IMPLEMENTASI ALGORITMA C4.5 DALAM DIAGNOSIS AUTISME PADA ANAK MENGGUNAKAN RUMUSAN *DIAGNOSTIC AND STATISTICAL MANUAL OF MENTAL DISORDERS V*

Ifnu Wisma Dwi Prastya[1)], Intan Yuniar Purbasari[2)], Basuki Rahmat[3)]
E-mail : [1)]ifnudwi867@gmail.com , [2)]intanyuniar.if@upnjatim.ac.id ,
[3)]basukirahmat.if@upnjatim.ac.id

[1,2,3)]Program Studi Informatika, Fakultas Ilmu Komputer, UPN "Veteran" Jawa Timur

Abstrak

Diagnosis autisme merupakan langkah pertama dalam proses penanganan autisme. Namun, permasalahannya banyak orang tua yang masih belum mengerti terkait gejala yang dialami oleh anaknya dan bagaimana cara penagannya. Masih banyak orang tua yang memilih untuk langsung berkonsultasi kedokter atupun tenaga medis. Sedangkan jumlah dokter atau tenaga medis dalam bidang gangguan perkembanagan mental dan otak masih sangant sedikit. Maka dari itu, dibutuhkan cara pendiagnosisan autisme secara mudah dan gampang diakses oleh orang tua, sehingga orang tua dapat dengan mudah mendiagnosis secara dini autisme pada anak. Algoritma C4.5 merupakan salah satu algoritma yang dapat memprediksi tingkat akurasi diagnosis autisme dan *Diagnostic and Statistical Manual of Mental Disorders* merupakan sebuah acuan yang digunakan untuk mendiagnosa suatu gangguan kejiwaan. Penelitian ini menggunakan 70 data, dengan pembagian data dengan komposisi 70% untuk data latih dan 30 % data uji, sehingga ditemukan 50 data untuk digunakan sebagai data latih dan 20 data untuk data uji. Pengujian dalam sistem ini menggunakan metode *Confusion Matrix*. Pohon keputusan yang terbangun dari sistem ini memiliki nilai akurasi sebesar 90%, dan menghasilkan nilai *precision* sebesar 93,33% dan nilai *recall* sebesar 93,33%.

Kata Kunci : Diagnosis, Autisme, Algoritma C4.5, DSM-V (*Diagnostic and Statistical Manual of Mental Disorders V*).

1. PENDAHULUAN

Diagnosis autisme pada anak merupakan langkah awal terbaik dalam proses penanganan secara dini penyakit autisme. Namun, pada kenyataannya kebanyakan para orang tua tidak mengetahui tentang gejala yang ditunjukan oleh anaknya dan bagaimana penangannya. Para orang tua lebih memilih untuk melakukan konsultasi ke dokter atau tenaga medis, namun para tenaga medis yang fokus pada bidang gangguan perkembangan otak khususnya autisme, masih belum mencukupi jumlahnya.

Seiring dengan perkembangan teknologi yang semakin pesat, membuat semua sektor kehidupan membutuhkan sentuhan teknologi. Baik dari bidang pendidikan, bisnis ataupun bahakna sektor kesehatan. Dalam sektor kesahatan teknologi banyak digunakan untuk membantu tenaga medis untuk memperlancar kerjanya.

Dalam penyelesaian masalah di atas, salah satu metode ataupun teknologi yang dapat digunakan untuk membuat sistem dan menyelesaikan masalah tersebut adalah *Data Mining*. *Data Mining* adalah proses untuk mengekstraksi atau menggali pengetahuan dari data yang berjumlah besar[1]. Dalam *Data Mining* terdapat beberapa cara ataupun proses penggalian data, dalam penelitian ini menggunakan proses klasifikasi. Salah satu algoritma yang bisa digunakan adalah Algoritma C4.5

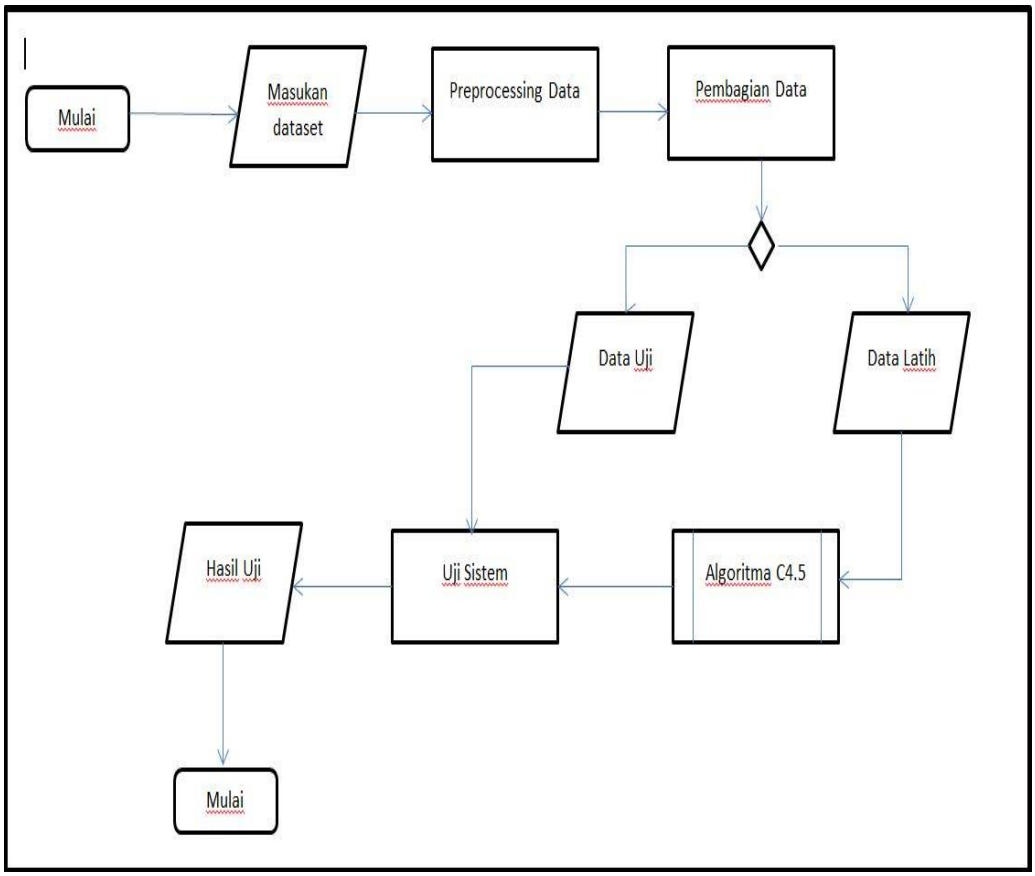
Namun di era yang semakin pesat kemajuannya di ranah digital, kolaborasi rumpun ilmu sangat dibutuhkan. Dalam menyelesaikan permasalahan diatas kolaborasi antara ilmu psikologi dan ilmu komputer dibutuhkan. DSM-V (*Diagnostic and Statistical Manual of Mental Disorders V*) merupakan sebuah pembaharuan dari *Diagnostic and Statistical Manual of Mental Disorders* pada tahun 2013. *Diagnostic and Statistical Manual of Mental Disorders* merupakan sebuah acuan yang digunakan untuk mendiagnosa suatu gangguan kejiwaan[2]. Kolaborasi antara DSM-V dan algoritma C4.5 diharapkan dapat menyelesaikan permasalahan diatas.

Pendeteksian dini sebuah penyakit dapat dilakukan oleh sistem yang menggunakan algoritma C4.5. Dengan algoritma C4.5 sebuah sistem dapat dibangun untuk menciptakan sebuah aturan keputusan yang dapat digunakan untuk mendiagnosis penyakit autisme yang terjadi pada anak. Algoritma C4.5 dapat menyelesaikan masalah secara sistematis dengan membentuk suatu *decision tree* dengan melalui beberapa tahapan yaitu memilih atribut sebagai akar, membuat cabang untuk masing-masing *record* dari atribut, membagi kasus ke dalam cabang dan mengulangi proses untuk masing-masing cabang sampai semua kasus pada cabang menghasilkan sebuah keputusan yang sesuai[3].

Hasil penelitian menggunakan algoritma C4.5 ini menghasilkan nilai akurasi sebesar 90%.

2. METODOLOGI

Adapun proses untuk perancangan aplikasi dalam sub-bab ini akan dibagi menjadi beberapa tahap, yaitu analisis data, data preprocessing, pembagian data, pembuatan sistem dan uji coba sistem.



Gambar 1. Diagram alur penelitian

2.1 Analisis Data

Dataset yang digunakan pada penelitian merupakan data anak penderita autisme yang didapat dari Rumah Anak Berkebutuhan Khusus Dinas Kesehatan Kota Surabaya dan data anak bukan penderita autisme(normal) dari penyebaran kuisioner masyarakat sekitar. Data tersebut berisi detail dari nama anak, usia anak, alamat anak, gejala-gejala yang dirasakan oleh penyandang autisme.

2.2 Data Preprocessing

Tahapan preprocessing data dilakukan untuk mendapatkan data yang siap digunakan. Pada proses klasifikasi terdapat beberapa tahap dalam melakukan preprocessing data, untuk pengambilan data yang relevan digunakan sebagai penelitian dan mengubah data numerik menjadi kategorik. Dari dataset terdapat sebelas atribut yaitu tidak mampu menjalin interaksi non verbal (kontak mata, ekspresi wajah, gesture), kesulitan bermain dengan teman sebaya, tidak ada empati dan minat, tidak ada timbal balik sosial/emosional, terlambatnya atau tidak ada perkembangan bicara, gangguan dalam memulai/pertahankan percakapan, menggunakan bahasa idiosinkartik secara stereotypic dan berulang, tidak ada permainan khayalan/imaginatif, mempertahankan satu minat/lebih dengan berlebihan, terpaku pada suatu rutinitas dan Gerakan gestur khas (manarisme motoric) stereotypic dan berulang (ex: jentikan jari) yang akan diubah menjadi data kategorik.

pengkodean atribut dilakukan seperti berikut:
A1 = Tidak mampu menjalin interaksi non verbal (kontak mata, ekspresi wajah, gesture).
A2 = Kesulitan bermain dengan teman sebaya.
A3 = Tidak ada empati dan minat
A4 = Tidak ada timbal balik sosial/emosional.
B1 = Terlambatnya atau tidak ada perkembangan bicara.
B2 = Gangguan dalam memulai/pertahankan percakapan.
B3 = Menggunakan bahasa idiosinkartik secara stereotypic dan berulang.
B4 = Tidak ada permainan khayalan/imaginatif.
C1 = Mempertahankan satu minat/lebih dengan berlebihan.
C2 = Terpaku pada suatu rutinitas.
C3 = Gerakan gestur khas (manarisme motoric) stereotypic dan berulang (ex: jentikan jari).

Tabel 1. Contoh header atribut

	A1	A2	A3	A4	B1	B2	B3	B4	C1	C2	C3

2.3 Pembagian Data

Dalam proses pengolahan data, sebelumnya dilakukan proses pembagian data. Proses pembagian data yang dimaksud adalah proses dimana membagi data uji dan data latih. Dalam penulisan ini data uji diambil sembanyak 20 data dan data latih sebanyak 50 data.

2.4 Algoritma C4.5

Algoritma C4.5 adalah program yang memberi kontribusi satu set data berlabel dan menghasilkan pohon keputusan sebagai keluaran. Adapun langkah-langkah dalam pembuatan algoritma C4.5 sebagai berikut

1. Pilih atribut sebagai akar.
2. Buat cabang untuk masing-masing record dari atribut.
3. Membagi kasus ke dalam cabang,
4. Ulangi proses untuk masing-masing cabang sampai semua kasus pada cabang menghasilkan suatu keputusanyang sesuai.

Dalam proses memilih suatu atribut menjadi akar maka perlu dilakukan proses perhitungan *entropy* dan *gain. Gain* dengan nilai tertinggi akan digunakan menjadi root atau akar awal. Untuk menghitung *gain* dapat dilakukan dengan menggunakan rumus dibawah ini

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

Dimana,
S = Himpunan Kasus
A = Atribut
n = Jumlah Sampel
|Si| = Jumlah Kasus pada Partisi ke-i
|S| = Jumlah Kasus dalam S
Sedangkan untuk perhitungan nilai *entropy* dapat dilakukan dengan rumus:

$$Entropy(S_1, S_2, \ldots, S_n) = \sum_{i=1}^{n} Pi * log_2(Pi) \quad \text{.....(2)}$$

Dengan,
S = Himpunan Kasus
n = Jumlah sampel
P*i* = proporsi kelas

Setelah perhitungan *entropy* sehingga ditemukan *gain*, hingga ditemukan *gain* dengan nilai terbesar selesai, maka dihasilkan suatu pohon keputusan yang dapat diubah menjadi suatu aturan yang dapat digunakan dengan menggunakan kaidah IF-THEN-ELSE.

3. HASIL DAN PEMBAHASAN

3.1 Basis Pengetahuan

Pada basis pengetahuan, akan dijelaskan tabel gejala autisme sesuai dengan rumusan DSM-V yang akan dijelaskan secara rinci sebagai berikut:

Tabel 2. Gejala Autisme

Kode Gejala	Nama Gejala
A1	Tidak mampu menjalin interaksi non verbal (kontak mata, ekspresi wajah, gestur).
A2	Kesulitan bermain dengan teman sebaya.
A3	Tidak ada empati dan minat.
A4	Tidak ada timbal balik sosial/emosional
B1	Terlambatnya atau tidak ada perkembangan bicara (tidak berusaha untuk melakukan komunikasi verbal).
B2	Gangguan dalam memulai/pertahankan percakapan.
B3	Menggunakan bahasa idiosinkartik secara stereotypic dan berulang.
B4	Tidak ada permainan khayalan/imaginatif.
C1	Mempertahankan 1 minat/lebih dengan berlebihan
C2	Terpaku pada suatu rutinitas.
C3	Gerakan gestur khas (manarisme motorik) stereotypic dan berulang (ex: jentikan jari)

3.2 Algoritma C4.5

Pada tahapan ini dilakukan analisa perhitungan terhadap kasus permasalahan, dalam penelitian ini menggunakan Algoritma C4.5. Akan dijelaskan secara rinci detail Algoritma C4.5 menjadi analisa perhitungan, rule-rule yang dihasilkan dan gambaran pohon keputusan (decision tree).

3.2.1 Analisis Perhitungan

Dalam penelitian ini akan dilakukan percobaan dengan mengklasifikasikan sebuah data dengan menggunakan metode klasifikasi pada salah satu jenis data mining. Algoritma C4.5 digunakan untuk melakukan proses klasifikasi dengan mengolah gejala-gejala penderita autisme. Setelah data diolah menggunakan Algoritma C4.5 dilakukan proses pengujian menggunakankan metode *confusion matrix.*

Tabel 3. Data latih

No	A1	A2	A3	A4	B1	B2	B3	B4	C1	C2	C3	Hasil
1	YA	YA	YA	TIDAK	YA	YA	YA	YA	YA	YA	YA	YA
2	YA	YA	TIDAK	TIDAK	YA	YA	YA	TIDAK	YA	YA	YA	YA
3	YA	TIDAK	YA	YA	YA	YA	YA	YA	YA	YA	YA	YA
4	TIDAK	YA	TIDAK	TIDAK	YA	YA	YA	TIDAK	TIDAK	TIDAK	YA	YA
5	YA	YA	YA	TIDAK	YA	YA	YA	TIDAK	YA	TIDAK	YA	YA
6	TIDAK	TIDAK	YA	YA	YA	YA	YA	TIDAK	YA	YA	TIDAK	YA
7	TIDAK	YA	YA	YA	YA	YA	TIDAK	YA	YA	YA	YA	YA
8	YA	YA	TIDAK	TIDAK	YA	YA	YA	YA	YA	TIDAK	TIDAK	YA
9	TIDAK	YA	YA	TIDAK	YA	YA	YA	TIDAK	TIDAK	TIDAK	TIDAK	YA
10	YA	YA	YA	YA	YA	YA	YA	TIDAK	TIDAK	YA	YA	YA
26	TIDAK	YA	YA	TIDAK	YA	YA	YA	YA	YA	YA	TIDAK	YA
27	YA	TIDAK	YA	TIDAK	YA	YA	YA	YA	TIDAK	TIDAK	YA	YA
28	TIDAK	YA	YA	TIDAK	YA	YA	YA	YA	YA	YA	YA	YA
29	YA	TIDAK	TIDAK	TIDAK	YA	YA	TIDAK	TIDAK	YA	YA	TIDAK	YA
30	YA	YA	YA	TIDAK	YA	YA	YA	TIDAK	YA	YA	YA	YA
31	YA	YA	YA	YA	YA	YA	YA	TIDAK	YA	YA	YA	YA
32	YA	YA	YA	YA	YA	YA	YA	TIDAK	YA	YA	YA	YA
33	YA	YA	YA	YA	TIDAK	YA	YA	TIDAK	TIDAK	TIDAK	TIDAK	YA
34	YA	YA	YA	YA	YA	YA	YA	TIDAK	TIDAK	YA	YA	YA
35	YA	YA	YA	YA	YA	YA	YA	TIDAK	YA	YA	YA	YA
36	TIDAK	YA	YA	YA	YA	YA	YA	TIDAK	YA	TIDAK	YA	YA
37	YA	YA	YA	TIDAK	YA	YA	YA	TIDAK	TIDAK	TIDAK	YA	YA
38	TIDAK	YA	YA	YA	YA	YA	YA	YA	YA	YA	YA	YA
39	TIDAK	YA	YA	TIDAK	YA	YA	YA	TIDAK	TIDAK	TIDAK	YA	YA
40	YA	YA	YA	TIDAK	YA	YA	YA	TIDAK	YA	YA	YA	YA
41	YA	YA	YA	YA	YA	YA	YA	YA	TIDAK	TIDAK	YA	YA
42	YA	YA	YA	YA	YA	YA	YA	TIDAK	YA	YA	YA	YA
43	TIDAK	YA	YA	TIDAK	TIDAK	YA	YA	YA	TIDAK	YA	TIDAK	YA
44	TIDAK	YA	YA	TIDAK	TIDAK	YA	TIDAK	TIDAK	YA	YA	TIDAK	YA
45	TIDAK	YA	YA	YA	TIDAK	TIDAK	TIDAK	YA	YA	TIDAK	YA	YA
46	TIDAK	YA	YA	YA	YA	TIDAK	TIDAK	YA	YA	YA	TIDAK	YA
47	TIDAK	YA	YA	TIDAK	TIDAK	YA	TIDAK	YA	YA	YA	TIDAK	YA
48	TIDAK	YA	YA	TIDAK	YA	YA	TIDAK	YA	YA	YA	TIDAK	YA
49	YA	YA	YA	YA	YA	YA	TIDAK	YA	YA	YA	TIDAK	YA
50	YA	YA	YA	YA	YA	YA	TIDAK	YA	YA	YA	TIDAK	YA
56	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK
57	TIDAK	TIDAK	TIDAK	YA	TIDAK	TIDAK	TIDAK	TIDAK	YA	YA	TIDAK	TIDAK
58	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK
59	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK
60	TIDAK	TIDAK	TIDAK	YA	TIDAK	YA	TIDAK	TIDAK	YA	YA	TIDAK	TIDAK
61	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	YA	YA	TIDAK	TIDAK
62	TIDAK	TIDAK	TIDAK	YA	TIDAK	TIDAK	TIDAK	TIDAK	YA	YA	TIDAK	TIDAK
63	TIDAK	TIDAK	YA	YA	TIDAK	TIDAK	TIDAK	TIDAK	YA	YA	TIDAK	TIDAK
64	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	YA	TIDAK	TIDAK
65	TIDAK	TIDAK	TIDAK	YA	TIDAK	TIDAK	TIDAK	TIDAK	YA	TIDAK	TIDAK	TIDAK
66	TIDAK	TIDAK	YA	TIDAK	TIDAK	TIDAK	TIDAK	YA	YA	TIDAK	TIDAK	TIDAK
67	TIDAK	TIDAK	TIDAK	TIDAK	YA	TIDAK	TIDAK	TIDAK	YA	TIDAK	TIDAK	TIDAK
68	TIDAK	TIDAK	TIDAK	YA	TIDAK	YA	TIDAK	TIDAK	TIDAK	YA	YA	TIDAK
69	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK
70	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK

Pada gambar diatas nilai atribut *Ya* menandakan kasus mengalami kondisi sesuai atribut sedangkan nilai atribut *Tidak* mengartikan bahwa kasus tidak mengalami kondisi sesuai atribut. Dan dalam nilai hasil dengan nilai *Ya* dapat diartikan bahwa kasus bernilai positif didiganosis autisme, sedangkan dengan nilai *Tidak* dapat diartikan bahwa kasus bernilai negatif autisme.

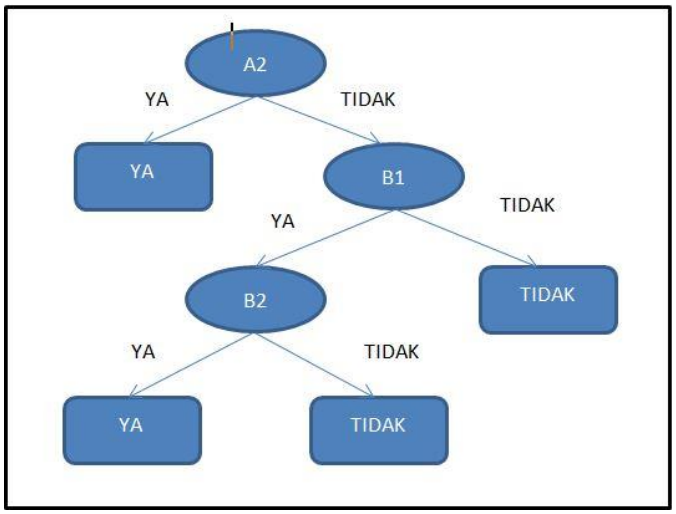
Dari gambar diatas dihasilkan sebuah nilai *gain* dan *entropy*. Berikut adalah hasil perhitungan *entropy* dan *gain* pada *node* pertama dari data diatas:

Tabel 4. Analisa Perhitungan

Node 1	Jumlah	Ya	Tidak	Entropi	Gain
Total	50	35	15	0.881290899	
A1					0.281290899
YA	20	20	0	0	
TIDAK	30	15	15	1	
A2					0.599145623
YA	31	31	0	0	
TIDAK	19	4	15	0.74248757	
A3					0.395969454
YA	33	31	2	0.32984607	
TIDAK	17	4	13	0.787126586	
A4					0.004505515
YA	23	17	6	0.828055725	
TIDAK	27	18	9	0.918295834	
B1					0.437863277
YA	31	30	1	0.205592508	
TIDAK	19	5	14	0.831474388	
B2					0.490140054
YA	35	33	2	0.315997133	
TIDAK	15	2	13	0.566509507	
B3					0.423162578
YA	26	26	0	0	
TIDAK	24	9	15	0.954434003	
B4					0.122525248
YA	17	16	1	0.322756959	
TIDAK	33	19	14	0.98337619	
C1					0.021596817
YA	33	25	8	0.799048521	
TIDAK	17	10	7	0.977417818	
C2					0.030363377
YA	31	24	7	0.770629069	
TIDAK	19	11	8	0.981940787	
C3					0.223136765
YA	23	22	1	0.258018669	
TIDAK	27	13	14	0.999010271	

Dari hasil perhitungan nilai entropy dan nilai *gain* yang terdapat pada gambar 3 terlihat bahwa atribut A2 (Kesulitan bermain dengan teman sebaya.) mempunyai nilai gain paling tinggi yaitu 0.599145623. Oleh karena itu A2 (Kesulitan bermain dengan teman sebaya) akan menjadi akar (node pertama) dari pohon keputusan yang terbentuk. Setelah didapatkan nilai gain yang tertinggi maka selanjutnya dihitung dengan cara yang sama dengan menggunakan persamaan entropy untuk mendapatkan nilai entropy dan persamaan gain untuk mendapatkan nilai gain.

Setelah dilakukan hasil perhitungan nilai entropy dan nilai gain, maka akan terbentuk pohon keputusan seperti gambar 4.



Gambar 4. Pohon keputusan

Setelah pohon keputusan (decision tree) terbentuk, selanjutnya dilakukan perubahan menjadi rule, berikut ini merupakan pembentukan rule tree.

Rule 1

IF A2 == Ya
THEN Ya (Positif Autisme)

Rule 2

IF A2 == Tidak
AND B1== Tidak
THEN Tidak (Negatif Autisme)

Rule 3

IF A2 == Tidak
AND B1 == Ya
AND B2 == Ya
THEN Ya (Positif Autisme)

Rule 4

IF A2 == Tidak
AND B1 == Ya
AND B2 == Tidak
THEN Tidak (Negatif Autisme)

Model yang telah dibentuk diuji tingkat akurasinya dengan memasukan atau uji yang berasal dari data uji dengan menggunakan *matrix confusion* pada untuk menguji tingkat akurasi. Dengan metode *confusion matrix* dihasilkan nilai akurasi, nilai *class recall* dan nilai *class precision* yang akan dijelaskan pada gambar dibawah ini:

Tabel 3. Perhitungan nilai akurasi

	True autisme	True not autisme
Pred. autisme	14	1
Pred. not autisme	1	4

Dari tabel diatas dihasilkan nilai akurasi sistem sebesar 90%, untuk nilai recall sebesar 93,33% dan nilai precision sebesar 93,33%.

4. KESIMPULAN DAN SARAN

4.1 Kesimpulan

Dari hasil penelitian yang sudah dilakukan, dapat disimpulkan bahwa diagnosis autisme pada anak dapat diimplementasikan menggunakan rumusan DSM-V dan algoritma C4.5. Setelah dilakukan pengujian dengan metode *confusin matrix* diperoleh hasil nilai akurasi sebesar 90%.

4.2 Saran

Dari hasil penelitian pembuatan program diagnosis autisme dengan rumusan Diagnostic and Statistical Manual of Mental Disorders V menggunakan algoritma C4.5 ini dapat diberikan beberapa saran diantaranya :

1. Dataset yang digunakan dapat lebih dikembangkan, dari segi jumlah data maupun parameternya, sehingga tidak hanya dapat mendeteksi autis atau tidak, namun dapat mengetahui terdiagnosis dengan jenis autisnya.
2. Sistem diagnosis autisme berbasis website menggunakan algoritma C4.5 ini dapat dikembangkan dalam bentuk platform lainnya, seperti mobile..
3. Diharapkan algoritma improvisasi apriori ini dapat dikombinasikan dengan algoritma lain untuk membantu mempercepat proses analisis serta memperkuat akurasi dari aturan asosiasi yang dihasilkan.

5. DAFTAR RUJUKAN

[1] Han, J., & Kamber, M. (2006). *Data Mining Concept and Techniques, 2nd edition*. USA: Elsevier, Inc.
[2] Assosiation, A. P. (2013). Diagnostic and Statistical Disorders Fifteeth Edition. Amerika: American Phisyciatric Assosiation.
[3] Bayu, D. B. (2018). Penerapan Algoritma C4.5 Untuk Deteksi Dini Autisme Pada Anak. SENTIKA.