

BAB II

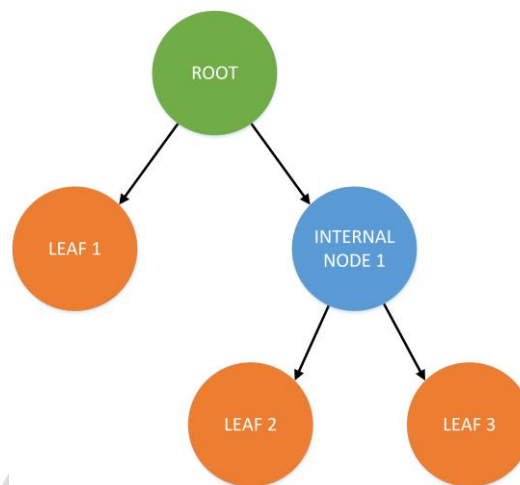
LANDASAN TEORI

2.1. *Data mining*

Terdapat beberapa definisi tentang *data mining* menurut pendapat para ahli. Definisi *data mining* menurut Turban, dkk. (2005) adalah proses yang menggunakan data statistik, matematika, kecerdasan buatan dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar. Sedangkan menurut Connolly dan Begg (2010), *data mining* adalah suatu proses ekstraksi atau penggalian data yang belum diketahui sebelumnya, namun dapat dipahami dan berguna dari database yang besar serta digunakan untuk membuat suatu keputusan bisnis yang sangat penting. Diungkapkan juga oleh Jayanthi Ranjan (2007), *data mining* sering kali dikenal dengan sebutan KDD (*Knowledge Discovery in Database*) yang memungkinkan suatu organisasi untuk melakukan kalkulasi dalam pengambilan keputusan melalui perakitan, akumulasi, analisis, dan pengaksesan data korporasi. Semua proses ini menggunakan beberapa macam *tool* seperti *tool* untuk *query* dan pelaporan, *tool* untuk proses analisis, dan *tool* untuk sistem pendukung keputusan.

2.2. *Klasifikasi Decision Tree*

Decision tree atau juga dapat disebut pohon keputusan, secara umum adalah gambaran pemodelan dari persoalan yang terdiri beberapa serangkaian keputusan yang mengarah pada solusi, dalam menyatakan keputusan dan memberikan sebuah solusi (Sunarko dan Pakaja, 2009). Konsep dari *decision tree* adalah mengubah sekumpulan data menjadi pohon keputusan yang merepresentasikan aturan-aturan dari sebuah keputusan.



Gambar 2.1. Konsep Dasar *Decision Tree*

Gambar 2.1 menggambarkan konsep dasar dari *decision tree*. *Decision tree* terdiri dari node untuk membentuk pohon berakar. Dari pohon tersebut diarahkan oleh *node* yang disebut akar (*root*) yang tidak mempunyai masukan (*input*) tetapi memiliki dua atau lebih keluaran (*output*). Jika *node* memiliki satu masukan (*input*) yang tepat dan *node* memiliki dua atau lebih keluaran (*output*), maka disebut dengan *internal node*. *Node* yang terakhir dan hanya memiliki masukan (*input*) disebut dengan *leaf node* atau biasanya disebut juga dengan *decision node* atau *terminal node*. *Leaf node* menunjukkan sebuah label atau hasil dari klasifikasi atau kelas keputusan (Tahir, 2019).

2.3. Algoritma ID3 (*Iterative Dichotomiser 3*)

ID3 (*Iterative Dichotomiser 3*) adalah algoritma *decision tree learning* (algoritma pembelajaran pohon keputusan) yang paling dasar. Algoritma ini melakukan pencarian secara menyeluruh (*greedy*) pada semua kemungkinan pohon keputusan. Salah satu algoritma induksi pohon keputusan yaitu ID3. ID3 dikembangkan oleh J. Ross Quinlan. Algoritma ID3 dapat diimplementasikan menggunakan fungsi *rekursif* (fungsi yang memanggil dirinya sendiri). Algoritma ID3 berusaha membangun *decision tree* (pohon keputusan) secara *top-down* (dari atas ke bawah) (David, 2004).

2.4. Entropy

Menurut Rokach dan Maimoon (2008), *information gain* atau biasa disebut *gain info* adalah kriteria pemisahan yang menggunakan pengukuran *entropy*. Untuk

mendapatkan *information gain* dari suatu atribut dibutuhkan *entropy* keseluruhan kelas atau $Entropy(S)$. Menurut Han et al. (2011), secara matematis *entropy* dirumuskan sebagai berikut :

$$Entropy(S) = \sum_{i=1}^c p_i \log_2 p_i \dots\dots(2.1)$$

Keterangan pada rumus 2.1 sebagai berikut :

S = himpunan kelas klasifikasi

c = banyaknya kelas klasifikasi

p_i = proporsi untuk kelas i

2.5. Information Gain

Setelah mendapatkan nilai *entropy*, maka dapat diukur efektivitas suatu atribut dalam mengklasifikasikan data yang disebut sebagai *information gain*. Secara matematis, *information gain* dari suatu atribut, dituliskan sebagai berikut :

$$Gain(S,A) = Entropy(S) - \sum_{v \in Values(A)} \frac{S_v}{S} Entropy(S_v) \dots\dots(2.2)$$

Keterangan pada rumus 2.2 sebagai berikut :

A = atribut

v = suatu nilai untuk atribut A

Values(A) = himpunan nilai untuk atribut A

S_v = sub-himpunan kelas klasifikasi

$Entropy(S_v)$ = *entropy* untuk sampel-sampel yang memiliki nilai v

Atribut dengan nilai *information gain* paling tinggi dibandingkan atribut yang lain, dipilih sebagai pemilah.

2.6. Contoh Perhitungan Metode Decision Tree

Pada contoh ini akan dilakukan prediksi “apakah bermain atau tidak ?” dengan jawaban “ya” atau “tidak” melalui 14 data *training* yang berisikan atribut cuaca, suhu, kelembapan, dan angin. Berikut adalah data yang digunakan :

Tabel 2.1. Data Training (Contoh Perhitungan)

| Cuaca | Suhu | Kelembapan | Angin | Bermain |
|---------|--------|------------|---------|---------|
| Cerah | Panas | Tinggi | Pelan | Tidak |
| Cerah | Panas | Tinggi | Kencang | Tidak |
| Mendung | Panas | Tinggi | Pelan | Ya |
| Hujan | Lembut | Tinggi | Pelan | Ya |
| Hujan | Dingin | Normal | Pelan | Ya |

| | | | | |
|---------|--------|--------|---------|-------|
| Hujan | Dingin | Normal | Kencang | Tidak |
| Mendung | Dingin | Normal | Kencang | Ya |
| Cerah | Lembut | Tinggi | Pelan | Tidak |
| Cerah | Dingin | Normal | Pelan | Ya |
| Hujan | Lembut | Normal | Pelan | Ya |
| Cerah | Lembut | Normal | Kencang | Ya |
| Mendung | Lembut | Tinggi | Kencang | Ya |
| Mendung | Panas | Normal | Pelan | Ya |
| Hujan | Lembut | Tinggi | Kencang | Tidak |

Tabel 2.1 berisikan data *training* yang akan digunakan dalam proses perhitungan metode *decision tree*. Selanjutnya adalah perhitungan *node* akar, diawali dengan menghitung jumlah masing-masing atribut lalu menghitung *entropy* dan *gain*.

Tabel 2.2. Perhitungan Atribut (Contoh Perhitungan)

| Node | Atribut | Detail | Jumlah | Ya | Tidak | Entropy | Gain |
|------|--------------|---------|-----------|----------|----------|---------|------|
| 1 | Total | | 14 | 9 | 5 | | |
| | Cuaca | Cerah | 5 | 2 | 3 | | |
| | | Mendung | 4 | 4 | 0 | | |
| | | Hujan | 5 | 3 | 2 | | |
| | Suhu | Dingin | 4 | 3 | 1 | | |
| | | Lembut | 6 | 4 | 2 | | |
| | | Panas | 4 | 2 | 2 | | |
| | Kelembapan | Normal | 7 | 6 | 1 | | |
| | | Tinggi | 7 | 3 | 4 | | |
| | Angin | Kencang | 6 | 3 | 3 | | |
| | | Pelan | 8 | 6 | 2 | | |

Tabel 2.2 berisikan data atribut yang sudah dikelompokkan dan dihitung jumlahnya sesuai dengan atribut bermain. Baris “total” di kolom “jumlah”, diisi dengan total data *training* yaitu 14 data, baris “total” di kolom “ya” diisi dengan total data *training* yang memiliki atribut “bermain” = “ya”, dan baris “total” di kolom “tidak” diisi dengan total data *training* yang memiliki atribut “bermain” = “tidak”. Dilanjutkan dengan menghitung atribut lainnya, contoh atribut “cuaca”, baris “cerah” di kolom “ya” diisi dengan total cuaca “cerah” yang memiliki atribut

“bermain” = ”ya”, baris “cerah” di kolom “tidak” diisi dengan total cuaca “cerah” yang memiliki atribut “bermain” = ”tidak”. Kemudian dilanjutkan menghitung atribut lainnya hingga semuanya terisi. Setelah semua sudah terisi, kemudian dilanjutkan dengan menghitung *entropy*. Berikut contoh perhitungan *entropy* pada atribut cuaca “cerah” :

$$\text{Entropy (cuaca cerah)} = - ((\text{cerah ya/cerah jumlah} * \text{LOG}(\text{cerah ya/cerah jumlah};2)) + (\text{cerah tidak/cerah jumlah} * \text{LOG}(\text{cerah tidak/cerah jumlah};2)))$$

$$\text{Entropy (cuaca cerah)} = - ((2 / 5 * \text{LOG}(2 / 5 ;2)) + (3 / 5 * \text{LOG}(3 / 5 ;2)))$$

$$\text{Entropy (cuaca cerah)} = 0,97095$$

Tabel 2.3. *Entropy Node 1* (Contoh Perhitungan)

| Node | Atribut | Detail | Jumlah | Ya | Tidak | Entropy | Gain |
|------|--------------|---------|-----------|----------|----------|---------|------|
| 1 | Total | | 14 | 9 | 5 | 0,9403 | |
| | Cuaca | Cerah | 5 | 2 | 3 | 0,9710 | |
| | | Mendung | 4 | 4 | 0 | 0,0000 | |
| | | Hujan | 5 | 3 | 2 | 0,9710 | |
| | Suhu | Dingin | 4 | 3 | 1 | 0,8113 | |
| | | Lembut | 6 | 4 | 2 | 0,9183 | |
| | | Panas | 4 | 2 | 2 | 1,0000 | |
| | Kelembapan | Normal | 7 | 6 | 1 | 0,5917 | |
| | | Tinggi | 7 | 3 | 4 | 0,9852 | |
| | Angin | Kencang | 6 | 3 | 3 | 1,0000 | |
| | | Pelan | 8 | 6 | 2 | 0,8113 | |

Hitung semua *entropy* tiap atribut, lalu dilanjutkan ke perhitungan *gain*. *Gain* dapat dihitung jika semua atribut sudah memiliki nilai *entropy*. Berikut contoh perhitungan *gain* pada atribut “cuaca” :

$$\text{Gain (cuaca)} = \text{entropy total} - ((\text{cerah jumlah/total jumlah} * \text{entropy cerah}) + (\text{mendung jumlah/total jumlah} * \text{entropy mendung}) + (\text{hujan jumlah/total jumlah} * \text{entropy hujan}))$$

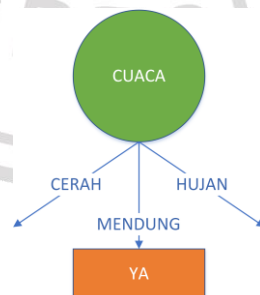
$$\text{Gain (cuaca)} = 0,94029 - ((5/14 * 0,97095) + (4/14 * 0) + (5/14 * 0,97095))$$

$$\text{Gain (cuaca)} = 0,2467$$

Tabel 2.4. *Gain Node 1* (Contoh Perhitungan)

| Node | Atribut | Detail | Jumlah | Ya | Tidak | Entropy | Gain |
|------|--------------|---------|-----------|----------|----------|---------|---------------|
| 1 | Total | | 14 | 9 | 5 | 0,9403 | |
| | Cuaca | Cerah | 5 | 2 | 3 | 0,9710 | 0,2467 |
| | | Mendung | 4 | 4 | 0 | 0,0000 | |
| | | Hujan | 5 | 3 | 2 | 0,9710 | |
| | Suhu | Dingin | 4 | 3 | 1 | 0,8113 | 0,0292 |
| | | Lembut | 6 | 4 | 2 | 0,9183 | |
| | | Panas | 4 | 2 | 2 | 1,0000 | |
| | Kelembapan | Normal | 7 | 6 | 1 | 0,5917 | 0,1518 |
| | | Tinggi | 7 | 3 | 4 | 0,9852 | |
| | Angin | Kencang | 6 | 3 | 3 | 1,0000 | 0,0481 |
| | | Pelan | 8 | 6 | 2 | 0,8113 | |

Dilanjutkan ke pemilihan *gain* tertinggi untuk dijadikan *node* akar. Pada tabel 2.4, diketahui nilai *gain* tertinggi dimiliki oleh atribut “cuaca”, maka atribut “cuaca” menjadi *node* akar. Karena kualifikasi “mendung” pada atribut “cuaca” memiliki nilai entropy 0 dengan hasil “ya” 4 data dan hasil “tidak” 0 data, maka kualifikasi “mendung” otomatis memiliki *leaf node* “ya”. Sedangkan kualifikasi “cerah” dan “hujan” yang memiliki nilai entropy tidak sama dengan 0, dilanjutkan perhitungannya ke *node* 2.

**Gambar 2.2.** *Pohon Keputusan Node 1* (Contoh Perhitungan)

Gambar 2.2 adalah pohon keputusan sementara berdasarkan hasil perhitungan *node* 1. Kemudian dilanjutkan dengan menghitung *node* 2 berdasarkan atribut “cuaca” kualifikasi “cerah” dan “hujan”.

Tabel 2.5. Data *Node* 2 Cuaca Cerah (Contoh Perhitungan)

| Node | Atribut | Detail | Jumlah | Ya | Tidak | Entropy | Gain |
|------|--------------|---------|----------|----------|----------|---------|------|
| 2 | Total | | 5 | 2 | 3 | | |
| | Suhu | Dingin | 1 | 1 | 0 | | |
| | | Lembut | 2 | 1 | 1 | | |
| | | Panas | 2 | 0 | 2 | | |
| | Kelembapan | Normal | 2 | 2 | 0 | | |
| | | Tinggi | 3 | 0 | 3 | | |
| | Angin | Kencang | 2 | 1 | 1 | | |
| | | Pelan | 3 | 1 | 2 | | |

Tabel 2.5 berisikan data atribut yang sudah dikelompokkan dan dihitung jumlahnya sesuai dengan atribut “cuaca” kualifikasi “cerah”. Baris “total” di kolom “jumlah”, diisi dengan total data atribut “cuaca” = “cerah” yaitu 5 data, baris “total” di kolom “ya” diisi dengan total data atribut “cuaca” = “cerah” yang memiliki atribut “bermain” = “ya” yaitu 2 data, dan baris “total” di kolom “tidak” diisi dengan total data atribut “cuaca” = “cerah” yang memiliki atribut “bermain” = “tidak” yaitu 3 data. Kemudian dilanjutkan menghitung atribut lainnya hingga semuanya terisi. Setelah semua sudah terisi, kemudian dilanjutkan dengan menghitung *entropy*. Berikut contoh perhitungan *entropy* pada atribut suhu “dingin” :

$$Entropy(\text{suhu dingin}) = - \left(\left(\frac{\text{dingin ya}}{\text{dingin jumlah}} * \text{LOG}\left(\frac{\text{dingin ya}}{\text{dingin jumlah}}; 2\right) \right) + \left(\frac{\text{dingin tidak}}{\text{dingin jumlah}} * \text{LOG}\left(\frac{\text{dingin tidak}}{\text{dingin jumlah}}; 2\right) \right) \right)$$

$$Entropy(\text{suhu dingin}) = - \left(\left(\frac{1}{1} * \text{LOG}\left(\frac{1}{1}; 2\right) \right) + \left(\frac{0}{1} * \text{LOG}\left(\frac{0}{1}; 2\right) \right) \right)$$

$$Entropy(\text{suhu dingin}) = 0$$

Tabel 2.6. *Entropy Node 2* Cuaca Cerah (Contoh Perhitungan)

| Node | Atribut | Detail | Jumlah | Ya | Tidak | Entropy | Gain |
|------|--------------|---------|----------|----------|----------|---------------|------|
| 2 | Total | | 5 | 2 | 3 | 0,9710 | |
| | Suhu | Dingin | 1 | 1 | 0 | 0,0000 | |
| | | Lembut | 2 | 1 | 1 | 1,0000 | |
| | | Panas | 2 | 0 | 2 | 0,0000 | |
| | Kelembapan | Normal | 2 | 2 | 0 | 0,0000 | |
| | | Tinggi | 3 | 0 | 3 | 0,0000 | |
| | Angin | Kencang | 2 | 1 | 1 | 1,0000 | |
| | | Pelan | 3 | 1 | 2 | 0,9183 | |

Hitung semua *entropy* tiap atribut, lalu dilanjutkan ke perhitungan *gain*. Berikut contoh perhitungan *gain* pada atribut “suhu” :

$$\text{Gain (suhu)} = \text{entropy total} - ((\text{dingin jumlah}/\text{total jumlah} * \text{entropy dingin}) + (\text{lembut jumlah}/\text{total jumlah} * \text{entropy lembut}) + (\text{panas jumlah}/\text{total jumlah} * \text{entropy panas}))$$

$$\text{Gain (suhu)} = 0,970951 - ((1/5*0) + (2/5*1) + (2/5*0))$$

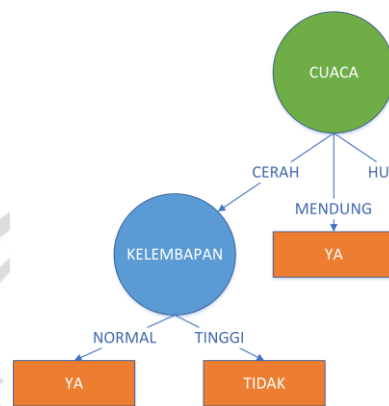
$$\text{Gain (suhu)} = 0,570951$$

Tabel 2.7. *Gain Node 2* Cuaca Cerah (Contoh Perhitungan)

| Node | Atribut | Detail | Jumlah | Ya | Tidak | Entropy | Gain |
|------|--------------|---------|----------|----------|----------|---------------|---------------|
| 2 | Total | | 5 | 2 | 3 | 0,9710 | |
| | Suhu | Dingin | 1 | 1 | 0 | 0,0000 | 0,5710 |
| | | Lembut | 2 | 1 | 1 | 1,0000 | |
| | | Panas | 2 | 0 | 2 | 0,0000 | |
| | Kelembapan | Normal | 2 | 2 | 0 | 0,0000 | 0,9710 |
| | | Tinggi | 3 | 0 | 3 | 0,0000 | |
| | Angin | Kencang | 2 | 1 | 1 | 1,0000 | 0,0200 |
| | | Pelan | 3 | 1 | 2 | 0,9183 | |

Hitung semua *gain* tiap atribut, lalu dilanjutkan ke pemilihan *gain* tertinggi. Pada tabel 2.7 diketahui nilai *gain* tertinggi dimiliki oleh atribut “kelembapan”, dengan kualifikasi “normal” dan “tinggi” memiliki *entropy* 0. Karena kedua kualifikasi

tersebut memiliki nilai *entropy* 0, sehingga keduanya menghasilkan *leaf node*. Kualifikasi “normal” dengan hasil “ya” 2 data dan hasil “tidak” 0 data menghasilkan *leaf node* “ya”. Kualifikasi “tinggi” dengan hasil “ya” 0 data dan hasil “tidak” 3 data menghasilkan *leaf node* “tidak”.



Gambar 2.3. Pohon Keputusan *Node 2* Kelembapan (Contoh Perhitungan)

Gambar 2.3 adalah pohon keputusan sementara berdasarkan hasil perhitungan *node 2* atribut “cuaca” kualifikasi “cerah”. Kemudian dilanjutkan dengan menghitung *node 2* berdasarkan atribut “cuaca” kualifikasi “hujan”.

Tabel 2.8. Data *Node 2* Cuaca Hujan (Contoh Perhitungan)

| Node | Atribut | Detail | Jumlah | Ya | Tidak | Entropy | Gain |
|------|--------------|---------|----------|----------|----------|---------|------|
| 2 | Total | | 5 | 3 | 2 | | |
| | Suhu | Dingin | 2 | 1 | 1 | | |
| | | Lembut | 3 | 2 | 1 | | |
| | | Panas | 0 | 0 | 0 | | |
| | Kelembapan | Normal | 3 | 2 | 1 | | |
| | | Tinggi | 2 | 1 | 1 | | |
| | Angin | Kencang | 2 | 0 | 2 | | |
| | | Pelan | 3 | 3 | 0 | | |

Tabel 2.8 berisikan data atribut yang sudah dikelompokkan dan dihitung jumlahnya sesuai dengan atribut “cuaca” kualifikasi “hujan”. Baris “total” di kolom “jumlah”, diisi dengan total data atribut “cuaca” = “hujan” yaitu 5 data, baris “total” di kolom “ya” diisi dengan total data atribut “cuaca” = “hujan” yang memiliki atribut “bermain” = “ya” yaitu 3 data, dan baris “total” di kolom “tidak” diisi

dengan total data atribut “cuaca” = “hujan” yang memiliki atribut “bermain” = “tidak” yaitu 2 data. Kemudian dilanjutkan menghitung atribut lainnya hingga semuanya terisi. Setelah semua sudah terisi, kemudian dilanjutkan dengan menghitung *entropy*. Berikut contoh perhitungan *entropy* pada atribut suhu “dingin” :

$$Entropy (\text{suhu dingin}) = - ((\text{dingin ya} / \text{dingin jumlah} * \text{LOG}(\text{dingin ya} / \text{dingin jumlah}; 2)) + (\text{dingin tidak} / \text{dingin jumlah} * \text{LOG}(\text{dingin tidak} / \text{dingin jumlah}; 2)))$$

$$Entropy (\text{suhu dingin}) = - ((1 / 2 * \text{LOG}(1 / 2 ; 2)) + (1 / 2 * \text{LOG}(1 / 2 ; 2)))$$

$$Entropy (\text{suhu dingin}) = 1$$

Tabel 2.9. *Entropy Node 2* Cuaca Hujan (Contoh Perhitungan)

| Node | Atribut | Detail | Jumlah | Ya | Tidak | Entropy | Gain |
|------|--------------|---------|----------|----------|----------|---------------|------|
| 2 | Total | | 5 | 3 | 2 | 0,9710 | |
| | Suhu | Dingin | 2 | 1 | 1 | 1,0000 | |
| | | Lembut | 3 | 2 | 1 | 0,9183 | |
| | | Panas | 0 | 0 | 0 | 0,0000 | |
| | Kelembapan | Normal | 3 | 2 | 1 | 0,9183 | |
| | | Tinggi | 2 | 1 | 1 | 1,0000 | |
| | Angin | Kencang | 2 | 0 | 2 | 0,0000 | |
| | | Pelan | 3 | 3 | 0 | 0,0000 | |

Lalu dilanjutkan ke perhitungan *gain*. *Gain* dapat dihitung jika semua atribut sudah memiliki nilai *entropy*. Berikut contoh perhitungan *gain* pada atribut “suhu” :

$$Gain (\text{suhu}) = entropy \text{ total} - ((\text{dingin jumlah} / \text{total jumlah} * entropy \text{ dingin}) + (\text{lembut jumlah} / \text{total jumlah} * entropy \text{ lembut}) + (\text{panas jumlah} / \text{total jumlah} * entropy \text{ panas}))$$

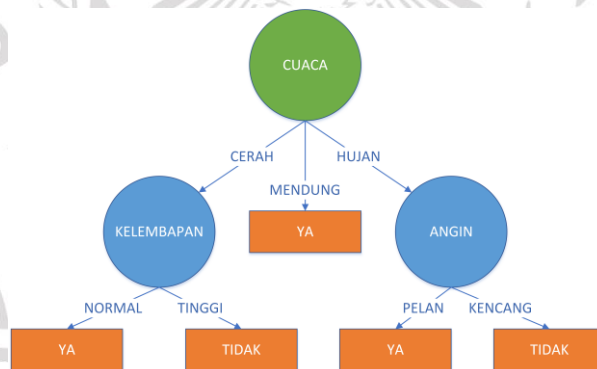
$$Gain (\text{suhu}) = 0,970951 - ((2/5 * 1) + (3/5 * 0,918296) + (0/5 * 0))$$

$$Gain (\text{suhu}) = 0,019973$$

Tabel 2.10. Gain Node 2 Cuaca Hujan (Contoh Perhitungan)

| Node | Atribut | Detail | Jumlah | Ya | Tidak | Entropy | Gain |
|------|------------|---------|--------|----|-------|---------|--------|
| 2 | Total | | 5 | 3 | 2 | 0,9710 | |
| | Suhu | Dingin | 2 | 1 | 1 | 1,0000 | 0,0200 |
| | | Lembut | 3 | 2 | 1 | 0,9183 | |
| | | Panas | 0 | 0 | 0 | 0,0000 | |
| | Kelembapan | Normal | 3 | 2 | 1 | 0,9183 | 0,0200 |
| | | Tinggi | 2 | 1 | 1 | 1,0000 | |
| | Angin | Kencang | 2 | 0 | 2 | 0,0000 | 0,9710 |
| | | Pelan | 3 | 3 | 0 | 0,0000 | |

Hitung semua *gain* tiap atribut, lalu dilanjutkan ke pemilihan *gain* tertinggi. Pada tabel 2.10 diketahui nilai *gain* tertinggi dimiliki oleh atribut “angin”, dengan kualifikasi “kencang” dan “pelan” memiliki *entropy* 0. Karena kedua kualifikasi tersebut memiliki nilai *entropy* 0, sehingga keduanya menghasilkan *leaf node*. Kualifikasi “kencang” dengan hasil “ya” 0 data dan hasil “tidak” 2 data menghasilkan *leaf node* “tidak”. Kualifikasi “pelan” dengan hasil “ya” 3 data dan hasil “tidak” 0 data menghasilkan *leaf node* “ya”.

**Gambar 2.4.** Pohon Keputusan Node 2 Angin (Contoh Perhitungan)

Gambar 2.4 adalah pohon keputusan berdasarkan hasil perhitungan *node* 2 atribut “cuaca” kualifikasi “hujan”. Karena tiap *node* telah memiliki *leaf node*, maka perhitungan *node* tidak dilanjutkan dan gambar 2.4 menjadi pohon keputusan akhir.

Pengujian *decision tree* dilakukan dengan menelusuri tiap data uji melalui pohon keputusan. Contoh pohon keputusan pada gambar 2.4, memiliki *node* akar cuaca, maka atribut cuaca yang ditelusuri pertama kali, kemudian dilanjutkan ke *node* lainnya.

Tabel 2.11. Contoh data uji *decision tree*

| Cuaca | Suhu | Kelembapan | Angin | Hasil Uji (Bermain ?) |
|-------|--------|------------|-------|-----------------------|
| Cerah | Panas | Tinggi | Pelan | Tidak |
| Hujan | Dingin | Normal | Pelan | Ya |

Pada perhitungan *node* metode *decision tree*, dapat juga ditemukan beberapa kondisi yang membuat *leaf node* bernilai kosong, untuk mengatasi hal tersebut dapat dilakukan dengan cara berikut ini :

1. Jumlah data detail atribut terpilih suatu *node* bernilai 0

Tabel 2.12. Contoh kondisi tertentu *decision tree* (1)

| Node | Atribut | Detail | Jumlah | Ya | Tidak | Entropy | Gain |
|------|--------------|--------|----------|----------|----------|---------------|--------|
| 1 | Total | | 5 | 3 | 2 | 0,9710 | |
| | Suhu | Dingin | 2 | 1 | 1 | 1,0000 | 0,0200 |
| | | Lembut | 3 | 2 | 1 | 0,9183 | |
| | | Panas | 0 | 0 | 0 | 0,0000 | |

Pada tabel 2.11, atribut suhu “panas” memiliki jumlah data 0, sehingga untuk menentukan *leaf node* diambil dari jumlah data mayoritas dari *node* tersebut. Dari 5 data, kelas “Ya” menjadi mayoritas data dengan 3 data. Sehingga atribut suhu “panas” memiliki *leaf node* “Ya”.

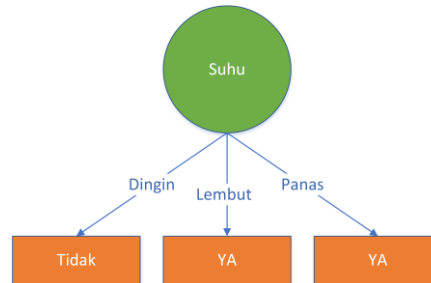
2. Jumlah data detail atribut terpilih suatu *node* bernilai sama

Tabel 2.13. Contoh kondisi tertentu *decision tree* (2)

| Node | Atribut | Detail | Jumlah | Ya | Tidak | Entropy | Gain |
|------|--------------|--------|----------|----------|----------|---------------|--------|
| 1 | Total | | 5 | 3 | 2 | 0,9710 | |
| | Suhu | Dingin | 2 | 1 | 1 | 1,0000 | 0,0200 |
| | | Lembut | 3 | 2 | 1 | 0,9183 | |
| | | Panas | 0 | 0 | 0 | 0,0000 | |

Pada tabel 2.12, atribut suhu “dingin” memiliki jumlah data 2 dengan kelas “Ya” 1 data dan kelas “Tidak” 1 data, untuk menentukan *leaf node* dapat diambil dari kelas apapun menurut pertimbangan masing-masing. Apabila memilih kelas “Tidak”, maka atribut suhu “dingin” memiliki *leaf node* “Tidak”.

Berdasarkan dua kondisi tersebut, dapat digambarkan pohon keputusan sebagai berikut.



Gambar 2.5. Contoh pohon keputusan (kondisi tertentu)

2.7. Penelitian Sebelumnya

Penulis melakukan analisis dari hasil penelitian sebelumnya untuk meningkatkan topik penelitian. Berikut merupakan hasil dari penelitian sebelumnya:

- a. Budi Setiadi, Bambang Lareno (2015) dengan judul “*Penerapan Algoritma Decision Tree Untuk Penilaian Agunan Pengajuan Kredit*”. Penelitian ini membahas tentang evaluasi metode *decision tree* untuk penilaian agunan pengajuan kredit. Data yang digunakan bersumber dari lembaga penilai kredit Kota Banjarmasin dengan atribut kategori wilayah, luas tanah, harga tanah, luas bangunan dan harga bangunan. Data tersebut akan diproses dengan metode *decision tree* C45 lalu dibandingkan hasilnya dengan data yang diproses menggunakan metode *k-nearest neighbor* dan *naive bayes*. Hasil dari penelitian menunjukkan bahwa metode *decision tree* dapat diterapkan dalam penilaian agunan kredit dengan akurasi 71%. *Decision tree* memprediksi lebih akurat dibandingkan *k-nearest neighbor* dengan akurasi 63,67% dan *naive bayes* dengan akurasi 61,67%, sehingga mampu memberikan solusi bagi pihak terkait dalam membuat penilaian agunan pengajuan kredit.
- b. Nurdiana Handayani, Herry Wahyono, Joko Trianto, Dwi Sidik Permana (2021) dengan judul “*Prediksi Tingkat Risiko Kredit dengan Data Mining Menggunakan Algoritma Decision Tree C.45*”. Penelitian ini membahas tentang analisa prediksi tingkat resiko kredit pada nasabah PT WOM Finance. Dalam proses analisa prediksi, penelitian ini menggunakan metode *decision*

tree C45 dengan 1.153 data training dan 91 data uji. Atribut data yang digunakan diantaranya jenis kelamin, status perkawinan, tenor, pekerjaan, pinjaman dan penghasilan. Proses analisa dijalankan melalui aplikasi WEKA milik perusahaan dengan hasil keterangan “lancar” dan “tidak lancar”. Hasil dari penelitian menunjukkan tingkat akurasi metode *decision tree* C45 mencapai 74% untuk pemakaian 6 atribut dan 79% untuk pemakaian 5 atribut. Penelitian ini juga menyimpulkan bahwa pemilihan atribut sangat mempengaruhi dalam pengolahan metode C45 karena keputusan bergantung pada atribut yang dipilih.

- c. Habibah, Khairuddin Nasution, Tasliyah Haramaini (2022) dengan judul “*Penerapan Iterative Dichotomizer 3 pada Aplikasi Penerima Bantuan Langsung Tunai untuk Pedagang Kaki Lima dan Warung pada Masyarakat Kecamatan Padang Sidimpuan Angkola Julu*”. Penelitian ini membahas tentang sistem pendukung keputusan yang dapat membantu dalam menentukan penerima BLT (Bantuan Langsung Tunai). Metode yang digunakan dalam penelitian ini adalah *decision tree* ID3. Data penelitian bersumber dari data pedagang kaki lima dan warung di Kecamatan Padang Sidimpuan Angkola Julu. Atribut yang digunakan diantaranya jumlah penghasilan, jumlah tanggungan, aset kepemilikan dan jenis usaha. Hasil dari penelitian ini menyimpulkan bahwa dalam penerapan ID3 pada aplikasi penerima BLT, membutuhkan *dataset* dan data peserta sehingga dapat menghasilkan pohon keputusan yang berfungsi sebagai *rules* dalam menentukan penerima BLT.
- d. Siti Nurul Afiah, Wahyu Dini Aula Nabila (2021) dengan judul “*Implementasi Iterative Dichotomiser 3 (ID3) untuk Penentuan Kelayakan Pemberian Kredit pada PT.BPR Ploso Saranaartha Jombang*”. Penelitian ini membahas tentang implementasi metode *decision tree* ID3 dalam penentuan kelayakan pemberian kredit di PT. BPR Ploso Saranaartha Jombang. Penelitian ini dilakukan karena belum adanya prosedur secara sistematis dalam menentukan kelayakan nasabah, selama ini perusahaan masih menggunakan metode seleksi debitur dalam menentukan kelayakan nasabah. Data yang digunakan sebanyak 300 data *training* yang terdiri dari 272 data dengan

kolektibilitas lancar dan 28 data dengan kolektibilitas macet. Untuk pengujian, digunakan 20 data uji. Dalam proses pengujian, sistem melakukan *preprocessing* dalam menentukan atribut, yaitu dengan memilih atribut yang digunakan dan tidak digunakan. Sehingga pada tiap pengujian, atribut yang digunakan dapat berbeda-beda. Hasil dari penelitian ini menunjukkan tingkat akurasi pengujian data mencapai 88,51%, sehingga disimpulkan bahwa metode ID3 dapat diterapkan dalam penentuan kelayakan pengajuan kredit.

- e. David, Erlin Elisa (2022) dengan judul “*Analisis Kelayakan Menerima Pinjaman Kredit dengan Algoritma C4.5 Pada PT.BPR Buana Arta Mulia*”. Penelitian ini membahas tentang analisis kelayakan menerima pinjaman kredit di PT. BPR Buana Arta Mulia. Metode yang digunakan adalah *decision tree* C45. Penelitian bertujuan untuk memperbaiki kualitas hasil dari proses analisa kredit yang selama ini digunakan. Data yang digunakan terdiri dari data kredit perusahaan mulai Januari sampai September 2021, dengan data ujinya sebanyak 40 data. Atribut yang digunakan diantaranya pendapatan, jumlah kredit, tujuan kredit dan nilai jaminan. Hasil dari penelitian ini menunjukkan bahwa atribut nilai jaminan dan pendapatan sebagai atribut yang paling berpengaruh dalam penentuan kelayakan kredit, yang pada akhirnya membuktikan bahwa metode *decision tree* C45 cukup akurat untuk penentuan kelayakan kredit.