

Prediksi Risiko Anemia Berdasarkan Gender Menggunakan Algoritma C4.5

Asilsani Silvana¹, Widiarina², Bambang Junadi³

^{1,2,3}Sistem Informasi, Universitas Bina Sarana Informatika

Jl. Lingkar Utara No. 8, Perwira, Kec. Bekasi Utara, Kota Bekasi, Jawa Barat 17124, Indonesia

e-mail: 1asilsans04@gmail.com, 2widiarina.wda@bsi.ac.id, 3bambang.bbj@bsi.ac.id

Artikel Info : Diterima : 06-05-2026 | Direvisi : 26-05-2026 | Disetujui : 28-05-2026

Abstrak - Anemia merupakan salah satu masalah kesehatan yang terjadi di berbagai negara dan telah dikenal luas oleh masyarakat. Kondisi ini dapat memberikan dampak yang signifikan terhadap kesehatan individu. Oleh karena itu, diperlukan suatu model prediksi yang mampu mendeteksi risiko anemia secara dini sehingga tindakan pencegahan dapat dilakukan lebih cepat dan tepat. Penelitian ini bertujuan untuk memanfaatkan tools RapidMiner dalam memprediksi risiko anemia menggunakan metode Decision Tree. Data yang digunakan dalam penelitian ini merupakan data publik yang diperoleh dari Kaggle. Proses preprocessing data, pemodelan, dan evaluasi dilakukan menggunakan RapidMiner. Hasil penelitian menunjukkan bahwa model Decision Tree yang dibangun mampu memprediksi risiko anemia dengan tingkat akurasi sebesar 81,66%. Faktor-faktor yang paling berpengaruh dalam prediksi risiko anemia meliputi usia, jenis kelamin, dan status gizi. Di antara faktor-faktor tersebut, jenis kelamin (gender) menunjukkan pengaruh yang cukup signifikan terhadap hasil prediksi. Model yang dihasilkan dapat digunakan sebagai alat bantu untuk mengidentifikasi individu yang memiliki risiko tinggi mengalami anemia sehingga dapat dilakukan tindakan pencegahan dan penanganan yang lebih tepat guna mengurangi dampak serta prevalensi anemia.

Kata Kunci : Prediksi risiko anemia, Decision tree, RapidMiner

Abstracts - Anemia is one of the health problems that occurs in many countries and is widely recognized by the public. This condition can have a significant impact on an individual's health. Therefore, a predictive model is needed to detect the risk of anemia at an early stage so that preventive measures can be implemented promptly and effectively. This study aims to utilize the RapidMiner tool to predict anemia risk using the Decision Tree method. The data used in this research were obtained from a publicly available dataset on Kaggle. The processes of data preprocessing, modeling, and evaluation were carried out using RapidMiner. The results show that the Decision Tree model developed in this study is capable of predicting anemia risk with an accuracy of 81.66%. The most influential factors in predicting anemia risk include age, gender, and nutritional status. Among these factors, gender was found to have a considerable influence on the prediction results. The proposed model can be used as a decision-support tool to identify individuals at high risk of developing anemia, enabling appropriate preventive and treatment measures to reduce the impact and prevalence of anemia.

Keywords : Anemia risk prediction, Decision tree, RapidMiner

PENDAHULUAN

Menurut World Health Organization (WHO) menyatakan “bahwa lebih dari 30% penduduk di dunia mengalami anemia. Persentase pada negara maju sebesar 4,3-20% dan pada negara berkembang sebesar 30-48% dengan anemia gizi besi. Secara global, sebesar 43%, diderita anak-anak, 38% ibu hamil, 29% wanita tidak hamil dan sebesar 29% semua wanita usia subur didiagnosa anemia” (Amir & Djokosujono, 2019).

Ketika jumlah hemoglobin (HB) dalam darah lebih sedikit dalam darah lebih sedikit dari biasanya, hal ini disebut anemia (Ermawati et al., 2024). Oleh karena itu, memprediksi risiko anemia secara dini sangatlah penting dalam upaya pengobatan dan perawatan agar dapat mengetahui risiko anemia sebelum kondisinya memburuk. Serta dapat melakukan pencegahan awal sebelum terkena penyakit anemia.

Mengenai prediksi penyakit ini sudah ada penelitian sebelumnya yang dilakukan dengan berbagai model yang digunakan, salah satunya seperti teorema bayes (Sulardi & Witanti, 2020), regresi logistik (Ermawati et al.,



2024), naive bayes dan certainty factor (Yanto et al., 2020), untuk memprediksi penyakit anemia. Namun dari berbagai model yang digunakan pada penelitian sebelumnya memerlukan waktu yang cukup lama.

Terkait permasalahan tersebut, oleh karena itu peneliti akan menggunakan alat bantu tools RapidMiner dengan metode decision tree dalam penelitian prediksi risiko anemia. Menerapkan metode decision tree yang cukup populer dengan bantuan tools ini yang bersifat open source dan biasa dijadikan sebagai solusi dalam hal memprediksi (Nofitri & Irawati, 2019). Karena pada penelitian ini penerapan metode decision tree pada tools RapidMiner dapat menghasilkan model prediksi yang mudah dipahami dan dapat memberikan hasil prediksi dengan cepat. Dengan demikian, peneliti berharap bahwa penelitian ini akan bermanfaat bagi bidang medis dan dengan prediksi risiko yang tepat akan meningkatkan pencegahan dan pengobatan pasien yang berisiko anemia.

METODE PENELITIAN

A. Data Mining

Proses yang menggunakan satu atau lebih teknik pembelajaran untuk secara otomatis menganalisis dan mengekstrak pengetahuan (Putri Setyadini, 2022). Proses pencarian informasi yang tersembunyi dalam kumpulan data yang sangat besar disebut dengan data mining (Lestari & Hafiz, 2020).

B. Algoritma C4.5

Diantara algoritma data mining yang digunakan untuk membangun pohon keputusan adalah algoritma c4.5. Proses yang dilakukan dalam pohon keputusan yaitu dengan memecah data yang besar menjadi record kecil atau catatan kecil yang mengacu pada baris data individual dalam kumpulan data. Pohon keputusan dapat menunjukkan suatu pola yang cukup sederhana untuk dimengerti (Ratnawati & Natalia, 2021)

C. Decision Tree

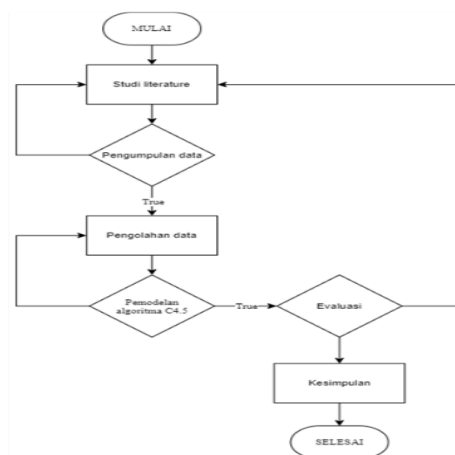
Pengambilan keputusan yang digambarkan dengan bentuk struktur pohon yang bercabang untuk memberikan hasil yang mungkin terjadi. Struktur pohon ini terdiri dari beberapa node yang terhubung dengan cabang-cabangnya (Abdul Majid et al., 2022). Pohon keputusan dianggap sebagai salah satu metodologi yang digunakan tugas klasifikasi dan prediksi karena kepraktisan dalam pemodelannya (Elfaladonna & Rahmadani, 2019).

D. RapidMiner

RapidMiner adalah tools data mining yang bersifat open source (terbuka) yang cukup populer dan canggih, dilengkapi dengan berbagai mesin data mining termasuk transformasi, pemodelan dan visualisasi data (Nofitri & Irawati, 2019). Salah satu model yang dapat dibuat dengan RapidMiner ini yaitu model decision tree yang akan diterapkan dalam penelitian ini.

E. Tahapan Penelitian

Menerapkan tools RapidMiner sebagai alat bantu dalam memprediksi risiko anemia dengan metode Algoritma C4.5. Decision tree yang sudah dikenal sebagai pemecahan solusi dalam memprediksi sesuatu. Dan dengan adanya tools RapidMiner ini akan lebih mempermudah dalam pemrosesan data dan hanya dengan waktu yang singkat dapat memberikan hasil prediksi dengan hasil yang akurat. Selain itu, metode Algoritma C4.5 mudah untuk dipahami karena pemodelan yang berbentuk seperti pohon yang bercabang-cabang.



Sumber: (Pambudi, 2023)

Gambar 1. Kerangka Pemikiran

Sebelum melakukan pengumpulan data dilakukan studi literatur terlebih dahulu yang berkaitan dengan penerapan metode algoritma c4.5 dalam memprediksi sebuah penyakit, selain itu juga studi literatur tentang masalah kesehatan yang berdampak buruk atau membahayakan seseorang. Setelah melakukan studi literatur

selanjutnya yaitu pengumpulan data yang akan digunakan dalam penelitian prediksi risiko anemia. Setelah memiliki data yang akan digunakan dalam penelitian, selanjutnya data akan melalui proses pengolahan data. Dalam pengolahan terdiri dari transformasi data dan pembagian data. Data yang sudah melakukan transformasi data dan juga split data, maka bisa melanjutkan proses selanjutnya. Proses selanjutnya yaitu proses pemodelan dimana peneliti akan membangun model decision tree dalam memprediksi risiko anemia. Setelah setelah melakukan proses pengolahan data maka dibuatlah evaluasi agar dapat mengetahui hasil dari prediksi risiko anemia. Tahap terakhir adalah membuat kesimpulan dalam penelitian yang sudah dilakukan.

F. Pemodelan Algoritma C4.5

Proses yang dilakukan dalam pohon keputusan yaitu dengan memecah data yang besar menjadi record kecil atau catatan kecil yang mengacu pada baris data individual dalam kumpulan data. Akar tertinggi dalam membuat pohon keputusan yaitu dari nilai gain tertinggi. Namun sebelum mencari nilai gain tertinggi maka, harus mengetahui nilai entropy setiap atribut. Dengan menggunakan rumus seperti dibawah ini.

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

Keterangan:

S = Total

P_i = jumlah status nilai atribut

Setelah mengetahui nilai entropy maka selanjutnya mencari nilai gain. Dalam menentukan nilai gain tertinggi ditentukan berdasarkan rumus di bawah ini.

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

Keterangan:

A = Atribut

S = Total

|S_i| = Jumlah kasus nilai atribut

S_i = nilai setiap atribut

G. Evaluasi

Setelah melakukan proses pengolahan data maka dibuatlah evaluasi agar dapat mengetahui hasil dari prediksi risiko anemia dengan rumus sebagai berikut.

$$Accuracy = \frac{Tp + Tn}{Tp + Tn + Fp + Fn} \times 100\%$$

HASIL DAN PEMBAHASAN

A. Pengumpulan Data

Peneliti memperoleh dataset ini dalam website <https://www.kaggle.com/datasets/ragishehab/anemia-data-set>. Data pada penelitian ini termasuk data publik. Dalam dataset ini terdapat 8.544 record dan memiliki 5 atribut dan 2 variabel penentu. 5 atribut dalam dataset terdiri dari, gender, hemoglobin, MCH, MCHC, MCV. Dan 2 variabel penentu yaitu anemic dan not anemic.

Tabel 1. Dataset Anemia

No.	Gender	Hemoglobin	MCH	MCHC	MCV	Result
1	0	15.2	29.9	33.4	89.3	0
2	0	11.9	31	32.5	95.4	1
3	0	17.2	31.1	34.3	90.5	0
.....
8.544	0	13.7	27.2	33.7	80.7	0

B. Transformasi Data

Melakukan perubahan dari data bentuk numerik diubah menjadi data bentuk nominal pada kolom atribut

hemoglobin, MCH, MCHC dan MCV agar dapat lebih mudah di proses dalam tools RapidMiner. Sedangkan untuk atribut gender dan results tidak diperlukan transformasi data karna sudah memiliki data bentuk nominal. Berikut ini gambar yang dihasilkan dari proses transformasi data.

Tabel 2. Hasil transformasi data

No.	Gender	Hemoglobin	MCH	MCHC	MCV	Result
1	male	normal	normal	normal	normal	not anemic
2	male	rendah	normal	normal	tinggi	anemic
3	male	normal	normal	normal	normal	not anemic
....
8.544	male	normal	normal	normal	rendah	not anemic

C. Split Data

Pembagian dataset menjadi data training dan testing. Rasio yang digunakan adalah 70:30 dengan artian 70% dari 8.544 dataset yaitu 5.981 data training dan 30% dari 8.544 dataset yaitu 2.563 data testing. Dibawah ini terdapat gambar hasil pembagian data training dan data testing.

Row No.	Result	prediction(R...	confidence(L...	confidence(R...	Gender	Hemoglobin	MCH	MCHC	MCV
1	not anemic	not anemic	1	0	male	normal	normal	normal	normal
2	not anemic	not anemic	1	0	male	normal	normal	normal	normal
3	not anemic	anemic	0.471	0.529	female	normal	normal	normal	normal
4	not anemic	not anemic	1	0	male	normal	normal	normal	normal
5	not anemic	not anemic	1	0	male	normal	rendah	normal	rendah
6	not anemic	not anemic	1	0	male	normal	normal	normal	rendah
7	anemic	anemic	0.471	0.529	female	normal	normal	normal	normal
8	anemic	anemic	0.471	0.529	female	normal	normal	normal	normal
9	not anemic	anemic	0.471	0.529	female	normal	normal	normal	normal
10	anemic	anemic	0	1	female	rendah	normal	normal	rendah
11	not anemic	not anemic	1	0	male	normal	normal	normal	normal
12	not anemic	not anemic	1	0	male	normal	normal	normal	normal
13	not anemic	anemic	0.471	0.529	female	normal	normal	normal	tinggi
14	anemic	anemic	0	1	female	rendah	rendah	rendah	rendah
15	anemic	anemic	0.124	0.876	female	normal	rendah	normal	rendah

Gambar 2. Hasil data testing

D. Proses Pemodelan

Langkah pertama adalah mengetahui jumlah kasus dalam setiap atribut yang tersedia dalam dataset sehingga akan menghasilkan seperti gambar dibawah ini.

Tabel 3. Hasil prediksi dalam nilai setiap atribut berdasarkan dataset

Atribut		Jumlah kasus (S)	Anemic (S1)	Not Anemic (S2)
Total		8544	2904	5640
Gender	male	28	2	26
	female	8	3	5
Hemoglobin	rendah	3	3	0
	normal	35	3	32
	tinggi	0	0	0
MCH	rendah	5	2	3
	normal	19	6	13
	tinggi	4	0	4
MCHC	rendah	0	0	0
	normal	40	8	32
	tinggi	0	0	0
MCV	rendah	6	2	4
	normal	28	5	23
	tinggi	5	1	4

Langkah selanjutnya adalah menghitung entropy dan gain. Yang pertama mencari nilai entropy total, berikut ini

cara mencari nilai entropy total.

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

$$Entropy (Total) = (- 2,904/8,544 * \log_2 (2,904/8,544)) + (- 5,640/8,544 * \log_2 (5,640/8,544)) = 0,9247$$

Langkah selanjutnya adalah mencari nilai entropy setiap nilai atribut masing-masing. Yang terdapat dalam atribut gender, hemoglobin, MCH, MCHC dan MCV. Berikut yang pertama nilai yang terdapat pada atribut gender.

$$Entropy (male) = (- 2/28 * \log_2 (2/28)) + (- 26/28 * \log_2 (26/28)) = 0,3712$$

$$Entropy (female) = (- 3/8 * \log_2 (3/8)) + (- 5/8 * \log_2 (5/8)) = 0,9544$$

Yang kedua, ada atribut hemoglobin yang terdiri dari rendah, normal dan tinggi.

$$Entropy (rendah) = (- 3/3 * \log_2 (3/3)) + (- 0/3 * \log_2 (0/3)) = 0$$

$$Entropy (normal) = (- 3/35 * \log_2 (3/35)) + (- 32/35 * \log_2 (32/35)) = 0,4220$$

$$Entropy (tinggi) = (- 0/0 * \log_2 (0/0)) + (- 0/0 * \log_2 (0/0)) = 0$$

Yang ketiga, ada atribut MCH (Mean Corpuscular Hemoglobin) yang terdiri dari rendah, normal dan tinggi.

$$Entropy (rendah) = (- 2/5 * \log_2 (2/5)) + (- 3/5 * \log_2 (3/5)) = 0,9710$$

$$Entropy (normal) = (- 6/19 * \log_2 (6/19)) + (- 13/19 * \log_2 (13/19)) = 0,4220$$

$$Entropy (tinggi) = (- 0/4 * \log_2 (0/4)) + (- 4/4 * \log_2 (4/4)) = 0$$

Yang keempat, ada atribut MCHC (Mean Corpuscular Hemoglobin Concentration) yang terdiri dari rendah, normal dan tinggi.

$$Entropy (rendah) = (- 0/0 * \log_2 (0/0)) + (- 0/0 * \log_2 (0/0)) = 0$$

$$Entropy (normal) = (- 8/40 * \log_2 (8/40)) + (- 32/40 * \log_2 (32/40)) = 0,7219$$

$$Entropy (tinggi) = (- 0/0 * \log_2 (0/0)) + (- 0/0 * \log_2 (0/0)) = 0$$

Yang terakhir, ada atribut MCV (Mean Corpuscular Volume) yang terdiri dari rendah, normal dan tinggi.

$$Entropy (rendah) = (- 2/6 * \log_2 (2/6)) + (- 0/0 * \log_2 (4/6)) = 0,9183$$

$$Entropy (normal) = (- 5/28 * \log_2 (5/28)) + (- 23/28 * \log_2 (23/28)) = 0,6769$$

$$Entropy (tinggi) = (- 1/5 * \log_2 (1/5)) + (- 4/5 * \log_2 (4/5)) = 0,7219$$

Setelah semua entropy masing-masing nilai atribut sudah diketahui, maka selanjutnya adalah menghitung gain pada baris atribut gender, hemoglobin, MCH dan MCHC. Berikut rumus perhitungan gain:

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

$$Gain (Gender) = 0,9247 - ((28/8,544) * 0,3712) + ((8/8,544) * 0,9544) = 0,9226$$

$$Gain (Hemoglobin) = 0,9247 - ((3/8,544) * 0) + ((35/8,544) * 0,4220) + ((0/8,544) * 0) = 0,9226$$

$$Gain (MCH) = 0,9247 - ((5/8,544) * 0,9710) + ((19/8,544) * 0,8997) + ((4/8,544) * 0) = 0,2833$$

$$Gain (MCHC) = 0,9247 - ((0/8,544) * 0) + ((40/8,544) * 0,7219) + ((0/8,544) * 0) = 0,8997$$

$$Gain (MCV) = 0,9247 - ((0/8,544) * 0) + ((40/8,544) * 0,7219) + ((0/8,544) * 0) = 0,5842$$

Tabel 4. Hasil perhitungan entropy dan gain

Atribut		Jumlah kasus (S)	Anemic (S1)	Not Anemic (S2)	Entropy	Gain
Total		8544	2904	5640	0.9247	
Gender						0.9226
	male	28	2	26	0.3712	
	female	8	3	5	0.9544	
Hemoglobin						0.9230
	rendah	3	3	0	0	
	normal	35	3	32	0.4220	
	tinggi	0	0	0	0	
MCH						0.2833
	rendah	5	2	3	0.9710	

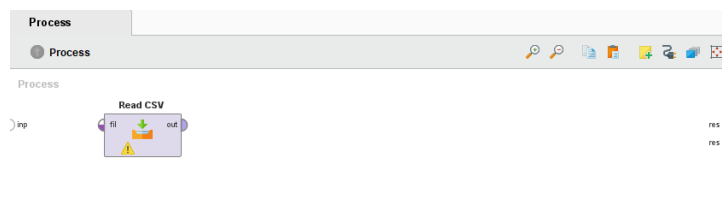
MCHC	normal	19	6	13	0.8997	
	tinggi	4	0	4	0	
						0.8997
	rendah	0	0	0	0	
MCV	normal	40	8	32	0.7219	
	tinggi	0	0	0	0	
						0.5842
	rendah	6	2	4	0.9183	
	normal	28	5	23	0.6769	
	tinggi	5	1	4	0.7219	

Berdasarkan hasil perhitungan entropy dan gain pada Tabel 4 diketahui nilai gain yang tertinggi yaitu hemoglobin dengan jumlah 0.9230, maka akar pertama yang digunakan dalam pembuatan pohon keputusan adalah hemoglobin. Dan dilanjutkan dengan akar kedua yaitu gender dengan gain sebesar 0.9226 dan seterusnya seperti itu.

E. Penerapan pada tools RapidMiner

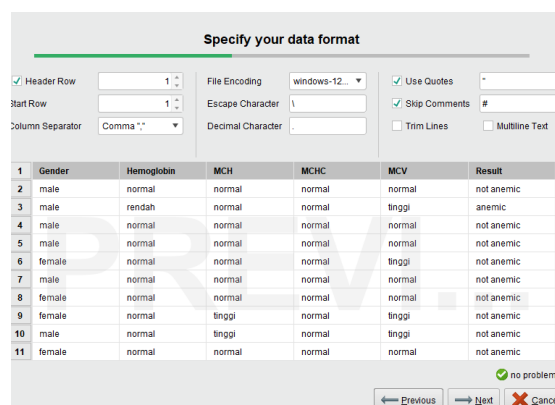
Dalam pemrosesan data dalam RapidMiner adalah melakukan pengimporan data kedalam tools RapidMiner. Langkah pertama yang dilakukan adalah melakukan tracking operator “read csv” halaman proses. Karena dataset yang akan diimport memiliki format csv. Maka dari itu sangat diperlukan operator “read csv”.

Sebelum mentracking operator, maka harus melakukan pencarian dahulu pada kolom operator. maka ketikkan pada kolom yang tersedia dan setelah itu klik pada ikon “read csv”.



Gambar 3. Sebelum pengimporan dataset

Gambar 3 menyajikan tampilan pada tools RapidMiner yang sudah terdapat operator “read csv” pada halaman proses. Tanda seru yang terdapat pada operator “read csv” dikarenakan belum melakukan pengimporan dataset kedalam operator “read csv”



Gambar 4. Tampilan saat melakukan pengimporan data

Setelah muncul tampilan pada Gambar 4 kita hanya perlu mengklik tombol next saja dan langsung lanjut ke proses selanjutnya. Proses selanjutnya adalah perubahan tipe data dari polynominal menjadi binomial, karena hanya memiliki 2 nilai saja yaitu (male/female) dan juga penrubahan tipe data pada atribut results diubah menjadi binominal karna hanya memiliki 2 nilai yaitu (anemic/not anemic). Sedangkan pada atribut hemoglobin, MCH, MCHC dan MCV memiliki 3 nilai (rendah, normal, tinggi). Jadi, tetap menggunakan tipe data polynominal karna tipe data polynominal pada RapidMiner diartikan dengan tipe data untuk atribut yang memiliki lebih dari 2 nilai.

Format your columns.

Date format: Replace errors with missing values

	Gender <i>binominal</i>	Hemoglobin <i>polynomial</i>	MCH <i>polynomial</i>	MCHC <i>polynomial</i>	MCV <i>polynomial</i>	Result <i>binominal</i>
1	male	normal	normal	normal	normal	not anemic
2	male	rendah	normal	normal	tinggi	anemic
3	male	normal	normal	normal	normal	not anemic
4	male	normal	normal	normal	normal	not anemic
5	female	normal	normal	normal	tinggi	not anemic
6	male	normal	normal	normal	normal	not anemic
7	female	normal	normal	normal	normal	not anemic
8	female	normal	tinggi	normal	tinggi	not anemic
9	male	normal	tinggi	normal	tinggi	not anemic
10	female	normal	normal	normal	normal	not anemic
11	male	normal	normal	normal	normal	not anemic
12	female	normal	rendah	normal	rendah	anemic

no problems.

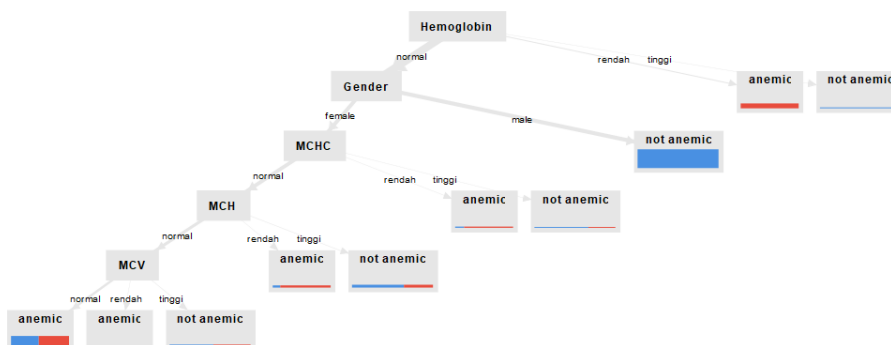
Gambar 5. Setelah melakukan perubahan tipe data

Setelah melakukan perubahan tipe data maka, tahap selanjutnya adalah menentukan atribut yang akan dijadikan sebagai label (target). Pada penelitian ini peneliti menjadikan atribut result sebagai label. Selanjutnya, melakukan split data terlebih dahulu. Dengan mentracking operator “split data” pada halaman proses dengan cara seperti yang dilakukan pada saat mentracking operator “read csv” dan Untuk membagi data menjadi testing dan training. Dalam pembagian data training dan data testing dalam penelitian ini menggunakan rasio 70:30 dan dapat dituliskan dengan 0.7 dan 0.3

ratio
0.7
0.3

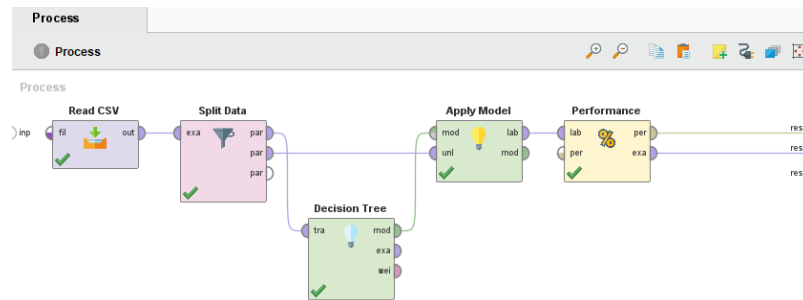
Gambar 6 Rasio split data

Untuk melihat hasil split data training, maka kita harus merunning dengan cara mengklik tombol start (segitiga biru) seperti yang terlihat pada diatas. Kemudian klik “ExampleSet (Split Data)” yang berada disebelah “Tree (Decision Tree)” dan untuk melihat hasil prediksi yang berbentuk pohon keputusan, klik “Tree (Decision Tree)” yang berada di sebelah “ExampleSet (Split Data)”. Maka akan muncul gambar seperti berikut.



Gambar 7. Hasil prediksi berupa pohon keputusan

Hasil prediksi risiko anemia ini menggambarkan bahwa wanita lebih tinggi berisiko terkena anemia dibandingkan pria. Karena walaupun kadar hemoglobin wanita itu normal tetap harus dilihat dari kadar MCHC, MCH dan MCVnya. Sedangkan pria jika hemoglobinnya normal maka dapat dinyatakan negatif berisiko anemia.



Gambar 8. Pemodelan dalam RapidMiner secara keseluruhan

Gambar 8 merupakan tahap terakhir setelah melalui berbagai proses yang sudah dilakukan yaitu mendapatkan hasil tingkat akurasi prediksi risiko anemia yang telah diproses oleh RapidMiner. Selanjutnya lakukan tracking “Apply Model” dan “Performance”.

The screenshot shows the 'PerformanceVector (Performance)' window with 'Table View' selected. It displays an accuracy of 81.66% and a confusion matrix table.

	true not anemic	true anemic	class precision
pred. not anemic	1338	116	92.02%
pred. anemic	354	755	68.08%
class recall	79.08%	86.68%	

Gambar 9. Hasil tingkat akurasi

Pada Gambar 9 menyajikan hasil dari tingkat akurasi pengujian dataset anemia yang sudah di proses oleh RapidMiner. Dalam tabel tersebut menyatakan bahwa yang prediksi anemia pada tools RapidMiner berjumlah 755 yang sesuai dengan dataset dan sedangkan prediksi not anemic berjumlah 1.338.

The screenshot shows the 'ExampleSet (Apply Model)' window with a detailed data table. The table has 15 rows and 10 columns.

Row No.	Result	prediction(R...	confidence(...	confidence(...	Gender	Hemoglobin	MCH	MCHC	MCV
1	anemic	anemic	0	1	male	rendah	normal	normal	tinggi
2	not anemic	anemic	0.478	0.522	female	normal	normal	normal	normal
3	not anemic	not anemic	0.641	0.359	female	normal	tinggi	normal	tinggi
4	not anemic	anemic	0.478	0.522	female	normal	normal	normal	normal
5	not anemic	anemic	0.478	0.522	female	normal	normal	normal	normal
6	not anemic	not anemic	1	0	male	normal	normal	normal	normal
7	not anemic	not anemic	1	0	male	normal	normal	normal	normal
8	not anemic	not anemic	1	0	male	normal	normal	normal	normal
9	not anemic	not anemic	1	0	male	normal	normal	normal	normal
10	not anemic	anemic	0.478	0.522	female	normal	normal	normal	normal
11	anemic	anemic	0	1	female	rendah	normal	normal	rendah
12	not anemic	not anemic	1	0	male	normal	normal	normal	normal
13	not anemic	not anemic	1	0	male	normal	normal	normal	normal
14	not anemic	not anemic	1	0	male	normal	normal	normal	normal
15	anemic	anemic	0	1	female	rendah	rendah	normal	rendah

Gambar 10. Tampilan data hasil

F. Evaluasi

Pada penelitian ini, menyatakan bahwa 755 data positif benar, 1338 data negatif benar, 354 data positif salah dan 116 data negatif salah. Tingkat akurasi yang dihasilkan dalam penelitian ini adalah 81.66%. Untuk memastikan nilai akurasi dapat dilihat dari perhitungan dibawah ini.

$$Accuracy = \frac{Tp+Tn}{Tp+Tn+Fp+Fn} \times 100\%$$

$$Accuracy = \frac{755+1338}{755+1338+354+116} \times 100\% = 81.66\%$$

KESIMPULAN

Berdasarkan hasil penelitian, dapat disimpulkan bahwa penggunaan algoritma C4.5 berhasil memprediksi risiko anemia dengan tingkat akurasi sebesar 81,66%. Hasil prediksi menunjukkan bahwa jenis kelamin merupakan salah satu faktor yang berpengaruh dalam menentukan risiko anemia. Pada hasil pengujian, seluruh responden laki-laki dengan kadar hemoglobin normal diklasifikasikan tidak berisiko mengalami anemia. Sementara itu, pada responden perempuan dengan kadar hemoglobin normal, penentuan risiko anemia masih dipengaruhi oleh nilai MCH (Mean Corpuscular Hemoglobin), MCHC (Mean Corpuscular Hemoglobin Concentration), dan MCV (Mean Corpuscular Volume). Selain jenis kelamin, kadar hemoglobin, MCH, MCHC, dan MCV juga merupakan faktor penting yang memengaruhi prediksi risiko anemia. Hasil penelitian menunjukkan bahwa perempuan memiliki risiko anemia yang lebih tinggi dibandingkan laki-laki. Model yang dihasilkan dalam penelitian ini dapat dimanfaatkan sebagai alat bantu deteksi dini anemia sehingga upaya pencegahan dan penanganan dapat dilakukan secara lebih cepat dan tepat. Untuk penelitian selanjutnya, disarankan untuk menambahkan faktor-faktor risiko lain yang berpotensi meningkatkan akurasi model prediksi. Selain itu, penelitian dapat dilakukan dengan jumlah data yang lebih besar dan lebih beragam agar model yang dihasilkan memiliki kemampuan generalisasi yang lebih baik. Penelitian berikutnya juga dapat membandingkan kinerja algoritma C4.5 dengan algoritma klasifikasi lainnya, seperti Support Vector Machine (SVM), Naïve Bayes, dan Neural Network. Lebih lanjut, model yang telah dikembangkan dapat diimplementasikan ke dalam bentuk aplikasi deteksi risiko anemia guna mendukung proses skrining dan pengambilan keputusan di bidang kesehatan.

REFERENSI

- Abdul Majid, M. B., Cani, Y. M., & Enri, U. (2022). Penerapan Algoritma K-Means dan Decision Tree Dalam Analisis Prestasi Siswa Sekolah Menengah Kejuruan. *Jurnal Sistem Komputer Dan Informatika (JSON)*, 4(2), 355. <https://doi.org/10.30865/json.v4i2.5299>
- Ahmad, I., Samsugi, S., & Irawan, Y. (2022). Implementasi Data Mining Sebagai Pengolahan Data. *Jurnal Teknoinfo*, 16(1), 46. <http://portaldata.org/index.php/portaldata/article/view/107>
- Amir, N., & Djokosujono, K. (2019). Faktor-Faktor yang Berhubungan dengan Konsumsi Tablet Tambah Darah (TTD) pada Remaja Putri di Indonesia: Literatur Review. *Jurnal Kedokteran Dan Kesehatan*, 15(2), 119. <https://doi.org/10.24853/jkk.15.2.119-129>
- Elfaladonna, F., & Rahmadani, A. (2019). Analisa Metode Classification-Decision Tree Dan Algoritma C.45 Untuk Memprediksi Penyakit Diabetes Dengan Menggunakan Aplikasi Rapid Miner. *SINTECH (Science and Information Technology) Journal*, 2(1), 10–17. <https://doi.org/10.31598/sintechjournal.v2i1.293>
- Ermawati, Ibbas, R., & Kurniawan, B. A. (2024). Klasifikasi Penderita Anemia Menggunakan Metode Regresi Logistik. *Jurnal MSA (Matematika Dan Statistika Serta Aplikasinya)*, 11(2), 93–101. <https://doi.org/10.24252/msa.v11i2.45083>
- Hana, F. M. (2020). Klasifikasi Penderita Penyakit Diabetes Menggunakan Algoritma Decision Tree C4.5. *Jurnal SISKOM-KB (Sistem Komputer Dan Kecerdasan Buatan)*, 4(1), 32–39. <https://doi.org/10.47970/siskom-kb.v4i1.173>
- Indriani, F., Rahayu, R. P., Diii, P., Teknologi, K. I., & Indragiri, B. (2023). Factors Related to Incidence of Anemia in Adolescent Girl Faktor-Faktor yang Berhubungan dengan Kejadian Anemia Pada Remaja Putri. *INCH : Journal of Infant And Child Healthcare*, 2(1), 33–39.
- Irawan, E., & Gunawan, I. (2019). Penerapan C4.5 pada Keaktifan Mahasiswa dalam Pengumpulan Berkas di Biro Akademik. *REMIK (Riset Dan E-Jurnal Manajemen Informatika Komputer)*, 3(2), 87. <https://doi.org/10.33395/remik.v3i2.10132>
- Lestari, A. F., & Hafiz, M. (2020). Penerapan Algoritma Apriori Pada Data Penjualan Barbar Warehouse. *INOVTEK Polbang - Seri Informatika*, 5(1), 96. <https://doi.org/10.35314/isi.v5i1.1317>
- Meila Azzahra Sofyan, F., Voutama, A., & Umaidah, Y. (2023). Penerapan Algoritma C4.5 Untuk Prediksi Penyakit Paru-Paru Menggunakan Rapidminer. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(2), 1409–1415. <https://doi.org/10.36040/jati.v7i2.6810>
- Nofitri, R., & Irawati, N. (2019). Integrasi Metode Neive Bayes Dan Software Rapidminer Dalam Analisis Hasil Usaha Perusahaan Dagang. *JURTEKSI (Jurnal Teknologi Dan Sistem Informasi)*, 6(1), 35–42. <https://doi.org/10.33330/jurteksi.v6i1.393>
- Pambudi, A. (2023). Penerapan Crisp-Dm Menggunakan Mlr K-Fold Pada Data Saham Pt. Telkom Indonesia (Persero) Tbk (Tlkm) (Studi Kasus: Bursa Efek Indonesia Tahun 2015-2022). *Jurnal Data Mining Dan Sistem Informasi*, 4(1), 1. <https://doi.org/10.33365/jdmsi.v4i1.2462>
- Putri Setyadini, N. (2022). Penerapan Data Mining Untuk Prediksi Hasil Produksi Karet Menggunakan Algoritma Decision Tree C4.5. *Informatika*, 2(7), 1–11.
- Ratnawati, R., & Natalia, F. (2021). Penerapan Algoritma C4.5 pada Pengaruh Iklan Tanpa Label Harga di Instagram. *Sistemasi*, 10(1), 172. <https://doi.org/10.32520/stmsi.v10i1.1131>
- Sibarani, A. J. P. (2020). Implementasi Data Mining Menggunakan Algoritma Apriori Untuk Meningkatkan Pola

- Penjualan Obat. *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)*, 7(2), 262–276. <https://doi.org/10.35957/jatisi.v7i2.195>
- Sulardi, N., & Witanti, A. (2020). Sistem Pakar Untuk Diagnosis Penyakit Anemia Menggunakan Teorema Bayes. *Jurnal Teknik Informatika (Jutif)*, 1(1), 19–24. <https://doi.org/10.20884/1.jutif.2020.1.1.12>
- Susanti, R., Anggraeni, I., & Kautsari, W. (2023). Prediksi Tingkat Keparahan Anemia Pada Ibu Hamil Trimester I Dengan Menggunakan Model Logit Regresi Logistik Ordinal. *Jurnal Keluarga Berencana*, 8(2), 79–86. <https://doi.org/10.37306/kkb.v8i2.119>
- Yanto, M., Khairiazaz, Y., & Kunci, K. (2020). *Komparasi Metode Naive Bayes dan Certainty Factor untuk Mendiagnosa Penyakit Anemia Pendahuluan Metode Penelitian*. 19, 1–8.