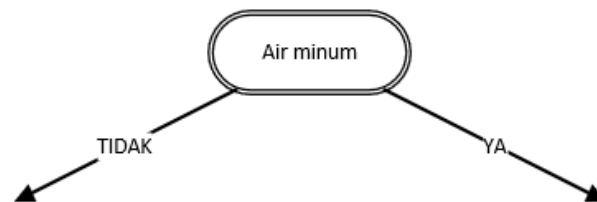
dilakukan adalah menentukan atribut yang akan digunakan sebagai akar (root) dengan menghitung nilai gain yang paling tinggi. Berikut adalah hasil perhitungan nilai entropy dan gain dari semua atribut yang terdapat pada tabel 3.2 di atas

Tabel 3.3 Hasil perhitungan nilai entropy dan gain dari tabel 3.1

Attribut	nilai	jumlah kasus	Berisiko stunting	Tidak berisiko	entropy	gain
Total		20	9	11	0,992774454	
Baduta						0,011000853
	Ya	8	3	5	0,954434003	
	Tidak	12	6	6	1	
Balita						0,0048532
	Ya	12	5	7	0,979868757	
	Tidak	8	4	4	1	
PUS						0,017285704
	Ya	6	2	4	0,918295834	
	Tidak	14	7	7	1	
PUS hamil						0,060023352
	Ya	8	5	3	0,954434003	
	Tidak	12	4	8	0,918295834	
Air minum						0,279111587
	Ya	13	3	10	0,779349837	
	Tidak	7	6	1	0,591672779	
Jamban						0,000721763
	Ya	7	3	4	0,985228136	
	Tidak	13	6	7	0,995727452	
PUS muda						0,066653707
	Ya	10	3	7	0,881290899	
	Tidak	10	6	4	0,970950594	
PUS tua						0,000721763
	Ya	7	3	4	0,985228136	
	Tidak	13	6	7	0,995727452	

PUS dekat						0,02667706
	Ya	9	5	4	0,99107606	
	Tidak	11	4	7	0,945660305	
PUS banyak						0,02667706
	Ya	11	4	7	0,945660305	
	Tidak	9	5	4	0,99107606	

Pada table diatas diketahui bahwa variabel air minum memiliki nilai gain tertinggi, sehingga node akar akan dilakukan dari data dengan atribut air minum yang mempunyai keterangan “Ya” masuk ke node kanan, sedangkan data dengan atribut air minum yang mempunyai keterangan “Tidak” masuk ke node kiri. Contoh pencabangannya dapat dilihat pada gambar berikut.



Gambar 3.3 Root node (akar percabangan) decision tree

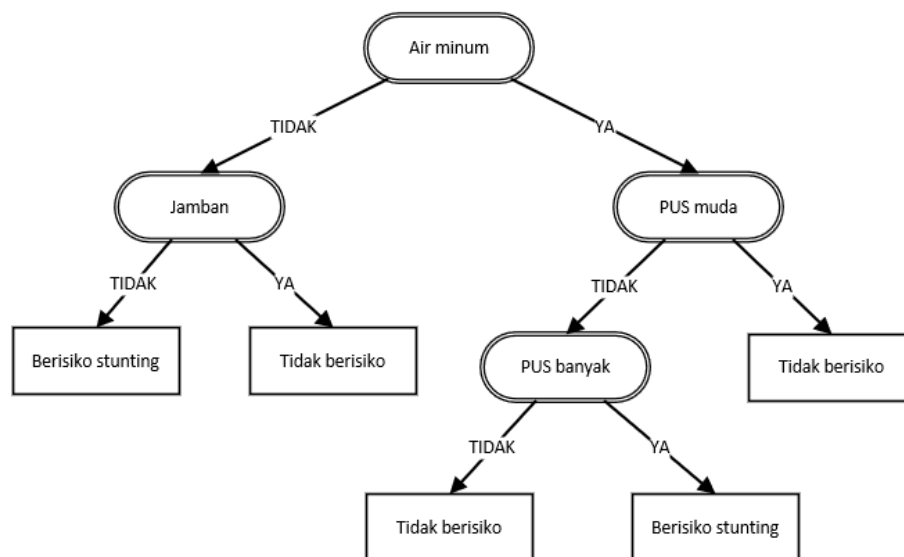
Setelah Setelah berhasil dilakukan node pertama maka dilakukan node berikutnya sampai menemukan nilai daun yang diinginkan. Dari hasil perhitungan gain di tabel 3.3, nilai gain terbesar akan dipilih untuk setiap atribut yang dihitung. Gain numerik yang terbesar akan dibandingkan dengan atribut kategorikal, dan nilai gain tertinggi dari semua atribut akan dijadikan node baru.

Tabel 3.4 Hasil perhitungan nilai entropy dan gain dari node atribut air minum dengan keterangan “Ya”

Attribut	nilai	jumlah kasus	yes	no	entropy	gain
Total		13	3	10	0,77935	
Baduta						0,495675528
	Ya	6	1	5	0,650022	
	Tidak	7	2	5	0,863121	
Balita						0,532611028
	Ya	8	1	7	0,543564	
	Tidak	5	2	3	0,970951	
PUS						0,48778118
	Ya	5	1	4	0,721928	
	Tidak	8	2	6	0,811278	
PUS hamil						0,532611028
	Ya	5	2	3	0,970951	
	Tidak	8	1	7	0,543564	
Air minum						0,48619706
	Ya	13	3	10	0,77935	
	Tidak	0	0	0	0	
Jamban						0,532611028
	Ya	5	2	3	0,970951	
	Tidak	8	1	7	0,543564	
PUS muda						0,692774454
	Ya	7	0	7	0	
	Tidak	6	3	3	1	
PUS tua						0,552129004
	Ya	3	0	3	0	
	Tidak	10	3	7	0,881291	
PUS dekat						0,48778118
	Ya	5	1	4	0,721928	
	Tidak	8	2	6	0,811278	
PUS banyak						0,48778118

	Ya	8	2	6	0,811278
	Tidak	5	1	4	0,721928

Dari tabel 3.4, dapat diketahui bahwa atribut PUS muda mempunyai nilai gain tertinggi, sehingga PUS muda menjadi node baru. Proses ini akan terus dilakukan sampai didapatkan hasil daun dengan ketentuan node mempunyai data noise terendah. Berikut adalah bentuk decision tree dari tabel 3.2 yang telah selesai dihitung setiap nodenya.



Gambar 3.4 Pembentukan decision tree yang telah selesai dari tabel 3.2

3.7 Evaluasi

Evaluasi hasil ini sesuai dengan penelitian yang dipakai sebagai acuan utama dari eksperimen (Okta Qomaruddin, 2020) Evaluasi dilakukan dengan menganalisa hasil identifikasi yang telah dilakukan oleh decision tree c4.5 dan dibandingkan dengan data asli untuk diukur keakurasian algoritma tersebut. Confusion matrix

digunakan untuk mengetahui tingkat akurasi dari model klasifikasi C4.5 yang telah dijalankan.

Confusion matrix atau error matrix adalah tabel yang dibuat untuk menunjukkan akurasi dari suatu algoritma yang terdiri dari kolom yang mewakili kelas data yang diprediksi dan baris yang mewakili kelas data asli atau sebaliknya (Machová dkk. 2023). Metode evaluasi dengan menggunakan matriks kebingungan menghasilkan nilai akurasi, presisi, dan recall. Akurasi mencerminkan persentase data yang diklasifikasikan dengan benar dari semua pengujian yang dilakukan, presisi adalah perbandingan kasus yang diprediksi positif yang benar-benar positif dalam data asli, dan recall adalah perbandingan kasus positif yang benar-benar diprediksi sebagai positif. (Powers dan Ailab, 2020.).

Tabel 3.5 Confusion matrix

		Predicted Class	
		Positive	Negative
Actual class	Positive	TP	FN
	Negative	FP	TN

Akurasi Setelah data uji dimasukkan ke dalam confusion matrix, nilai-nilai ini dihitung untuk mengevaluasi kinerja algoritma dengan menggunakan accuracy, precision, dan recall. Accuracy mengukur seberapa akurat keseluruhan model dengan membandingkan prediksi yang benar dengan total prediksi. Precision mengukur seberapa tepat prediksi positif yang sebenarnya positif. Precision memberikan gambaran seberapa baik model dalam mengklasifikasikan hasil positif. Recall mengukur seberapa banyak dari kelas positif yang berhasil diprediksi dengan benar oleh model.

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \quad (3.3)$$

$$Precision = \frac{TP}{TP + FP} \quad (3.4)$$

$$Recall = \frac{TP}{TP + FN} \quad (3.5)$$

Keterangan:

TP : jumlah true positive

FP : jumlah false positive

TN : jumlah true negative

FN : jumlah false negative

Berdasarkan penelitian yang telah dilakukan, didapatkan bahwa dari 20 data keluarga yang ada dengan model decision tree c4.5 diprediksi bahwa sebanyak 7 keluarga berisiko stunting dan 13 keluarga tidak berisiko stunting. Tetapi kenyatannya pada data latih ada 9 keluarga berisiko stunting dan 11 keluarga tidak berisiko stunting. Setelah diketahui hasil output dari data asli dan prediksi model, maka dibentuklah matrix confusion untuk mengukur keakurasian model perhitungan terhadap data asli.

Tabel 3.6 Confusion matrix data latih

		Predicted Class	
		Positive	Negative
Actual class	Positive	7	2
	Negative	2	9

Setelah didapatkan data confusion matrixnya, maka akan dapat dihitung nilai accuracy, precision dan recall.

$$Nilai Accuracy = \frac{7 + 9}{7 + 2 + 2 + 9} = 80\%$$

$$Nilai precision = \frac{7}{7 + 2} = 77,7\%$$

$$Nilai Recall = \frac{7}{7 + 2} = 77,7\%$$

Dari perhitungan di atas maka diperoleh nilai akurasi sebesar 80%, nilai precision sebesar 77,7% dan nilai recall sebesar 77,7%.

3.8 Skenario Pengujian

Penelitian ini melakukan skenario uji coba beberapa kali untuk mendapatkan hasil akurasi yang terbaik. Uji coba ini dilakukan dengan bertujuan untuk membandingkan keakurasian perhitungan klasifikasi keluarga berisiko stunting dengan menggunakan metode C4.5 dengan data latih yang ada. Melakukan uji akurasi pada hasil klasifikasi membantu untuk mengidentifikasi kesalahan dalam klasifikasi, sehingga dapat ditentukan persentase klasifikasi yang akurat (Bashit dkk. 2019).

Data dipecah menjadi dua bagian, yaitu data pelatihan (training data) dan data pengujian (test data), dengan pembagian ini dilakukan dalam lima partisi untuk menghasilkan lima skenario pengujian yang berbeda. Pembagian ini bertujuan untuk menemukan skenario pengujian yang dapat memberikan akurasi terbaik bagi model pembelajaran mesin. Akurasi adalah persentase data yang diprediksi dengan benar dari seluruh data. Beragam skenario pengujian ini dapat memengaruhi kinerja model pembelajaran mesin, karena jumlah dan variasi data pelatihan dan pengujian dapat memengaruhi kemampuan model untuk belajar dan menggeneralisasi. Distribusi persentase data pelatihan dan data pengujian dalam setiap skenario pengujian dapat dilihat pada Tabel 3.7 berikut.

Tabel 3.7 Skenario pembagian *dataset*

Skenario	Rasio data latih dan test data
1	90:10
2	80:20
3	70:30

4	60:40
5	50:50

Langkah berikutnya adalah menjalankan perhitungan klasifikasi menurut skenario masing-masing. Berikut adalah contoh perhitungan dari skenario 1, yaitu dengan skenario pembagian data latih dan data uji sebesar 90:10

Tabel 3.8 Confusion matrix skenario 1

		Prediksi Class	
		Positive	Negative
Actual class	Positive	11	1
	Negative	1	8

Berdasarkan skenario pertama yang telah dilakukan, didapatkan bahwa dari 20 data keluarga yang, diprediksi bahwa sebanyak 12 keluarga berisiko stunting dan 8 keluarga tidak berisiko stunting. Tetapi kenyatannya pada ada 11 keluarga berisiko stunting dan 9 keluarga tidak berisiko stunting. Setelah diketahui hasil dari prediksi tersebut, maka dapat dihitung nilai accuracy, recall, dan precisionnya

$$\text{Nilai Accuracy} = \frac{11 + 1}{11 + 1 + 1 + 8} = 90\%$$

$$\text{Nilai precision} = \frac{11}{11 + 1} = 91\%$$

$$\text{Nilai Recall} = \frac{11}{11 + 1} = 91\%$$

Pada skenario pertama, diperoleh hasil akurasi yang baik; jika prediksi dapat dilakukan dengan sukses, maka precision dan recall akan seimbang. Penelitian ini melibatkan lima skenario pengujian yang dilaksanakan. Skenario 1 hingga skenario 5 mengikuti prosedur yang sama dengan skenario awal, namun menggunakan data uji dan data latih yang bervariasi.

Tabel 3.9 Hasil performa scenario uji coba 1 - 5

Skenario	Accuracy	Precision	Recall
1	0,904761905	0,916666667	0,916666667
2	0,666666667	0,764705882	0,764705882
3	0,904761905	0,928571429	0,928571429
4	0,904761905	0,923076923	0,923076923
5	0,818181818	0,857142857	0,857142857

Terlihat pada Tabel 3.5, setelah selesai melakukan pengujian setiap skenario, maka seluruh skenario dari pertama sampai akhir dirata-ratakan tingkat akurasinya adalah 0,83982684, hasil precision rata-ratanya adalah 0,87803275 dan hasil recall rata – ratanya adalah 0,87803275.

BAB IV

UJI COBA DAN PEMBAHASAN

4.1 Pengujian

Tahap pengujian dilakukan dengan menguraikan proses uji coba yang dilakukan untuk mengumpulkan data yang dibutuhkan untuk memastikan efektivitas dan efisiensi dari metode yang diusulkan dalam penelitian ini. Pengujian ini dilakukan untuk menghitung performa dari algoritma C4.5 dalam mengklasifikasikan keluarga berisiko stunting dengan data rekapitulasi keluarga berisiko stunting yang terdiri dari 5912 record, dan 10 atribut. Dataset kemudian dipisahkan menjadi data latih dan data uji sehingga dapat mengevaluasi kinerja model secara objektif dengan diujikan terhadap data baru. Berikut adalah 5 skenario pembagian data yang dijalankan pada tahap pengujian ini.

Tabel 4.1 Skenario uji coba pembagian data latih dan data uji

Skenario	Rasio data latih dan test data
1	90:10
2	80:20
3	70:30
4	60:40
5	50:50

4.2 Implementasi

Pada bagian ini akan dilakukan pemodelan pada decision tree C4.5 dengan dimulai dari memasukan data, normalisasi, sampai dengan pembentukan decision tree.

4.2.1 Input Data

Data yang digunakan berupa rekapitulasi data keluarga sasaran yang memiliki resiko untuk melahirkan anak stunting di wilayah Kecamatan Poncokusumo pada tahun 2022. Dataset ini berisi 5912 record dan 11 atribut yang dimana salah satu atribut adalah tujuan atau class label. Untuk memasukan dataset diperlukan library pandas, yang selanjutnya dimasukan syntax pada gambar berikut.

```
#import package
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn import preprocessing
from google.colab import files
stunting = files.upload()
#upload file 20data.csv
```

Gambar 4.1 Input data menggunakan library pandas

4.2.2 Normalisasi Dataset

Data yang sudah di masukan selanjutnya akan di normalisasi sehingga data awal akan berubah menjadi data integer dengan rentang 0-1 seperti pada gambarr berikut.

```
#ubah semua tipe data keint
from sklearn.preprocessing import LabelEncoder
le = LabelEncoder()

for column in df:
    if df[column].dtypes == object or df[column].dtypes == float:
        df[column]=le.fit_transform(df[column])
```

Gambar 4.2 Convert tipe data menjadi integer

Dataset akan dipisah dari kelas label atau kelas tujuan yaitu "Outcome" menjadi data atr dataset dan cls_dataset untuk kelas label Selanjutnya akan dibagi

menjadi data training dan data testing menggunakan train test split dari library sklearn.

4.2.3 Pembagian Data

```
from sklearn.model_selection import train_test_split
xtrain, xtest, ytrain, ytest = train_test_split (
    atr_df, cls_df, test_size=0.3, random_state=100
)

#test size = 0.3 berarti pembagian data latih 70% dan data uji 30%
```

Gambar 4.3 Pembagian data train_test_split

Dataset akan dipisah dari kelas label atau kelas tujuan menjadi data atr dataset dan cls_dataset untuk kelas label. Selanjutnya akan dibagi training data (data uji) dan test data (test data) menggunakan train test split dari library sklearn pada gambar berikut.

4.2.4 Modeling Decision Tree C4.5

Algoritma C4.5 membuat pohon keputusan dengan memilih atribut optimal pada setiap tahapnya menggunakan metrik seperti gain ratio atau information gain. Proses pembentukan pohon keputusan ini sangat krusial karena menghasilkan visualisasi yang membantu dalam memahami logika pengambilan keputusan di dalam pohon.

```
tree_df = DecisionTreeClassifier(
    random_state=0,max_depth=None,
    min_samples_split=10, min_samples_leaf=10,
    min_weight_fraction_leaf=0,
    max_leaf_nodes=None, min_impurity_decrease=0)

tree_df.fit(xtrain, ytrain)
```

Gambar 4.4 Modeling decision tree menggunakan DecisionTreeClassifier

4.3 Hasil Uji Coba

Tingkat keberhasilan Algoritma C4.5 dalam sistem ini ditentukan dengan membandingkan nilai antara nilai y_{test} dan y_{pred} . Nilai y_{test} adalah kelas sebenarnya dari data uji yang diberikan, sedangkan nilai y_{pred} adalah kelas yang diprediksi oleh model C4.5 yang telah dilatih dengan data latih. Setelah kedua nilai tersebut didapatkan maka dapat dihitung nilai accuracy, precision dan recallnya.

Untuk menampilkan hasil klasifikasi, penulis akan memakai confusion matrix untuk evaluasi keakurasian algoritmanya. Confusion matrix adalah tabel yang menunjukkan bagaimana data uji didistribusikan berdasarkan kelas asli dan kelas prediksi. Dengan menggunakan metode ini, penulis dapat menilai keakurasian sistem dalam mengklasifikasikan keluarga berisiko stunting.

4.3.1 Hasil Uji Coba Skenario 1

Pada skenario pertama dilakukan pengujian pembagian data dengan perbandingan rasio sebesar 90:10 yang artinya 90% dari total keseluruhan data adalah data training, sedangkan 10% dari total keseluruhan data adalah data testing. Berikut adalah table hasil perbandingan dari data latih dengan hasil klasifikasi model sesuai dengan skenario 1.

Tabel 4.2 Perbandingan hasil data latih dengan model C4.5 skenario 1

No	Hasil Data Latih	Hasil Klasifikasi Model C4.5
1	1	1
2	1	1
3	1	1
4	0	0
5	1	1
6	1	1
7	1	1
8	0	0
9	1	1
...
5912	1	0

Hasil prediksi hasil klasifikasi uji coba di atas kemudian akan dihitung nilai accuracy, precision dan recallnya dengan menggunakan confusion matrix.

Tabel 4.3 Confusion matrix skenario 1

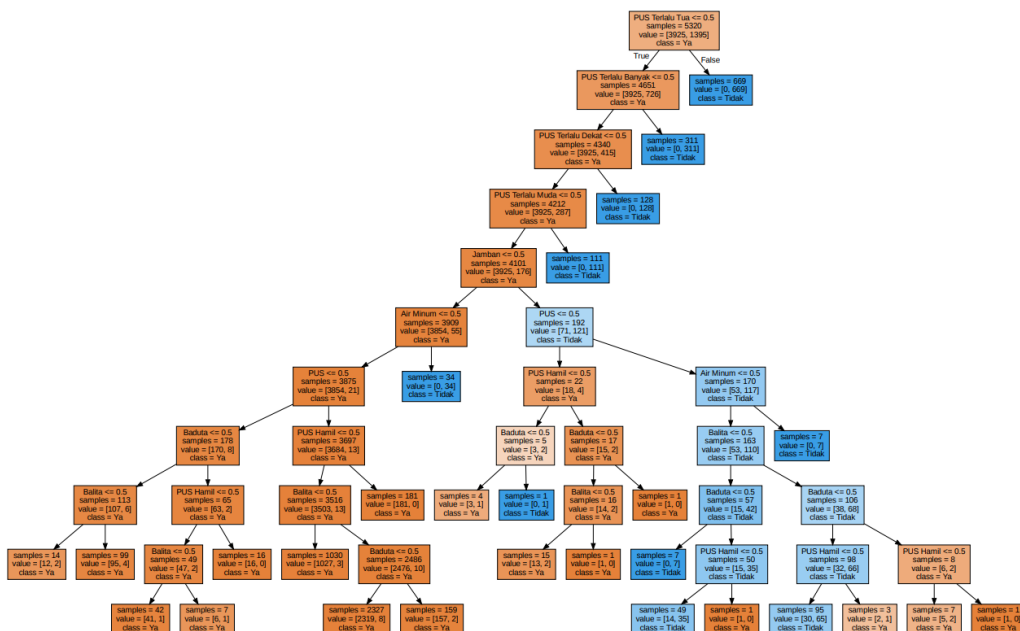
		Prediction Class	
		Positive	Negative
Actual Class	Positive	422	11
	Negative	2	388

Selanjutnya adalah menghitung precision value dari pengujian klasifikasi antara lain, perhitungan accuracy, precision dan recall menggunakan rumus berikut:

$$Accuracy (sk1) = \frac{422+388}{422+388+11+2} = 0,978040541$$

$$Precision (sk1) = \frac{422}{422+11} = 0,974595843$$

$$Recall (sk1) = \frac{422}{422+2} = 0,995283019$$



Gambar 4.5 Hasil pembentukan decision tree C4.5 pada skenario 1

4.3.2 Hasil Uji Coba Skenario 2

Pada uji coba skenario kedua dilakukan pengujian pembagian data dengan perbandingan rasio sebesar 80:20 yang artinya 80% dari total keseluruhan data adalah data training, sedangkan 20% dari total keseluruhan data adalah data testing. Berikut adalah table hasil perbandingan dari data latih dengan hasil klasifikasi model sesuai dengan skenario 2.

Tabel 4.4 Perbandingan hasil data latih dengan model C4.5 skenario 2

No	Hasil Data Latih	Hasil Klasifikasi Model C4.5
1	1	1
2	1	1
3	1	1
4	0	0
5	1	1
6	1	1
7	1	1
8	0	0
9	1	1
...
5912	1	1

Hasil prediksi hasil klasifikasi uji coba di atas kemudian akan dihitung nilai accuracy, precision dan recallnya dengan menggunakan confusion matrix.

Tabel 4.5 Confusion matrix skenario 2

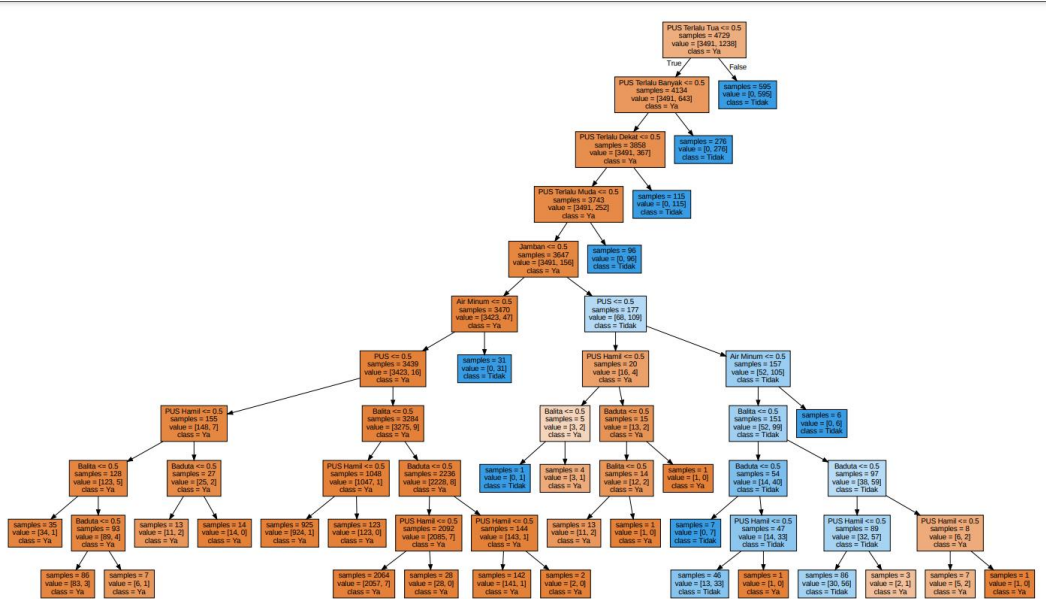
		Prediction Class	
		Positive	Negative
Actual Class	Positive	855	12
	Negative	7	309

Selanjutnya adalah menghitung precision value dari pengujian klasifikasi antara lain, perhitungan accuracy, precision dan recall menggunakan rumus berikut:

$$Accuracy (sk2) = \frac{855+309}{855+309+12+7} = 0,984956453$$

$$Precision (sk2) = \frac{855}{855+12} = 0,98615917$$

$$Recall (sk2) = \frac{855}{855+7} = 0,99187935$$



Gambar 4.6 Hasil pembentukan decision tree C4.5 pada skenario 2

4.3.3 Hasil Uji Coba Skenario 3

Pada uji coba skenario kedua dilakukan pengujian pembagian data dengan perbandingan rasio sebesar 70:30 yang artinya 70% dari total keseluruhan data adalah data training, sedangkan 30% dari total keseluruhan data adalah data testing. Berikut adalah table hasil perbandingan dari data latih dengan hasil klasifikasi model sesuai dengan skenario 3.

Tabel 4.6 Perbandingan hasil data latih dengan model C4.5 skenario 3

No	Hasil Data Latih	Hasil Klasifikasi Model C4.5
1	1	1
2	1	1
3	1	1
4	0	0
5	1	1
6	1	1
7	1	1
8	0	0
9	1	1
...
5912	1	0

Hasil prediksi hasil klasifikasi uji coba di atas kemudian akan dihitung nilai accuracy, precision dan recallnya dengan menggunakan confusion matrix.

Tabel 4.7 Confusion matrix skenario 3

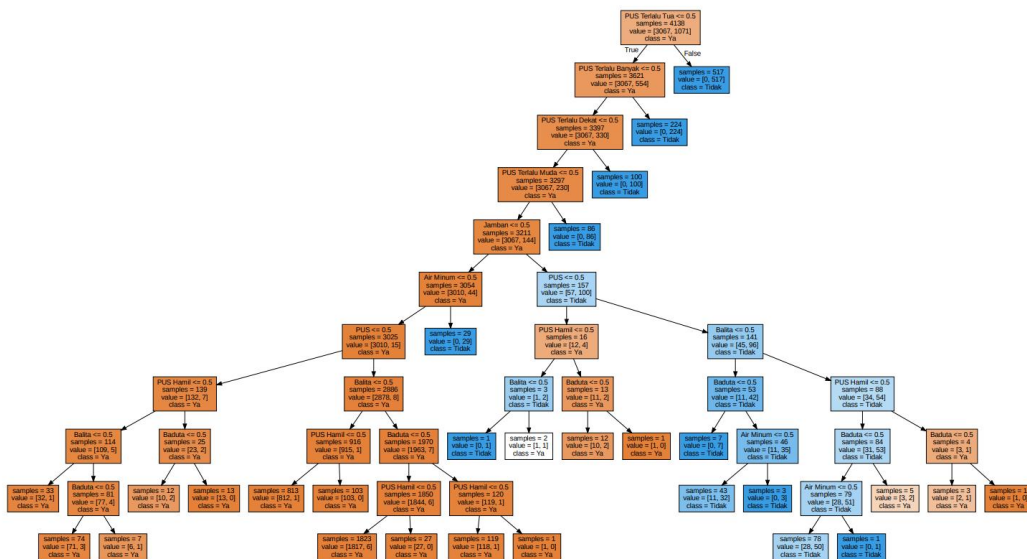
		Prediction Class	
		Positive	Negative
Actual Class	Positive	1274	17
	Negative	8	475

Selanjutnya adalah menghitung precision value dari pengujian klasifikasi antara lain, perhitungan accuracy, precision dan recall menggunakan rumus berikut:

$$Accuracy (sk3) = \frac{1274+475}{1274+475+17+8} = 0,985907554$$

$$Precision (sk3) = \frac{1274}{1274+17} = 0,986831913$$

$$Recall (sk3) = \frac{1274}{1274+8} = 0,99375975$$



Gambar 4.7 Hasil pembentukan decision tree C4.5 pada skenario 3

4.3.4 Hasil Uji Coba Skenario 4

Pada uji coba skenario kedua dilakukan pengujian pembagian data dengan perbandingan rasio sebesar 60:40 yang artinya 60% dari total keseluruhan data

adalah data training, sedangkan 40% dari total keseluruhan data adalah data testing. Berikut adalah table hasil perbandingan dari data latih dengan hasil klasifikasi model sesuai dengan skenario 4.

Tabel 4.8 Perbandingan hasil data latih dengan model C4.5 skenario 4

No	Hasil Data Latih	Hasil Klasifikasi Model C4.5
1	1	1
2	1	1
3	1	1
4	0	0
5	1	1
6	1	1
7	1	1
8	0	0
9	1	1
...
5912	1	0

Hasil prediksi hasil klasifikasi uji coba di atas kemudian akan dihitung nilai accuracy, precision dan recallnya dengan menggunakan confusion matrix.

Tabel 4.9 Confusion matrix skenario 4

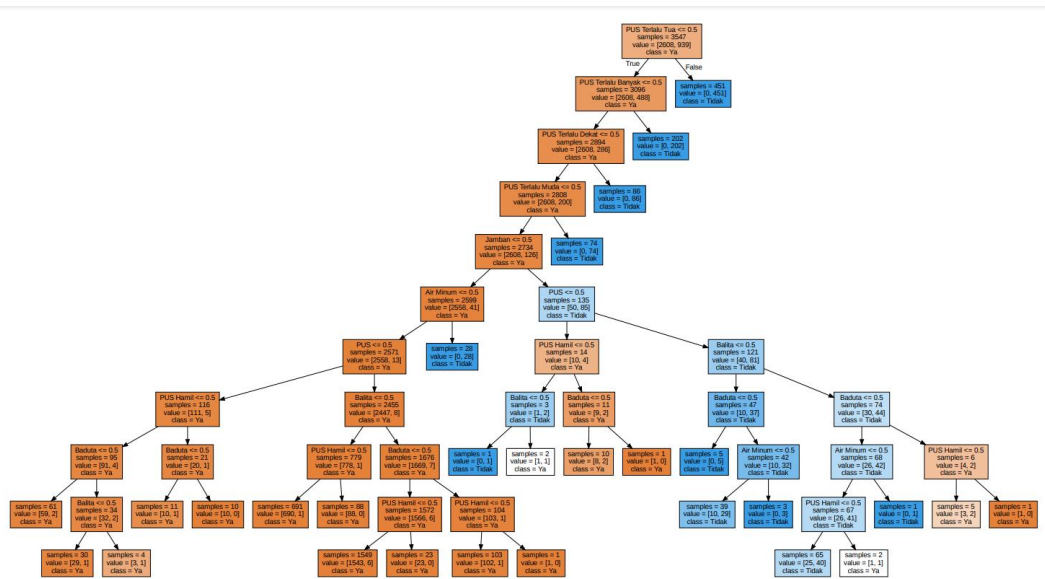
		Prediction Class	
		Positive	Negative
Actual Class	Positive	1729	21
	Negative	10	605

Selanjutnya adalah menghitung precision value dari pengujian klasifikasi antara lain, perhitungan accuracy, precision dan recall menggunakan rumus berikut:

$$Accuracy (sk4) = \frac{1729 + 605}{1729 + 605 + 21 + 10} = 0,986892178$$

$$Precision (sk4) = \frac{1729}{1729 + 21} = 0,988$$

$$Recall (sk4) = \frac{1729}{1729 + 10} = 0,994249569$$



Gambar 4.8 Hasil pembentukan decision tree C4.5 pada skenario 4

4.3.5 Hasil Uji Coba Skenario 5

Pada uji coba skenario kedua dilakukan pengujian pembagian data dengan perbandingan rasio sebesar 50:50 yang artinya 50% dari total keseluruhan data adalah data training, sedangkan 50% dari total keseluruhan data adalah data testing. Berikut adalah table hasil perbandingan dari data latih dengan hasil klasifikasi model sesuai dengan skenario 5.

Tabel 4.10 Perbandingan hasil data latih dengan model C4.5 skenario 5

No	Hasil Data Latih	Hasil Klasifikasi Model C4.5
1	1	1
2	1	1
3	1	1
4	0	0
5	1	1
6	1	1
7	1	1
8	0	0
9	1	0
...
5912	1	0

Hasil prediksi hasil klasifikasi uji coba di atas kemudian akan dihitung nilai accuracy, precision dan recallnya dengan menggunakan confusion matrix.

Tabel 4.11 Confusion mterix skenario 5

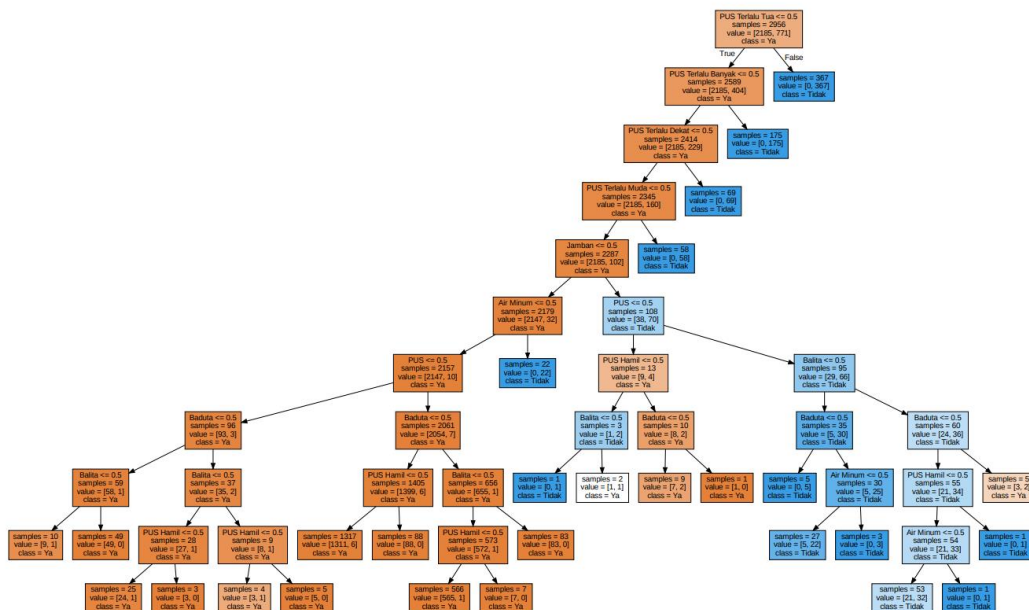
		Prediction Class	
		Positive	Negative
Actual Class	Positive	2140	33
	Negative	13	770

Selanjutnya adalah menghitung precision value dari pengujian klasifikasi antara lain, perhitungan accuracy, precision dan recall menggunakan rumus berikut:

$$Accuracy (sk5) = \frac{2140+770}{2140 + 770 + 33 + 13} = 0,98443843$$

$$Precision (sk5) = \frac{2140}{2140 + 33} = 0,984813622$$

$$Recall (sk5) = \frac{2140}{2140 + 13} = 0,993961914$$



Gambar 4.9 Hasil pembentukan decision tree C4.5 pada skenario 5

4.4 Pembahasan

Pembahasan pada sub bab ini merupakan hasil yang didapat dari hasil uji coba yang telah dilakukan menggunakan metode algoritma C4.5 dengan bebrapa

skenario yang berbeda. Uji coba ini bertujuan untuk mengevaluasi kinerja algoritma dalam berbagai kondisi dan parameter yang berbeda, sehingga dapat menentukan konfigurasi yang paling optimal untuk menyelesaikan masalah yang diberikan.

Tabel 4.12 Hasil pengujian skenario 1 - 5

No	Pengujian	Akurasi C4.5	Precision	Recall
1	Skenario 1 (90% - 10%)	0,978040541	0,974595843	0,995283019
2	Skenario 2 (80% - 20%)	0,984956453	0,98615917	0,99187935
3	Skenario 3 (70% - 30%)	0,985907554	0,986831913	0,99375975
4	Skenario 4 (60% - 40%)	0,986892178	0,988	0,994249569
5	Skenario 5 (50% - 50%)	0,98443843	0,984813622	0,993961914
	Rata - rata	0,984047031	0,98408011	0,99382672

Dapat diketahui pada table 4.11 diatas, pengujian yang telah dilakukan bahwa pembagian data tidak berpengaruh banyak terhadap keakurasian dari algoritma C4.5. Pembagian data pada umumnya adalah langkah yang penting untuk menilai stabilitas dan kinerja generalisasi model di berbagai kumpulan data dengan menghasilkan variabilitas data yang memadai. Hal ini dikarenakan bahwa data yang digunakan dalam penelitian ini bersifat homogen, yang berarti tidak banyak variabilitas di dalam datasetnya. Faktor lain yang mungkin mempengaruhi keakurasian adalah tree pruning, yaitu langkah pemangkasan untuk mengurangi ukuran pohon yang berfungsi untuk membantu mencegah overfitting. Pruning biasanya didasarkan pada metrik kinerja yang dievaluasi pada set validasi terpisah dan tidak dipengaruhi oleh distribusi data pelatihan atau pengujian. Meskipun decision tree C4.5 pada umumnya akurat dalam pemisahan antara data pelatihan

dan pengujian, penting untuk diperhatikan bahwa performa model secara keseluruhan masih dapat dipengaruhi oleh faktor-faktor seperti kualitas data pelatihan, pilihan parameter dan kompleksitas tree.

Algoritma C4.5 tetap menghasilkan keputusan yang konsisten meskipun data dibagi-bagi dengan skenario yang berbeda karena penggunaan atribut yang berlimpah dalam proses pengambilan keputusan. Oleh karena itu, diperlukan upaya pengurangan atribut untuk menyusutkan data dan mempermudah prediksi dengan menggunakan atribut yang lebih sedikit.

Salah satu metode yang pengurangan data yang digunakan adalah filtrasi manual. Metode ini melibatkan peninjauan dan pemilihan fitur secara langsung berdasarkan pengetahuan domain atau kriteria tertentu, seperti korelasi antar fitur, nilai informasi, atau kontribusi terhadap tujuan analisis. Proses filtrasi manual umumnya dimulai dengan eksplorasi dataset untuk memahami struktur dan sifatsifatnya. Analisis kemudian mengidentifikasi fitur yang mungkin tidak relevan atau redundant, seperti fitur yang memiliki banyak nilai yang tidak konsisten. Meskipun filtrasi manual dapat memakan waktu dan bergantung pada penilaian subjektif, metode ini memungkinkan fleksibilitas dan penyesuaian berdasarkan kebutuhan spesifik dari analisis yang dilakukan. (Tang, Dai, and Xiang 2019)

Tabel 4.13 Akurasi hasil pengujian skenario 1 – 5 dengan pengurangan data

No	Jumlah Atribut	Rasio data latih dan uji	Akurasi Data
1	10	90:10	98%
2	9	90:10	94%
3	8	90:10	91%
4	7	90:10	78%
5	6	90:10	76%

Terlihat dari tabel 4.12 akurasi data telah terbukti berkurang jika beberapa atribut dihilangkan dari dataset yang digunakan. Dalam skenario 1 dilakukan pengujian menggunakan seluruh atribut dengan akurasi sebesar 98%. Dalam skenario 2 dilakukan pengujian tanpa menggunakan atribut “PUS terlalu banyak” dengan akurasi sebesar 94%, akurasi berkurang 4% dibandingkan dengan skenario 1. Dengan demikian pengurangan atribut “PUS terlalu banyak” tidak berpengaruh besar terhadap akurasi hasil klasifikasi. Dalam skenario 3 dilakukan pengujian tanpa menggunakan atribut “PUS terlalu banyak” dan “PUS terlalu dekat”, akurasi berkurang 7% dibandingkan dengan skenario 1 dan berkurang sebesar 3% dibandingkan dengan skenario 2. Dengan demikian pengurangan atribut “PUS terlalu dekat” tidak berpengaruh besar terhadap akurasi hasil klasifikasi.

Dalam skenario 4 dilakukan pengujian tanpa menggunakan atribut “PUS terlalu banyak”, “PUS terlalu dekat”, dan “PUS terlalu tua”, akurasi berkurang 20% dibandingkan dengan skenario 1 dan berkurang sebesar 13% dibandingkan dengan skenario 3. Dapat diketahui pengurangan atribut “PUS terlalu tua” berpengaruh besar terhadap hasil keakurasian klasifikasi. Dalam skenario 5 dilakukan pengujian tanpa menggunakan atribut “PUS terlalu banyak”, “PUS terlalu dekat”, “PUS terlalu tua”, dan “PUS terlalu muda”, akurasi berkurang 22% dibandingkan dengan skenario 1 dan berkurang sebesar 2% dibandingkan dengan skenario 4. Dapat diketahui pengurangan atribut “PUS terlalu muda” tidak berpengaruh besar terhadap hasil keakurasian klasifikasi.

Dengan demikian dapat disimpulkan bahwa atribut “PUS terlalu tua” berpengaruh paling besar terhadap keakurasian klasifikasi, yaitu sebesar 13%.

Sedangkan hasil pengurangan dengan attribut yang lain tidak terlalu berpengaruh besar terhadap total akurasi dari klasifikasi.

4.5 Integrasi AI -Quran dan Hadits

Alquran sangat menekankan pentingnya memberikan asupan yang halal dan baik untuk anak-anak sebagaimana yang tertera dalam Q.S. Al-Baqarah [2]: 168.

وَأَنْفِقُوا فِي سَبِيلِ اللَّهِ وَلَا تُلْفُوا بِأَيْدِيكُمْ إِلَى التَّهْلُكَةِ وَأَحْسِنُوا إِنَّ اللَّهَ يُحِبُّ الْمُحْسِنِينَ

“Wahai manusia! Makanlah dari (makanan) yang halal dan baik yang terdapat di bumi, dan janganlah kamu mengikuti langkah-langkah setan. Sungguh, setan itu musuh yang nyata bagimu.” (QS. Al-Baqarah: 168)

Dalam perspektif Al-Qur'an, ayat ini mengajarkan pentingnya memilih dengan hati-hati jenis makanan yang dikonsumsi. Dalam konteks pencegahan stunting, ayat ini menyoroti bahwa manusia harus memilih makanan yang halal dan baik, yang memberikan nutrisi yang seimbang dan memadai untuk tubuh. Ayat ini juga menegaskan agar manusia tidak tergoda oleh syaitan yang bisa mendorong perilaku merugikan, termasuk memilih makanan tidak sehat yang berpotensi menyebabkan stunting. Dengan demikian, ayat ini mendorong manusia untuk memperhatikan jenis makanan yang mereka konsumsi untuk menjaga kesehatan tubuh dan mencegah stunting.

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Penelitian ini bertujuan untuk mengevaluasi performa algoritma klasifikasi C4.5 dalam konteks pengklasifikasikan keluarga berisiko stunting berdasarkan data rekapitulasi keluarga. Setelah menguji dataset menggunakan algoritma C4.5 dengan bahasa pemrograman Python, dan melakukan perhitungan terhadap dataset, diperoleh hasil performa yang dapat digunakan untuk mengukur seberapa baik program dalam mengklasifikasikan keluarga yang berisiko stunting atau tidak berdasarkan atribut-atribut yang telah ditetapkan. Penerapan Algoritma C4.5 pada klasifikasi keluarga berisiko stunting menghasilkan accuracy rata – rata sebesar 98,404%. precision rata – rata sebesar 98,408%, dan recall rata – rata sebesar 99,382%,

Algoritma C4.5 menunjukkan kemampuan tinggi dalam mengklasifikasikan data pelatihan dengan akurasi yang baik. Namun, perbedaan kinerja yang signifikan antara data pelatihan dan data pengujian mengindikasikan terjadinya overfitting. Model ini cenderung terlalu cocok dengan data pelatihan, sehingga tidak mampu menggeneralisasi data pengujian dengan baik. Meskipun data telah dibagi menjadi beberapa skenario dengan rasio pembagian yang beragam, metode ini tidak cukup efektif untuk mengurangi overfitting dalam penelitian ini. Penelitian ini menyimpulkan bahwa hanya dengan membagi data, masalah overfitting pada algoritma C4.5 tidak dapat diatasi sepenuhnya. Disarankan untuk menggunakan dataset yang lebih bervariasi untuk mengatasi overfitting pada algoritma C4.5.

5.2 Saran

Berdasarkan penelitian yang telah dilakukan, saran bagi penelitian selanjutnya adalah dengan menguji algoritma pada dataset yang lebih beragam. Dengan memperluas cakupan dataset, hasil yang diperoleh akan lebih representatif dan dapat memberikan gambaran yang lebih menyeluruh tentang performa algoritma.

Selain itu, penggunaan data yang bervariasi juga memungkinkan evaluasi yang lebih komprehensif terhadap keakuratan algoritma C4.5 yang digunakan. Hal ini penting untuk memastikan bahwa algoritma tidak hanya efektif dalam kondisi tertentu tetapi juga dapat diandalkan dalam berbagai situasi yang berbeda. Peneliti berikutnya diharapkan dapat mengeksplorasi variabel lain yang berpotensi mempengaruhi kejadian stunting pada balita dan meningkatkan jumlah sampel di area yang lebih luas.

Dengan sistem yang telah dikembangkan, diharapkan dapat membantu mengidentifikasi keluarga-keluarga yang berisiko stunting ke dalam kategori yang sudah tersedia. Hal ini diharapkan dapat menjadi pendukung hipotesis bagi pihak puskesmas dalam upaya mengatasi prevalensi stunting pada balita di Kecamatan Ponsokusumo

DAFTAR PUSTAKA

- Achmad Sopian, dan I Lina Widyastuti. 2022. "Pemutakhiran, Verifikasi Dan Validasi Data Sasaran Keluarga Berisiko Stunting." *Pusat Pendidikan Dan Pelatihan Kependudukan BKKBN*.
- Bashit, Nurhadi, Yudo Prasetyo, Andri Suprayogi, dan Penulis Korespondensi. 2019. "Klasifikasi Berbasis Objek untuk Pemetaan Penggunaan Lahan menggunakan Citra SPOT 5 di Kecamatan Ngaglik." *TEKNIK* 40 (2): 122–28. <https://doi.org/10.14710/teknik.v40n2.23050>.
- Chowdhury, Mohammad Rocky Khan, Mohammad Shafiur Rahman, Mohammad Mubarak Hossain Khan, Mohammad Nazrul Islam Mondal, Mohammad Mosiur Rahman, dan Baki Billah. 2016. "Risk factors for child malnutrition in Bangladesh: A multilevel analysis of a nationwide population-based survey." *Journal of Pediatrics* 172 (Mei):194-201.e1. <https://doi.org/10.1016/j.jpeds.2016.01.023>.
- Danaei, Goodarz, Kathryn G. Andrews, Christopher R. Sudfeld, Günther Fink, Dana Charles McCoy, Evan Peet, Ayesha Sania, Mary C. Smith Fawzi, Majid Ezzati, dan Wafaie W. Fawzi. 2016. "Risk Factors for Childhood Stunting in 137 Developing Countries: A Comparative Risk Assessment Analysis at Global, Regional, and Country Levels." *PLoS Medicine* 13 (11). <https://doi.org/10.1371/journal.pmed.1002164>.
- Fakhrurriqfi, Muhammad, dan Retantyo Wardoyo. 2013. "Perbandingan Algoritma Nearest Neighbour, C4.5 dan LVQ untuk Klasifikasi Kemampuan Mahasiswa." *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)* 7 (2): 145. <https://doi.org/10.22146/ijccs.3353>.
- Hermanto, Bambang, dan Azhari SN. 2017. "Klasifikasi Nilai Kelayakan Calon Debitur Baru Menggunakan Decision Tree C4.5." *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)* 11 (1): 43. <https://doi.org/10.22146/ijccs.15946>.
- Machová, Kristína, Martina Szabóová, Ján Paralič, dan Ján Mičko. 2023. "Detection of emotion by text analysis using machine learning." *Frontiers in Psychology* 14. <https://doi.org/10.3389/fpsyg.2023.1190326>.
- Nazir, Alwis, Amany Akhyar, Yusra Yusra, dan Elvia Budianita. 2022. "Toddler Nutritional Status Classification Using C4.5 and Particle Swarm Optimization." *Scientific Journal of Informatics* 9 (1): 32–41. <https://doi.org/10.15294/sji.v9i1.33158>.

- Okta Qomaruddin Aziz. (2020). Performance improvement in Resampling Based Clustering. *MATICS: Jurnal Ilmu Komputer Dan Teknologi Informasi (Journal of Computer Science and Information Technology)*, 12(1).
- Powers, D M W, dan Ailab. t.t. "Evaluation: From Precision, Recall And F-Measure To Roc, Informedness, Markedness & Correlation."
- Prasetya, Tio, Irfan Ali, Cep Lukman Rohmat, dan Odi Nurdiawan. 2020. "Klasifikasi Status Stunting Balita Di Desa Slangit Menggunakan Metode K-Nearest Neighbor." *Informatics For Educators And Professionals* 4 (2): 93–104.
- Quinlan, J. R. 2014. C4. 5: Programs for machine learning. Elsevier.
- Sadikin, Mujiono, dan Fahri Alfiandi. 2018. "Comparative Study of Classification Method on Customer Candidate Data to Predict its Potential Risk." *International Journal of Electrical and Computer Engineering (IJECE)* 8 (6): 4763. <https://doi.org/10.11591/ijece.v8i6.pp4763-4771>.
- Wijaya, Yusup Indra. 2018. "Klasifikasi Tingkat Pengetahuan Ibu Dalam Pemberian Asi Eksklusif Menggunakan Algoritma Decision Tree C4.5 Berbasis Pso." *Technologia: Jurnal Ilmiah* 9 (2): 120. <https://doi.org/10.31602/Tji.V9i2.1377>.
- Tang, Xiaochuan, Yuanshun Dai, and Yanping Xiang. 2019. "Feature Selection Based on Feature Interactions with Application to Text Categorization." *Expert Systems with Applications* 120 (April):207–16. <https://doi.org/10.1016/j.eswa.2018.11.018>.
- UNICEF. (2019). The State of the World's Children 2019: Children, food and nutrition – Growing well in a changing world. United Nations Children's Fund.
- World Health Organization. (2019). Levels and trends in child malnutrition: UNICEF/WHO/The World Bank Group Joint Child Malnutrition Estimates.
- Zhang, Yiyang, Yiyang Liu, Xiaoyan Guo, Zhu Liu, Xiankun Zhang, dan Kun Liang. 2022. "A BiLSTM-Based DDoS Attack Detection Method for Edge Computing." *Energies* 15 (21). <https://doi.org/10.3390/en15217882>.