



IMPLEMENTASI CRISP-DM MODEL MENGGUNAKAN METODE DECISION TREE DENGAN ALGORITMA CART UNTUK PREDIKSI LILA IBU HAMIL BERPOTENSI GIZI KURANG

¹⁾Dita Anies Munawwaroh, ²⁾Arum Handini Primandari

¹⁾ Teknik Mesin, Politeknik Negeri Semarang

²⁾ Prodi Statistika, Universitas Islam Indonesia

primandari.arum@uii.ac.id

Abstract

LILA is measured in pregnant women to monitor nutritional levels during pregnancy. The classifications in the LILA's measurement are the good nutrition category if the LILA measurement is more or equal to 23.5 cm and the undernutrition category if the LILA measurement is less than 23.5 cm. This study aims to classify LILA based on age, height, weight, blood pressure, hemoglobin level, blood sugar, gestational age, and hip circumference by employing the CRISP-DM methodology. The data used is based on 150 pregnant women taken in May - June 2022 at the Sumber Health Center, Sumber District, Rembang Regency. The decision tree method (decision tree) with the CART algorithm is worked to classify LILA in either the good and poor category. The data is divided into training and testing data by a ratio of 80%:20%. The decision tree method can classify all training data correctly. While evaluating the method with data testing produces values of accuracy, precision, recall, and f1-score, respectively, are 90%, 96%, 92%, and 94%.

Keywords: CRISP-DM, Decision Tree, Upper Arm Circumference

Abstrak

Pengukuran LILA pada ibu hamil dilakukan untuk memantau tingkat gizi pada masa kehamilan. Terdapat klasifikasi pada pengukuran lila yaitu kategori gizi baik apabila pengukuran LILA lebih dari sama dengan 23,5 cm, dan kategori gizi kurang apabila pengukuran LILA kurang dari 23,5 cm. Pada penelitian ini bertujuan untuk mengklasifikasi LILA berdasarkan pada umur, tinggi badan, berat badan, tekanan darah, kadar hemoglobin, gula darah, usia kehamilan, dan lingkaran panggul dengan menggunakan metodologi CRISP-DM. Data yang digunakan berdasarkan 150 ibu hamil yang diambil pada bulan Mei-Juni 2022 di Puskesmas Sumber, Kecamatan Sumber, Kabupaten Rembang. Metode pohon keputusan (*decision tree*) dengan algoritma CART digunakan untuk melakukan klasifikasi LILA pada kategori baik dan kurang. Sebanyak 80% data digunakan sebagai training dan 20% data digunakan sebagai testing. Metode *decision tree* dapat mengklasifikasikan semua data training dengan benar. Sementara evaluasi metode dengan data testing, menghasilkan nilai akurasi, *precision*, *recall*, dan *f1-score* berturut-turut yaitu 90%, 96%, 92%, dan 94%.

Kata Kunci: CRISP-DM, Decision Tree, LILA

Article History

Received

24 Juni 2022

Accepted

7 Juli 2022

Published

16 Agustus 2022

This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



1. Pendahuluan

Berdasarkan Rencana Pembangunan Jangka Panjang Nasional Tahun 2005-2025 telah menempatkan Angka Kematian Ibu (AKI) dan Angka Kematian Bayi (AKB) menjadi indikator derajat kesehatan dan keberhasilan penyelenggaraan pembangunan kesehatan. Program percepatan penurunan kematian ibu ditetapkan menjadi proyek prioritas strategis dalam prioritas pembangunan nasional. Angka Kematian Ibu (AKI) merupakan salah satu indikator untuk melihat keberhasilan upaya kesehatan ibu. AKI diperoleh dari rasio kematian ibu selama masa kehamilan, persalinan, dan nifas yang

disebabkan oleh kehamilan, persalinan, dan nifas atau pengelolannya (bukan karena sebab lain) yang dihitung pada setiap 100.000 kelahiran hidup. Data yang dihimpun selama tahun 1991 hingga tahun 2020 telah mengalami angka penurunan sebesar 1,8 persen per tahunnya. Meskipun mengalami penurunan AKI masih belum mencapai SDGs tahun 2030 yaitu kurang dari 70 per 100.000 kelahiran hidup (Badan Keahlian Sekretariat Jendral DPR RI, 2021). Sehingga, AKI menjadi sasaran/ indikator pembangunan kesehatan yang merupakan investasi dalam meningkatkan kualitas sumber daya manusia.

Pembangunan kesehatan bertujuan untuk meningkatkan derajat kesehatan masyarakat dengan sasaran menurunnya angka kematian ibu dan angka kematian bayi. Salah satu penyebab tidak langsung dari AKI adalah masalah gizi. Disisi lain, penyebab kematian bayi pada masa perinatal adalah berat bayi lahir rendah (BBLR). Hal tersebut dapat disebabkan saat masa kehamilan terjadi anemia, lahir kurang bulang, dan kurangnya asupan gizi selama kehamilan. Sehingga, AKI dan AKB merupakan hal yang tidak terpisahkan.

Pada tahun 2011-2012 di kabupaten Banjar telah dilakukan pemantauan terhadap hubungan status gizi ibu hamil dengan kejadian bayi lahir mati. Metode yang dipakai pada penelitian observasional analitik dengan pendekatan *case control*, Teknik purposive sampling, kemudian data dianalisis dengan uji *Fischer Exact* dan *Chi Square*. Hasilnya pemantaun tersebut terdapat ibu hamil yang memiliki tinggi badan <154 cm sebesar 6,67%, penambahan berat selama hamil < 10 kg sebesar 96,6 %, kadar hemoglobin < 11 gr/dl sebesar 76,67 %, dan Lingkar Lengan Atas (LILA) < 23,5 cm sebesar 50 %. Hasil pengolahan data menunjukkan bahwa tidak ada hubungan tinggi badan ibu hamil ($p=0,671$) dan penambahan berat badan selama hamil ($p=0,195$) dengan kajadian bayi lahir mati. Sedangkan, terdapat hubungan antara kadar Hemoglobin ibu hamil ($p=0,001$, OR=7,76) dan Lingkar Lengan Atas (LILA) ibu hamil ($p=0,001$, OR=14) terhadap kejadian bayi lahir mati. Hal tersebut dapat disimpulkan bahwa terdapat hubungan yang signifikan antara Lingkar Lengan Atas (LILA) dan kadar hemoglobin terhadap kejadian bayi lahir mati (Mariyatul et al., 2014).

Telah dilakukan penelitian dengan metode penelitian yaitu observasional analitik dengan pendekatan retrospektif yang dilaksanakan di Kabupaten Karanganyar pada bulan Oktober hingga Desember 2016. Hasil penelitian dianalisis dengan *chi square* sedemikian sehingga dapat ditarik kesimpulan bahwa terdapat hubungan yang positif dan signifikan antara LILA ibu hamil dengan berat badan bayi lahir, namun tidak terdapat hubungan peningkatan berat badan ibu selama hamil dengan berat badan bayi lahir (Putri &

Hastutik, 2017). Selanjutnya, dilakukan penelitian dengan metode penelitian yaitu observasional analitik dengan pendekatan *cross sectional* yang dilaksanakan di Kabupaten Karanganyar pada bulan Mei-Juni 2018. Data penelitian dianalisis dengan menggunakan analisis *chi square* sedemikian sehingga dapat ditarik kesimpulan yang menyatakan bahwa tidak ada hubungan yang signifikan antara usia ibu dengan peningkatan berat badan selama hamil lahir (Putri & Hastutik, 2018).

Lebih lanjut, terdapat istilah kekurangan energi kronis (KEK), yaitu suatu keadaan yang terjadi apabila seorang ibu hamil mengalami kekurangan nutrisi dalam jangka waktu lama atau kronik yang ditandai dengan LILA < 23,5 cm. Faktor faktor yang mempengaruhi terjadinya kekurangan energi kronik pada ibu hamil adalah asupan gizi, umur, jarak kehamilan, pendapatan keluarga, paritas dan Pendidikan, dll (Pratama, 2020). Hal tersebut juga dilakukan penelitian dengan metode penelitian data Riskesdas 2007 di provinsi Papua dan Papua barat dengan pendekatan *cross sectional*. Tujuan penelitian ini untuk mengetahui hubungan usia, tingkat Pendidikan, status ekonomi, pekerjaan dan asupan zat gizi makri dengan status gizi. Data penelitian dianalisa dengan menggunakan regresi logistik sedemikian sehingga diperoleh kesimpulan bahwa terdapat hubungan usia ibu hamil, dan asupan energi terhadap status gizi ibu hamil di Provinsi Papua dan Papua Barat (Auliana et al., 2016).

Terdapat penelitian mengenai salah satu penyebab anemia adalah defisiensi zat besi yang biasanya pada ibu hamil dihubungkan dengan status gizi yang kurang memadai. Hal tersebut menjadi motivasi untuk mengetahui korelasi antara status gizi dan kadar hemoglobin pada kejadian anemia ibu hamil trimester III. Metode penelitian yang digunakan adalah *purposive sampling* pada 30 sampel ibu hamil trimester III di wilayah kerja puskesmas Ponre Kecamatan Gantarang, Kabupaten Bulukumba. Data dianalisa dengan uji *Spearman's rho* sedemikian sehingga diperoleh kesimpulan bahwa ibu hamil trimester III yang berstatus gizi (IMT) kurang, mengalami anemia ringan sebesar 10 %, anemia sedang sebesar 3,3 %, dan kadar Hemoglobin normal kurang dari 3,3%. Sedangkan, responden dengan IMT normal terdapat 3,3 % tidak mengalami anemia, anemia ringan 40 %, dan tidak ditemui anemia sedang. Serta terdapat tambahan bahwa untuk IMT obesitas tetap dapat mengalami anemia ringan sebesar 30 % (Kamaruddin et al., 2019).

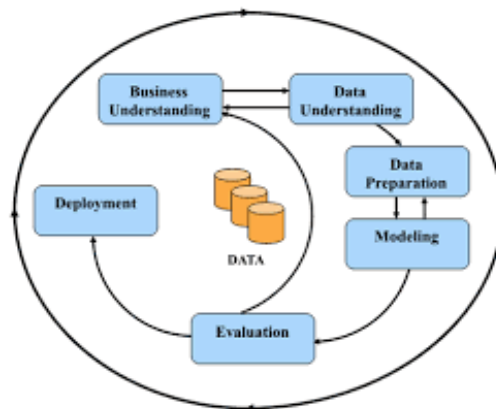
Pada penelitian yang dilakukan oleh (Lubis, 2003) untuk mengetahui status gizi ibu hamil dengan cara memantau pertambahan berat badan selama hamil, mengukur lingkar lengan atas (LILA), dan mengukur kadar hemoglobin. Pemantauan biasanya

dilakukan di puskesmas atau ditempat pelayanan kesehatan ibu hamil. Tujuan pengukuran tersebut untuk mengetahui apakah seseorang menderita kurang energi kronis (KEK) yang dapat dilihat dari ukuran LILA, sedangkan apakah seseorang menderita anemia gizi yang dapat dilihat dari kadar hemoglobin.

Pada penelitian yang dilakukan untuk mengklasifikasikan masalah gizi menggunakan metode *Decision Tree* dengan menggunakan Algoritma Iterative Dichotomiser 3 (IDP3) pada studi kasus status gizi balita Indonesia Tahun 2017. Pertimbangan pemilihan analisis klasifikasi untuk dapat menemukan kriteria utama yang mempengaruhi suatu kasus (Ardiyanti, 2019). Oleh karena itu, penelitian ini akan dilakukan analisis klasifikasi dengan metode *Decision Tree* untuk dapat menemukan kriteria utama yang mempengaruhi LILA ibu hamil dengan gizi baik dan gizi kurang. Kriteria yang digunakan dalam penelitian ini adalah umur, tinggi badan, berat badan, tekanan darah, kadar hemoglobin, gula darah, usia kehamilan, dan lingkar panggul.

2. Metode Penelitian

Pada penelitian ini digunakan metodologi data mining CRISP-DM untuk mendapatkan alur pemecahan masalah. Metodologi ini terdiri dari 6 tahapan yaitu *Business Understanding*, *Data Understanding*, *Data preparation*, *Modelling*, *Evaluation*, dan *Deployment*. Proses tersebut dapat dijelaskan dengan bagan berikut.



Gambar 1. Metodologi CRIPS-DM

a. *Business Understanding* (Pemahaman Bisnis)

Tahap pertama adalah memahami kebutuhan serta tujuan yang ingin dicapai pada penelitian. Sehingga dapat dirumuskan masalah, akar masalah, variabel yang akan digunakan sebagai variabel bebas yang diduga mempengaruhi variabel terikat.

b. *Data Undertstanding* (Pemahaman Data)

Pada tahapan ini akan dilakukan pengumpulan data, pendeskripsian data, serta evaluasi kualitas data. Data dideskripsikan untuk mendapatkan wawasan (*insight*) potensi

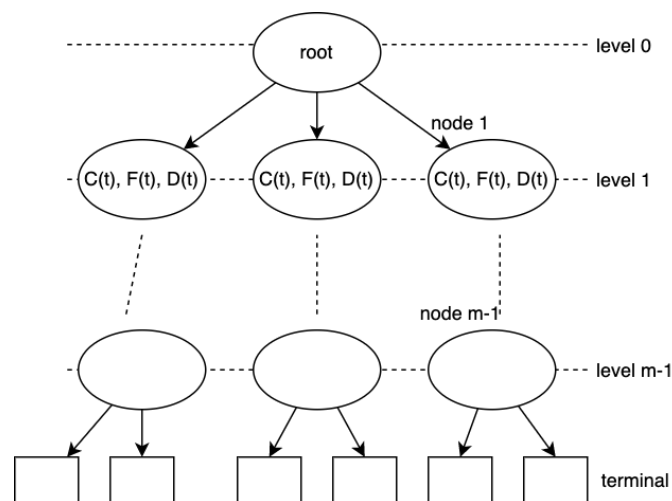
masalah. Visualisasi data, umumnya, menjadi metode yang sangat membantu dalam memahami data. Selain itu, ringkasan numerik (summary) dimanfaatkan untuk mengetahui ukuran pemusatan maupun penyebaran data.

c. *Data Preparation* (Persiapan Data)

Pada tahapan ini data mentah akan dibuat sebagai dataset yang akan dilakukan pembersihan data (*Data Cleaning*), melakukan pemilihan data (*Data Selection*), *record* dan atribut-atribut dan juga melakukan transformasi data (*Data Transformation*) untuk menjadikan masukan dalam tahap pemodelan.

d. *Modelling* (Pemodelan)

Tahapan ini akan secara langsung melibatkan *Machine Learning* untuk penentuan Teknik data *mining*, alat batu dan algoritma yang digunakan. Pada penelitian ini akan digunakan model klasifikasi *decision tree* dengan menggunakan algoritma CART. *Decision tree* secara khusus termasuk pada *supervised learning* yaitu ketika pembelajaran mesin memiliki *output*. Seperti halnya pohon, *decision tree* terdiri dari simpul akar (*root node*), cabang (*branches*), dan simpul daun (*leaf nodes*). Simpul akar (*root node*) adalah simpul yang menjadi orang tua (*parent*) untuk semua simpul pada *decision tree*. Setiap simpul dalam pohon keputusan menunjukkan variabel (*feature/attribute*), cabang merupakan keputusan, sementara daun (*leaf*) merupakan hasil (*outcome*) (Patel & Prajapati, 2018).



Gambar 2. Ilustrasi *Decision Tree*

Dengan $C(t)$ merupakan subset dari kelas, $F(t)$ merupakan variabel pada simpul tersebut, dan $D(t)$ merupakan keputusan pada cabang tersebut (Priyanka & Kumar, 2020).

Algoritma yang digunakan dalam klasifikasi pohon keputusan adalah ID3, C4.5, CART, dan lainnya. CART (*Classification and Regression Tree*) terdiri 1) membagi

kumpulan data secara rekursif untuk membangun pohon biner; 2) memangkas dengan fungsi kerugian minimal untuk menghindari pemasangan berlebihan (Sarkar et al., 2019).

Algoritma CART menggunakan *gini index* untuk menentukan variabel yang optimal yang digunakan untuk keputusan percabangan disebut juga sebagai *index* kemurnian (*purity/impurity*). Diketahui K kelas dalam klasifikasi, probabilitas suatu titik sampel termasuk ke dalam kelas k adalah p_k . *Gini index* untuk p_k adalah

$$Gini(p) = 1 - \sum_{k=1}^K p_k^2 \quad (1)$$

Untuk sejumlah sampel D , *gini index* didefinisikan:

$$Gini(p) = 1 - \sum_{k=1}^K \left(\frac{|C_k|}{|D|} \right)^2 \quad (2)$$

dengan C_k merupakan sampel bagian (*subset*) D yang termasuk ke kelas k dan K adalah banyaknya kelas (Tang & Zhang, 2020).

Jika sampel D tersegmentasi menjadi D_1 dan D_2 berdasarkan variabel ke- i , *gini index* untuk himpunan D didefinisikan sebagai:

$$Gini(D, i) = \frac{|D_1|}{|D|} Gini(D_1) + \frac{|D_2|}{|D|} Gini(D_2) \quad (3)$$

Selain *gini index*, proses percabangan dapat memanfaatkan nilai *entropy* yang didefinisikan sebagai berikut.

$$entropy = \sum_{k=1}^K -p_k \log_2(p_k) \quad (4)$$

Untuk sejumlah sampel D , didefinisikan:

$$entropy = \sum_{k=1}^K - \left(\frac{|C_k|}{|D|} \right) \log_2 \left(\frac{|C_k|}{|D|} \right) \quad (5)$$

CART akan membagi pohon biner menjadi dua anak simpul secara berulang. Algoritma yang digunakan:

1. Tentukan masing-masing variabel terbaik untuk pembagian. Untuk setiap variabel dengan K nilai yang berbeda, terdapat $K - 1$ kemungkinan untuk melakukan pembagian. Tentukan pembagian yang memaksimalkan kriteria pembagian. Hasil dari pembagian mengandung pembagi terbaik (satu untuk setiap variabel).
2. Tentukan *node* terbaik untuk pembagian. Diantara pembagi terbaik dari langkah pertama, pilih satu yang memaksimalkan kriteria pembagian. Kriteria tersebut menggunakan *gini index* atau *entropy*.
3. Bagi *node* menggunakan *node* pembagi terbaik dari langkah 2. Ulangi langkah 1 hingga mencapai kriteria berhenti.

e. *Evaluation* (Pengujian)

Tahapan ini untuk melihat tingkat performansi dari pola yang dihasilkan oleh metode klasifikasi. Suatu *confusion matrix* merupakan tabel yang digunakan untuk mengevaluasi performa klasifikasi (Ting, 2017)(Singh et al., 2021).

Tabel 1. Confusion Matrix

		Data prediksi	
		Kelas positif	Kelas negatif
Data aktual	Kelas positif	TP	FN
	Kelas negatif	FP	TN

dengan TP adalah *true positive* (benar dikategorikan ke kelas positif), TN adalah *true negative* (benar dikategorikan ke kelas negatif), FN adalah *false negative* (salah dikategorikan ke kelas negatif), dan FP adalah *false positive* (salah dikategorikan ke kelas positif). Berdasarkan *confusion matrix* tersebut, dapat ditentukan beberapa evaluasi seperti akurasi, *precision*, *recall*, dan *f1-score*.

$$akurasi = \frac{TP+TN}{TP+TN+FP+FN} \quad (6)$$

$$precision = \frac{TP}{TP+FP} \quad (7)$$

$$recall = \frac{TP}{TP+FN} \quad (8)$$

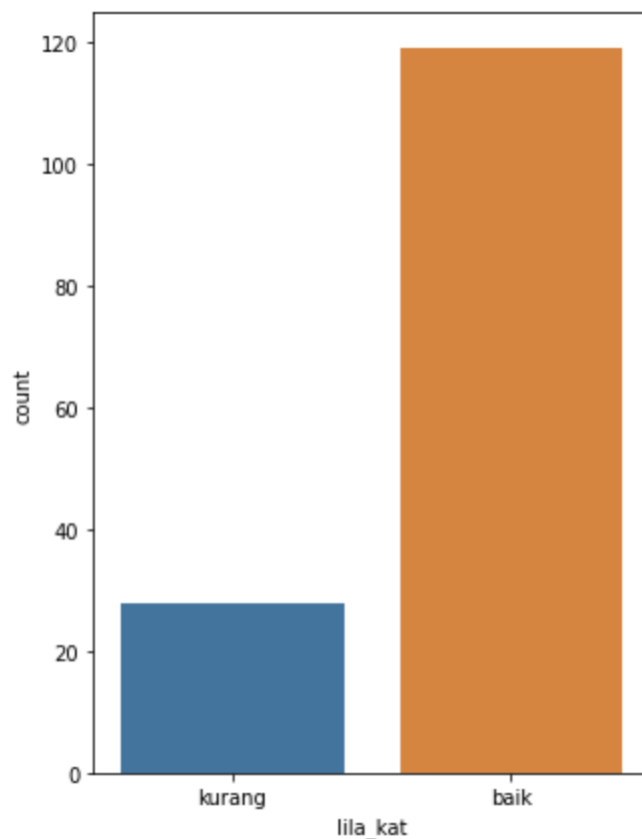
$$f1 - score = 2 * \frac{precision \times recall}{precision + recall} \quad (9)$$

f. *Deployment* (Penyebaran)

Pada tahapan ini akan dilakukan pembuatan laporan dan jurnal dengan menggunakan hasil yang telah didapatkan.

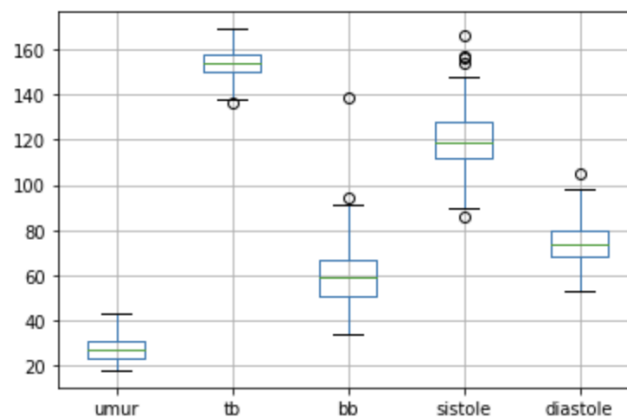
3. Hasil dan Pembahasan

Ibu hamil dengan pengukuran LILA pada kategori kurang jauh lebih sedikit daripada ibu hamil dengan LILA yang baik. Sebesar 19% (28 kasus dari 147) ibu hamil memiliki LILA yang kurang, sedangkan sisanya sebesar 81% (119 kasus dari 147) memiliki LILA yang baik. Kondisi tersebut divisualisasikan pada Gambar 3.



Gambar 3. Grafik Batang Frekuensi LILA

Boxplot pada Gambar 4 digunakan untuk mengetahui deskripsi variabel independen berupa ada atau tidaknya outlier, kesimetrisan distribusi, dan rentang nilai.



Gambar 4. Boxplot Variabel Independen

Berdasarkan Gambar 4., empat dari lima variabel memiliki pencilan yang digambarkan dengan titik-titik di luar boxplot. Terdapat satu ibu dengan tinggi badan rendah (pencilan), dua ibu memiliki berat badan yang terlalu tinggi (kecenderungan mengalami obesitas), dan tiga ibu memiliki risiko tekanan darah tinggi.

Sebanyak 147 data ibu hamil dibagi menjadi data training dan data testing dengan berturut-turut sebanyak 80%:20%. Dengan kata lain terdapat 117 data digunakan untuk membangun model decision tree dan 30 data digunakan untuk menguji performa model.

Fungsi obyektif dalam optimalisasi klasifikasi LILA adalah akurasi. Model terbaik dipilih berdasarkan akurasi terbesar dengan kombinasi parameter-parameter yang ditunjukkan pada Tabel 2.

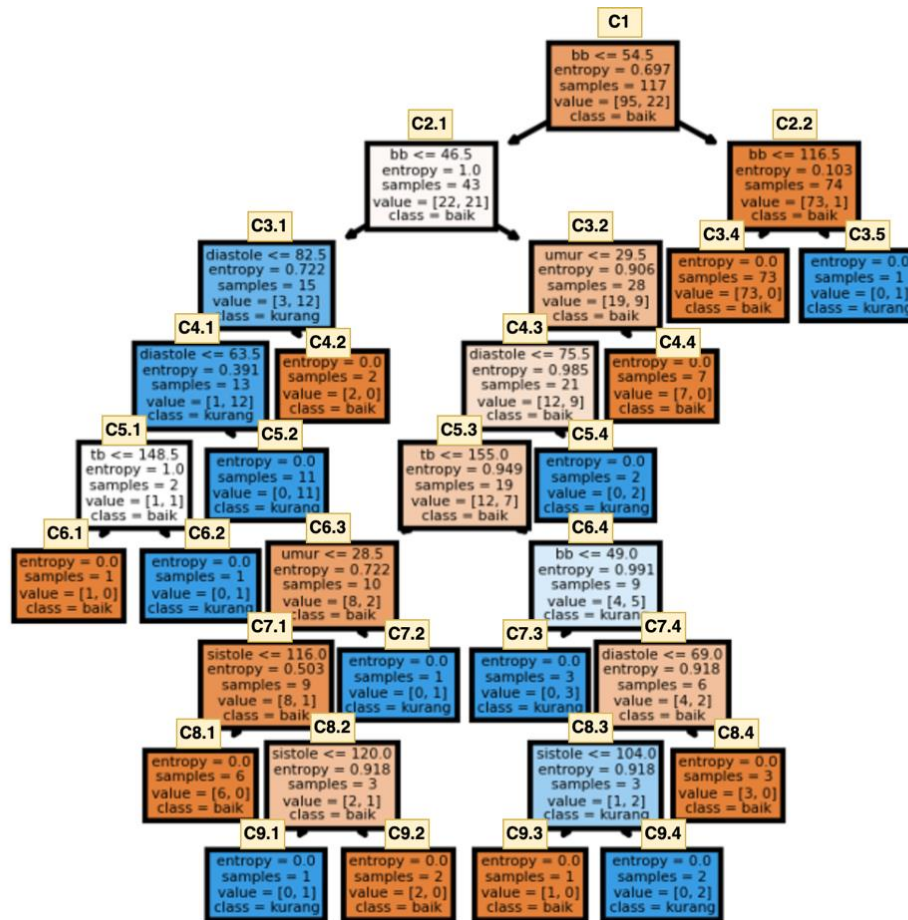
Tabel 2. Parameter Decision Tree

No	Parameter	Nilai	Nilai Terbaik
1.	Kriteria	Gini, <i>entropy</i>	<i>Entropy</i>
2.	Minimum sample untuk membagi simpul (<i>splitting</i>)	2, 3, 4, 5	2
3.	Jumlah minimum sampel yang diperlukan untuk berada pada simpul daun (<i>leaf node</i>).	1, 2, 3	1
4.	Strategi yang digunakan untuk memilih split pada setiap node	<i>Best, random</i>	<i>best</i>

Dengan menggunakan model decision tree semua kelas diklasifikasikan dengan benar, sehingga akurasi untuk data training mencapai 100%. Demikian juga untuk skor precision and recall, mencapai 100%. Model tersebut digambarkan dalam bentuk node dan cabangnya seperti pada Gambar 5.

Dalam setiap node di Gambar 5, terdapat informasi mengenai nilai entropy (entropy), jumlah sampel (sample), banyaknya data setiap kategori (value), dan kategorinya (class).

1. Node pembagi (split) pertama (C1) merupakan variabel *bb* (Berat Badan) dengan banyak sampel yang akan diklasifikasi sebesar 117, terbagi dalam kategori “baik” sebanyak 95 dan kategori “kurang” sebanyak 22. Nilai entropy pada node pertama yaitu 0.697. Jika $bb \leq 54.5$ maka diarahkan ke cabang C2.1, sebaliknya ke cabang C2.2.
2. Di node C2.1, pembagi masih menggunakan informasi mengenai berat badan. Di node C2.1 ini terdapat 43 sampel yang memenuhi kriteria $bb \leq 54.5$. Sementara di node C2.2, terdapat 74 sampel yang memenuhi negasinya ($bb > 54.5$). Pada node C2.1, terdapat 22 sampel kategori “baik” dan 21 sampel kategori “kurang”. Node ini melakukan pembagian dengan nilai $bb \leq 46.5$. Jika sampel memenuhi kondisi tersebut, maka akan masuk ke percabangan ketiga C3.1. Jika tidak, maka akan masuk ke percabangan ketiga C3.2.



Gambar 5. Model Decision Tree

3. Node C3.2 menggunakan nilai variabel $\text{diastole} \leq 82.5$ untuk melakukan pembagian. Terdapat sebanyak 15 sampel dalam node ini (sampel tersebut yang memenuhi syarat pada cabang sebelumnya yaitu $bb \leq 46.5$). Pada node ini, terdapat 3 data dengan kategori “baik” dan 12 data dengan kategori “kurang”. Jika sampel memenuhi kriteria hingga sampai pada node ini, maka LILA dikategorikan “kurang” (mayoritas kelas).
4. Pada percabangan keempat node C4.2, terdapat kondisi yang tidak mungkin dilakukan percabangan lagi yaitu tidak memenuhi $\text{diastole} \leq 82.5$ memiliki kategori baik. Jika diurut dari awal: jika sampel memenuhi $bb < 54.5$ dan $bb < 46.5$, serta $\text{diastole} > 82.5$, maka kategori “baik”. Demikian model pohon tersebut dibentuk hingga mencapai cabang akhir yaitu 9.

Berdasarkan Gambar 5, aturan terpendek yaitu ketika berat badan lebih besar daripada 54.5 dan lebih lebih besar daripada 116.5, maka akan dikategorikan LILA “kurang”. Sebaliknya, jika berat badan lebih besar daripada 54.5 dan kurang dari 116.5, maka akan dikategorikan LILA “baik”.

Model pohon tersebut diuji dengan menggunakan data testing sebanyak 30. Confusion matrix yang diperoleh pada pengujian ditunjukkan pada Tabel 3.

Tabel 3. Confusion Matrix

Kategori	Prediksi LILA “baik”	Prediksi LILA “kurang”
Aktual LILA “baik”	22	2
Aktual LILA “kurang”	1	5

Berdasarkan Tabel 3., banyaknya LILA kategori baik yang benar diprediksi sebesar 22 (True Positive) dan banyaknya LILA kategori “kurang” yang benar diprediksi sebesar 5 (True Negative). LILA kategori “baik” yang diprediksi “kurang” sebanyak 2 (False Negative). LILA kategori “kurang” yang diprediksi baik sebanyak 1 (False Positive). Jika diambil kelas positif adalah LILA kategori baik, maka:

$$\text{akurasi} = \frac{22 + 5}{30} = 0.9$$

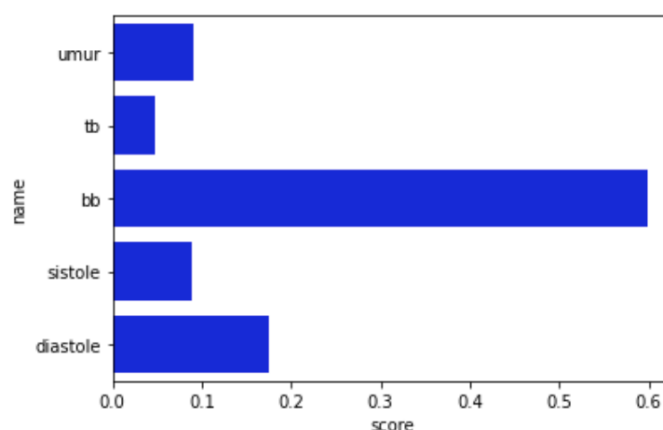
$$\text{precision} = \frac{22}{22 + 1} = 0.96$$

$$\text{recall} = \frac{22}{22 + 2} = 0.92$$

$$\text{f1 - score} = 2 * \frac{0.96 \times 0.92}{0.96 + 0.92} = 0.94$$

Akurasi merupakan kemampuan model untuk memprediksi secara benar ke kategori masing-masing. Proporsi prediksi benar pada kategori positif terhadap keseluruhan sampel yang diprediksi positif. Recall yang disebut juga sensitivitas klasifikasi merupakan proporsi yang benar diprediksi ke kelas positif terhadap keseluruhan data positif. Nilai f1-score merupakan rata-rata harmonis dari precision dan recall. Keempat nilai akurasi yang tinggi menunjukkan model baik dalam melakukan klasifikasi.

Variabel yang signifikan dalam melakukan klasifikasi ditunjukkan dengan skor variable importance. Berat badan merupakan variabel terpenting dalam mengkategorikan LILA, disusul dengan diastole, umur, sistole, dan tinggi badan.



Gambar 6. Skor Variable Importance

4. Kesimpulan dan Saran

Berdasarkan akurasi yang tinggi, metode decision tree dapat mengklasifikasikan LILA dengan baik. Sementara tingginya nilai recall dan precision menjadi indikasi bahwa ketidakseimbangan kategori tidak menimbulkan kecenderungan klasifikasi ke kelas mayor (LILA baik). Berdasarkan aturan yang terbentuk, ibu hamil yang mengalami obesitas (berat badan lebih dari 116.5 kg) memiliki LILA yang kurang. Variabel yang paling signifikan dalam mengkategorikan LILA adalah berat badan. Hal ini mendukung anggapan secara umum bahwa LILA mungkin dipengaruhi berat badan.

Evaluasi model decision tree dengan kategori tidak berimbang untuk studi kasus kategorisasi LILA sudah sangat baik. Penambahan jenis kategori LILA (menjadi kurang, sedang, dan baik) dapat membantu identifikasi awal LILA ibu hamil, sehingga dapat melakukan pencegahan LILA kurang sejak dini. Apabila model LILA ingin dimanfaatkan oleh pengguna/pihak lain, maka perlu dibangun user interface berbasis website atau aplikasi smartphone.

5. Ucapan Terima Kasih

Terima kasih kepada dr. Prisa Surya Hadiningtyas yang membantu proses pengambilan data di Puskesmas Sumber, Kecamatan Sumber, Kabupaten Rembang.

Pustaka

- Ardiyanti, M. (2019). *Klasifikasi Masalah Gizi Menggunakan Metode Decision Tree Algoritma Iterative Dichotomiser 3 (ID3) (Studi Kasus Status Gizi Balita Indonesia Tahun 2017)*. Universitas Brawijaya.
- Auliana, U., Iskari, N., & Tiurma, H. (2016). Hubungan Usia, Tingkat Pendidikan, Status Ekonomi, Pekerjaan, dan Asupan Zat Gizi Makro dengan Status Gizi Ibu Hamil di Provinsi Papua dan Papua Barat. *Nutrire Diaita*, 8(1), 9–17.

- Badan Keahlian Sekretariat Jendral DPR RI. (2021). Analisis RKP dan Pembicaraan Pendahuluan APBN. In *Pusat Kajian Agggaran Badan Keahlian DPR RI*. https://www.neliti.com/id/publications/218225/kemajuan-teknologi-informasi-dan-komunikasi-dalam-industri-media-di-indonesia%0Ahttp://leip.or.id/wp-content/uploads/2015/10/Della-Liza_Demokrasi-Deliberatif-dalam-Proses-Pembentukan-Undang-Undang-di-Indonesia
- Kamaruddin, M., Hasrawati, Usmia, S., Jusni, Misnawaty, & Handayani, I. (2019). Korelasi Antara Status Gizi Dan Kadar Hemoglobin Pada Kejadian Anemia Ibu Hamil Trimester Iii. *Medika Alkhairaat: Jurnal Penelitian Kedokteran Dan Kesehatan*, 1(3), 82–88. <https://doi.org/10.31970/ma.v1i3.32>
- Lubis, Z. (2003). Hubungan Pengetahuan Gizi dan Asupan Zat Gizi (Protein,Vitamin C,Zat Besi dengan Kejadian Anemia Pada Ibu Hamil. In *Program Pascasarjana Institut Pertanian Bogor*. <http://ojs.yapenas21maros.ac.id/index.php/jika/article/view/1>
- Mariyatul, Triawanti, & Noor, M. S. (2014). Hubungan Statsu Gizi Ibu Hamil dengan Kejadian Bayi Lahir Mati di Kabupaten Banjar Periode 2011-2012. *JPKMI Jurnal Publikasi Kesehatan Masyarakat Indonesia*, 1(1), 52–59. <https://ppjp.ulm.ac.id/journal/index.php/JPKMI/article/view/601>
- Patel, H. H., & Prajapati, P. (2018). Study and Analysis of Decision Tree Based Classification Algorithms. *International Journal of Computer Sciences and Engineering*, 6(10), 74–78. <https://doi.org/10.26438/ijcse/v6i10.7478>
- Pratama, G. R. (2020). *Studi Literatur: Faktor-Faktor yang Mempengaruhi Terjadinya Kurang Energi Kronis (KEK) Pada Ibu Hamil* [Universitas Bhakti Kencana]. <https://doi.org/10.1016/j.tmaid.2020.101607%0Ahttps://doi.org/10.1016/j.ijisu.2020.02.034%0Ahttps://onlinelibrary.wiley.com/doi/abs/10.1111/cjag.12228%0Ahttps://doi.org/10.1016/j.ssci.2020.104773%0Ahttps://doi.org/10.1016/j.jinf.2020.04.011%0Ahttps://doi.o>
- Priyanka, & Kumar, D. (2020). Decision tree classifier: a detailed survey. *Int. J. Information and Decision Sciences*, 12(3), 246–269.
- Putri, N. K. S. E., & Hastutik. (2017). Hubungan Ukuran Lingkar Lengan Atas (LiLA) Ibu Dan Peningkatan Berat Badan Selama Kehamilan Dengan Berat Badan Lahir Bayi Di Kabupaten Karanganyar. *Jurnal Maternal*, 2(1), 1–6.
- Putri, N. K. S. E., & Hastutik. (2018). Hubungan Usia Ibu dengan Peningkatan Berat Badan selama Hamil di Kabupaten Karanganyar. *Ejurnal.Stikesmhk.Ac.Id*, II(4),

255–258.

https://ejurnal.stikesmhk.ac.id/index.php/jurnal_ilmiah_maternal/article/download/704/623

- Sarkar, S., Raj, R., Vinay, S., Maiti, J., & Pratihari, D. K. (2019). An optimization-based decision tree approach for predicting slip-trip-fall accidents at work. *Safety Science*, 118(May), 57–69. <https://doi.org/10.1016/j.ssci.2019.05.009>
- Singh, P., Singh, N., Singh, K. K., & Singh, A. (2021). Diagnosing of disease using machine learning. *Machine Learning and the Internet of Medical Things in Healthcare*, 89–111. <https://doi.org/10.1016/B978-0-12-821229-5.00003-3>
- Tang, R., & Zhang, X. (2020). CART Decision Tree Combined with Boruta Feature Selection for Medical Data Classification. *2020 5th IEEE International Conference on Big Data Analytics, ICBDA 2020*, 80–84. <https://doi.org/10.1109/ICBDA49040.2020.9101199>
- Ting, K. M. (2017). Confusion Matrix. In G. I. Webb (Ed.), *Encyclopedia of Machine Learning and Data Mining* (pp. 260–261). Springer, Boston, MA. <https://doi.org/10.1007/978-1-4899-7687-1>