

## Klasifikasi Gangguan Autisme Pada Anak Menggunakan Algoritma C4.5 Dengan Teknik *Random Forest*

Andre Eko Cahyo<sup>1\*</sup>, Agung Nilogiri<sup>1\*</sup>

<sup>1</sup>Universitas Muhammadiyah Jember

email: [andreekocahyo06@gmail.com](mailto:andreekocahyo06@gmail.com)

DOI: <https://doi.org/10.32528/nms.v1i6.241>

\*Correspondensi: Andre Eko Cahyo

Email: [andreekocahyo06@gmail.com](mailto:andreekocahyo06@gmail.com)

Published: November, 2022



**Copyright:** © 2022 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

**Abstrak:** Autisme merupakan jenis gangguan yang terjadi pada otak, gangguan tersebut menyebabkan beberapa area di bagian otak tidak mampu bekerjasama. Gangguan tersebut bisa diklasifikasikan secara akurat dengan ensemble method. Ensemble method adalah metode yang digunakan untuk meningkatkan akurasi algoritma dan pengklasifikasian dengan membangun beberapa classifier dari data training. Berdasarkan hasil penelitian, pada perhitungan yang telah dilakukan dengan 15 kali percobaan sebelum menggunakan teknik Ensemble didapatkan akurasi terbaik pada k-fold 6 percobaan ke-4 dengan akurasi sebesar 83,33%, dimana nilai positif gangguan interaksi sosial memiliki hasil 80%, nilai positif pada gangguan komunikasi sebesar 100%, dan pada nilai positif gangguan memiliki perilaku presisi 100%. Kemudian dilakukan perhitungan dengan 15 kali percobaan sesudah menggunakan teknik Ensemble Random Forest didapatkan akurasi terbaik pada k-fold 4 percobaan ke-3 dengan akurasi sebesar 88,89%. Nilai positif gangguan interaksi sosial memiliki hasil 80,00%, nilai positif pada gangguan komunikasi sebesar 100%, dan pada nilai positif gangguan memiliki perilaku presisi 100%. Dengan demikian pada penelitian yang telah dilakukan mendapatkan peningkatan akurasi sebesar 5,56%.

**Keywords:** *Autisme, Otak, Algoritma C4.5, Ensemble Method, Random Forest*

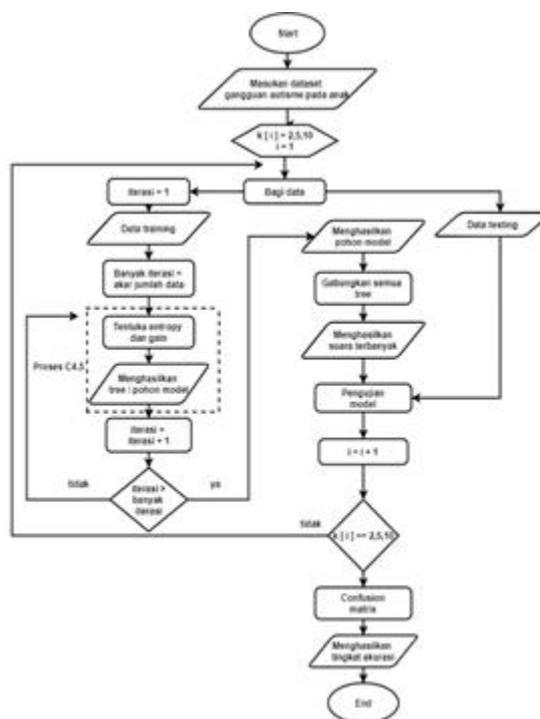
### PENDAHULUAN

Autisme adalah jenis gangguan yang terjadi pada otak seseorang (penderita), yang menyebabkan beberapa area di otak tidak mampu bekerjasama secara normal, sehingga penderita autisme sulit berkomunikasi dan berhubungan sosial dengan orang lain. Penyebab autisme terdiri dari beberapa faktor, namun pada umumnya terjadi karena faktor genetik dan lingkungan, gambaran umum seorang anak yang didiagnosis mengalami gangguan autisme menunjukkan kurang adanya respon terhadap orang lain, seperti mengalami kendala dalam kemampuan berbahasa dan berkomunikasi, serta memunculkan respon yang aneh terhadap lingkungan sekitarnya. Anak yang mengalami gangguan autisme juga kurang responstif terhadap emosi orang lain, kurang mampu mengendalikan perilaku dalam konteks sosial, kurang mampu menggunakan isyarat sosial seperti tertawa, senyum, dan melakukan kontak mata, hal ini berbeda dengan apa yang dilakukan pada seseorang yang normal pada umumnya, untuk menangani permasalahan mengenai masalah autisme perlu diadakannya tindakan oleh pihak dokter agar permasalahan autisme dapat segera ditangani dengan cara atau metode penanganan yang akan dilakukan oleh dokter. Semakin berkembangnya teknologi dimasa kini dapat membantu kegiatan medis dalam menangani suatu penyakit agar dapat segera teratasi seperti antara lain, *machine learning* menjadi salah satu banyak diminati di bidang komputer. *Machine learning* dapat juga digunakan dibidang kesehatan

untuk memprediksi dan juga mengklasifikasikan suatu penyakit dari data yang dialami pasien. Algoritma klasifikasi data mining bermanfaat untuk membantu diagnosa suatu penyakit, salah satu contohnya adalah gangguan autisme pada anak. *Ensemble method* adalah menggabungkan beberapa klasifikasi tree untuk menghasilkan kinerja prediksi yang lebih baik daripada klasifikasi tree tunggal. Sehingga meningkatkan akurasi model. Salah satunya adalah random forest yang merupakan peningkatan dari algoritma C4.5, oleh karena itu pemilihan metode yang digunakan membuktikan kelebihan pada random forest mendapatkan akurasi yang lebih baik dibandingkan dengan metode klasifikasi biasa, dengan demikian pada penelitian ini akan dilakukan klasifikasi algoritma C4.5 dengan metode *Random Forest*.

### METODE PENELITIAN

Metode penelitian yang digunakan terdiri dari: *Teknik Random Forest* (menyiapkan data training, membagi data, dan perhitungan *random forest* untuk data 1) dan Skenario Uji (*Tool Rapid Miner studio* versi 9.5). Pada gambar 1 berikut menunjukkan bagan alir penelitian yang akan dilakukan dengan *Teknik Random Forest*.



**Gambar 1** Diagram Alir Penelitian Teknik *Random Forest*  
[Sumber: Hasil Perhitungan]

#### A. TEKNIK *RANDOM FOREST*

##### a. Menyiapkan Data Training

Data training yang digunakan dalam penelitian ini menggunakan data gejala gangguan autisme pada anak. Metode dalam pemilihan data dilakukan secara acak (random).

##### b. Membagi Data

Pada tahap ini data training akan dibagi menjadi beberapa data dengan menggunakan rumus  $\sqrt{\text{jumlah data}}$

##### c. Perhitungan *Random Forest* Untuk Data 1

Pertama tentukan akar berdasarkan pada nilai Gain tertinggi dari atribut-atribut yang ada. Sedangkan untuk mendapatkan nilai Gain harus menentukan nilai Entropi terlebih dahulu.

## B. SKENARIO UJI

Pengujian yang akan dilakukan menggunakan *Tool Rapid Miner* studio versi 9.5 sebagai pengaplikasiannya. Sebelum digunakan pada *Tool Rapid Miner*, data di bagi menjadi data *training* dan data *testing*. Dimana dengan jumlah data yang berbeda sesuai dengan *k-fold cross validation*. Data akan dibagi menjadi k sama dengan 2, 3, 4, dan 6 seperti yang diperlihatkan sebagai berikut:

Tabel 2. Skenario K-Fold

skenario	K-fold percobaan ke-	Range Data Training	%	Range Data Testing	%
1	2fold percobaan ke1	1 - 18	50%	19 - 36	50%
2	2fold percobaan ke2	19 - 36	50%	1 - 18	50%
3	3fold percobaan ke1	1 - 24	66%	25 - 36	34%
4	3fold percobaan ke2	1 - 12 & 25 - 36	66%	13 - 24	34%
5	3fold percobaan ke3	13 - 36	66%	1 - 12	34%
6	4fold percobaan ke1	1 - 27	75%	28 - 36	25%
7	4fold percobaan ke2	1 - 18 & 28 - 36	75%	19 - 27	25%
8	4fold percobaan ke3	1 - 9 & 19 - 36	75%	10 - 18	25%
9	4fold percobaan ke4	10 - 36	75%	1 - 9	25%
10	6fold percobaan ke1	1 - 30	83%	31 - 36	17%
11	6fold percobaan ke2	1 - 24 & 31 - 36	83%	25 - 30	17%
12	6fold percobaan ke3	1 - 18 & 25 - 36	83%	19 - 24	17%
13	6fold percobaan ke4	1 - 12 & 19 - 36	83%	13 - 18	17%
14	6fold percobaan ke5	1 - 6 & 13 - 36	83%	7 - 12	17%
15	6fold percobaan ke6	7 - 36	83%	1 - 6	17%

[Sumber: Hasil Perhitungan]

## HASIL DAN PEMBAHASAN

### 1. Hasil Penelitian

Pada tahap penelitian uji 2-fold diperoleh hasil akurasi dan presisi percobaan ke-2 yang ditunjukkan pada gambar 2 sebagai berikut:

#### a. Hasil Uji 2-Fold

**Gambar 2.** Hasil Akurasi Dan Presisi Percobaan Ke-2

accuracy: 73.68%

	true Gangguan Komuni...	true Gangguan Interaks...	true Gangguan Perilaku	class precision
pred. Gangguan Komu...	4	0	0	100.00%
pred. Gangguan Interak...	0	4	0	100.00%
pred. Gangguan Perilaku	3	2	6	54.55%
class recall	57.14%	66.67%	100.00%	

[Sumber: Hasil Perhitungan]

Akurasi yang diperoleh sebesar 73,68% dan presisi yang didapatkan sebanyak 3, dimana nilai positif gangguan `interaksi sosial memiliki hasil 100%, nilai positif pada gangguan komunikasi sebesar 100%, dan pada nilai positif gangguan memiliki perilaku presisi 54.55%.

b. Hasil Uji 3-Fold

Pada gambar 3 menunjukkan hasil akurasi dan presisi percobaan ke-3.

accuracy: 75.00%

	true Gangguan Perilaku	true Gangguan Interaksi...	true Gangguan Komuni...	class precision
pred. Gangguan Perilaku	2	0	0	100.00%
pred. Gangguan Interak...	1	3	0	75.00%
pred. Gangguan Komun...	2	0	4	66.67%
class recall	40.00%	100.00%	100.00%	

**Gambar 3** Hasil Akurasi Dan Presisi Percobaan Ke-3

[Sumber: Hasil Perhitungan]

Akurasi yang didapatkan sebesar 75,00% dan presisi yang didapatkan sebanyak 3, dimana nilai positif gangguan `interaksi sosial memiliki hasil 75,00%. Jika nilai positif pada gangguan komunikasi sebesar 66,67%, dan pada nilai positif gangguan memiliki perilaku presisi 100%.

c. Hasil Uji 4-Fold

Pada gambar 4 menunjukkan hasil akurasi dan presisi percobaan ke-4.

accuracy: 88.89%

	true Gangguan Perilaku	true Gangguan Komuni...	true Gangguan Interaks...	class precision
pred. Gangguan Perilaku	1	0	0	100.00%
pred. Gangguan Komu...	0	3	0	100.00%
pred. Gangguan Interak...	1	0	4	80.00%
class recall	50.00%	100.00%	100.00%	

**Gambar 4** Hasil Akurasi Dan Presisi Percobaan Ke-4

[Sumber: Hasil Perhitungan]

Akurasi yang didapatkan sebesar 88,89% dan presisi yang didapatkan sebanyak 3, dimana nilai positif gangguan interaksi sosial memiliki hasil 80%, nilai positif pada gangguan komunikasi sebesar 100%, dan pada nilai positif gangguan memiliki perilaku presisi 100%.

d. Hasil Uji 6-Fold

Pada gambar 5 menunjukkan hasil akurasi dan presisi percobaan ke-6.

accuracy: 83.33%

	true Gangguan Interak...	true Gangguan Komuni...	true Gangguan Perilaku	class precision
pred. Gangguan Interak...	3	0	1	75.00%
pred. Gangguan Komu...	0	2	0	100.00%
pred. Gangguan Perilaku	0	0	0	0.00%
class recall	100.00%	100.00%	0.00%	

Gambar 5 Hasil Akurasi Dan Presisi Percobaan Ke-6  
[Sumber: Hasil Perhitungan]

Akurasi yang diperoleh sebesar 83,33% dan presisi yang didapatkan sebanyak 3, dimana nilai positif gangguan `interaksi sosial memiliki hasil 75.00%, nilai positif pada gangguan komunikasi sebesar 100%, dan pada nilai positif gangguan memiliki perilaku presisi 0.00%%

2. Pembahasan

Pengujian data yang sudah dilakukan dengan menghitung jumlah keseluruhan data sebanyak 15 kali menggunakan teknik *cross validation* yang digunakan untuk mengevaluasi kinerja model dimana data dibagi menjadi data latih (training) dan data uji (testing). Pada tabel 3 menunjukkan data hasil akurasi dan presisi pada k-2, k-3, k-4, dan k-6. Hasil jumlah k rata-rata diperoleh nilai yang berbeda-beda pada tiap pengujian nilai k. Namun juga terdapat mayoritas hasil yang sama, sehingga diambil nilai akurasi dan presisi yang tertinggi. Setiap k-fold terdapat nilai yang tertinggi, maka hasil dari uji skenario sebanyak 4 kali. Jumlah keseluruhan data dihitung 15 kali menghasilkan nilai akurasi yang tertinggi pada k-fold 4 percobaan ke-3 yaitu nilai akurasi 88,89% dan presisi yang didapatkan sebanyak 3, dimana nilai positif gangguan `interaksi sosial memiliki hasil 80,00%, nilai positif pada gangguan komunikasi sebesar 100%, dan pada nilai positif gangguan memiliki perilaku presisi 100%.

Pengujian Ke-	K=2						K=3						K=4						K=6					
	A	P			A	P			A	P			A	P			A	P						
		GIS	GK	GP																				
1	73,6 8%	100 %	100 5%	54,5 0%	75,0 7%	66,6 %	100 7%	66,6 7%	66,6 %	100 %	100 0%	40,0 7%	66,6 %	100 %	100 %	33,3 3%								
2	72,2 2%	71,43 %	75,0 0%	66,6 7%	61,5 4%	80,0 0%	100 3%	33,3 8%	77,7 %	100 0%	66,6 7%	66,6 7%	50,0 %	100 %	100 %	100 %								
3					75,0 0%	75,0 0%	66,6 7%	100 %	88,8 9%	80,0 0%	100 %	100 %	50,0 0%	100 %	100 %	40,0 0%								
4									66,6 7%	66,6 7%	60,0 0%	100 %	83,3 3%	75,0 0%	100 %	100 0%								
5													66,6 7%	50,0 0%	100 %	50,0 0%								
6													50,0 0%	50,0 0%	50,0 0%	0%								

Articulate Writ

Tabel 3. Hasil Akurasi Dan Presisi k-2, k-3, k-4, dan k-6  
[Sumber: Hasil Perhitungan]

---

## KESIMPULAN

Berdasarkan hasil penelitian dan pembahasan pada bab sebelumnya, maka diperoleh kesimpulan sebagai berikut :

1. Berdasarkan hasil penelitian 36 record data dengan 15 kali percobaan, dimana nilai akurasi tertinggi diperoleh pada k-4 pada percobaan 3. Nilai akurasi yang diperoleh sebesar 88,89% dan presisi yang diperoleh sebanyak 3, dimana nilai positif gangguan interaksi sosial memiliki hasil 80%. Jika nilai positif pada gangguan komunikasi sebesar 100%, dan pada nilai positif gangguan memiliki perilaku presisi 100%.
2. Pada perhitungan yang telah dilakukan sebelum menggunakan teknik *Ensemble Method* diperoleh akurasi terbaik pada k-6 fold percobaan ke-4 dengan akurasi sebesar 83,33%, dimana nilai positif gangguan interaksi sosial memiliki hasil 80%, nilai positif pada gangguan komunikasi sebesar 100%, dan pada nilai positif gangguan memiliki perilaku presisi 100%.
3. *Ensemble method* lebih baik dalam meningkatkan akurasi dan presisi, khususnya pada teknik Random Forest dibandingkan dengan menggunakan metode algoritma C4.5 yang telah dibuktikan pada hasil penelitian yang diperoleh.

## REFERENSI

- Biau, G. 2012. Analysis of A Random Forest Model. Journal of Machine Learning Research 13. Paris : Universite Pierre et Marie Curie.
- Kusrini, luthfi taufiq Emha, (2009), Algoritma Data Mining, Penerbit Andi, Yogyakarta.
- Peters, T. (2004). Autisme: Hubungan Pengetahuan Teoritis dan Intervensi Pendidikan Bagi Penyandang. Jakarta: Dian Rakyat.
- Refaeilzadeh, P., Tang, L., & Liu, H. (2009). Cross-Validation. In L. LIU & M. T. ÖZSU (Eds.), Encyclopedia of Database Systems (pp. 532–538).[https://doi.org/10.1007/978-0-387-39940-9\\_565](https://doi.org/10.1007/978-0-387-39940-9_565).
- Sugara, B., Widyatmoko, D., Prakoso, B. S., & Saputro, D. M. (2018). Penerapan algoritma c4.5 untuk deteksi dini gangguan autisme pada anak. 2018(Sentika).
- Yuwono, J. 2009. Memahami Anak Autistik. Bandung: CV Alfabeta B. Sugara, D. Widyatmoko, B. S. Prakoso, and D. M. Saputro, “Penerapan Algoritma C4.5 untuk Deteksi Dini Autisme Pada Anak,” Semin. Nas. Teknol. Inf. dan Komun., vol. 2018, no. Sentika, pp. 87–96, 2018.