

# Klasifikasi Jenis Pekerjaan bagi Penderita Asma Dengan Algoritma C4.5

Ririn Restu Aria  
Universitas Bina Sarana Informatika  
Jakarta, Indonesia

[ririn.rra@bsi.ac.id](mailto:ririn.rra@bsi.ac.id)

\*Penulis Korespondensi

Diajukan : 31/07/2025

Diterima : 07/08/2025

Dipublikasi : 13/08/2025

## ABSTRAK

Penyakit asma yang dimiliki oleh seseorang sangat mempengaruhi ketika akan melakukan sebuah pekerjaan yang akan dilakukan, permasalahan akan muncul ketika penderita melakukan sebuah pekerjaan dilokasi yang dapat menyebabkan penyakit asmanya kambuh sehingga mengganggu produktifitas yang dimiliki. Maka diperlukan solusi bagaimana bisa menentukan jenis pekerjaan yang sesuai dengan kriteria yang dimiliki oleh penderita berdasarkan lokasi pekerjaan yang akan dilakukan. Penelitian ini dilakukan dengan menerapkan algoritma C4.5 dengan 5 atribut yang digunakan yaitu jenis kelamin, status merokok, alergi, polusi udara dan kegiatan fisik yang dilakukan sedangkan untuk label dipilih atribut jenis pekerjaan dengan jumlah data yang diolah sebanyak 300 penderita asma. Berdasarkan hasil perhitungan didapatkan penentuan jenis pekerjaan yang akan dipilih dipengaruhi oleh polusi udara yang memiliki *Gain Ratio* terbesar dengan nilai 1,317. Hasil pengujian validasi menggunakan *confusion matrix* tingkat *accuracy* dalam memprediksi capaian pembelajaran daring rata-rata sebesar 81,67%, *precision* sebesar 88.38%, *recall* sebesar 84,54%. Dengan hasil nilai *accuracy* yang didapat tersebut maka dapat dikatakan termasuk klasifikasi yang baik.

**Kata Kunci:** Algoritma C4.5, *Data Mining*, *Gain Ratio*

## I. PENDAHULUAN

Asma adalah salah satu penyakit kronis pada saluran napas yang ditandai dengan peradangan dan penyempitan pada saluran napas. Saluran napas menjadi sensitif terhadap berbagai rangsangan seperti *alergen*, udara dingin, polusi udara atau aktivitas fisik. Penyakit Asma dapat disebabkan oleh beberapa faktor yaitu faktor genetik dan faktor lingkungan. Untuk faktor genetik dengan riwayat keluarga asma dapat meningkatkan resiko seseorang memiliki penyakit ini jika salah satu atau kedua orang tua memiliki asma maka anak mengalami asma juga tinggi sedangkan faktor lingkungan terjadi karena paparan terhadap zat – zat iritan tertentu dapat memicu reaksi asma, *alergen* umum yang dapat memicu asma yaitu serbuk sari, bulu binatang, debu rumah, tungau debu, jamur dan serbuk kayu dan juga polusi udara, asap rokok dan bahan kimia tertentu (Kemenkes, 2024). Menurut Kementerian Kesehatan RI, Penyakit asma merupakan salah satu penyakit yang saat ini paling banyak diidap oleh masyarakat Indonesia sebanyak 4,5% dari total jumlah penduduk Indonesia (Jakarta, 2023). Asma dapat terjadi karena munculnya inflamasi kronis di jalan napas karena hiperaktivitas bronkus dan obstruksi jalan napas. Penyakit ini mengganggu terbawanya oksigen ke paru-paru dan rongga dada (Ansyari et al., 2023)

Klasifikasi jenis pekerjaan bagi penderita asma sangat penting sebagai salah satu langkah penting dalam mendeteksi dini, pengobatan dan pencegahan penyakit asma. Untuk itu penulis menggunakan metode *data mining* yang dapat mengklasifikasikan jenis pekerjaan bagi penderita asma. *Data mining* digunakan dengan menggunakan serangkaian proses yang memanfaatkan satu atau lebih teknik pembelajaran mesin secara otomatis dalam menganalisis dan mengekstrak

informasi dari kumpulan data. Salah satu pengklasifikasi *data mining* adalah algoritma C4.5 yang digunakan untuk mengklasifikasi jenis pekerjaan bagi penderita penyakit asma sehingga mendapatkan akurasi yang diperoleh dari hasil proses evaluasi yang dilakukan. Tahapan-tahapan ini akan diulangi untuk setiap cabang sampai semua kasus pada cabang habis terpartisi. Dari penyelesaian tersebut maka akan didapatkan beberapa *rule*. Diharapkan penelitian ini dapat membantu penderita asma dalam menentukan jenis pekerjaan yang akan mereka lakukan dalam bidang yang dikerjakan saat ini.

## II. STUDI LITERATUR

### 2.1 Penelitian Terdahulu

Penelitian pertama dilakukan oleh Achmad Ridwan dengan judul “Penerapan Algoritma C4.5 Untuk Klasifikasi Penyakit Diabetes Mellitus” dengan *dataset* yang diperoleh dari *dataset UCI Machine Learning* yaitu *Early Stage Diabetes Risk* pada tahun 2020 dan terdiri dari 17 atribut. Analisis penelitian ini antara lain *Data Preprocessing*, *Modeling*, dan *Evaluasi*. Model klasifikasi penelitian memakai Algoritma C4.5. Hasil Klasifikasi pada riset ini didapatkan akurasi sebesar 94.23% dan nilai AUC nya yaitu 0,952 (Ridwan, 2022).

Penelitian kedua oleh Farhan Muzakki, Iqbal Ubaydillah, Nur Rahma Assyiami, Sugiyarti Soleha dengan judul “Penerapan Algoritma C4.5 Untuk Prediksi Penyakit Jantung Menggunakan Rapidminer” dengan hasil penelitian menggunakan metode *Decision Tree* melalui algoritma C4.5 hasil yang diperoleh mulai dari nilai *accuracy*, *precision*, dan *recall* dengan membandingkan *split* data pada data *training* dan data *testing* yang terbesar adalah *split* data 90 –10 yaitu *accuracy* = 65.25%, *recall* = 70.87% , dan *precision* = 62.46%. Dengan menggunakan algoritma C4.5 ini diharapkan dapat menghasilkan akurasi yang tepat untuk memprediksi penyakit jantung (Muzakki et al., 2024)

Penelitian ketiga oleh Adinda Dwi Putri, Fitriana Sholekhah, Eric Dadynata, Lusiana Efrizoni, Rahmaddeni, Nur Sapina dengan judul “Penerapan Algoritma *Decision Tree* C4.5 untuk Memprediksi Tingkat Kelangsungan Hidup Pasien Kanker Tiroid dengan memanfaatkan pengolahan bahasa alami (NLP) dengan *Count Vectorizer* untuk mengubah teks menjadi data numerik. Dalam penilaian keakuratan prediksi, evaluasi dilakukan dengan matriks kebingungan (*confusion matrix*) untuk mengukur kinerja model dalam klasifikasi. Selain itu, *Area Under Curve* (AUC) juga dihitung untuk mengevaluasi *performa* model. Hasil eksperimen menunjukkan bahwa metode ini memberikan prediksi yang akurat tentang tingkat kelangsungan hidup pasien dengan kanker tiroid, mencapai akurasi sebesar 97% dan AUC sebesar 0.95, menunjukkan kinerja yang sangat baik. Penelitian ini penting untuk memperdalam pemahaman tentang penerapan *Decision Tree* dalam konteks medis dan potensi algoritma ini dalam mendukung pengambilan keputusan klinis di masa depan (Putri et al., 2024).

Penelitian keempat oleh Jefri Junifer Pangaribuan, Romindo, Mirza Ilhami, Segar Napitupulu, Wenripin Chandra dengan judul “Prediksi Kanker Payudara Melalui Penerapan Algoritma C4.5” dengan tujuan penelitian menggunakan penerapan algoritma prediksi C4.5 adalah dapat menghasilkan pohon keputusan yang memiliki tingkat akurasi yang dapat diterima dan efisien dalam menangani atribut yang bertipe diskrit atau numerik. Hasil penelitian menunjukkan bahwa penerapan algoritma *Data Mining Decision Tree* C4.5 digunakan untuk melakukan prediksi penyakit kanker payudara menggunakan data *training* rekam medis sebanyak 116 data dapat menghasilkan sebanyak 4 *rule*. Tingkat keakuratan dari algoritma *Data Mining Decision Tree* C4.5 cukup akurat karena dari 10 percobaan terdapat 9 percobaan yang mendapatkan hasil prediksi yang tepat sesuai *rule* yang ada (Pangaribuan et al., 2023)

Beberapa penelitian sebelumnya menunjukkan bahwa penelitian yang dilakukan berkaitan dengan berbagai penyakit namun untuk penyakit asma belum ada. Oleh karena itu peneliti melakukan pengklasifikasian jenis pekerjaan yang akan dilakukan berdasarkan tempat atau kondisi bagi penderita asma dengan menggunakan algoritma C4.5 sehingga bisa menghasilkan perhitungan yang akurat.

### 2.2 Data Mining

*Data mining* dapat disebut *Knowledge discovery in Database (KDD)* merupakan kegiatan pengumpulan, pemakaian data historis menemukan pola atau hubungan dalam data yang berukuran besar (Naldy & Andri, 2021)(Mardiansa et al., 2023). *Data mining* adalah proses kegiatan ekstraksi data dari *database* yang besar yang belum diketahui hal – hal sebelumnya sehingga bisa digunakan dalam pengambilan keputusan dimasa depan (Apriyadi et al., 2022)

**2.3 Algoritma C4.5**

Algoritma C4.5 adalah salah satu algoritma untuk membuat *decision tree* berdasarkan data *training* yang sudah ada (Ridwan, 2022). Algoritma C4.5 adalah algoritma yang pada umumnya digunakan guna untuk membentuk *Decision Tree* (Pangaribuan et al., 2023). Algoritma C4.5 metode yang digunakan untuk klasifikasi yang menghasilkan model berupa pohon keputusan. Algoritma ini memiliki input yaitu *training samples* dan *samples* (RahmaSari et al., 2024).

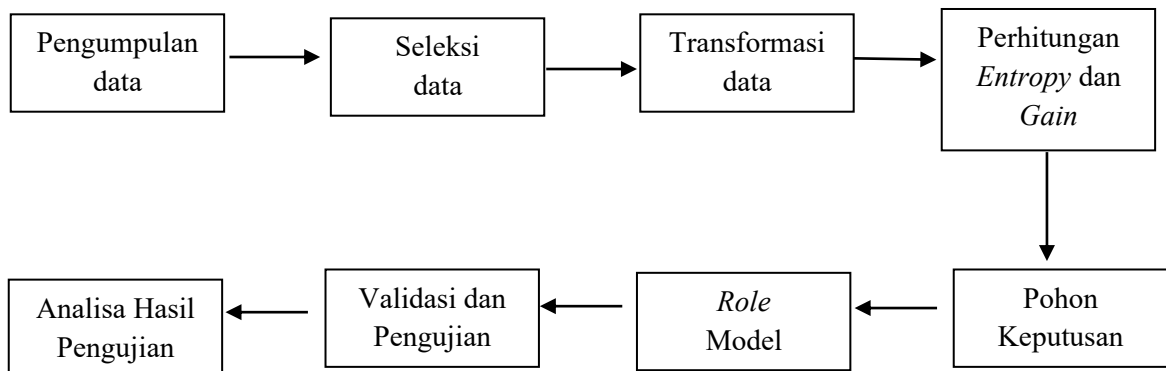
Secara umum algoritma C4.5 untuk membangun pohon keputusan adalah (Purwanto et al., 2020):

- a. Pemilihan atribut sebagai akar
- b. Buat cabang untuk tiap-tiap nilai
- c. Bagi kasus dalam cabang
- d. Ulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama

**III. METODE**

Penelitian ini menggunakan tahapan KDD (*Knowledge Discovery in Database*) dengan mencari informasi yang memiliki nilai, mudah dipahami dan baru dari penyimpanan data yang besar dan kompleks untuk menghasilkan sebuah informasi melalui data yang sudah ada.

**3.1 Tahapan Metodologi Penelitian**



Gambar 1. Tahapan Penelitian

**A. Pengumpulan Data**

Dalam penelitian ini, proses pengambilan data dilakukan dengan mengambil data sekunder yang sudah tersedia dari salah satu penyedia *dataset* yaitu *Kaggle*. *Dataset* dapat diakses melalui link <https://www.kaggle.com/datasets/miadul/asthma-synthetic-medical-dataset>.

**B. Seleksi Data**

Seleksi data dilakukan dengan memilih data yang akan digunakan dalam algoritma C4.5 ini dengan menggunakan 6 atribut dan 300 data yang dipilih.

Tabel 1. Atribut yang digunakan

No	Atribut Terpilih	Keterangan
1	Jenis Kelamin	Atribut
2	Merokok	Atribut
3	Alergi	Atribut
4	Polusi Udara	Atribut

5	Aktivitas Fisik	Atribut
6	Jenis Pekerjaan	Label

Atribut yang dipilih ada 6 yaitu jenis kelamin, Merokok, Alergi, Polusi Udara, Aktivitas Fisik dan Jenis Pekerjaan. Hal ini digunakan untuk diolah menggunakan algoritma C4.5.

Tabel 2.

Dataset Penderita Asma

No	Jenis Kelamin	Merokok	Alergi	Polusi Udara	Kegiatan Fisik	Jenis Pekerjaan
1	Wanita	Pernah	Tidak	Sedang	Jarang	Luar ruangan
2	Pria	Pernah	Debu	Rendah	Rata-rata	Dalam ruangan
3	Wanita	Tidak Pernah	Tidak	Sedang	Rata-rata	Dalam ruangan
4	Pria	Tidak Pernah	Alergi lebih dari satu	Tinggi	Jarang	Luar ruangan
5	Pria	Tidak Pernah	Tidak	Sedang	Aktif	Dalam ruangan
6	Other	Tidak Pernah	Serbuk Sari	Sedang	Rata-rata	Luar ruangan
7	Wanita	Pernah	Serbuk Sari	Sedang	Jarang	Luar ruangan
8	Pria	Tidak Pernah	Binatang	Tinggi	Jarang	Dalam ruangan
298	Pria	Tidak Pernah	Debu	Sedang	Rata-rata	Dalam ruangan
299	Pria	Tidak Pernah	Binatang	Rendah	Jarang	Dalam ruangan
300	Pria	Perokok	Tidak	Sedang	Rata-rata	Luar ruangan

C. Transformasi Data

Transformasi data adalah proses mentransformasi atau mengubah data ke dalam bentuk yang sesuai, agar dapat di proses dengan perhitungan algoritma C4.5 agar bisa diproses, salah satunya dengan menggunakan aplikasi RapidMiner dengan mengubah tipe data yang digunakan pada tabel penelitian.

D. Perhitungan Entropy dan Information Gain

Perhitungan pertama adalah mengukur tingkat keragaman nilai yang ada pada sebuah kolom terhadap kriteria pada label hasil

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

Keterangan:

- S: Jumlah seluruh data
- n: Jumlah kumpulan data
- pi: Nilai atribut dibagi label hasil
- log2: Logaritma biner / basis dua



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

Keterangan:

- S: Jumlah seluruh data
- A: *Atribut* atau kolom
- n: Jumlah partisi S
- |S<sub>i</sub>|: Jumlah kasus pada partisi ke-i
- |S|: Jumlah kasus dalam S

E. Pohon Keputusan

Merupakan hasil dari proses perhitungan yang telah dilakukan dengan menghitung *Entropy* dan *Information Gain* dimana nilai *information gain* tertinggi akan dijadikan sebagai akar.

F. Rule Model

Adalah aturan – aturan yang akan mengurai penjelasan yang akan mempresentasikan pohon keputusan berdasarkan data yang akan diolah dari penentuan jenis pekerjaan bagi penderita asma.

G. Validasi dan Pengujian

Validasi dan pengujian adalah Pengujian yang dilakukan untuk mengetahui semua fungsi bekerja dengan baik atau tidak. Validasi dilakukan dengan *Ten-fold Cross Validation*. *Ten-fold Cross Validation* adalah validasi yang dilakukan dengan cara membagi suatu set data menjadi sepuluh segmen yang berukuran sama besar dengan cara melakukan pengacakan data. Validasi dan pengujian dilakukan untuk mengetahui tingkat akurasi, presisi, dan recall dari hasil prediksi klasifikasi. Akurasi adalah *persentase* dari catatan yang diklasifikasikan dengan benar dalam pengujian *dataset*. *Presisi* adalah *persentase* data yang diklasifikasikan sebagai model baik yang sebenarnya juga baik. *Recall* adalah pengukuran tingkat pengenalan positif sebenarnya.

H. Analisa hasil pengujian

Analisa yang dilakukan untuk memastikan bahwa hasil pengujian benar-benar sesuai dengan pembahasan. Analisa dilakukan dengan melakukan perhitungan kembali hasil validasi dan pengujian (akurasi, *presisi*, dan *recall*) secara manual. Apakah perhitungan yang dilakukan akan menghasilkan nilai yang sama atau tidak, dibantu dengan *Confusion matrix*. *Confusion Matrix* adalah model yang akan membentuk *matrix* yang terdiri dari *true* positif atau tupel positif dan *true* negatif atau *tupel* negatif. *Confusion matrix* berisi informasi aktual (*actual*) dan prediksi (*predicted*) pada sistem klasifikasi. *Confusion matrix* dapat memvisualisasi kinerja algoritma klasifikasi.

IV. HASIL DAN PEMBAHASAN

Berdasarkan dataset penderita asma yang terdapat pada tabel 2 dapat diklasifikasikan menjadi 5 tabel berdasarkan jenis kelamin, merokok, alergi, polusi udara dan kegiatan fisik.

Tabel 3. Jenis Pekerjaan

Jenis Kelamin	Jenis Pekerjaan	
	Dalam Ruang	Luar Ruang
Pria	106	41
Wanita	101	52

Tabel 5. Polusi Udara

Polusi	Jenis Pekerjaan	
	Dalam Ruang	Luar Ruang
Rendah	66	19
Sedang	92	58
Tinggi	49	19

Tabel 4. Merokok

Merokok	Jenis Pekerjaan	
	Dalam Ruang	Luar Ruang

Tabel 6. Kegiatan Fisik



Tidak Pernah	116	53
Pernah	63	27
Perokok	28	13

Kegiatan Fisik	Jenis Pekerjaan	
	Dalam Ruangan	Luar Ruangan
Jarang	80	42
Rata – Rata	88	37
Aktif	39	14

Tabel 7. Alergi

Alergi	Jenis Pekerjaan	
	Dalam Ruangan	Luar Ruangan
Tidak	64	26
Debu	54	32
Serbuk Sari	42	17
Binatang	28	12
Alergi lebih dari 1	19	6

#### 4.1 Perhitungan Algoritma C4.5

##### A. Melakukan perhitungan *Entropy* dan *Gain*

Tabel 8. Hasil Perhitungan *Entropy* dan *Gain*

Node	Keterangan	Total	Dalam Ruangan	Luar Ruangan	<i>Entropy</i>	<i>Gain</i>
Total		300	207	93	0.893173	
Jenis Kelamin	Wanita	153	101	52	0.967023	0.852126
	Pria	147	106	41	0.922724	
Merokok	Pernah	90	63	27	0.785477	1.275485
	Tidak pernah	169	116	53	0.971883	
	Perokok	41	28	13	0.515566	
Alergi	Tidak	90	64	26	0.781274	1.10719
	Debu	86	54	32	0.789715	
	Alergi lebih dari 1	25	19	6	0.365	
	Serbuk Sari	59	42	17	0.631787	
	Binatang	40	28	12	0.505091	
Polusi Udara	Rendah	85	66	19	0.732697	1.317607
	Sedang	150	92	58	0.981307	
	Tinggi	65	49	16	0.652508	
Kegiatan Fisik	Jarang	122	80	42	0.905614	1.000364
	Kadang-kadang	125	88	37	0.891409	
	Aktif	53	39	14	0.588979	

Berdasarkan tabel diatas untuk mencari

### 1. Menghitung *Entropy*

- a. *Entropy* (Total)  

$$= ((-207/300) * \log_2(207/300)) + ((-93/300) * \log_2(93/300)) = 0,893173$$
- b. *Entropy* Jenis Kelamin (Wanita)  

$$= ((-101/300) * \log_2(101/300)) + ((-52/300) * \log_2(52/300)) = 0,967023$$
- c. *Entropy* Jenis Kelamin (Pria)  

$$= ((-106/300) * \log_2(106/300)) + ((-41/300) * \log_2(41/300)) = 0,922724$$
- d. *Entropy* Merokok (Pernah)  

$$= ((-63/300) * \log_2(63/300)) + ((-27/300) * \log_2(27/300)) = 0,785477$$
- e. *Entropy* Merokok (Tidak Pernah)  

$$= ((-116/300) * \log_2(116/300)) + ((-53/300) * \log_2(53/300)) = 0,971883$$
- f. *Entropy* Merokok (Perokok)  

$$= ((-28/300) * \log_2(28/300)) + ((-13/300) * \log_2(13/300)) = 0,515566$$
- g. *Entropy* Alergi (Tidak)  

$$= ((-64/300) * \log_2(64/300)) + ((-26/300) * \log_2(26/300)) = 0,781274$$
- h. *Entropy* Alergi (Debu)  

$$= ((-54/300) * \log_2(54/300)) + ((-32/300) * \log_2(32/300)) = 0,789715$$
- i. *Entropy* Alergi (Alergi lebih dari 1)  

$$= ((-19/300) * \log_2(19/300)) + ((-6/300) * \log_2(6/300)) = 0,365$$
- j. *Entropy* Alergi (Serbuk Sari)  

$$= ((-42/300) * \log_2(42/300)) + ((-17/300) * \log_2(17/300)) = 0,631787$$
- k. *Entropy* Alergi (Binatang)  

$$= ((-28/300) * \log_2(28/300)) + ((-12/300) * \log_2(12/300)) = 0,505091$$
- l. *Entropy* Polusi Udara (Rendah)  

$$= ((-66/300) * \log_2(66/300)) + ((-19/300) * \log_2(19/300)) = 0,732697$$
- m. *Entropy* Polusi Udara (Sedang)  

$$= ((-92/300) * \log_2(92/300)) + ((-58/300) * \log_2(58/300)) = 0,981307$$
- n. *Entropy* Polusi Udara (Tinggi)  

$$= ((-49/300) * \log_2(49/300)) + ((-16/300) * \log_2(16/300)) = 0,652508$$
- o. *Entropy* Kegiatan Fisik (Jarang)  

$$= ((-80/300) * \log_2(80/300)) + ((-42/300) * \log_2(42/300)) = 0,905614$$
- p. *Entropy* Kegiatan Fisik (Kadang – kadang)  

$$= ((-88/300) * \log_2(88/300)) + ((-37/300) * \log_2(37/300)) = 0,891409$$
- q. *Entropy* Kegiatan Fisik (Aktif)  

$$= ((-39/300) * \log_2(39/300)) + ((-14/300) * \log_2(14/300)) = 0,588979$$

### 2. Menghitung *Gain*

- a. *Gain* Jenis Kelamin  

$$= (0,893173) - ((153/300) * 0,967023) + ((147/300) * 0,922724) = 0,852126$$
- b. *Gain* Merokok  

$$= (0,893173) - ((90/300) * 0,785477) - ((169/300) * 0,971883) - ((41/300) * 0,515566) = 1,275485$$
- c. *Gain* Alergi  

$$= (0,893173) - ((90/300) * 0,781274) - ((86/300) * 0,789715) - ((25/300) * 0,365 - ((59/300) * 0,631787) - ((40/300) * 0,505091) = 1,10719$$
- d. *Gain* Polusi Udara  

$$= (0,893173) - ((85/300) * 0,732697) - ((150/300) * 0,981307) - ((65/300) * 0,652508) = 1,317607$$
- e. *Gain* Kegiatan Fisik  

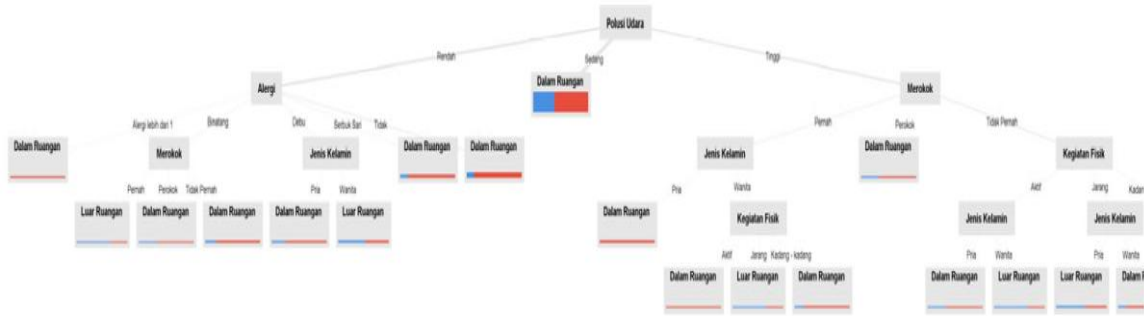
$$= (0,893173) - ((122/300) * 0,905614) - ((125/300) * 0,891409) - ((53/300) * 0,588979) = 1,000364$$

Dari perhitungan yang telah dilakukan diperoleh *Gain* tertinggi adalah polusi udara sebesar

1,317607 maka polusi udara dapat menjadi *node* Akar, selanjutnya jika dilihat untuk polusi udara memiliki 3 nilai yaitu polusi udara rendah, polusi udara sedang dan polusi udara tinggi maka perlu dilakukan proses perhitungan lagi. Metode perhitungan untuk *node* selanjutnya sama dengan perhitungan menentukan *node* yang ada. Perbedaannya adalah *atribut* dengan nilai terbesar pada perhitungan sebelumnya tidak dihitung lagi. *Atribut* total diganti dengan total dari nilai *atribut* terbesar pada perhitungan sebelumnya.

B. Mengolah data dengan RapidMiner

1. Pohon keputusan dengan 5 klasifikasi yang dimiliki



Gambar 1. Pohon Keputusan dengan akar Polusi Udara

Berdasarkan perhitungan yang dilakukan dengan menggunakan aplikasi Rapidminer dan juga perhitungan secara manual dapat dievaluasi model bahwa 2 proses tersebut memiliki hasil yang sama dalam membuat pohon keputusan / *tree* yang digunakan dalam pengklasifikasian jenis pekerjaan yang akan dilakukan oleh penderita asma.

2. Penentuan *Rule* untuk aturan pembuatan pohon keputusan.

I. *Tree*

Polusi Udara = Rendah

| Alergi = Alergi lebih dari 1: Dalam Ruangan {Luar Ruangan=0, Dalam Ruangan=6}

| Alergi = Binatang

| | Merokok = Pernah: Luar Ruangan {Luar Ruangan=2, Dalam Ruangan=1}

| | Merokok = Perokok: Dalam Ruangan {Luar Ruangan=1, Dalam Ruangan=2}

| | Merokok = Tidak Pernah: Dalam Ruangan {Luar Ruangan=2, Dalam Ruangan=8}

| Alergi = Debu

| | Jenis Kelamin = Pria: Dalam Ruangan {Luar Ruangan=2, Dalam Ruangan=6}

| | Jenis Kelamin = Wanita: Luar Ruangan {Luar Ruangan=6, Dalam Ruangan=5}

| Alergi = Serbuk Sari: Dalam Ruangan {Luar Ruangan=2, Dalam Ruangan=12}

| Alergi = Tidak: Dalam Ruangan {Luar Ruangan=4, Dalam Ruangan=26}

Polusi Udara = Sedang: Dalam Ruangan {Luar Ruangan=58, Dalam Ruangan=92}

Polusi Udara = Tinggi

| Merokok = Pernah

| | Jenis Kelamin = Pria: Dalam Ruangan {Luar Ruangan=0, Dalam Ruangan=12}

| | Jenis Kelamin = Wanita

| | | Kegiatan Fisik = Aktif: Dalam Ruangan {Luar Ruangan=0, Dalam Ruangan=3}

| | | Kegiatan Fisik = Jarang: Luar Ruangan {Luar Ruangan=2, Dalam Ruangan=1}

| | | Kegiatan Fisik = Kadang - kadang: Dalam Ruangan {Luar Ruangan=1, Dalam Ruangan=5}

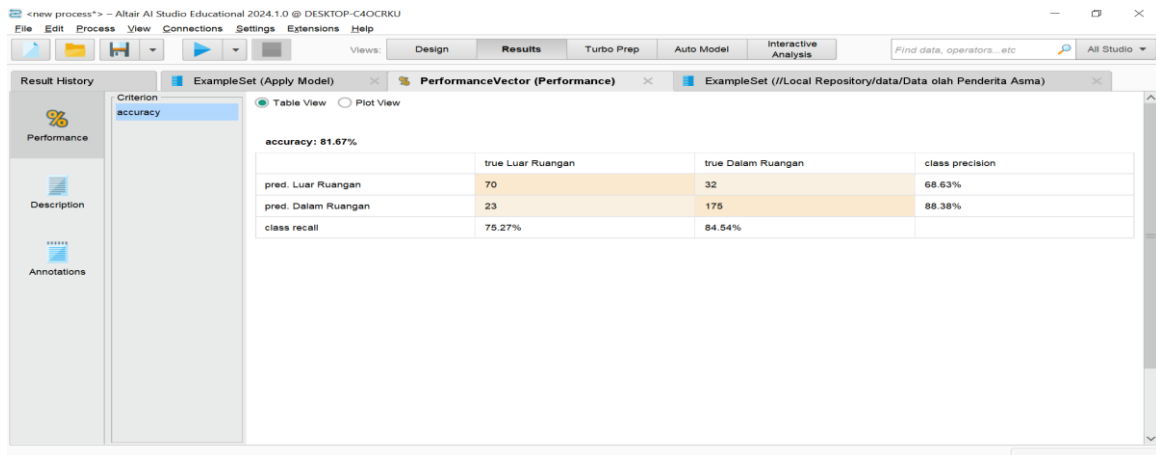
| Merokok = Perokok

| | Kegiatan Fisik = Aktif: Dalam Ruangan {Luar Ruangan=1, Dalam Ruangan=3}

- | | Kegiatan Fisik = Kadang - kadang: Luar Ruangan {Luar Ruangan=1, Dalam Ruangan=1}
- | Merokok = Tidak Pernah
- | | Kegiatan Fisik = Aktif
- | | | Jenis Kelamin = Pria: Dalam Ruangan {Luar Ruangan=1, Dalam Ruangan=2}
- | | | Jenis Kelamin = Wanita: Luar Ruangan {Luar Ruangan=2, Dalam Ruangan=1}
- | | Kegiatan Fisik = Jarang
- | | | Jenis Kelamin = Pria: Luar Ruangan {Luar Ruangan=4, Dalam Ruangan=3}
- | | | Jenis Kelamin = Wanita: Dalam Ruangan {Luar Ruangan=1, Dalam Ruangan=6}
- | | | Kegiatan Fisik = Kadang - kadang: Dalam Ruangan {Luar Ruangan=3, Dalam Ruangan=12}

Rule diatas memberikan gambaran tentang bagaimana sebuah keputusan itu akan diperoleh berdasarkan sebuah syarat tertentu, seperti Polusi Udara = Rendah maka akan dicek kembali apakah penderita memiliki Alergi = Alergi lebih dari 1: Dalam Ruangan {Luar Ruangan=0, Dalam Ruangan=6} maka berdasarkan data diatas jumlah yang diperoleh luar ruangan = 0 penderita dan dalam ruangan ada 6 penderita. Maka hal yang sama akan dilakukan berdasarkan *atribut* yang berbeda dengan persyaratan dan perhitungan yang sama.

3. Nilai Akurasi yang diperoleh dengan menggunakan *performace*  
 Untuk bisa mengetahui apakah data yang digunakan dan perhitungan yang dilakukan memiliki pengklasifikasian yang baik atau tidak. Hasil pengujian validasi menggunakan *confusion matrix* tingkat *accuracy* dalam memprediksi capaian jenis pekerjaan rata-rata sebesar 81,67%, *precision* sebesar 88.38%, *recall* sebesar 84,54%



Gambar 2. Hasil pengujian validasi menggunakan *confusion matrix*

## V. KESIMPULAN

Berdasarkan hasil penelitian yang sudah dilakukan oleh peneliti bahwa: Implementasi algoritma C4.5 dapat digunakan dengan baik dan cocok dalam proses mengklasifikasi jenis pekerjaan bagi penderita asma dan perhitungan lebih mudah dalam menentukan hasil tersebut yaitu menggunakan aturan-aturan yang dihasilkan dari pohon keputusan. Hasil penerapan algoritma C4.5 sudah didapatkan nilai *gain* yang tertinggi adalah atribut Polusi udara yaitu 1,317607 yang menunjukkan bahwa atribut polusi udara adalah atribut yang paling berpengaruh dalam menentukan hasil pohon keputusan dan jenis pekerjaan yang akan dilakukan. Hasil pengujian validasi menggunakan *confusion matrix* tingkat *accuracy* dalam memprediksi capaian pembelajaran daring rata-rata sebesar 81,67%, *precision* sebesar 88.38%, *recall* sebesar 84,54%

## VI. REFERENSI

- Ansyari, M., Riduansyah, M., Ariani, M., & Fetriyah, U. H. (2023). PENGALAMAN KELUARGA DALAM MERAWAT ANAK DENGAN ASMA DI UGD. *Jurnal Ilmiah Permas: Jurnal Ilmiah STIKES Kendal*, 13(3), 1083–1088.
- Apriyadi, Lubis, M. R., & Damanik, B. E. (2022). PENERAPAN ALGORITMA C5.0 DALAM MENENTUKAN TINGKAT PEMAHAMAN MAHASISWA TERHADAP PEMBELAJARAN DARING. *KOMPUTA: Jurnal Ilmiah Komputer Dan Informatika*, 11(1), 11–20.
- Jakarta, D. (2023). *Penyebab, Gejala dan Cara Mencegah Penyakit Asma*. Dinkes Jakarta.
- Kemendes. (2024). *Asma*. Kemendes. <https://ayosehat.kemkes.go.id/topik-penyakit/lansia-penyakit-pernapasan-kronis/asma>
- Mardiansa, Sari, H. L., & Prahasti. (2023). Penerapan Data Mining Untuk Mengetahui Minat Siswa Pada Pelajaran IPA Menggunakan Metode K-Means Clustering. *Jurnal Multidisiplin Dehasen*, 2(4), 693–702.
- Muzakki, F., Ubaydillah, I., Assyiami, N. R., & Soleha, S. (2024). Penerapan Algoritma C4.5 Untuk Prediksi Penyakit Jantung Menggunakan Rapidminer. *Jurnal Komputer Antartika*, 2(2), 71–79. <https://ejournal.mediaantartika.id/index.php/jka/article/view/304/71>
- Naldy, E. T., & Andri. (2021). Penerapan Data Mining Untuk Analisis Daftar Pembelian Konsumen Dengan Menggunakan Algoritma Apriori Pada Transaksi Penjualan Toko Bangunan MDN. *Jurnal Nasional Ilmu Komputer*, 2(2), 89–101.
- Pangaribuan, J. J., Romindo, Ilhami, M., Napitupulu, S., & Chandra, W. (2023). Prediksi Kanker Payudara Melalui Penerapan Algoritma C4.5. *JURNAL MANAJEMEN SISTEM INFORMASI DAN TEKNOLOGI*, 13(2), 165–173.
- Purwanto, A., Primajaya, A., & Voutama, A. (2020). Penerapan Algoritma C4.5 dalam Prediksi Potensi Tingkat Kasus Pneumonia di Kabupaten Karawang. *Jurnal Sistem Informasi Dan Teknologi*, 8(4), 390–396.
- Putri, A. D., Sholekhah, F., Dadynata, E., Lusiana Efrizoni, R., & Sapina, N. (2024). The Application of C4.5 Decision Tree Algorithm for Predicting the Survival Rate of Thyroid Cancer Patients. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 4(4), 1485–1495.
- RahmaSari, Z. D., JasmirJasmir, & Arvita, Y. (2024). Penerapan Data Mining Untuk Prediksi Penyakit Diabetes Menggunakan Algoritma C4.5. *Jurnal Informatika Dan Rekayasa Komputer (JAKAKOM)*, 4(1). <https://doi.org/https://doi.org/10.33998/jakakom.v4i1>
- Ridwan, A. (2022). PENERAPAN ALGORITMA C4.5 UNTUK KLASIFIKASI PENYAKIT DIABETES MELLITUS. *Jurnal Bisnis Digital Dan Sistem Informasi*, 41–48. <https://jurnal2.umku.ac.id/index.php/BIDISFO/article/viewFile/1841/1096>