JIKB
Jurnal Ilmu Komputer dan Bisnis
STMIK DHARMAPALA RIAU

Jurnal Ilmu Komputer dan Bisnis (JIKB)
Desember-2022, Vol. XIII, No.2a, hal.67-77
ISSN(P): 2087-3921; ISSN(E): 2598-9715

Implementasi Data Mining C4.5 Untuk Klasifikasi Faktor Resiko Kesehatan Pada Ibu Hamil

[1]Monica Dessy Setyaningsih, [2]Ayu Wahyuni, [3]Antonius Yadi Kuntoro
[1,2,3] Universitas Nusa Mandiri

Alamat Surat
Email: mdessy818@gmail.com, ayuwahyuni0223@gmail.com, antonius.aio@nusamandiri.ac.id

Article History:
Diajukan: 10 Oktober 2022; Direvisi: 21 November 2022; Diterima: 29 November 2022

ABSTRAK

Kasus kematian seorang ibu dapat dinyatakan menjadi peristiwa yang cukup kompleks. Faktor resiko sangat penting untuk diidentifikasi secara dini guna mengembangkan strategi yang komprehensif untuk pencegahan komplikasi terkait kehamilan. Faktor resiko dapat diklasifikasikan berdasarkan data rekam medis dari pasien ibu hamil. Faktor yang menjadi acuan terdiri usia, tekanan darah, gula darah serta detak jantung. Dari hal tersebut dapat diklasifikasikan resiko kesehatan bagi seorang ibu. Ketika diketahui rekam medis pasien serta resiko yang ada maka dapat dilakukan antisipasi secara medis bagi pasien ibu hamil. Untuk proses klasifikasi yang manual kurang efektif dikarenakan akan cukup lama waktu yang dibutuhkan. Untuk mendukung manganalisa faktor resiko kesehatan pada ibu hamil maka digunakan data mining klasifikasi berbasis komputer agar dapat menggali informasi dari dataset faktor resiko kesehatan bagi ibu hamil. Maka dari penulis mengkaji dari data sekunder yang didapat dengan menggunakan Algoritma C4.5 dengan langkah yang pertama yaitu mengidentifikasi dan merumuskan masalah, menentukan tujuan penelitian, mempelajari buku dan jurnal terkait, pengumpulan data, pengolahan data, dan yang terakhir pengujian metode Algoritma C45 dan didapatlah hasil bahwa penerapan metode algoritma C4.5 pada analisa faktor resiko kesehatan pada ibu hamil dapat diambil kesimpulan yaitu pohon keputusan yang dihasilkan dari 100 data pasien memiliki akar *(root) Systolic BP* menjadi faktor resiko paling utama dibandingkan atribut lainnya, dalam penerapan Algoritma C4.5 menghasilkan nilai akurasi 83,33% dengan pembagian data training dan data testing 70:30, dan dalam penggunaan metode Algoritma C4.5 menjadi salah satu metode yang tepat untuk klasifikasi faktor resiko kesehatan pada ibu hamil.

Kata kunci: Kesehatan kehamilan, Algoritma C4.5, Data Mining, Rapid Miner

ABSTRACT

The case of a mother's death can be expressed as a fairly complex event. It is very important to identify risk factors early in order to develop a comprehensive strategy for the prevention of pregnancy-related complications. Risk factors can be classified based on medical record data from pregnant women. Factors that became the reference consisted of age, blood pressure, blood sugar and heart rate. From this, it can be classified as a health risk for a mother. When it is known the patient's medical record and the risks that exist, medical anticipation can be made for pregnant women patients. The manual classification process is less effective because it will take quite a long time. To support the analysis of health risk factors in pregnant women, a computer-based classification data mining is used in order to extract information from a dataset of health risk factors for pregnant women. So from the authors review of secondary data obtained using the C4.5 Algorithm with the first step, namely identifying and formulating problems, determining research objectives, studying related books and journals, data collection, data processing, and finally testing the C45 Algorithm method and obtained the results that the application of the C4.5 algorithm method in the analysis of

https://doi.org/10.47927/jikb.v13i2a.376

health risk factors in pregnant women can be concluded that the decision tree generated from 100 patient data has a Systolic BP root to be the most important risk factor compared to other attributes, in the application of the C4.5 algorithm produces an accuracy value of 100% with the distribution of training data and testing data of 90:10, and the use of the C4.5 Algorithm method is one of the appropriate methods for classifying health risk factors in pregnant women.

Keywords: Pregnancy health, C4.5 Algorithm, Data Mining, Rapid Miner

1. PENDAHULUAN

Teknologi informasi yang berkembang pada saat ini memberikan kegunaan bagi semua orang. Teknologi informasi yang berkembang pada saat ini memberikan kegunaan bagi semua orang. Bagai kehidupan sehari–hari teknologi informasi menjadi sebuah kebutuhan yang sangat penting. Dalam bidang pekerjaan teknologi informasi juga memberikan kemudahan, sehingga teknologi informasi dapat diterapkan tidak hanya dalam satu bidang salah satunya dunia kedokteran dan kesehatan. Kesehatan menjadi hal sangat penting terutama untuk kesehatan seorang ibu. (Muzakir & Wulandari, 2016)

Kasus kematian seorang ibu dapat dinyatakan menjadi peristiwa yang cukup kompleks yang memiliki penyebab dan dibedakan menjadi determinan dekat, antara dan jauh. Determinan dekat yang berhubungan secara langsung yaitu gangguan obstetrik seperti pendarahan, preeklamsi/eklamsi, dan penyakit atau infeks yang selama atau sebelum kehamilan sudah diderita oleh ibu sehingga mampu memperburuk keadaan seperti jantung, malaria, tubekulosis, ginjal dan *acquired immunodeficiency syndrome.* Determinan antara yang berhubungan yaitu faktor kesehatan, status kesehatan ibu, status reproduksi, akses terhadap pelayanan kesehatan, serta perilaku pengguanaan fasilitas kesehatan. Sedangkan untuk determinan jauh berhubungan dengan faktor demografi dan sosiokultural. Selain itu juga terdapat kebijakan yang berhubungan secara tidak langsung antara lain rendahnya kesadaran masyarakat tentang kesehatan ibu, kurang baiknya pemberdayaan perempuan, latar belakang pendidikan, sosial ekonomi keluarga, serta lingkungan masyarakat dan politik. (Astuti et al., 2017)

Faktor resiko sangat penting untuk diidentifikasi secara dini guna mengembangkan strategi yang komprehensif untuk pencegahan komplikasi terkait kehamilan. Pemantuan faktor resiko yang telah terbukti berhubungan dengan kematian ibu akan mampu mengurangi angka kematian ibu. (Bauserman et al., 2015)

Faktor resiko dapat diklasifikasikan berdasarkan data rekam medis dari pasien ibu hamil. Faktor yang menjadi acuan terdiri usia, tekanan darah, gula darah serta detak jantung. Dari hal tersebut dapat diklasifikasikan resiko kesehatan bagi seorang ibu. Ketika diketahui rekam medis pasien serta resiko yang ada maka dapat dilakukan antisipasi secara medis bagi pasien ibu hamil. Untuk proses klasifikasi yang manual kurang efektif dikarenakan akan cukup lama waktu yang dibutuhkan. Untuk mendukung manganalisa faktor resiko kesehatan pada ibu hamil maka digunakan data mining klasifikasi berbasis komputer agar dapat menggali informasi dari daset faktor resiko kesehatan bagi ibu hamil.

Klasifikasi merupakan proses mencari suatu himpunan model yang mampu mendiskripsikan dan membedakan kelas – kelas data atau konsep dengan tujuan mampu menggunakan model tersebut untuk memprediksi kelas suatu objek yang mana kelasnya belum diketahui. (Nurdiana & Algifari, 2020)

Berdasarkan latar belakang diatas maka penulis akan melakukan penelitian dengan metode algoritma C.45 karena metode ini belum pernah digunakan untuk menganalisa faktor resiko kesehatan ibu hamil. Karena sebelumnya hanya digunakan untuk prediksi penyakit seperti stroke, diabetes, tuberculosis serta kesehatan balita. Maka dari itu penulis mengambil judul "Implementasi Data Mining Algoritma C4.5 Untuk Klasifikasi Faktor Resiko Kesehatan Pada Ibu Hamil".

1.1. Kajian Pustaka

Data Mining menurut Gartner Group ialah merupakan proses penemuan hubungan baru yang memiliki pola, arti, serta kebiasaan dengan memisah - misah data yang sebagian besar tersimpan di tempat penyimpanan dengan memanfaatkan teknologi yang mengenalkan pola seperti teknik statistik dan matematika. Data mining adalah penggabungan beberapa disiplin ilmu dengan penyatuan teknik dari pembelajaran mesin, pengenalan pola, statistik, database, dan visualisasi yang diguankan menangani masalah pengambilan informasi dari database yang besar. (Mardi, 2017)

Menurut David Hand, Heikki Mannila, dan Padhraic Smyth dari MIT Data mining yaitu analisis dari suatu data yang merupakan data yang memiliki ukuran besar guna penemuan hubungan yang jelas dan penarikan kesimpulannya yang belum diketahui sebelumnya dengan cara saat ini menjadi lebih mudah untuk memahami serta berguna bagi yang memiliki data tersebut. (Mardi, 2017)

Data Mining merupakan suatu proses yang memanfaatkan *statistics, mathematics, artificial intelligence* dan *machine learning* guna mengekstrasi dan mengindentifikasi informasi yang mempunyai manfaat serta pengetahuan yang memiliki hubungan dengan banyak macam *database* dengan ukuran besar. *Data mining* disebut juga suatu rangkaian proses guna menggali nilai lebih dari suatu kumpulan data yang berupa ilmu pengetahuan yang selama ini tidak dikenali secara manual. (Mardi, 2017)

Data Mining merupakan bidang yang sudah cukup lama. Ada hal yang sulit dalam mendefinisikan data mining yaitu bahwa ternyata *data mining* mewarisi aspek yang lebih banyak dan teknik dari bidang – bidang ilmu terdahulu yang sudah mapan lebih dulu. Data mining mempunyai akar yang panjang dari bidang ilmu yang berbeda seperti kecerdasan buatan (*artificial intelegent), machine learning, statistik, database*, dan juga *information retrieval*.

Klasifikasi ialah suatu proses penempatan suatu obyek atau konsep kedalam suatu set kategori berdasarkan objek atau konsep yang bersangkutan. Klasifikasi dimanfaatkan untuk membantu serta mengelompokkan data. Klasifikasi merupakan cabang dari *discovery data mining*. (Andriani, 2013)

Klasifikasi data mining merupakan suatu metode pembelajaran guna memprediksi nilai dari kumpulan atribut dalam menggambarkan serta membedakan kelas data atau konsep yang bertujuan guna memprediksi kelas dari *objek* yang label kelasnya belum dapat ditemukan. (Saputra, 2014)

Penggambaran klasifikasi seperti berikut ini. Data input disebut juga sebagai training set, yang terdiri atas banyak contoh diberi sebuah label kelas khusus. Bertujuan untuk menganalisis data input serta mengembangkan deskripsi atau model akurat untuk tiap kelas menggunakan fitur pada data. (Widiastiwi & Ernawati, 2021)

Algoritma C45 merupakan pengembangan dari ID3 yang salah satu algotima pohon keputusan. Algoritma C45 ialah salah satu solusi pemecahan kasus yang sering digunakan dalam pemecahan masalah teknik klasifikasi. Hasil luaran dari algoritma C45 ini adalah sebuah *decision tree* layaknya teknik klasifikasi yang lainnya. Pohon keputusan ialah suatu struktur yang bisa dimanfaatkan untuk membagi himpunan data yang besar menjadi kumpulan record yang kecil dengan menerapkan serangkaian aturan pohon keputusan. (Sunge & Aditasari.Ana Angelia, 2018)

Menurut buku *The Top Ten Algorithms in Data Mining* karangan Xindong Wu dan Vipin Kumar menjelaskan bahwa Algoritma C45 menjadi salah satu algoritma yang sangat terkenal yang digunakan oleh sebagian besar peneliti di dunia. Algoritma C45 diciptakan oleh J.Rose Quinland sebagai pengembangan dari algoritma ID3. Kelebihan dari Algoritma C45 adalah mampu menghasilkan model yang pohon.

Membangun pohon keputusan dari algoritma C45 secara umum sebagai berikut:

a. Memilih atribut yang sebagai akar

b. Membuat cabang untuk masing – masing nilai
c. Membagi kasus dalam cabang
d. Mengulangi proses untuk masing – masing cabang hingga semua kasus pada cabang memiliki kelas yang sama.

Algoritma C45 memiliki pembenahan dari algoritma yang sebelumnya yaitu dalam hal pemangkasan. Yang mengalami perbaikan ialah:

a. Perhitungan gain ratio pada algoritma C45 pada masing – masing atribut serta atribut yang memiliki nilai yang paling tinggi akan dipilih menjadi simpul. Penggunaan gain ratio ini memperbaiki kekurangan dari ID3 yang menggunakan information gain
b. Pemangkasan bisa dilakukan pada saat pohon dibangun ataupun pada proses pohon selesai dibangun.
c. Dapat menangani continous atribut
d. Dapat menangani missing data
e. Mampu membangkitkan rule dari sebuah pohon

Untuk memilih atribut sebagai akar didasarkan pada nilai *gain* tertinggi dari atribut atribut yang ada. Rumus menghitung *gain* sebagai berikut:

$$Gain\ (\mathrm{S, A}) = Entropy\ (\mathrm{S}) - \sum_{i=1}^{n} \frac{|Si|}{|S|} * Entropy(Si)$$

Keterangan:

S : Himpunan Kasus
A : Atribut
n : Jumlah Partisi atribut A
|Si| : Jumlah Kasus pada partisi ke-i
|S| : Jumlah Kasus pada S

Sedangkan untuk mencari nilai entropy digunakan sebagai persamaan berikut:

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

Keterangan:

S : Himpunan Kasus
n : Jumlah partisi S
pi : proporsi dari Si terhadap S

Kehamilan merupakan suatu rangkaian proses yang dilalui oleh wanita yang dimulai dengan pertemuan antara sel telur dan sel sperma dalam indung telur wanita, selanjutnya ke pembentukan zigot, pelekatan atau penempelan di dinding rahim, pembentukan plasenta serta pertumbuhan dan perkembangan hasil konsepsi sampai cukup waktu. Normalitas kehamilan ialah kehamilan dimana ibu hamil dalam keadaan sehat tidak memiliki riwayat *obstetric* buruk, ukuran uterus sama/sesuai usia kehamilan serta hasil pemeriksaan fisik dan laboratorium normal. (Hikmatulloh et al., 2019)

2. METODE

2.1. Instrumen Penelitian

Dalam penelitian ini dibutuhkan beberapa yang terdiri dari perangkat keras dn perangkat lunak. Adapun perangkat keras yang digunakan dalam penelitian ini yaitu Laptop dengan spesifikasi *Processor Core i3- 5005U 2.0 GHz, Ram 4Gb, Hardisk, Mouse, Printer*. Sedangkan perangkat luank yang digunakan *Operating System Windows 10* sebagai sistem operasi serta *Software Rapid Miner* 9.10 sebagai pengujian data.

Berdasarkan permasalahan yang ada tentang faktor resiko kesehatan ibu hamil peneliti membutuhkan informasi tentang kehamilan yang didapatkan dari studi pustaka, literatur yang berkaitan dengan permasalahan tersebut.

2.2. Pengumpulan Data

1. Jenis Data

Dalam penelitian ini penulis menggunakan data sekunder. Data sekunder ialah data yang sudah dikumpulkan oleh suatu lembaga pengumpul data serta telah dipublikasikan kepada masyarakat yang menggunakan data. (Paramita et al., 2021) Penulis menggunakan

data yang dipublikasikan di UCI Machine Learning yang berupa data faktor resiko kesehatan ibu hamil di Bangladesh.

2. Teknik Pengumpulan Data

   Teknik pengumpulan data yang tepat dengan mempertimbangkan penggunaan berdasarkan sumber dan jenis data. Untuk penelitian ini menggunakan teknik pengumpulan data sebagai berikut:

   a. Observasi

      Observasi ini dilakukan untuk mengamati objek yang akan diteliti. Pada penelitian ini observasi dilakukan dengan teknik observasi tidak langsung yaitu teknik dokumenter data sekunder. Dilakukan dengan pengambilan data Age, Systolic BP, Blood Sugar, Diatolic BP, Body Temp, dan Heart Rate yang merupakan data rekam medis dari rumah sakit, klinik bersalin di Bangladesh.

   b. Studi Pustaka

      Yang dilakukan ialah dengan mencari bahan yang mendukung dalam pendefinisian permasalahan. Penulis mengumpulkan data melalui buku – buku, literatur serta jurnal yang berkaitan dengan objek permasalahan.

2.3. Pengolahan Data

1. *Data Cleaning* (Pembersihan Data)

   Data Cleaning merupakan yang proses yang dijalankan untuk penghilangan noise pada data yang tidak konsisten atau dapat dikatakan tidak relevan. (Muslim et al., 2019) Data yang diperoleh dari hasil eksperimen yang telah ada tidak semua mempunyai isi data yang sempurna terdapat data yang hilang, data yang tidak valid dapat juga data yang salah ketik. Maka dari itu dilakukan pembersihan data karena berpengaruh terhadap performa teknik data mining.

2. *Data Selection* (Seleksi Data)

   Tidak semua data yang ada digunakan semuanya, dikarenakan hanya yang sesuai saja yang diambil serta dianalisis. Pada proses ini dilakukan sistem sampel pada pengambilan data yang akan dianalisis.

3. *Data Transformation* (Transformasi Data)

   Transformasi data ialah proses pengubahan data serta penggabungan data ke suatu format tertentu. Data mining membutuhkan format tertentu dan khusus sebelum diaplikasikan.

2.4. Model yang diusulkan

Data yang akan di uji dibagi menjadi data training dan data testing sesuai dengan kebutuhan metode yang akan digunakan yaitu metode algoritma C45.

2.5. Pengujian dan Eksperimen Model

Pengujian dilakukan dengan cara manual dengan menghitung gain dan entropy untuk menentukan akar (*root*) dan dilakukan dengan *software* RapidMiner.

2.6. Evaluasi Hasil

Evaluasi hasil dilakukan setelah pengujian dengan menganalisa hasil dari algoritma C45 yang berupa pohon keputusan. Selain itu juga dapat menganalisa nilai *accuracy* dari RapidMiner.

3. HASIL DAN PEMBAHASAN

3.1. *Dataset*

Dataset yang digunakan dalam penelitin ini diambil dari data publik yaitu dari UCI Machine Learning yang dapat diakses melalui halaman: https://archive.ics.uci.edu/ml/index.php Data tersebut diambil dari rumah sakit dan klinik bersalin di Bangladesh. Data terebut berfokus pada faktor resiko kesehatan pada ibu hamil. Indikator yang digunakan dalam data tersebut yaitu *Age, Systolic BP, Diastolic BP, Blood Sugar, Body Temp, Heart Rate*. Dan yang menjadi atribut tujuan yaitu level resiko dari kesehatan ibu hamil. Tabel atribut dapat dilihat sebagai berikut:

Tabel 1. Tabel Atribut

No	Atribut	Keterangan
1	*Age*	Usia ketika seorang wanita mengalami kehamilan
2	*Systolic BP*	Nilai atas tekanan darah dalam satuan mmHg
3	*Diastolic BP*	Nilai bawah tekanan darah dalam mmHg
4	*Blood Sugar*	Kadar gula darah dalam konsentrasi molar mmol/L
5	*Body Temp*	Suhu Tubuh dalam satuan Fahrenheit
6	*Heart Rate*	Detak jantung istirahat normal denyut / menit

3.2.Pengujian Algoritma C45

Langkah dalam pengujian algoritma C45:

1. Menyiapkan data yang akan diolah dengan algoritma C45
2. Memilih atribut sebagai akar
   Pemilihan atribut sebagai akar (node) berdasarkan pada nilai gain tertinggi dari atribut yang ada. Menghitung gain menggunakan rumus dibawah ini:

   $$\text{Gain (S, A)} = \text{Entropy (S)} - \sum_{i=1}^{n} \frac{|S_i|}{|S|} \; Entropy\ (S_i)$$

   Keterangan:
   S : Kasus
   A : Atribut
   N : Jumlah Partisi dalam atribut
   $|S_i|$: Jumlah kasus dalam partisi ke i
   $|S|$: Jumlah Kasus dalam S

   Untuk perhitungan entropy dengan rumus sebagai berikut:

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

   Keterangan:
   S : Himpunan Kasus
   n : Jumlah partisi S
   pi : proporsi dari Si terhadap S

3. Mengulangi langkah menghitung entropy dan gain hingga semua record terpartisi
4. Proses partisi akan terhenti apabila sudah tidak ada atribut dalam record untuk dipartisi lagi dan sudah tidak ada record yang kosong pada cabang yang kosong.
   Sebelum perhitungan Gain yang harus dilakukan yaitu menghitung entropy total. Perhitungan Entropy Total sebagai berikut:

   $$\text{Entropy Total} = \left(-\frac{26}{100}\ x\ \log_2 \frac{26}{100}\right) + \left(-\frac{49}{100}\ x\ \log_2 \frac{49}{100}\right) + \left(-\frac{25}{100}\ x\ \log_2 \frac{25}{100}\right)$$

   = 1, 569569

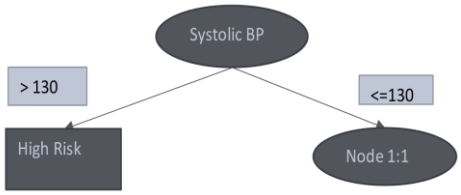
Untuk perhitungan keseluruhan dalam penentuan node akar dapat dilihat dalam tabel di bawah ini:

Tabel 2. Perhitungan Node

Atribut	Nilai	Jumlah Data	Jumlah (high)	Jumlah (mid)	Jumlah (low)	Entropy	Gain	Split Info	Gain Ratio
Total (S)		100	26	49	25	1,50957			
Age (year)	< =25	31	15	4	12	1,4179734	0,188044885	0,893173458	0,210535684
	>25	69	11	45	13	1,2781932			
Systolic BP (mmHg)	< =130	86	12	49	25	1,3770004	0,325349674	0,584238812	0,556877886
	>130	14	14	0	0	0			
Diastolic BP (mmHg)	<=90	87	15	48	24	1,4231658	0,170896055	0,557438185	0,306574002
	>90	13	11	1	1	0,7732283			
Blood Sugar (mmol/L)	<=7,8	78	12	41	25	1,429305	0,394712131	0,760167503	0,519243626
	>7,8	22	14	8	0	0			
	<=99	72	17	32	23	1,5375807	0,056614479	0,855450811	0,06618087

Atribut	Nilai	Jumlah Data	Jumlah (high)	Jumlah (mid)	Jumlah (low)	Entropy	Gain	Split Info	Gain Ratio
Body Temp (F)	>99	28	9	17	2	1,235348			
Heart Rate (denyut /menit)	<=77	33	8	9	16	1,5132025	0,110584413	0,914926373	0,120867008
	>77	67	18	40	9	1,3427295			

Berdasarkan perhitungan *root* (akar) nilai gain tertinggi pada atribut Systolic BP yaitu 0,556877886. Sehingga pohon keputusan node1 digambarkan sebagai berikut:



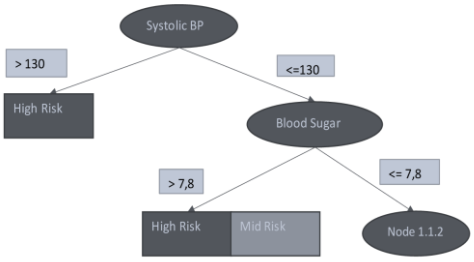
Gambar 1. Pohon Keputusan Menentukan Akar

Setelah akar ditentukan maka langkah selanjutnya ialah menentukan cabang pohon dengan melakukan studi kasus pada atribut Systolic serta mengulang kembali perhitungan entropy dan gain. Hasil yang didapatkan seperti pada tabel di bawah ini:

Tabel 3. Perhitungan cabang pohon 1

Atribut	Nilai	Jumlah Data	Jumlah (high)	Jumlah (mid)	Jumlah (low)	Entropy	Gain	Split Info	Gain Ratio
Systolic BP (<=130)		86	12	49	25	1,3770004			
Age (year)	< =25	21	5	4	12	1,409975	0,142433008	0,801932502	0,177612215
	>25	65	7	45	13	1,1778972			
Diastolic BP (mmHg)	<=90	84	12	48	24	1,3787835	0,030281609	0,159350063	0,190031984
	>90	2	0	1	1	0			
Blood Sugar (mmol/L)	<=7,8	69	3	41	25	1,1735835	0,435404287	0,717252478	0,607044661
	> 7,8	17	9	8	0	0			
Body Temp (F)	<=99	59	4	32	23	1,271763	0,122010728	0,897684493	0,135917161
	>99	27	8	17	2	1,2183368			
Heart Rate (denyut /menit)	<=77	30	5	9	16	1,4355917	0,130236477	0,933025295	0,139585151
	>77	56	7	40	9	1,1456061			

Dari tabel 3 warna kuning merupakan perhitungan gain tertinggi yang bisa untuk dijadikan cabang berikutnya. Maka dapat ditentukan hasil pohon keputusan sebagai berikut:



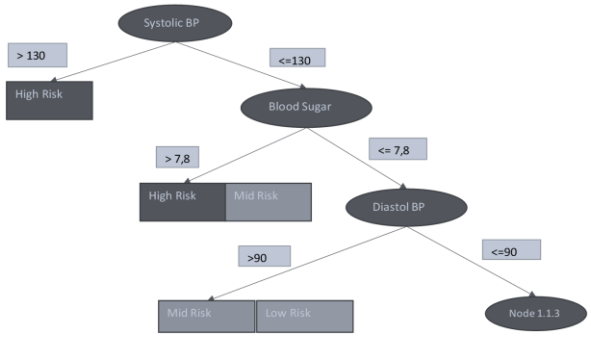
Gambar 2. Pohon Keputusan Cabang 1

Tahap selanjutnya adalah melakukan perhitungan entropy dan gain untuk cabang ke 2. Hasil perhitungannya seperti tabel berikut ini:

Tabel 4. Perhitungan cabang pohon 2

Atribut	Nilai	Jumlah Data	Jumlah (high)	Jumlah (mid)	Jumlah (low)	Entropy	Gain	Split Info	Gain Ratio
Blood Sugar (<=7,8)		69	3	41	25	1,1735835			
Age (year)	< =25	16	1	3	12	1,0140977	0,559809849	0,88177307	0,634868391
	>25	53	2	38	13	1,0198646			
Diastolic BP (mmHg)	<=90	67	3	40	24	1,1754756	0,461222876	0,406794097	1,133799335
	>90	2	0	1	1	0			
Body Temp (F)	<=99	53	1	29	23	1,1067247	0,497501995	0,88177307	0,564206384
	>99	16	2	12	2	1,0612781			
Heart Rate (denyut /menit)	<=77	27	3	8	16	1,3195213	0,962732055	1,029685741	0,934976582
	>77	42	0	33	9	0			

Berdasarkan hasil perhitungan dari tabel 4 nilai gain tertinggi pada angka yang berwarna kuning. Pohon keputusan yang dibuat sebagai berikut:



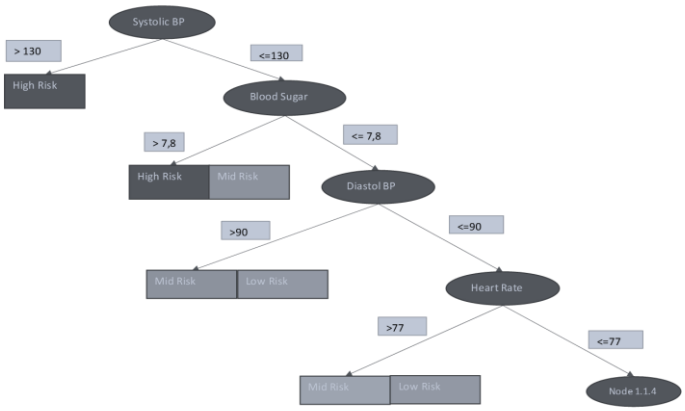
Gambar 3. Pohon Keputusan Cabang 2

Langkah selanjutnya menentukan cabang ke 3 dengan menghitung kembali Entropy dan gain seperti pada proses sebelumnya. Hasil perhitungan seperti tabel berikut ini:

Tabel 5. Perhitungan pohon keputusan cabang 3

Atribut	Nilai	Jumlah Data	Jumlah (high)	Jumlah (mid)	Jumlah (low)	Entropy	Gain	Split Info	Gain Ratio
Diastolic BP (<=90)		67	3	40	24	1,1754756			
Age (year)	< =25	14	1	2	11	0,9463729	0,59441847	0,856709134	0,693839305
	>25	53	2	38	13	1,0198646			
Body Temp (F)	<=99	51	1	28	22	1,1094168	0,521643288	0,898442397	0,580608495
	>99	16	2	12	2	1,0612781			
Heart Rate (denyut /menit)	<=77	25	3	7	15	1,323467	0,992271604	1,023093047	0,969874252
	>77	42	0	33	9	0			

Berdasarkan perhitungan entropy dan gain untuk pohon keputusan cabang 3, nilai gain tertingi pada *Heart Rate*. Maka dari itu heart rate akan menjadi cabang selanjutnya. Tampilan pohon keputusan seperti tampilan berikut ini:



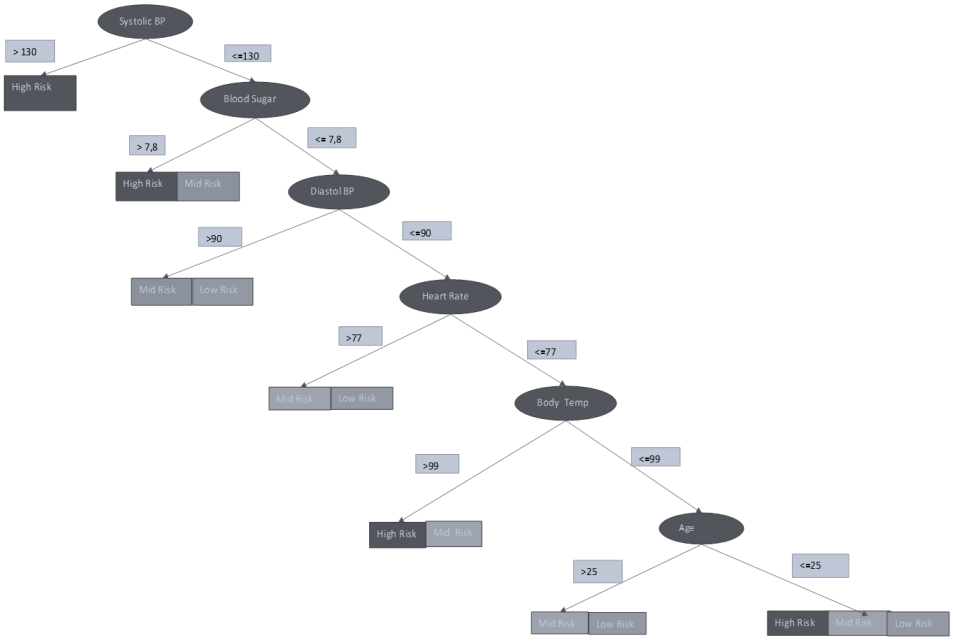
Gambar 4. Pohon Keputusan Cabang 3

Setelah mendapatkan pohon keputusan cabang 3 selanjutnya menentukan cabang 4 dengan menghitung entropy dan gain seperti cara sebelumnya. Adapun hasil perhitungan seperti pada tabel dibawah ini:

Tabel 6. Perhitungan Pohon Keputusan Cabang 4

Atribut	Nilai	Jumlah Data	Jumlah (high)	Jumlah (mid)	Jumlah (low)	Entropy	Gain	Split Info	Gain Ratio
Heart Rate (<=77)		25	3	7	15	1,323467			
Age (year)	< =25	9	1	1	7	0,9864267	1,012255421	0,792177836	1,27781336
	>25	16	2	6	8	1,4056391			
Body Temp (F)	<=99	20	1	4	15	0,9917601	1,146358475	0,728004843	1,574657759
	>99	5	2	3	0	0			

Dari perhitungan entropy dan gain pada tabel 6 nilai gain tertinggi pada atribut *Body Temp* dan dipastikan atribut *age* dengan nilai gain terendah. Dari perhitungan secara manual maka didapat pohon keputusan seperti gambar berikut:



Gambar 5. Pohon Keputusan Perhitungan Manual

3.3.Hasil Analisa

Perbandingan untuk data training dan data testing pada penelitian ini yaitu 70:30 setelah dilakukan dengan pengujian Rapidminer menghasilkan nilai accuracy sebesar 83,33%. Seperti terlihat pada tabel berikut ini.

Tabel 7. Accuracy Data Testing

accuracy: 83.33%

	true high risk	true low risk	true mid risk	class precision
pred. high risk	7	0	1	87.50%
pred. low risk	1	6	2	66.67%
pred. mid risk	0	1	12	92.31%
class recall	87.50%	85.71%	80.00%	

4. SIMPULAN DAN SARAN

Berdasarkan hasil pengujian dan pembahasan yang telah dilakukan maka dapat disimpulkan bahwa Pohon keputusan yang dihasilkan dari 100 data pasien memiliki akar (root) Systolic BP. Sehingga Systolic BP menjadi faktor resiko paling utama dibandingkan atribut yang lainnya. Penerapan Algoritma C45 menghasilkan nilai akurasi 83,33% dengan pembagian data training dan data testing 70:30. Penggunaan metode Algoritma C45 menjadi salah satu metode yang tepat untuk klasifikasi faktor resiko kesehatan ibu hamil.

Saran yang dapat diberikan bagi pengembangan penelitian selanjutnya adalah perlunya dilakukan penelitian selanjutnya dengan melakukan perbandingan terhadapa metode yang lainnya, seperti Naïve Baiyes, SVM, k-NN, dan lain -lain agar bisa melihat nilai akurasi yang paling baik. Penelitian selanjutnya dapat menambahkan atribut atau jumlah data agar diperoleh nilai akurasi yang lebih baik.

5. DAFTAR PUSTAKA

Andriani, A. (2013). Sistem prediksi penyakit diabetes berbasis decision tree. *Jurnal Bianglala Informatika*, *I*(1), 1–10.

Astuti, S. K., Aziz, M. A., Farisa, I., & Arya, D. (2017). *Artikel asli Faktor Risiko Kematian Ibu di RSUP Dr . Hasan Sadikin Bandung Tahun 2009ÿ2013*. *5*(2), 52–56.

Bauserman, M., Lokangaka, A., Thorsten, V., Tshefu, A., Goudar, S. S., Esamai, F., Saleem, S., Pasha, O., Patel, A., Berrueta, M., Kodkany, B., Chomba, E., Liechty, E. A., Hambidge, K. M., Krebs, N. F., Derman, R. J., Hibberd, P. L., Althabe, F., Carlo, W. A., … Bose, C. L. (2015). *Faktor risiko kematian dan tren angka kematian ibu di negara-negara berpenghasilan rendah dan menengah ibu : analisis prospektif kohort longitudinal*. *12*(Suppl 2), 1–9.

Hikmatulloh, H., Rahmawati, A., Wintana, D., & Ambarsari, D. A. (2019). Penerapan Algoritma Iterative Dichotomiser Three (Id3) Dalam Mendiagnosa Kesehatan Kehamilan. *Klik - Kumpulan Jurnal Ilmu Komputer*, *6*(2), 116. https://doi.org/10.20527/klik.v6i2.189

Jollyta, D., Siddik, M., Mawengkang, H., & Efendi, S. (2021). *Teknik Evaluasi Cluster Solusi Menggunakan Python dan Rapidminer*. Deepublish.

Mardi, Y. (2017). Data Mining : Klasifikasi Menggunakan Algoritma C4.5. *Edik Informatika*, *2*(2), 213–219. https://doi.org/10.22202/ei.2016.v2i2.1465

Muslim, M. A., Prasetiyo, B., Mawarni, E. L. H., Herowati, A. J., Mirqotussa'adah, Rukmana, S. H., & Nurzahputra, A. (2019). *Data Mining Algoritma C45*.

Muzakir, A., & Wulandari, R. A. (2016). Model Data Mining sebagai Prediksi Penyakit Hipertensi Kehamilan dengan Teknik Decision Tree. *Scientific Journal of Informatics*, *3*(1), 19–26. https://doi.org/10.15294/sji.v3i1.4610

Nurdiana, N., & Algifari, A. (2020). Studi Komparasi Algoritma Id3 Dan Algoritma Naive Bayes

Untuk Klasifikasi Penyakit Diabetes Mellitus. *INFOTECH Journal*, *6*(2), 18–23. https://ejournal.unma.ac.id/index.php/infotech/article/view/816

Paramita, R. W. D., Rizal, N., & Sulistyan, R. B. (2021). *Metode penelitian kuantitatif*. Widya Gama Press.

Saputra, R. A. (2014). Komparasi Algoritma Klasifikasi Data Mining Untuk Memprediksi Penyakit Tuberculosis (Tb): Studi Kasus Puskesmas Karawang. *Seminar Nasional Inovasi Dan Tren (SNIT)*, *April*, 1–8.

Sunge, A. S., & Aditasari.Ana Angelia. (2018). Penerapan Algoritma C4.5 Pada Klasifikasi Kelahiran Bayi Prematur Di Desa Setia Mekar. *Jurnal Teknologi Pelita Bangsa*, *8*.

Widiastiwi, Y., & Ernawati, I. (2021). *Klasifikasi Penyakit Batu Ginjal Menggunakan Algoritma Decision Tree C4 . 5 Dengan Membandingkan Hasil Uji Akurasi*. *5*(2), 128–135.