



Hak cipta dan penggunaan kembali:

Lisensi ini mengizinkan setiap orang untuk menggubah, memperbaiki, dan membuat ciptaan turunan bukan untuk kepentingan komersial, selama anda mencantumkan nama penulis dan melisensikan ciptaan turunan dengan syarat yang serupa dengan ciptaan asli.

Copyright and reuse:

This license lets you remix, tweak, and build upon work non-commercially, as long as you credit the origin creator and license it on your new creations under the identical terms.

BAB II

LANDASAN TEORI

Beberapa telaah literatur yang diperlukan dalam penelitian ini adalah sebagai berikut.

2.1 Sistem Pakar

Sistem pakar atau *expert system* merupakan cabang keilmuan dari *artificial intelligence* (AI), yang memiliki konsep berupa menyalurkan pengetahuan spesifik suatu bidang dari manusia ke komputer (Liao, 2005). Pengetahuan tersebut disimpan di dalam komputer dan digunakan pengguna sesuai kebutuhan. Sistem dapat membuat keputusan dan akhirnya mencapai sebuah kesimpulan (Liao, 2005).

Sebuah sistem pakar terdiri dari 3 bagian yang dapat dijabarkan sebagai berikut (Copley, 2011).

1. Tampilan antarmuka (*user interface*)

User interface memungkinkan pengguna biasa untuk memberikan *query* kepada sistem pakar, dan menerima saran dari sistem pakar. *Query* yang dimaksud dapat berupa pertanyaan kepada sistem pakar, ataupun *input* jawaban dari pertanyaan oleh sistem pakar. *User interface* dirancang sangat sederhana sehingga mudah digunakan.

2. *Knowledge base*

Knowledge base merupakan kumpulan aturan dan fakta yang dibuat berdasarkan informasi yang disediakan oleh pakar.

3. *Inference engine*

Inference engine berfungsi seperti *search engine*, yaitu mengecek *knowledge base* untuk informasi yang sesuai dengan *query* pengguna.

2.2 Decision Tree

Pohon keputusan atau *decision tree* adalah sebuah *tree* dimana tiap *branch* merepresentasikan pilihan antara sejumlah alternatif, dan tiap *leaf node* merepresentasikan sebuah keputusan. Algoritma untuk membangun pohon keputusan yang lazim digunakan adalah ID3, C4.5, dan CART (Lakshmi, et al., 2013). *Decision tree* umumnya digunakan dalam menyelesaikan kasus berikut (Song & Lu, 2015).

1. *Variable selection*. Banyak variabel yang tersimpan memiliki hubungan yang kecil dan tidak seharusnya digunakan dalam pelaksanaan data *mining*. Metode *decision tree* dapat digunakan untuk memilih variabel input yang paling relevan yang harus digunakan dalam membentuk model pohon keputusan (*decision tree*). Menilai kepentingan relatif dari variabel. Biasanya, kepentingan suatu variabel dihitung berdasarkan reduksi dari akurasi model saat variabel dihilangkan.
2. Menangani *value* yang hilang. Pohon keputusan dapat menangani *missing* data melalui dua cara: dengan mengklasifikasikan *value* yang hilang ke dalam satu kategori tersendiri dan dapat dianalisis menggunakan nilai variabel lain, atau dengan membangun *tree model* yang membuat variabel dengan *missing value* sebagai *target* untuk membuat prediksi dan mengubah nilai yang hilang menjadi hasil prediksi.

3. Prediksi. Prediksi merupakan penggunaan model pohon keputusan yang paling penting. Menggunakan model pohon yang dibentuk dari data historis sebelumnya, mempermudah untuk memprediksi hasil yang akan didapatkan.
4. Manipulasi data. Pohon keputusan dapat membantu memutuskan cara terbaik untuk menghilangkan variabel numerik menjadi sejumlah kategori yang lebih mudah diatur.

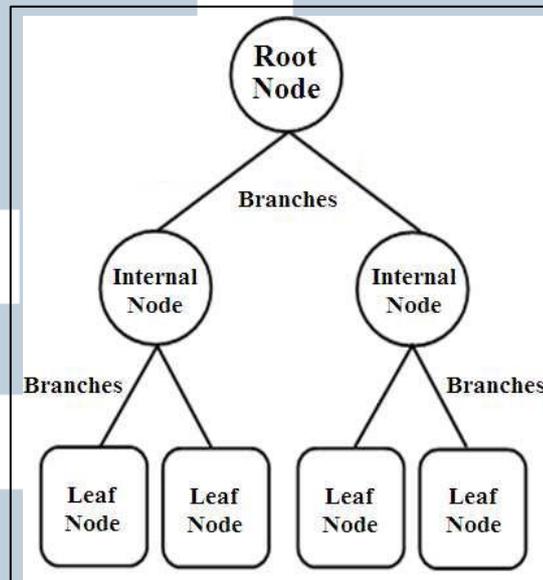
Kelebihan *decision tree* adalah sebagai berikut (Saxena, 2017).

1. *Decision tree* mudah dimengerti.
2. *Decision tree* mengikuti cara yang menyerupai manusia dalam mengambil keputusan.
3. Interpretasi visualisasi dari model pohon keputusan (*decision tree models*) yang rumit dapat disimplifikasi.
4. Jumlah *hyper-parameter* yang harus diperbaiki mendekati *null*.

Sedangkan, kekurangan dari *decision tree* adalah sebagai berikut (Saxena, 2017).

1. Besar kemungkinan terjadinya *overfitting* pada *decision tree*. *Overfitting* adalah
2. Secara umum, *decision tree* memberikan hasil akurasi yang lebih rendah untuk sebuah *dataset* dibandingkan algoritma *machine learning* lain.
3. *Information gain* pada *decision tree* untuk variabel kategorikal memberikan hasil yang bias untuk atribut dengan jumlah kategori lebih besar.
4. Kalkulasi dapat menjadi rumit ketika terdapat banyak *class labels*.

Komponen utama pada model pohon keputusan (*decision tree models*) adalah *nodes* dan *branches*, serta tahap penting dalam membangun pohon keputusan adalah *splitting*, *stopping*, dan *pruning* (Song & Lu, 2015).



Gambar 2.1 Komponen Utama Model Pohon Keputusan (Sá, Almeida, Rocha, & Mota, 2011)

a. Node

Terdapat tiga macam *node* pada model pohon keputusan dengan penjabaran sebagai berikut.

1. *Root node* atau *decision node*, menggambarkan pilihan yang menghasilkan pembagian data menjadi dua atau lebih *subset* yang bersifat *mutually exclusive*.
2. *Internal node* atau *chance node*, menggambarkan salah satu pilihan yang tersedia pada titik tertentu didalam *decision tree*.
3. *Leaf node* atau *end node*, menggambarkan hasil keputusan dari kombinasi dari *decisions* dan *events*.

b. Branch

Branch merepresentasikan kemungkinan suatu kejadian yang berasal dari *root node* dan *internal node*.

c. Splitting

Sebelum membangun sebuah model pohon keputusan, variabel *input* harus didefinisikan terlebih dahulu. Lalu, data dibelah pada *root node* atau *internal node* selanjutnya, menjadi dua atau lebih '*bin*' berdasarkan status variabel tersebut.

d. Stopping

Semakin kompleks sebuah pohon keputusan, maka semakin rendah reliabilitas pohon tersebut dalam melakukan prediksi. Oleh sebab itu, pembuatan model pohon keputusan harus dapat berhenti pada keadaan tertentu.

e. Pruning

Pruning merupakan cara alternatif untuk membangun pohon keputusan apabila aturan *stopping* tidak bekerja dengan baik. *Pruning* dapat dilakukan dengan dua cara, yaitu *pre-pruning* dan *post-pruning* (Bramer, 2016). *Pre-pruning* mencegah terbentuknya *branch* yang tidak signifikan, salah satunya dengan memberikan jumlah *subset* minimal untuk membuat *branch* baru (Bramer, 2016). Jumlah subset minimal yang umumnya digunakan adalah 5 atau 10 (Bramer, 2016). Sedangkan, *post-pruning* dilakukan dengan membangun pohon keputusan, kemudian menghilangkan *branch* yang tidak signifikan (Bramer, 2016).

2.3 Attention Deficit Hyperactivity Disorder

Attention Deficit Hyperactivity Disorder atau ADHD adalah sebuah pola inatensi atau hiperaktivitas/impulsivitas yang persisten, yang mengganggu aktifitas

atau perkembangan seseorang (American Psychiatric Association, 2013). Inatensi adalah ketidakmampuan untuk menyelesaikan pekerjaan dengan teliti, memperhatikan, memikirkan, mendengarkan, atau menyaksikan seseorang atau sesuatu (CDC, 2014). Hiperaktivitas adalah keadaan dimana seseorang memiliki tingkat aktivitas atau kegembiraan yang tinggi dan tidak biasa (CDC, 2014). Impulsivitas adalah bertindak sesuai keinginan, pemikiran, dan perasaan secara tiba-tiba, bukan dari pemikiran secara hati-hati (CDC, 2014). ADHD telah memengaruhi sebelas persen anak-anak usia sekolah (Visser, et al., 2014). Gangguan ini dapat diidentifikasi dari hal-hal berikut (American Psychiatric Association, 2013).

1. Inatensi atau gangguan konsentrasi, yang ditandai dengan gejala berikut.
 - a. Seringkali gagal memberikan perhatian kepada *detail*
 - b. Sering membuat kesalahan kecil di sekolah, tempat kerja, atau selama aktivitas lainnya
 - c. Sering kesulitan dalam memusatkan perhatian pada sebuah tugas atau aktivitas tertentu
 - d. Seringkali terlihat tidak mendengarkan jika berbicara langsung
 - e. Seringkali tidak mengikuti instruksi dan gagal menyelesaikan pekerjaan rumah, tugas, atau pekerjaan di tempat kerja
 - f. Sering kesulitan dalam mengorganisir tugas dan aktivitas
 - g. Seringkali menghindari, tidak menyukai, ataupun enggan melakukan pekerjaan yang mengharuskan usaha mental secara terus-menerus
 - h. Sering kehilangan barang
 - i. Perhatiannya mudah terganggu oleh rangsangan luar

- j. Seringkali menjadi pelupa dalam aktivitas sehari-hari
2. Hiperaktivitas dan Impulsivitas, yang ditandai dengan gejala berikut.
- a. Sering terlihat gelisah atau menggoyangkan tangan, kaki, atau gelisah di tempat duduk
 - b. Selalu bergerak pada situasi dimana mengharuskan seseorang untuk tetap diam
 - c. Sering berlarian atau bersemangat pada situasi yang tidak tepat
 - d. Seringkali tidak dapat berpartisipasi pada kegiatan santai dan hening.
 - e. Seringkali bertindak seakan dikendalikan oleh sensor motorik (mesin)
 - f. Terlalu banyak berbicara
 - g. Seringkali memberikan jawaban saat pertanyaan belum selesai ditanyakan
 - h. Seringkali kesulitan menunggu giliran
 - i. Sering mengganggu orang lain

2.4 Child Behaviour Checklist (CBCL)

Child Behaviour Checklist (CBCL) adalah alat ukur yang umum digunakan untuk mendeteksi masalah perilaku dan emosi pada anak (Gregory, 2016). CBCL pertama kali dikembangkan oleh Achenbach untuk mendeteksi perilaku bermasalah pada anak melalui kuesioner yang mendeskripsikan dan mendeteksi perilaku tersebut (Gregory, 2016). Kuesioner terdiri dari 112 pertanyaan, dimana masing-masing pertanyaan pada kuesioner terbagi menjadi delapan kategori, yang berfokus pada delapan aspek perilaku yang berbeda. Delapan aspek perilaku tersebut yaitu *social withdrawal*, masalah somatik, masalah depresi, masalah sosial, masalah

pikiran, masalah konsentrasi, perilaku membangkang, dan perilaku agresif (Gregory, 2016). Tabel 2.1 menunjukkan pertanyaan yang terdapat pada kuesioner Child Behaviour Checklist.

Tabel 2.1 Child Behaviour Checklist

0	1	2	No.	Pertanyaan
			1.	Bertingkah laku kekanak-kanakkan untuk usianya
			2.	Meminum minuman beralkohol tanpa izin orangtua (deskripsikan):
			3.	Banyak membantah
			4.	Gagal menyelesaikan kegiatan-kegiatan yang dimulai
			5.	Hanya menikmati sedikit hal
			6.	Buang air besar di luar toilet
			7.	Membual, membanggakan diri
			8.	Sulit berkonsentrasi atau memusatkan perhatian untuk jangka waktu yang lama
			9.	Tidak dapat mengusir sesuatu dari pikirannya, terobsesi (deskripsikan):
			10.	Tidak dapat duduk tenang, tidak bisa diam, hiperaktif
			11.	Terlalu bergantung pada orang dewasa
			12.	Mengeluh kesepian
			13.	Terlihat bingung
			14.	Sering menangis
			15.	Kejam pada binatang
			16.	Kejam, menggertak atau jahat kepada orang lain
			17.	Melamun atau banyak pikiran
			18.	Melukai diri dengan sengaja atau mencoba bunuh diri
			19.	Menuntut banyak perhatian
			20.	Menghancurkan barang miliknya sendiri
			21.	Menghancurkan barang milik keluarganya atau orang lain
			22.	Tidak patuh di rumah
			23.	Tidak patuh di sekolah
			24.	Tidak makan dengan baik
			25.	Tidak akur dengan anak-anak lain
			26.	Tidak kelihatan bersalah setelah berperilaku buruk
			27.	Gampang cemburu atau iri
			28.	Tidak mematuhi peraturan di rumah, sekolah, atau tempat lain
			29.	Takut terhadap binatang, situasi, atau tempat tertentu, selain sekolah (deskripsikan):
			30.	Takut pergi ke sekolah
			31.	Takut dirinya telah melakukan hal buruk

Tabel 2.2 Child Behaviour Checklist (Lanjutan)

0	1	2	No.	Pertanyaan
			32.	Merasa diri harus sempurna
			33.	Merasa atau mengeluh bahwa tidak ada yang mencintainya
			34.	Merasa orang lain berniat buruk padanya
			35.	Merasa tidak bernilai atau inferior
			36.	Sering terluka atau terkena kecelakaan
			37.	Sering berkelahi
			38.	Sering diejek
			39.	Berteman dengan orang lain yang sering bermasalah
			40.	Mendengar bunyi atau suara yang tidak ada (deskripsikan):
			41.	Impulsif atau bertingkah laku tanpa berpikir terlebih dahulu
			42.	Lebih suka menyendiri
			43.	Berbohong atau mencontek
			44.	Mengigit kuku jari
			45.	Cemas, tegang
			46.	Menampilkan tingkah laku cemas atau gelisah
			47.	Mimpi buruk
			48.	Tidak disukai anak lain
			49.	Konstipasi
			50.	Terlalu penakut atau cemas
			51.	Merasa pusing atau pening
			52.	Merasa terlalu bersalah
			53.	Terlalu banyak makan
			54.	Terlalu letih tanpa alasan
			55.	Berat badan berlebihan
			56.	Masalah fisik tanpa alasan medis yang diketahui
				a. Nyeri atau sakit (bukan di perut atau kepala)
				b. Sakit kepala
				c. Mual
				d. Masalah dengan mata (yang tidak dibenarkan dengan kacamata)
				e. Ruam atau masalah kulit lainnya
				f. Sakit perut
				g. Muntah
				h. Lainnya (deskripsikan):
			57.	Menyerang orang secara fisik
			58.	Bermain-main dengan anggota badan seperti hidung, kulit atau lainnya
			59.	Bermain-main dengan alat kelamin sendiri di public
			60.	Bermain-main dengan alat kelamin sendiri secara berlebihan
			61.	Pekerjaan sekolah tergolong buruk

Tabel 2.3 Child Behaviour Checklist (Lanjutan)

0	1	2	No.	Pertanyaan
			62.	Koordinasi gerak yang buruk atau ceroboh
			63.	Lebih suka berada dengan anak-anak yang lebih tua
			64.	Lebih suka berada dengan anak-anak yang lebih muda
			65.	Menolak berbicara
			66.	Mengulangi kegiatan tertentu terus-menerus (deskripsikan):
			67.	Lari dari rumah
			68.	Banyak berteriak
			69.	Suka menyimpan rahasia
			70.	Melihat hal-hal yang tidak ada (deskripsikan):
			71.	Sadar diri atau mudah malu
			72.	Suka membakar
			73.	Masalah seksual (deskripsikan):
			74.	Pamer atau membadut
			75.	Terlalu malu atau segan
			76.	Tidur lebih sedikit daripada anak-anak lain
			77.	Tidur lebih banyak daripada anak-anak lain selama siang hari dan/atau malam hari (deskripsikan):
			78.	Alpa atau mudah dialihkan perhatiannya
			79.	Masalah bicara (deskripsikan):
			80.	Pandangan mata kosong
			81.	Mencuri di rumah
			82.	Mencuri di luar rumah
			83.	Menyimpan terlalu banyak barang yang tidak dibutuhkan (deskripsikan):
			84.	Perilaku aneh (deskripsikan):
			85.	Ide-ide aneh (deskripsikan):
			86.	Keras kepala, cemberut atau mudah marah
			87.	Suasana hati atau perasaan berubah tiba-tiba
			88.	Sering merajuk
			89.	Sering curiga
			90.	Menyumpah atau berkata kotor
			91.	Berbicara tentang bunuh diri
			92.	Mengigau atau berjalan dalam tidur
			93.	Berbicara terlalu banyak
			94.	Banyak mengejek
			95.	Mudah marah
			96.	Terlalu banyak berpikir tentang seks
			97.	Mengancam orang-orang
			98.	Mengisap jempol
			99.	Merokok, mengunyah atau menghirup tembakau
			100.	Masalah tidur (deskripsikan):
			101.	Bolos dari sekolah
			102.	Tidak aktif, lamban atau kurang energi
			103.	Sedih atau depresi

Tabel 2.4 Child Behaviour Checklist (Lanjutan)

0	1	2	No.	Pertanyaan
			104.	Bising
			105.	Menggunakan obat-obatan untuk keperluan non-medis (tidak termasuk alkohol atau tembakau) (deskripsikan):
			106.	Perusakan
			107.	Mengompol saat tidak tidur
			108.	Mengompol saat tidur
			109.	Mengeluh
			110.	Berharap berjenis kelamin yang berlainan dengan dirinya
			111.	Tidak berhubungan dengan orang lain
			112.	Khawatir
			113.	Mohon tuliskan masalah yang dihadapi anak yang tidak tertulis di atas:

Keterangan:

0 = Tidak benar (sejauh pengamatan)

1 = Terkadang benar

2 = Benar atau seringkali benar

2.5 Connors' Parent Rating Scale

Connors' Parent Rating Scale adalah sebuah alat untuk melakukan riset dan klinis yang banyak digunakan untuk mendapatkan laporan orangtua mengenai masalah perilaku anak (Connors, Sitarenios, Parkez, & Epstein, 1998). Connors' Parent Rating Scale dapat digunakan untuk mendeteksi tujuh model perilaku berbeda, yaitu masalah kognitif, masalah oposisional, masalah hiperaktif-impulsif, masalah gelisah-pemalu, perfeksionisme, masalah sosial, dan masalah psikosomatis (Connors, Sitarenios, Parkez, & Epstein, 1998). Tabel 2.2 menunjukkan pertanyaan yang terdapat pada kuesioner Connors' Parent Rating Scale.

Tabel 2.5 Conners' Parent Rating Scale

No.	Pertanyaan	Tidak sama sekali	Hanya sedikit	Seringkali	Sangat sering
1.	Bermain-main dengan benda-benda (kuku, jari tangan, rambut, pakaian).				
2.	Bersikap lancang kepada orang dewasa.				
3.	Memiliki kesulitan dalam membuat teman.				
4.	Bersemangat, impulsif.				
5.	Ingin mengoperasikan benda-benda.				
6.	Mengunyah atau mengisap (jempol; pakaian; selimut).				
7.	Mudah atau seringkali menangis.				
8.	Menyimpan dendam, sering marah, merasa diperlakukan tidak adil.				
9.	Melamun.				
10.	Kesulitan dalam belajar.				
11.	Gelisah seperti 'cacing' kepanasan				
12.	Penakut (terhadap situasi, orang atau tempat baru; terhadap sekolah).				
13.	Tidak bisa diam, selalu bersemangat dan bergerak				
14.	Menghancurkan barang.				
15.	Berbohong.				
16.	Pemalu.				
17.	Lebih bermasalah daripada anak lain yang seumuran.				
18.	Berbicara dengan cara yang berbeda daripada anak lain yang seumuran (berbicara seperti bayi; terbata-bata; sulit dimengerti).				
19.	Menolak mengakui kesalahan atau menyalahkan orang lain.				
20.	Sering bertengkar.				
21.	Cemberut dan merajuk.				
22.	Mencuri.				
23.	Tidak patuh atau patuh tetapi tidak ikhlas.				

Tabel 2.6 Conners' Parent Rating Scale (Lanjutan)

No.	Pertanyaan	Tidak sama sekali	Hanya sedikit	Seringkali	Sangat sering
24.	Lebih mengkhawatirkan banyak hal daripada anak-anak lain (tentang kesendirian, penyakit atau kematian).				
25.	Gagal dalam menyelesaikan tugas.				
26.	Perasaannya mudah tersakiti.				
27.	Menggertak yang lain.				
28.	Tidak dapat menghentikan aktivitas yang berulang-ulang.				
29.	Kejam.				
30.	Kekanak-kanakan (meminta bantuan yang tidak diperlukan; menempel; butuh diyakinkan berulang).				
31.	Mudah terganggu dan sulit memusatkan perhatian.				
32.	Sakit kepala.				
33.	Suasana hati mudah berubah dengan cepat dan drastis.				
34.	Tidak menyukai atau tidak mengikuti peraturan dan batasan.				
35.	Sering berkelahi.				
36.	Tidak akur dengan saudara.				
37.	Mudah merasa frustrasi dalam berusaha.				
38.	Mengganggu anak-anak lain.				
39.	Secara umum tidak bahagia.				
40.	Bermasalah dengan makan (tidak nafsu makan; berkeliaran di antara suapan).				
41.	Sakit perut.				
42.	Bermasalah dengan tidur (tidak bisa tidur; bangun terlalu cepat; bangun di tengah malam).				
43.	Sakit atau nyeri lainnya.				
44.	Muntah atau mual.				
45.	Merasa dikhianati di dalam keluarga.				
46.	Membual dan menyombongkan diri.				

Tabel 2.7 Connors' Parent Rating Scale (Lanjutan)

No.	Pertanyaan	Tidak sama sekali	Hanya sedikit	Seringkali	Sangat sering
47.	Membiarkan diri disuruh-suruh.				
48.	Masalah pencernaan (buang air tidak teratur; konstipasi).				

2.6 Algoritma C4.5

Algoritma C4.5 adalah algoritma yang umum digunakan untuk membentuk pohon keputusan. Algoritma C4.5 menyediakan banyak keuntungan, seperti menangani data *training* dengan atribut nilai yang hilang (*missing value*), menangani atribut *cost* yang berbeda-beda, melakukan *pruning* terhadap pohon keputusan, dan menangani atribut dengan nilai diskrit dan kontinu (Adhatrao, et al., 2013). Pada algoritma ini, pemilihan atribut yang akan ditempatkan pada suatu simpul dilakukan dengan menghitung entropi informasi (*information entropy*) dan mencari nilai yang paling minimum. Algoritma ini memilih atribut yang menghasilkan perolehan informasi terbesar (*highest information gain*) dalam menghasilkan subpohon (*subtree*) untuk mengklasifikasikan sampel (Adhatrao, et al., 2013).

Menurut Larose (2005), tahapan dalam membuat pohon keputusan dengan algoritma C4.5 adalah sebagai berikut.

1. Mempersiapkan data *training*, yaitu data yang berasal dari *history* yang sebelumnya sudah terjadi dan sudah diklasifikasikan ke dalam kelas-kelas tertentu.

2. Menghitung *node* akar (*root node*). *Root node* ditentukan dengan menghitung nilai *information gain* dari masing-masing atribut, dan mengambil nilai *gain* tertinggi. *Information gain* didapatkan melalui:

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \quad \dots(2.1)$$

Keterangan:

S: *case set*

A: atribut

n: jumlah partisi atribut *A*

S_i: *case* ke-*i*

$|S_i|$: proporsi *S_i* terhadap *S*

$|S|$: jumlah *case* pada *S*

Sedangkan, *information entropy* dihasilkan dengan rumus berikut

$$Entropy(S) = \sum_{i=1}^n -p_i * \log_2 p_i \quad \dots(2.2)$$

Keterangan:

S: *case set*

n: jumlah partisi *S*

S_i: *case* ke-*i*

p_i: proporsi dari *S_i* terhadap *S*

3. Ulangi langkah 2 sampai seluruh data terpartisi.
4. Proses akan selesai jika:

- a. Semua data pada simpul *n* telah mendapat kelas
- b. Tidak ada atribut di dalam data yang dipartisi lagi
- c. Tidak ada data di dalam cabang yang kosong

Untuk memperbaiki informasi dari *gain* dapat menggunakan *gain ratio*.

$$GainRatio(S) = \frac{Gain(S,A)}{SplitInfo(S,A)} \quad \dots(2.3)$$

Keterangan:

S: sampel data yang digunakan untuk *training*

A: atribut

SplitInfo adalah nilai *split information* pada atribut.

$$SplitInfo(S,A) = \sum_{i=1}^n - \left(\frac{S_i}{S}\right) \log_2 \left(\frac{S_i}{S}\right) \quad \dots(2.4)$$

Keterangan:

S_i: case ke-*i*

S: case set

A: atribut

Atribut dengan nilai kontinu ditangani dengan melakukan diskretisasi nilai, yaitu mengkonversi tiap nilai kontinu menjadi sejumlah kategori (Bramer, 2016).

Diskretisasi untuk nilai kontinu dilakukan dengan menentukan *threshold* berdasarkan titik tengah dari tiap *interval* data (Quinlan, 1993). Langkah-langkah diskretisasi menurut Bramer (2016) adalah

1. Urutkan data secara *ascending*
2. Jika terdapat *n value* (v_1, v_2, \dots, v_n) berbeda, hitung nilai *information gain* untuk setiap *pseudo-attribute* ($A < v_2, A < v_3, \dots$) hingga *n-1*
3. Cari nilai *information gain* tertinggi

2.7 Confusion Matrix

Confusion matrix merupakan salah satu metode yang dapat digunakan untuk mengukur kinerja suatu metode klasifikasi. Pada dasarnya, *confusion matrix* mengandung informasi yang membandingkan hasil klasifikasi yang dilakukan oleh sistem dengan hasil klasifikasi yang seharusnya (Prasetyo, 2012). Tabel 2.8 menunjukkan gambaran *confusion matrix*.

Tabel 2.8 Confusion Matrix (Asch, 2013)

	True label A	True not A
Predicted label A	TP (True Positive)	FP (False Positive)
Predicted not A	FN (False Negative)	TN (True Negative)

Pada pengukuran kinerja menggunakan *confusion matrix*, terdapat empat istilah sebagai representasi hasil proses klasifikasi. Keempat istilah tersebut adalah *True Positive* (TP), *True Negative* (TN), *False Positive* (FP) dan *False Negative* (FN). Penjelasan mengenai istilah-istilah tersebut adalah sebagai berikut (Indriani, 2014).

- TP merupakan jumlah dokumen dari kelas A yang benar diklasifikasikan sebagai kelas A.
- TN merupakan jumlah dokumen yang bukan merupakan kelas A yang benar diklasifikasikan sebagai bukan kelas A.
- FP merupakan jumlah dokumen yang bukan merupakan kelas A yang salah diklasifikasikan sebagai kelas A.
- FN merupakan jumlah dokumen dari kelas A yang salah diklasifikasikan sebagai bukan kelas A.

Berdasarkan keempat nilai tersebut, dapat diperoleh nilai akurasi, presisi, dan *recall* (Sokolova & Lapalme, 2009). Nilai akurasi menggambarkan seberapa akurat

sistem dapat mengklasifikasikan data secara benar. Nilai presisi menggambarkan jumlah data yang diklasifikasikan secara benar dibagi dengan total data yang diklasifikasi sebagai kelas A. *Recall* menunjukkan berapa persen data kelas A yang terklasifikasikan dengan benar oleh sistem. Untuk klasifikasi dengan jumlah kelas lebih dari dua (*multi-class*), nilai akurasi, presisi, dan *recall* dihitung dengan cara mencari nilai rata-rata dari nilai akurasi, presisi, dan *recall* dari masing-masing kelas (Sokolova & Lapalme, 2009). Akurasi, presisi, dan *recall* untuk klasifikasi *multi-class* dapat dirumuskan sebagai berikut (Sokolova & Lapalme, 2009).

$$Akurasi = \frac{\sum_{i=1}^l \frac{TP_i + TN_i}{TP_i + TN_i + FP_i + FN_i}}{l} * 100\% \quad \dots(2.5)$$

$$Presisi = \frac{\sum_{i=1}^l TP_i}{\sum_{i=1}^l (TP_i + FP_i)} * 100\% \quad \dots(2.6)$$

$$Recall = \frac{\sum_{i=1}^l TP_i}{\sum_{i=1}^l (TP_i + FN_i)} * 100\% \quad \dots(2.7)$$

Keterangan:

l = jumlah kelas

