

## REKOMENDASI KESEHATAN JANIN DENGAN PENERAPAN ALGORITMA C5.0 MENGGUNAKAN *CLASSIFYING CARDIOTOGRAPHY DATASET*

Muhamad Rian Santoso<sup>1)</sup>, Purnawarman Musa<sup>2)</sup>

<sup>1,2</sup> Magister Manajemen Sistem Informasi, Universitas Gunadarma

Jl. Margonda Raya No. 100, Depok 16424, Jawa Barat

E-mail : <sup>1</sup>mrians21@gmail.com, <sup>2</sup>p\_musa@staff.gunadarma.ac.id

### ABSTRAK

Memiliki buah hati yang sehat tidak luput dari faktor kesehatan sang ibu dan kondisi janin di dalam rahim, sehingga butuh analisa terhadap kesehatan janin pada setiap ibu hamil. Penelitian ini mengusulkan Algoritma C5.0 memanfaatkan *dataset Cardiotocography* terkait kondisi janin. *Dataset Cardiotocography* terdiri dari 2.126 *record* dengan, dimana setiap *record* memiliki 22 kolom atribut dan terdapat 3 kelas klasifikasi, yaitu; *normal*, *suspect*, dan *pathological*. Dengan menghitung *entropy*, *information gain*, *split information*, dan *gain ratio*, serta menggunakan *confusion matrix* sebagai perbandingan akurasi dari data yang diteliti. Kelompok *record* dibagi menjadi data *training* dan data *testing* dengan variasi 60%, 70%, 80%, dan 90% untuk data *training* dan dengan 40%, 30%, 20%, 10% untuk data *testing*. Hasil dari rekomendari pada variasi pembagian 90% data *training* dan 10% data *testing* menghasilkan akurasi sebesar 93,40% dengan jumlah aturan sebanyak 257. Sedangkan variasi 80% data *training* dan 20% data *testing* menghasilkan akurasi sebesar 91,29% dengan jumlah aturan sebanyak 239. Pada variasi 70% data *training* dan 30% data *testing* menghasilkan akurasi sebesar 88,23% dengan jumlah aturan sebanyak 220. Dan variasi pembagian terkecil 60% data *training* dan 40% data *testing* menghasilkan akurasi sebesar 88,12% dengan jumlah aturan sebanyak 204. Berdasarkan variasi tersebut, maka dapat disimpulkan semakin besarnya data *training* akan menyebabkan akurasi menjadi semakin baik dengan jumlah aturan-aturan yang dapat berguna untuk dijadikan sebagai sistem penunjang keputusan.

**Kata kunci :** *Algoritma C5.0, Confusion Matrix, Dataset Cardiotocography, Janin, Klasifikasi.*

### ABSTRACT

Having a healthy baby cannot be separated from the health factors of the mother and the condition of the fetus in the womb, so it is necessary to analyze the fetal health of every pregnant woman. This study proposes the C5.0 Algorithm utilizing the Cardiotocography dataset related to fetal conditions. The Cardiotocography dataset consists of 2,126 records, where each record has 22 attribute columns, and there are three classification classes: *normal*, *suspect*, and *pathological*. By calculating *entropy*, *information gain*, *split information*, and *gain ratio*, and using *confusion matrix* to compare the accuracy of the data under study. The record group is divided into training data and testing data with variations of 60%, 70%, 80%, and 90% for training data and with 40%, 30%, 20%, 10% for testing data. The recommendations on the variation 90% of training data and 10% of testing data, which accuracy of 93,40% with 257 rules. While the variation of 80% of training data and 20% of testing data resulted in an accuracy of 91,29% with 239 rules. 70% of training data and 30% of testing data yields an accuracy of 88,23% with several rules of 220. Furthermore, the most negligible distribution variation of 60% training data and 40% testing data results in an accuracy of 88,12% with several rules of 204. Based on these variations, it can be is concluded that the greater the training data, the better the accuracy with the number of rules that can be used as a decision support system.

**Keywords:** *C5.0 Algorithm, Cardiotocography Dataset, Classification, Confusion Matrix, Fetus.*

## PENDAHULUAN

Memiliki seorang buah hati kondisi sehat merupakan salah satu impian besar bagi seorang ibu. Kondisi ibu yang tidak sehat dapat mengancam kesehatan janin di dalam rahim. Sebaliknya, janin yang tidak sehat dapat mengancam kesehatan ibu yang sedang mengandung, sehingga sang ibu memiliki kemungkinan mengalami keguguran, bahkan kematian pada saat melahirkan.

Berdasarkan analisa [1], angka kematian ibu hamil tahun 2010 berjumlah 346 jiwa, dan tahun 2015 berjumlah 305 jiwa dari 100.000 kelahiran hidup (KH). Sedangkan angka kematian bayi tahun 2012 adalah 32 jiwa dan tahun 2017 berjumlah 24 jiwa dari 1.000 kelahiran hidup (KH). Menurut POGI (Persatuan Obstetri Ginekologi Indonesia) terdapat 15% hingga 20% mengalami keguguran dari jumlah total ibu hamil. Bahkan beberapa ibu yang mengandung sering mengalami keguguran lebih dari 2 (dua) kali. Umumnya ibu yang sedang hamil tidak mengetahui kesehatan dan kondisi janin yang sedang dikandungnya. Faktor keguguran lainnya disebabkan ibu hamil kurang pengetahuan kesehatan janin, sehingga menimbulkan janin mengalami keguguran. Untuk mencegah keguguran, ibu hamil dapat melakukan pemeriksaan melalui laboratorium dengan metode *Cardiotography*.

Metode *Cardiotography* (CTG) proses pemeriksaan kesehatan janin pada ibu yang sedang hamil. CTG adalah alat untuk memantau aktivitas, denyut jantung janin, kontraksi rahim dan beberapa elemen yang dibutuhkan saat janin berada di dalam kandungan. CTG merupakan metode terbuka untuk pasien yang dianjurkan dokter berdasarkan indikasi medis secara berkala saat mengalami kondisi seperti demam tinggi, hipertensi atau diabetes. Hasil Laboratorium akan didiagnosa oleh dokter untuk mengevaluasi kondisi janin apakah dalam keadaan sehat baik sebelum dan selama persalinan. Misalnya, ketika kesehatan janin dari seorang ibu hamil mengalami perubahan pada denyut jantung janin

maupun kontraksi rahim pada ibu hamil, dengan menggunakan alat CTG petugas (dokter dan bidan) dapat mendeteksi dini kesehatan janin. Proses pemantauan kesehatan janin, serta kewaspadaan dan pada kondisi tertentu melakukan tindakan medis menghindari keguguran, bahkan dapat mengurangi resiko kematian bagi ibu dan janin [2].

Namun hasil uji laboratorium dari CTG terhadap kesehatan janin memerlukan seorang pakar kesehatan janin. Pemanfaatan sistem pendukung keputusan terhadap hasil uji CTG, maka hasil pemeriksaan CTG dapat dievaluasi secara sistem terkait kondisi janin dan merekomendasi secara dini dari sistem pendukung keputusan terhadap kondisi kesehatan janin.

Penelitian terkait kesehatan janin secara pengujian dan analisa laboratorium atau melakukan dengan penerapan sistem pendukung keputusan yang berbasis teknologi informasi telah dilakukan oleh para peneliti terdahulu. Penelitian oleh [3] menggunakan Algoritma C5.0 dalam mengklasifikasi risiko kehamilan. Hasil penelitian tersebut menerapkan proses pemangkasan yang menghasilkan lebih sedikit set aturan (*rule sets*) dan akurasi metode C5.0 lebih baik dari C4.5.

Menurut [4], menyimpulkan pohon keputusan dapat melakukan *Cardiotocography* fitur seleksi dan mengklasifikasikan risiko kehamilan dengan tingkat akurasi yang baik. Fitur seleksi terbukti cukup baik dan mampu meningkatkan akurasi hasil hingga 98,7%. Tahapan-tahapan pada penelitian tersebut adalah menghitung angka *information gain* tertinggi untuk menentukan akar (*root*), selanjutnya menentukan cabang (*branch*) menggunakan perhitungan *information gain* tertinggi setelah akar (*root*). *Node* cabang yang diklasifikasikan pada kelas tertentu akan menghasilkan simpul terminal dan klasifikasi benar atau salah menggunakan matriks konfusi.

Menurut [5], pohon keputusan dengan C4.5 menyatakan *Cardiotocography* dapat mengidentifikasi janin kekurangan oksigen (hipoksia) menggunakan *public dataset* pada repositori "The Data Mining

Repository of University of California Irvine (UCI)", terbagi dua fase, yaitu fase *training* (membuat pohon keputusan) dan fase klasifikasi. Hasil pohon keputusan Algoritma C4.5 memilih atribut sesuai jumlah *entropy*. Hasil analisa dan akurasi data *Cardiotography* sebesar 92%.

Pada penelitian ini bertujuan mendiagnosa suatu ketidakpastian terkait kondisi kesehatan janin yang dikandung seorang ibu hamil memanfaatkan *Cardiotography* menggunakan Algoritma C5.0 sebagai solusi merekomendasi kondisi kesehatan janin. Metode C5.0 menghemat lebih banyak memori dalam membentuk atau merancang pohon keputusan dibandingkan metode lain. Algoritma C5.0 merupakan metode klasifikasi variabel respon berdasarkan variabel *predictor*-nya dan Algoritma C5.0 dikenal lebih akurat daripada Algoritma C4.5. Pengembangan dan pengambilan keputusan untuk aturan-aturan proses pengklasifikasian pohon keputusan memanfaatkan Algoritma C5.0 meminimalisi tingkat keguguran dan resiko kematian janin yang terjadi di dunia terutama di Indonesia.

Perancangan klasifikasi pada kasus kesehatan janin mendapatkan data melalui CTG, dimana menggunakan Algoritma C5.0 berdasarkan pohon keputusan dan aturan terkait kondisi janin menggunakan persamaan (1-4). *Entropy* dapat menentukan seberapa informatif atribut yang akan digunakan. Jika nilai *entropy* besar maka nilai *information gain* akan rendah. Nilai dari *information gain* tertinggi yang dijadikan akar pohon keputusan. Namun penelitian ini, *gain ratio* tertinggi yang akan dijadikan akar pohon keputusan dan berfungsi mengatasi masalah atribut bertipe kategorikal. Selain itu *Gain ratio* sebagai peningkatan akurasi dibandingkan dengan pemilihan akar pohon keputusan berdasarkan *information gain*.

## METODE

Metode penelitian yang akan digunakan terdiri dari beberapa proses yang dapat dilihat pada Gambar 1.



Gambar 1. Langkah metode penelitian

### 1. Data collection.

Data yang dikumpulkan/dikoleksi harus sesuai dengan metode terkait yang akan digunakan, hal ini berpengaruh terhadap akurasi dan kualitas informasi yang diperoleh. Data dapat diperoleh dengan berbagai jenis, antara lain; Data Primer yang didapatkan secara langsung dengan melakukan teknik wawancara, maupun melakukan survey ke lapangan. Sedangkan Data Sekunder yang didapatkan melalui buku, jurnal, maupun media lain di internet. Kedua jenis data tersebut berupa *dataset* kondisi kesehatan janin yang telah diekstrasi menggunakan *Cardiotocograms* yang dapat diperoleh pada situs kaggle.com.

### 2. Data preprocessing.

Data yang diperoleh/dikoleksi sebelum diolah pada algoritma C5.0, maka perlu dilakukan *preprocessing*. Data *preprocessing* perlu dilakukan ketika data mempunyai nilai kosong, data mempunyai ketidaklengkapan atribut, data mempunyai nilai yang salah, dan data mengandung perbedaan kode atau nama

### 3. Splitting data.

Data yang sudah melewati tahap *preprocessing* akan dibagi dengan 4 (empat) uji coba, yaitu dengan membagi 60%, 70%, 80%, dan 90% data training dan 40%, 30%, 20%, 10% data testing.

### 4. Algoritma C5.0 implementation.

Algoritma C5.0 akan diimplementasikan dengan menghitung nilai *entropy*, *information gain*, *split information*, dan *gain ratio*, sehingga hasil pada tahap ini akan berupa pohon keputusan (*decision tree*)

Nilai *entropy* merupakan perhitungan *entropy* S adalah himpunan kasus, jumlah partisi S pada k dan  $p_j$  adalah probabilitas dari jumlah (Normal/Suspect/Pathological) dibagi Total Kasus persamaan 1 [6].

$$Ent(S) = \sum_{j=1}^k -p_j \log_2 p_j \quad (1)$$

Nilai *gain information* merupakan perhitungan *gain information*, dimana S adalah himpunan kasus, A adalah atribut,

$n$  adalah jumlah partisi atribut  $A$ ,  $|S_i|$  adalah jumlah kasus pada partisi ke- $i$ , dan  $|S|$  adalah jumlah kasus dalam  $S$  ditunjukkan pada persamaan 2 [6].

$$Gain(S, A) = Ent(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Ent(S_i) \quad (2)$$

Nilai *split information* merupakan perhitungan *split information* dapat digunakan ketika nilai  $S$  adalah ruang (data) sample yang digunakan untuk training,  $A$  adalah Atribut, dan  $S_i$  adalah jumlah sample untuk atribut  $i$  pada persamaan 3 [6].

$$SplitInf(S, A) = - \sum_{i=1}^n \frac{S_i}{S} * \log_2 \frac{S_i}{S} \quad (3)$$

Nilai *gain ratio* yang digunakan pada persamaan 4 [6], dimana  $A$  adalah Atribut,  $Gain(A)$  adalah *Information Gain* pada atribut  $A$  dan  $Split(A)$  adalah *Split Information* pada atribut  $A$ .

$$GainRatio(A) = \frac{Gain(A)}{Split(A)} \quad (4)$$

*Evaluation with confusion matrix.*

Adalah hasil evaluasi yang didapat pada tabel *matrix* yang dapat dilihat pada Tabel 1 di bawah ini.

Tabel 1. Contoh tabel matrix.

Prediksi	Aktual Positif	Aktual Negatif
Positif	TP	FP
Negatif	FN	TN

Keterangan Tabel 1:

- $TP$  (*True Positive*) adalah total kasus positif yang diklasifikasi benar.
- $TN$  (*True Negative*) adalah total kasus negatif yang diklasifikasi benar.
- $FP$  (*False Positive*) adalah total kasus positif yang diklasifikasi salah.
- $FN$  (*False Negative*) adalah total kasus negatif yang diklasifikasi salah.

Untuk menghitung nilai akurasi dalam *confusion matrix* maka dapat digunakan persamaan 5 di bawah ini [7].

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (5)$$

Keterangan pada persamaan 5 di atas dapat dilihat pada keterangan Tabel 1.

## HASIL DAN PEMBAHASAN

Pembahasan dilakukan dengan urutan metode penelitian yang telah dijelaskan di atas.

### a. Data collection.

Pengumpulan data yang digunakan untuk melakukan klasifikasi kesehatan janin dengan memanfaatkan metode Algoritma C5.0 diperoleh dari *public dataset* yang terdapat pada situs *kaggle.com*. *Dataset* tersebut memiliki 2.126 *record* dengan jumlah 22 kolom sebagai atribut. Berikut ini adalah penjelasan setiap kolom dari data *record* yang telah dikumpulkan akan dijadikan atribut untuk memprediksi kesehatan janin menggunakan Algoritma C5.0.

1. **Atribut *baseline\_value*** berisi data denyut jantung janin atau FHR (*Fetal Health Rate*) dengan satuan numerik. FHR pada janin normal umumnya berkisar antara 110 sampai dengan 160. Sedangkan FHR di bawah 110 diklasifikasikan *Fetal Bradycardia*. Jika di atas 160 diklasifikasikan *Fetal Tachcardia* [8]. *Fetal Tachcardia* adalah peningkatan detak jantung janin yang tidak normal. *Fetal Bradycardia* adalah rendahnya detak jantung janin secara tidak normal.
2. **Atribut *accelerations*** berisi data jumlah akslerasi janin per detik dengan satuan numerik. Pada kasus ini, atribut *accelerations* akan dibagi menjadi 2 (dua) kategori, antara lain bernilai *True* jika janin melakukan akselerasi lebih dari 0 detik, dan *False* jika janin tidak melakukan akselerasi saat dilakukan *Cardiotocography*.
3. **Atribut *fetal\_movement*** berisi data jumlah gerakan janin per detik dengan satuan numerik. Pada kasus ini, atribut *fetal\_movement* akan dibagi menjadi 2 (dua) kategori, bernilai *True* jika janin melakukan pergerakan lebih dari 0 detik, dan *False* jika janin tidak melakukan pergerakan saat pengujian *Cardiotocography*.
4. **Atribut *uterine\_contractions*** berisi data jumlah kontraksi uterus per detik dengan satuan numerik. Atribut *uterine\_contractions* dibagi menjadi 2 (dua) kategori, antara lain bernilai *True* jika terdapat kontraksi *uterine* pada janin lebih dari 0 detik, *False* jika janin tidak terdapat kontraksi *uterine* saat uji *Cardiotocography*.

5. **Atribut *light decelerations*** berisi data jumlah LD per detik dengan satuan numerik. Pada atribut dibagi 2 (dua) kategori, antara lain bernilai *True* jika terdapat tingkat deselerasi ringan pada janin lebih dari 0 detik, *False* jika janin tidak terdapat tingkat deselerasi ringan saat dilakukan *Cardiotocography*.
  6. **Atribut *severe decelerations*** berisi data jumlah SD per detik dengan satuan numerik. Atribut terdiri 2 (dua) kategori, antara lain bernilai *True* jika terdapat tingkat deselerasi yang berat pada janin lebih dari 0 detik, dan *False* jika janin tidak terdapat tingkat deselerasi yang berat saat dilakukan *Cardiotocography*.
  7. ***prolongued decelerations*** berisi data jumlah PD per detik dengan satuan numerik dan dibagi menjadi 2 (dua) kategori, bernilai *True* jika terdapat tingkat deselerasi yang berkepanjangan pada janin lebih dari 0 detik, dan *False* jika janin tidak terdapat tingkat deselerasi yang berkepanjangan saat dilakukan *Cardiotocography*.
  8. ***Abnormal short term variability*** Atribut berisi data persentase waktu dengan variabilitas jangka Panjang yang abnormal dalam satuan numerik.
  9. ***Mean value of short term variability*** Atribut berisi data rata-rata variabilitas jangka pendek dalam satuan numerik
  10. ***Percentage of time with abnormal long term variability*** Atribut berisi data persentase variabilitas jangka panjang yang abnormal dalam satuan numerik.
  11. ***Mean value of long term variability*** Atribut berisi data nilai rata-rata variabilitas jangka panjang dalam satuan numerik.
  12. **Atribut *histogram width*** berisi data lebar histogram yang dibuat menggunakan semua nilai dari sebuah *record* dalam satuan numerik.
  13. **Atribut *histogram min*** berisi data nilai minimum histogram dalam satuan numerik.
  14. **Atribut *histogram max*** berisi data nilai maksimum histogram dalam satuan numerik.
  15. ***Histogram number of peaks*** Atribut berisi tentang data jumlah puncak dalam pengujian histogram dalam satuan numerik.
  16. ***Histogram number of zeroes*** Atribut berisi tentang data jumlah angka nol dalam pengujian histogram dalam satuan numerik.
  17. **Atribut *histogram mode*** berisi tentang data mode histogram dalam satuan numerik.
  18. **Atribut *histogram mean*** berisi tentang data rata-rata histogram dalam satuan numerik.
  19. **Atribut *histogram median*** berisi tentang data nilai median histogram dalam satuan numerik.
  20. **Atribut *histogram variance*** berisi tentang data perbedaan histogram dalam satuan numerik.
  21. **Atribut *histogram tendency*** berisi tentang data kecenderungan histogram dalam satuan numerik.
  22. **Atribut *fetal health*** berisi tentang data kesehatan janin dalam satuan numerik.
- Pada atribut nomor 8 sampai nomor 20, nilai suatu data tidak dikategorikan khusus dikarenakan tidak adanya aturan klasifikasi tertentu pada nilai-nilai tersebut. Namun kelebihan Algoritma C5.0 untuk menangani suatu atribut jika terdapat atribut kontinyu atau jika terdapat *missing value* pada suatu atribut dalam *record* dengan cara mencari nilai tengah atau rata-rata. Jadi pengkategorian akan dilakukan berdasarkan nilai rata-rata dari sebuah *record*.

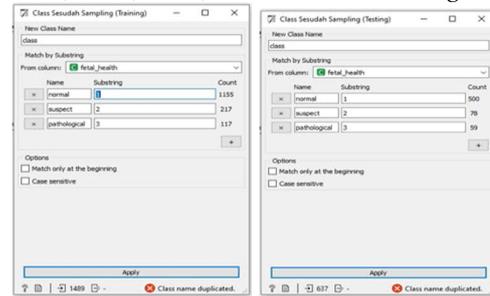
#### **b. Data preprocessing.**

*Public dataset* dengan jumlah dalam 22 kolom tersebut hanya mempunyai nilai atribut numerik, sedangkan untuk mengimplementasikan klasifikasi dengan menggunakan algoritma C5.0 data atribut numerik harus diubah ke dalam sebuah kategori data sehingga data akan menjadi bentuk yang paling sesuai dalam

mengimplementasikan algoritma C5.0. Hal tersebut menggunakan teknik umum pada data *preprocessing* yaitu data *transformation*.

**c. Splitting data.**

Penggunaan *public dataset* yang akan diteliti 2.126 data terdiri dari 1.655 data kelas yang bernilai normal, 295 data kelas yang bernilai *suspect*, dan 176 data kelas bernilai *pathological*. *Public dataset* diuji sebanyak 4 (empat) kali percobaan yaitu 60%, 70%, 80%, dan 90% data *training*, sedangkan data testing sebesar 40%, 30%, 20%, dan 10% data *testing*.



a). Data training                      b). Data testing

Gambar 2. Data klasifikasi dengan tools orange

Pada penelitian ini, misal proses *splitting* data yang digunakan adalah 70% untuk data *training* dan 30% untuk data *testing* (lihat Gambar 2). Hasil sampling sebesar 70% pada data *training* dengan tools Orange v-3.28 dengan tipe sampling *fixed proportion of data* menghasilkan 1.489 data dari 2.126 *dataset*. Gambar 2.a merupakan hasil total kelas normal sebanyak 1.155 data, total kelas *suspect* sebanyak 217 data, dan total kelas yang *pathological* sebanyak 117 data dari sampel *training* sebesar 1.489 data.

Sedangkan hasil sampling sebesar 30% pada data *testing* berjumlah 637 data dari 2.126 *dataset*. Pada Gambar 2.b menunjukkan hasil dari total kelas yang bernilai normal sebanyak 500 data, total kelas *suspect* sebanyak 78 data, dan total kelas *pathological* sebanyak 59 data dari sampel testing sebesar 637 data.

**d. Algoritma C5.0 implementation.**

Pertama menghitung nilai *entropy* menggunakan persamaan 1, yaitu :

**Langkah 1 :**

Berdasarkan hasil yang telah diolah sebelumnya pada proses split data training dan testing, maka langkah pertama adalah mendapatkan *entropy* dengan menghitung perbandingan dari total kasus *normal*, total kasus *suspect*, dan total kasus *pathological* dengan :

- Perbandingan normal → jumlah kasus normal / jumlah total kasus. Hasilnya adalah 1.155/1.489 = 0,776.
- Perbandingan *suspect* → jumlah kasus *suspect* / jumlah total kasus. Hasilnya adalah 217/1.489 = 0,146.
- Perbandingan *pathological* → jumlah kasus *pathological*/jumlah total kasus. Hasilnya 117/1.489 = 0,078.

**Langkah 2 :**

Berdasarkan persamaan 1, maka hasil yang dapat dilihat pada Tabel 2;  $Entropy_{[Total]} = (-0,776 * \log_2(0,776)) + (-0,146 * \log_2(0,146)) + (-0,078 * \log_2(0,078)) = 0,9775$ .

Tabel 2. Perhitungan entropy.

A	Nilai Atribut	N	S	P	Ent
total	-	1.155	217	117	0,9775
	Normal	1.152	217	117	0,9788
	Fetal	0	0	0	0
1	Bradycardia	3	0	0	0
	Fetal Tachardia	807	33	22	0,4043
2	False	348	184	95	1,4030
	True	415	103	60	1,1259
3	False	740	114	57	0,8690
	True	1.049	126	83	0,8098
4	False	106	91	34	1,4520
	True	508	40	70	0,8440
5	False	647	177	47	1,0131
	True	1	0	4	0
6	False	1.154	217	113	0,9706
	True	53	10	63	1,3156
7	False	1.102	207	54	0,8454
	Above	463	193	106	1,3344
8	Average Below	692	24	11	0,3217
	Average Above	517	14	66	0,6580
9	Average Below	638	203	51	1,0679
	Average Above	183	169	39	1,3674
10	Average Below	972	48	78	0,6241
	Average Above	548	74	15	0,6749
11	Average Below	607	143	102	1,1473
	Average Above	587	58	73	0,8661

	Below Average	568	159	44	1,0303
	Above Average	539	159	39	1,0318
13	Below Average	616	58	78	0,8600
	Above Average	556	85	47	0,8856
14	Below Average	599	132	70	1,0495
	Above Average	455	56	57	0,9187
15	Below Average	700	161	60	0,9974
	Above Average	301	36	28	0,8431
16	Below Average	854	181	89	1,0151
	Above Average	592	184	21	0,9451
17	Below Average	563	33	96	0,8468
	Above Average	595	192	24	0,9702
18	Below Average	560	25	93	0,7965
	Above Average	583	182	21	0,9481
19	Below Average	572	35	96	0,8498
	Above Average	369	17	64	0,8135
20	Below Average	786	200	53	0,9811
	0	622	98	63	0,9316
21	1	458	106	25	0,9210
	-1	75	13	29	1,2623

Deskripsi pada Tabel 2 yang terdiri dari 6 kolom. Kolom pertama adalah A yang merupakan atribut, atribut adalah nama atribut yang digunakan pada algoritma C5.0 yang telah disebutkan dengan nomor urut pada tahap data *collection* sebelumnya. Kolom kedua adalah Nilai Atribut yang digunakan pada algoritma C5.0 yang sudah ditransformasikan kedalam bentuk kategori.

Kolom ketiga adalah N yang merupakan jumlah kasus pada kelas data yang bernilai normal yang akan diklasifikasikan jumlahnya berdasarkan atribut dan nilai atributnya. Kolom keempat adalah S, dimana S merupakan jumlah kasus pada kelas data yang bernilai *suspect* yang akan diklasifikasikan jumlahnya berdasarkan atribut dan nilai atributnya. Kolom kelima adalah P, Nilai dari kolom P merupakan jumlah kasus pada kelas data bernilai *pathological* dan diklasifikasikan jumlahnya kepada atribut dan nilai atributnya. Kolom keenam

adalah Ent sebagai hasil nilai *entropy* memanfaatkan persamaan 1.

**Kedua** Proses menghitung suatu nilai *information gain* menggunakan persamaan 2, yaitu :

**Langkah 1 :**

Sebelum mendapatkan nilai *information gain*, maka tahap ke-1 terlebih dahulu menentukan *entropy* seperti yang didapat pada Tabel 2, maka hasil olah data didapat nilai-nilai adalah sebagai berikut;  $Entropy_{[Total]} = 0,9775$ ;  $Entropy_{[baseline\_value-Normal]} = 0,9788$ ;  $Entropy_{[baseline\_value-Fetal\ Tachcardia]} = 0$ ; dan  $Entropy_{[baseline\_value-Fetal\ Bradycardia]} = 0$ .

**Langkah 2 :**

Tahap ke-2 adalah menghitung nilai *information gain* setiap atribut, dimana formulanya adalah  $Gain[] = ((Total\ kasus\ pada\ atribut / total\ kasus\ keseluruhan) * entropy\ pada\ atribut)$  Pada Tabel 2 total kasus pada atribut *baseline\_value* dan nilai atribut Normal adalah 1.486 dari total kasus keseluruhan yang bernilai 1.489, serta nilai *entropy* atribut tersebut adalah 0,9788, maka :

1.  $Gain_{[baseline\_value-Normal]} = ((1.486/1.489) * 0,9788) = 0,9768$
2.  $Gain_{[baseline\_value-Fetal\ Tachcardia]} = ((0/1.489) * 0) = 0$
3.  $Gain_{[baseline\_value-Fetal\ Bradycardia]} = ((3/1.489) * 0) = 0$

Untuk menentukan nilai *information gain* yang lain, dapat menggunakan cara yang sama seperti cara di atas.

**Langkah 3 :**

Tujuan dari tahap ini mendapatkan nilai *baseline\_value* dengan cara mengurangi nilai *entropy* dari total kasus keseluruhan dengan nilai total kasus keseluruhan dengan nilai total *information gain*. Nilai *entropy* total dari kasus keseluruhan pada Tabel 2 adalah 0,9775 dan total nilai *information gain* adalah 0,9768, maka :

$$Gain_{[baseline\_value]} = 0,9775 - 0,9768 = 0,0007$$

Sehingga untuk nilai *information gain* pada atribut *baseline\_value* adalah sebesar 0,0007. Hasil dapat dilihat pada Tabel 3.

Tabel 3. Perhitungan *information gain*.

A	Nilai Atribut	N	S	P	Gain
total	-	1.155	217	117	-
	Normal	1.152	217	117	0,0007
1	Fetal Bradycardia	0	0	0	0,0007
	Fetal Tachardia	3	0	0	0,0007
2	True	807	33	22	0,1527
	False	348	184	95	0,1527
3	True	415	103	60	0,0088
	False	740	114	57	0,0088
4	True	1.049	126	83	0,0681
	False	106	91	34	0,0681
5	True	508	40	70	0,0346
	False	647	177	47	0,0346
6	True	1	0	4	0,0102
	False	1.154	217	113	0,0102
7	True	53	10	63	0,0923
	False	1.102	207	54	0,0923
8	Above Average	463	193	106	0,1375
	Below Average	692	24	11	0,1375
9	Above Average	517	14	66	0,0323
	Below Average	638	203	51	0,0323
10	Above Average	183	169	39	0,1582
	Below Average	972	48	78	0,1582
11	Above Average	548	74	15	0,0323
	Below Average	607	143	102	0,0323
12	Above Average	587	58	73	0,0264
	Below Average	568	159	44	0,0264
13	Above Average	539	159	39	0,0325
	Below Average	616	58	78	0,0325
14	Above Average	556	85	47	0,0037
	Below Average	599	132	70	0,0037
15	Above Average	455	56	57	0,0101
	Below Average	700	161	60	0,0101
16	Above Average	301	36	28	0,0046
	Above Average	854	181	89	0,0046
17	Above Average	592	184	21	0,0781
	Below Average	563	33	96	0,0781
18	Above Average	595	192	24	0,0864
	Below Average	560	25	93	0,0864
19	Above Average	583	182	21	0,0758
	Below Average	572	35	96	0,0758
20	Above Average	369	17	64	0,0471
	Below Average	786	200	53	0,0471

	0	622	98	63	0,0241
21	1	458	106	25	0,0241
	-1	75	13	29	0,0241

Pada Tabel 3, penggunaan kolom-kolom sama seperti Tabel 2, namun perbedaannya terdapat pada kolom keenam yaitu Gain yang merupakan nilai dari information gain pada atribut dan nilai atribut tersebut yang sudah dihitung menggunakan persamaan 2.

1. Nilai *split information* dapat dihitung menggunakan persamaan 3, terdapat beberapa tahapan dalam menghitung *split information*, dimana menghitung *split information* dari masing-masing atribut. Pada Tabel 3 terdapat total kasus pada atribut *baseline\_value* dan nilai atribut normal sebesar 1.486 dari total kasus keseluruhan sebesar 1.489, maka:

- $SplitInformation_{[baseline\_value-Normal]} = -((1.486/1.489) * (\log_2(1.486/1.489))) = 0,0028$
- $SplitInformation_{[baseline\_value-Fetal Tachardia]} = -((0/1.489) * (\log_2(0/1.489))) = 0$
- $SplitInformation_{[baseline\_value-Fetal Bradycardia]} = -((3/1.489) * (\log_2(3/1.489))) = 0,0181$

Hasil dapat dilihat pada Tabel 4, dimana menjumlahkan masing-masing *split information* dari setiap atribut dan nilai atribut menjadi :

$$SplitInformation_{[baseline\_value]} = 0,0028 + 0 + 0,0181 = 0,0209.$$

Tabel 4. Perhitungan split infromation.

A	Nilai Atribut	N	S	P	Split Inf
total	-	1.155	217	117	-
	Normal	1.152	217	117	0,0209
1	Fetal Bradycardia	0	0	0	0,0209
	Fetal Tachardia	3	0	0	0,0209
2	True	807	33	22	0,9820
	False	348	184	95	0,9820
3	True	415	103	60	0,9636
	False	740	114	57	0,9636
4	True	1.049	126	83	0,6225
	False	106	91	34	0,6225
5	True	508	40	70	0,9791
	False	647	177	47	0,9791
6	True	1	0	4	0,0324
	False	1.154	217	113	0,0324
7	True	53	10	63	0,4183
	False	1.102	207	54	0,4183
8	Above Average	463	193	106	0,9996

9	Below Average	692	24	11	0,9996
	Above Average	517	14	66	0,9715
10	Below Average	638	203	51	0,9715
	Above Average	183	169	39	0,8306
11	Below Average	972	48	78	0,8306
	Above Average	548	74	15	0,9849
12	Below Average	607	143	102	0,9849
	Above Average	587	58	73	0,9991
13	Below Average	568	159	44	0,9991
	Above Average	539	159	39	0,9999
14	Below Average	616	58	78	0,9999
	Above Average	556	85	47	0,9958
15	Below Average	599	132	70	0,9958
	Above Average	455	56	57	0,9591
16	Below Average	700	161	60	0,9591
	Above Average	301	36	28	0,8035
17	Below Average	854	181	89	0,8035
	Above Average	592	184	21	0,9964
18	Below Average	563	33	96	0,9964
	Above Average	595	192	24	0,9942
19	Below Average	560	25	93	0,9942
	Above Average	583	182	21	0,9978
20	Below Average	572	35	96	0,9978
	Above Average	369	17	64	0,8840
21	0	786	200	53	0,8840
	1	622	98	63	1,3052
	-1	458	106	25	1,3052
		75	13	29	1,3052

Pada Tabel 4, penggunaan kolom-kolom sama seperti tabel 2 dan Tabel 3, namun perbedaannya terdapat pada kolom ke-enam yaitu Split Inf yang merupakan nilai dari split information pada atribut dan nilai atribut tersebut yang sudah dihitung menggunakan persamaan 3.

2. Nilai *gain ratio* dapat dihitung menggunakan persamaan 4 seperti yang sudah dibahas. Pada Tabel 3, terdapat nilai *information gain* atribut *baseline\_value* adalah 0,0007, dan pada Tabel 4 terdapat nilai *split*

*information* pada atribut *baseline\_value* adalah 0,0209. Maka,  $GainRatio_{[baseline\_value]} = 0,0007 / 0,0209 = 0,0335$ .

Jadi *gain ratio* untuk atribut *baseline\_value* adalah 0,0335 dan hasil lain dapat dilihat pada Tabel 5.

Tabel 5. Hasil perhitungan gain ratio.

A	Nilai Atribut	N	S	P	Gain Ratio
total	-	1.155	217	117	-
1	Normal	1.152	217	117	0,0335
	Fetal Bradycardia	0	0	0	0,0335
2	Fetal Tachardia	3	0	0	0,0335
	True	807	33	22	0,1555
3	False	348	184	95	0,1555
	True	415	103	60	0,0091
4	False	740	114	57	0,0091
	True	1.049	126	83	0,1094
5	False	106	91	34	0,1094
	True	508	40	70	0,0353
6	False	647	177	47	0,0353
	True	1	0	4	0,3148
7	False	1.154	217	113	0,3148
	True	53	10	63	0,2207
8	False	1.102	207	54	0,2207
	Above Average	463	193	106	0,1376
9	Below Average	692	24	11	0,1376
	Above Average	517	14	66	0,0761
10	Below Average	638	203	51	0,0761
	Above Average	183	169	39	0,1905
11	Below Average	972	48	78	0,1905
	Above Average	548	74	15	0,0328
12	Below Average	607	143	102	0,0328
	Above Average	587	58	73	0,0264
13	Below Average	568	159	44	0,0264
	Above Average	539	159	39	0,0325
14	Below Average	616	58	78	0,0325
	Above Average	556	85	47	0,0037
15	Below Average	599	132	70	0,0037
	Above Average	455	56	57	0,0105
16	Below Average	700	161	60	0,0105
	Above Average	301	36	28	0,0057
17	Below Average	854	181	89	0,0057
	Above Average	592	184	21	0,0784
18	Below Average	563	33	96	0,0784
	Above Average	563	33	96	0,0784

18	Above Average	595	192	24	0,0869
	Below Average	560	25	93	0,0869
19	Above Average	583	182	21	0,0760
	Below Average	572	35	96	0,0760
20	Above Average	369	17	64	0,0533
	Below Average	786	200	53	0,0533
21	0	622	98	63	0,0185
	1	458	106	25	0,0185
	-1	75	13	29	0,0185

Pada Tabel 5 di atas, pengertian dari penggunaan kolom-kolom sama seperti tabel 2, tabel 3 dan Tabel 4, namun perbedaannya terdapat pada kolom keenam yaitu Gain Ratio yang merupakan nilai dari gain ratio pada atribut dan nilai atribut tersebut yang sudah dihitung menggunakan persamaan 4.

Seperti yang sudah diketahui, atribut yang dijadikan akar adalah atribut *severe\_decelerations* dengan nilai Gain Ratio sebesar 0,3148. Didalam atribut *severe\_decelerations* terdapat 2 (dua) nilai atribut. Nilai atribut tersebut adalah True dan False yang total kasus keseluruhannya dapat dilihat pada Tabel 6.

Tabel 6. Atribut *severe\_decelerations*.

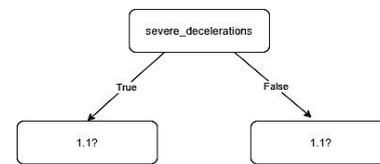
Nilai Atribut	Normal	Suspect	Pathological
True	1	0	4
False	1.154	217	113

Pada Tabel 6, nilai atribut True memiliki total kasus sebanyak 5 (lima) dengan jumlah kasus normal sebanyak 1, jumlah kasus *suspect* sebanyak 0, dan jumlah kasus *pathological* sebanyak 4. Sedangkan pada nilai atribut False memiliki total kasus sebanyak 1.484 dengan jumlah kasus normal sebanyak 1.154, jumlah kasus *suspect* sebanyak 217, dan jumlah kasus *pathological* sebanyak 113.

Untuk atribut *severe\_decelerations* dengan nilai atribut True, keputusan belum dapat dihasilkan dikarenakan jumlah kasus normal atau jumlah kasus *pathological* masih belum bernilai 0 (nol) sehingga harus dibuat *node* baru. Begitupun juga nilai atribut False, keputusan belum dapat dihasilkan dikarenakan jumlah kasus normal, jumlah kasus *suspect*, atau jumlah kasus

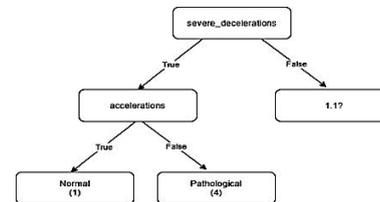
*pathological* masih belum bernilai 0 (nol) sehingga harus dibuat *node* baru. Pohon keputusan dapat dilihat pada Gambar 3 di bawah ini.

Gambar 3 menunjukkan rancangan pohon keputusan, dimana jika ingin mengetahui *node* 1.1 pada nilai atribut apakah *True* dan *False*. Maka hasil *node* 1.1 dilakukan perhitungan *entropy*, *gain*, *split information*, dan *gain ratio* seperti yang sudah dibahas di atas. Perbedaannya adalah atribut *severe\_decelerations* tidak perlu dihitung kembali, dan hanya perlu menghitung pada nilai atribut *severe\_decelerations* dengan nilai atribut *True* maupun *False*.



Gambar 3. Akar pohon keputusan

Gain Ratio tertinggi dari hasil dari perhitungan tersebut akan dijadikan sebagai *node* 1.1. Sehingga pohon keputusan untuk *node* 1.1 pada nilai atribut True dapat dilihat pada Gambar 4 di bawah ini.



Gambar 4. Pohon keputusan *node* 1.1 pada nilai atribut true.

Sedangkan untuk mengetahui *node* 1.1 pada nilai atribut false, dapat digunakan dengan cara mengulangi perhitungan seperti langkah di atas.

Setelah pohon keputusan sudah berada dalam tahap final, serta semua keputusan sudah diketahui dan dapat diklasifikasi secara sempurna, terdapat 220 aturan atau *rule set* dalam Algoritma C5.0 dalam mengklasifikasikan *dataset* Kesehatan janin dengan *Cardiotography* yang dapat dilihat pada Tabel 7 di bawah ini.

Tabel 7. Rule set pohon keputusan kesehatan janin algoritma C5.0

No	Rule	Keputusan
1	If (severe_decelerations = TRUE) & (accelerations = TRUE)	Normal
2	if (severe_decelerations = TRUE) & (accelerations = FALSE)	Pathological
3	if (severe_decelerations = FALSE) & (prolongued_decelerations = TRUE) & (percentage_of_time_with_abnormal_long_term_variability = Below Average) & histogram_mode = Below Average) & (abnormal_short_term_variability = Below Average) & light_decelerations = TRUE) & (mean_value_of_short_term_variability = Below Average)	Suspect
4	if (severe_decelerations = FALSE) & (prolongued_decelerations = TRUE) & (percentage_of_time_with_abnormal_long_term_variability = Below Average) & (histogram_mode = Below Average) & (abnormal_short_term_variability = Below Average) & (light_decelerations = TRUE) & (mean_value_of_short_term_variability = Above Average) & (histogram_variance = Below Average)	Suspect
-	-	-
2	if (severe_decelerations = FALSE) & (prolongued_decelerations = FALSE) & (percentage_of_time_with_abnormal_long_term_variability = Above Average) & (abnormal_short_term_variability = Above Average) & (uterine_contractions = FALSE) & (accelerations = FALSE) & (histogram_mean = Above Average) & (mean_value_of_long_term_variability = Above Average) & (histogram_number_of_peaks = Above Average)	Suspect

Hasil pada Tabel 7 di atas merupakan suatu rancangan kondisi aturan yang terbentuk berdasarkan pohon keputusan. Penerapan aturan tersebut digunakan sebagai rekomendasi terhadap kesehatan janin dalam memberikan validasi terkait kondisi janin.

**e. Evaluation with confusion matrix**

Setelah dilakukan implementasi Algoritma C5.0 dengan perhitungan *entropy*, *information gain*, *split information*, dan *gain ratio* hingga merancang suatu aturan yang telah terbentuk dalam pohon keputusan, maka akan dilakukan evaluasi terhadap data *testing*.

Pada tahapan *splitting data*, dimana disebutkan bahwa data *testing* yang berjumlah 637 data dengan perincian

data kelas normal sebanyak 500 data, data kelas *suspect* sebanyak 78 data, dan data kelas *pathological* sebanyak 59 data. Keseluruhan dari 637 data diperoleh dari total keseluruhan 2.126 data yang diantaranya terdiri dari 1.655 data kelas yang bernilai normal, 295 data kelas yang bernilai *suspect*, dan 176 data kelas bernilai *pathological* yang sudah di *split* menggunakan rasio 70% data *training* dan 30% data *testing*.

Berdasarkan pengolahan data untuk rasio 70% untuk data training dan 30% untuk data testing, maka hasil dibutuhkan suatu metode *confusion matrix* untuk menghitung dan mendapatkan nilai akurasi dari Algoritma C5.0 yang telah diimplementasi dalam pohon keputusan. Hasil *Confusion Matrix* dapat dilihat pada Tabel 8.

Tabel 8. Confusion matrix.

Predicted	Actual		
	Normal	Suspect	Pathological
Normal	470	24	3
Suspect	22	48	12
Pathological	8	6	44

Berdasarkan Tabel 8 di atas, maka akurasi dapat dihitung dengan persamaan 5 yang telah dibahas di atas, nilai dari persamaan tersebut dilihat di bawah ini.

$$= \frac{470 + 48 + 44}{470 + 48 + 44 + 22 + 8 + 24 + 32 + 12} = 0.88226059654 = 88,23\%$$

Pengujian *dataset* pada penelitian ini, menggunakan data sekunder sebanyak 2.126 *dataset* berasal dari kaggle.com, dimana atribut berjumlah 22, dan terdapat 3 kelas yang diklasifikasi adalah normal, *suspect*, dan *pathological*.

Hasil dari evaluasi dengan variasi *training* dan *testing* data, dimana variasi *splitting* data dapat dilihat pada Tabel 9.

Tabel 9. Hasil perbandingan evaluasi.

Training Data	Testing Data	Akurasi
90%	10%	93,40%
80%	20%	91,29%
70%	30%	88,23%
60%	40%	88,12%

Berdasarkan Tabel 9, maka pengujian dilakukan untuk data training terdiri dari 60%, 70%, 80%, dan 90%, sedangkan data *testing* terdiri dari 40%, 30%, 20%, 10%. Hasil pengujian untuk 90% data *training* dan 10% data *testing*

mendapatkan akurasi sebesar 93,40% dengan jumlah aturan sebanyak 257. Pengujian untuk 80% data training dan 20% data *testing* menghasilkan akurasi sebesar 91,29% dengan jumlah aturan sebanyak 239. Penulis juga melakukan pengujian menggunakan 70% data *training* dan 30% data *testing* menghasilkan akurasi sebesar 88,23% dengan jumlah aturan sebanyak 220. Sedangkan pengujian untuk 60% data *training* dan 40% data *testing* menghasilkan akurasi sebesar 88,12% dengan jumlah aturan sebanyak 204. Maka hasil akurasi semakin baik hingga mendekati 100% dan jumlah aturan-aturan yang dihasilkan lebih banyak dibandingkan data training dibawahnya.

### SIMPULAN

Sistem penunjang keputusan yang dapat mendiagnosa, memprediksi dan memberikan informasi terkait kondisi janin yang dikandung dengan penerapan Algoritma C5.0 dapat mengklasifikasikan kesehatan janin yang didapat dari alat CTG dengan menghitung *entropy*, *information gain*, *split information*, dan *gain ratio* berdasarkan aturan yang dirancang pada pohon keputusan.

Hasil akurasi berdasarkan variasi 90% data *training* dan 10% data *testing* sebesar 93,40%. Sedangkan variasi 80% data *training* dan 20% data *testing* menghasilkan akurasi sebesar 91,29%. Pada variasi 70% data *training* dan 30% data *testing* hasil akurasi sebesar 88,23%. Dan variasi terkecil, dimana 60% data *training* dan 40% data *testing* menghasilkan akurasi sebesar 88,12%.

Berdasarkan beberapa variasi disimpulkan semakin besar data *training*, maka hasil akurasi menjadi semakin baik sebagai sistem penunjang keputusan.

### SARAN

Hasil rekomendasi terhadap kesehatan janin dari sistem pendukung keputusan dalam memberikan suatu rekomendasi dengan metode Algoritma C5.0 sebaiknya memerlukan verifikasi dan akurasi dari seorang pakar terkait kesehatan janin dan kandungan.

### DAFTAR PUSTAKA

- [1] M. K. drg. Rudy Kurniawan, M. S. Yudianto, SKM, M. Boga Hardhana, S.Si, and M. K. Tanti Siswanti, SKM, *HEALTH STATISTICS*. Kementerian Kesehatan Republik Indonesia, 2019.
- [2] dr. K. Adrian, "Kapan Ibu Hamil Perlu Melakukan Cardiotocography?," 2020. [Online]: <https://www.alodokter.com/kapan-ibu-hamil-perlu-melakukan-cardiotocography-CTG> (accessed May 28, 2021).
- [3] Y. Azhar and R. Afdian, "Feature Selection on Pregnancy Risk Classification Using C5.0 Method," *Kinet. Game Technol. Inf. Syst. Comput. Network, Comput. Electron. Control*, vol. 3(4), 345–350, 2018, doi: 10.22219/kinetik.v3i4.703.
- [4] M. Z. Arif, R. Ahmed, U. H. Sadia, M. S. I. Tultul, and R. Chakma, "Decision Tree Method Using for Fetal State Classification from Cardiotography Data," *J. Adv. Eng. Comput.*, 2020, doi: 10.25073/jaec.202041.273.
- [5] E. M. Karabulut and T. Ibriki, "Analysis of Cardiotocogram Data for Fetal Distress Determination by Decision Tree Based Adaptive Boosting Approach," *J. Comput. Commun.*, vol. 2, pp. 32–37, 2014.
- [6] K. Madadipouya, "a New Decision Tree Method for Data Mining in Medicine," *Adv. Comput. Intell. An Int. J.*, vol. 2, no. 3, pp. 31–37, 2015, doi: 10.5121/acii.2015.2304.
- [7] J. Novakovic, A. Veljovi, S. Iiic, Z. Papic, and M. Tomovic, "Evaluation of Classification Models in Machine Learning," *Theory Appl. Math. Comput. Sci.*, vol. 7(1), 39-46, 2017, [Online]: <https://uav.ro/applications/se/journal/index.php/TAMCS/article/view/158>.
- [8] A. Sandlin, "Fetal Heart Rate Baseline Abnormalities: Bradycardia," *Pearls of Excellence*, 2020, doi: 10.1097/AOG.0b013e3182004fa9. Initial.