

# MATERI 6 - ALGORITMA

## ITERATIVE DICHOTOMIZER THREE (ID3)

### PENGERTIAN

ID3 singkatan dari *Iterative Dichotomiser Three*. Ada juga yang menyebut *Induction of Decision Tree*. ID3 adalah suatu algooritma matematika yang digunakan untuk menghasilkan suatu pohon keputusan yang mampu mengklasifikasi suatu obyek. ID3 diperkenalkan pertama kali oleh Ross Quinlan (1979). ID3 merepresentasi konsep-konsep dalam bentuk pohon keputusan. Aturan-aturan yang dihasilkan oleh ID3 mempunyai relasi yang hirarkis seperti suatu pohon (mempunyai akar, titik, cabang, dan daun). Beberapa peneliti menyebut struktur model yang dihasilkan ID3 sebagai pohon keputusan (*decision tree*) sementara peneliti yang lain menyebutnya pohon aturan (*rule tree*).

### CONTOH KASUS

14 Minggu Permainan Tennis pada Setiap Sabtu Pagi

Minggu	Ramalan_Cuaca	Suhu	Kelembaban	Angin	Bermain_Tenis
M1	Cerah	Panas	Tinggi	Lemah	Tidak
M2	Cerah	Panas	Tinggi	Kuat	Tidak
M3	Mendung	Panas	Tinggi	Lemah	Ya
M4	Hujan	Sejuk	Tinggi	Lemah	Ya
M5	Hujan	Dingin	Normal	Lemah	Ya
M6	Hujan	Dingin	Normal	Kuat	Tidak
M7	Mendung	Dingin	Normal	Kuat	Ya
M8	Cerah	Sejuk	Tinggi	Lemah	Tidak
M9	Cerah	Dingin	Normal	Lemah	Ya
M10	Hujan	Sejuk	Normal	Lemah	Ya
M11	Cerah	Sejuk	Normal	Kuat	Ya
M12	Mendung	Sejuk	Tinggi	Kuat	Ya
M13	Mendung	Panas	Normal	Lemah	Ya
M14	Hujan	Sejuk	Tinggi	Kuat	Tidak

Atribut Tujuan adalah Bermain Tennis yang memiliki value ya atau tidak.  
Atribut adalah Ramalan\_Cuaca, Suhu, Kelembaban, dan Angin.

# ALGORITMA DAN FLOWCHART

Entropy adalah formula untuk menghitung homogenitas dari sebuah sample/contoh.  
Solusi menggunakan entropy dari contoh kasus di atas :

S adalah koleksi dari 14 contoh dengan 9 contoh positif dan 5 contoh negatif, ditulis dengan notasi [9+,5-].

Positif di sini maksudnya value Bermain\_Tenis = Ya sedangkan negatif sebaliknya.

Entropy dari S adalah :

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

$$p_i = \frac{Z_i}{N}$$

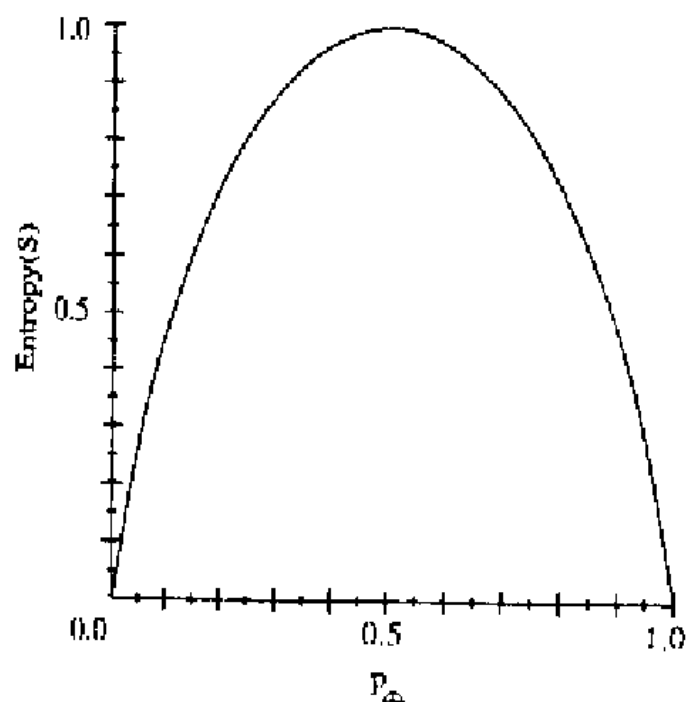
$Z_i$  = contoh positif + contoh negatif

$N$  = jumlah data

$$\begin{aligned} \text{Entropy}([9+,5-]) &= - (9/14) \log_2 (9/14) - (5/14) \log_2 (5/14) \\ &= - (0.6429) ((\log (9/14))/\log 2) - (0.3571) ((\log (5/14))/\log 2) \\ &= - (0.6429) (-0.1919/0.3010) - (0.3571) (-0.4472/0.3010) \\ &= - (0.6429) (-0.6375) - (0.3571) (-1.4857) \\ &= 0.4098 + 0.5305 \\ &= 0.94029 \end{aligned}$$

Catatan :

- Entropy(S) = 0, jika semua contoh pada S berada dalam kelas yang sama.
- Entropy(S) = 1, jika jumlah contoh positif dan jumlah contoh negatif dalam S adalah sama.
- $0 < \text{Entropy}(S) < 1$ , jika jumlah contoh positif dan jumlah contoh negatif dalam S tidak sama.



**Gain(S,A)** adalah *Information Gain* dari sebuah atribut A pada koleksi contoh S :

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

1. Values(Angin) = Lemah, Kuat

$$S_{\text{Lemah}} = [6+, 2-]$$

$$S_{\text{Kuat}} = [3+, 3-]$$

$$\begin{aligned} \text{Gain}(S, \text{Angin}) &= \text{Entropy}(S) - (8/14)\text{Entropy}(S_{\text{Lemah}}) - (6/14)\text{Entropy}(S_{\text{Kuat}}) \\ &= 0.94029 - (8/14)0.81128 - (6/14)1.0000 \\ &= 0.04813 \end{aligned}$$

2. Values(Kelembaban) = Tinggi, Normal

$$S_{\text{Tinggi}} = [3+, 4-]$$

$$S_{\text{Normal}} = [6+, 1-]$$

$$\begin{aligned} \text{Gain}(S, \text{Kelembaban}) &= \text{Entropy}(S) - (7/14)\text{Entropy}(S_{\text{Tinggi}}) - \\ &\quad (7/14)\text{Entropy}(S_{\text{Normal}}) \\ &= 0.94029 - (7/14)0.98523 - (7/14)0.59167 \\ &= 0.15184 \end{aligned}$$

3. Values(Suhu) = Panas, Sejuk, Dingin

$$S_{\text{Panas}} = [2+, 2-]$$

$$S_{\text{Sejuk}} = [4+, 2-]$$

$$S_{\text{Dingin}} = [3+, 1-]$$

$$\begin{aligned} \text{Gain}(S, \text{Suhu}) &= \text{Entropy}(S) - (4/14)\text{Entropy}(S_{\text{Panas}}) - \\ &\quad (6/14)\text{Entropy}(S_{\text{Sejuk}}) - (4/14)\text{Entropy}(S_{\text{Dingin}}) \\ &= 0.94029 - (4/14)1.00000 - (6/14)0.91830 - \\ &\quad (4/14)0.81128 \\ &= 0.02922 \end{aligned}$$

4. Values(Ramalan\_Cuaca) = Cerah, Mendung, Hujan

$$S_{\text{Cerah}} = [2+, 3-]$$

$$S_{\text{Mendung}} = [4+, 0-]$$

$$S_{\text{Hujan}} = [3+, 2-]$$

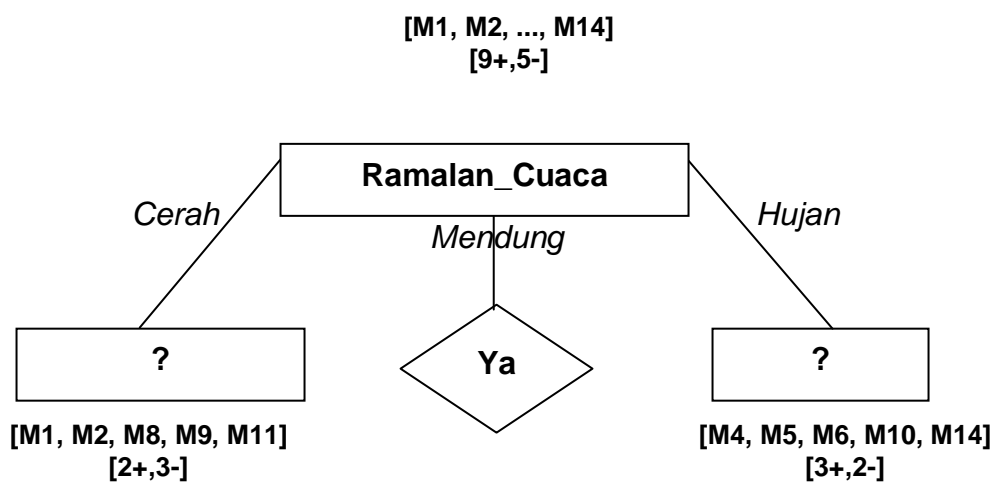
$$\begin{aligned} \text{Gain}(S, \text{Ramalan\_Cuaca}) &= \text{Entropy}(S) - (5/14)\text{Entropy}(S_{\text{Cerah}}) - \\ &\quad (4/14)\text{Entropy}(S_{\text{Mendung}}) - (5/14)\text{Entropy}(S_{\text{Hujan}}) \end{aligned}$$

$$\begin{aligned}
 &= 0.94029 - (5/14)0.97075 - (4/14)1.00000 - \\
 &\quad (5/14)0.97075 \\
 &= \underline{\underline{0.24675}}
 \end{aligned}$$

Jadi, information gain untuk 3 atribut yang ada adalah :

$$\begin{aligned}
 \text{Gain}(S,\text{Angin}) &= 0.04813 \\
 \text{Gain}(S,\text{Kelembaban}) &= 0.15184 \\
 \text{Gain}(S,\text{Suhu}) &= 0.02922 \\
 \text{Gain}(S,\text{Ramalan\_Cuaca}) &= \underline{\underline{0.24675}}
 \end{aligned}$$

Tampak bahwa attribute *Ramalan\_Cuaca* akan menyediakan prediksi terbaik untuk target attribute *Bermain\_Tenis*.



Untuk node cabang *Ramalan\_Cuaca* = *Cerah*,  
*S*<sub>Cerah</sub> = [M1, M2, M8, M9, M11]

Minggu	Ramalan_Cuaca	Suhu	Kelembaban	Angin	Bermain_Tenis
M1	Cerah	Panas	Tinggi	Lemah	Tidak
M2	Cerah	Panas	Tinggi	Kuat	Tidak
M8	Cerah	Sejuk	Tinggi	Lemah	Tidak
M9	Cerah	Dingin	Normal	Lemah	Ya
M11	Cerah	Sejuk	Normal	Kuat	Ya

$$\begin{aligned}
 1. \text{ Values}(\text{Suhu}) &= \text{Panas, Sejuk, Dingin} \\
 S_{\text{Panas}} &= [0+,2-] \\
 S_{\text{Sejuk}} &= [1+,1-] \\
 S_{\text{Dingin}} &= [1+,0-] \\
 \text{Gain}(S_{\text{Cerah}},\text{Suhu}) &= \text{Entropy}(S_{\text{Cerah}}) - (2/5)\text{Entropy}(S_{\text{Panas}}) - \\
 &\quad (2/5)\text{Entropy}(S_{\text{Sejuk}}) - (1/5)\text{Entropy}(S_{\text{Dingin}})
 \end{aligned}$$

$$\begin{aligned}
 &= 0.97075 - (2/5)0.00000 - (2/5)1.00000 - \\
 &\quad (1/5)0.00000 \\
 &= 0.57075
 \end{aligned}$$

2. Values(Kelembaban)

= Tinggi, Normal

$S_{Tinggi}$

= [0+,3-]

$S_{Normal}$

= [2+,0-]

Gain( $S_{Cerah}$ ,Kelembaban)

= **Entropy( $S_{Cerah}$ ) - (3/5)Entropy( $S_{Tinggi}$ ) -**

**(2/5)Entropy( $S_{Normal}$ )**

= 0.97075 - (3/5)0.00000 - (2/5)0.00000

= **0.97075**

3. Values(Angin)

= Lemah, Kuat

$S_{Lemah}$

= [1+,2-]

$S_{Kuat}$

= [1+,1-]

Gain( $S_{Cerah}$ ,Angin)

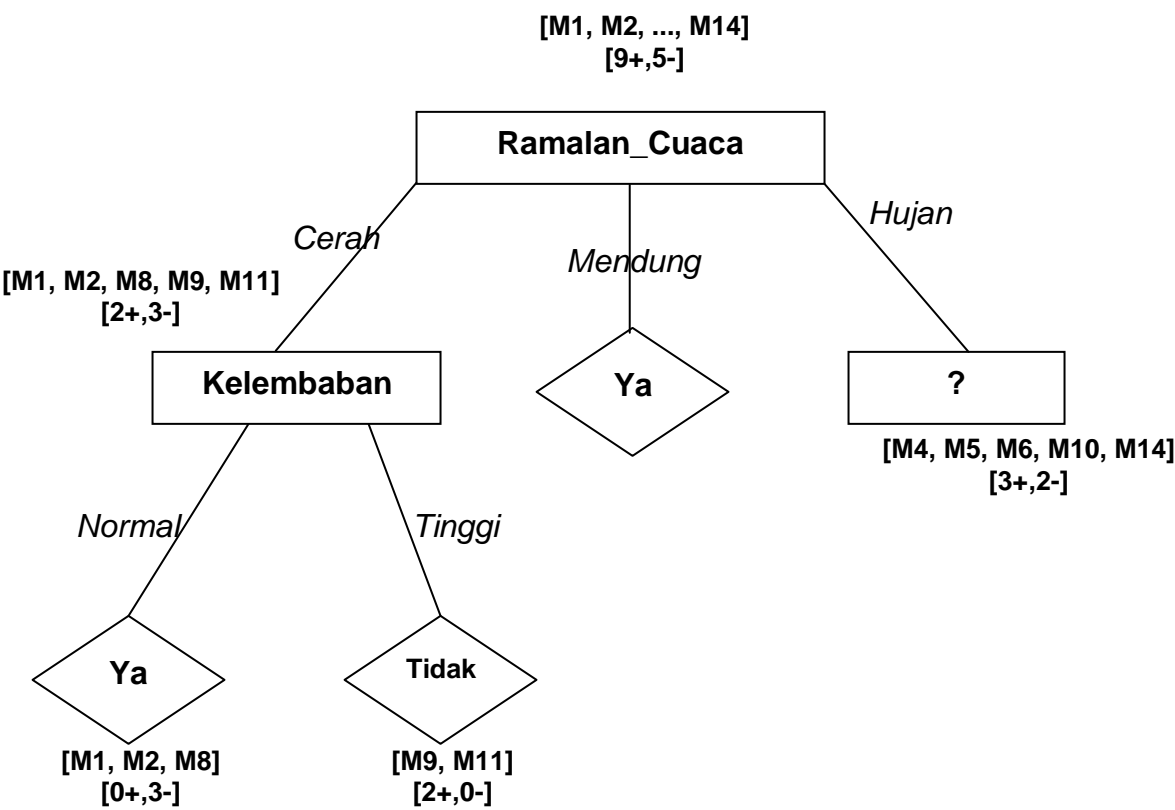
= **Entropy( $S_{Cerah}$ ) - (3/5)Entropy( $S_{Lemah}$ ) -**

**(2/5)Entropy( $S_{Kuat}$ )**

= 0.97075 - (3/5)0.91830 - (2/5)1.00000

= 0.01997

Atribut *Kelembaban* menyediakan prediksi terbaik pada level ini.



Untuk node cabang Ramalan\_Cuaca = Hujan,  
 $S_{Hujan} = [M4, M5, M6, M10, M14]$

Minggu	Ramalan_Cuaca	Suhu	Kelembaban	Angin	Bermain_Tenis
M4	Hujan	Sejuk	Tinggi	Lemah	Ya
M5	Hujan	Dingin	Normal	Lemah	Ya
M6	Hujan	Dingin	Normal	Kuat	Tidak
M10	Hujan	Sejuk	Normal	Lemah	Ya
M14	Hujan	Sejuk	Tinggi	Kuat	Tidak

1. Values(Suhu) = Sejuk, Dingin (Tidak ada **suhu = panas** saat ini)

$S_{Sejuk} = [2+, 1-]$

$S_{Dingin} = [1+, 1-]$

$Gain(S_{Hujan}, Suhu) = Entropy(S_{Hujan}) - (3/5)Entropy(S_{Sejuk}) - (2/5)Entropy(S_{Dingin})$

$= 0.97075 - (3/5)0.91830 - (2/5)1.00000$

$= 0.01997$

2. Values(Kelembaban) = Tinggi, Normal

$S_{Tinggi} = [1+, 1-]$

$S_{Normal} = [2+, 1-]$

$Gain(S_{Hujan}, Kelembaban) = Entropy(S_{Hujan}) - (2/5)Entropy(S_{Tinggi}) - (3/5)Entropy(S_{Normal})$

$= 0.97075 - (2/5)1.00000 - (3/5)0.91830$

$= 0.01997$

3. Values(Angin) = Lemah, Kuat

$S_{Lemah} = [3+, 0-]$

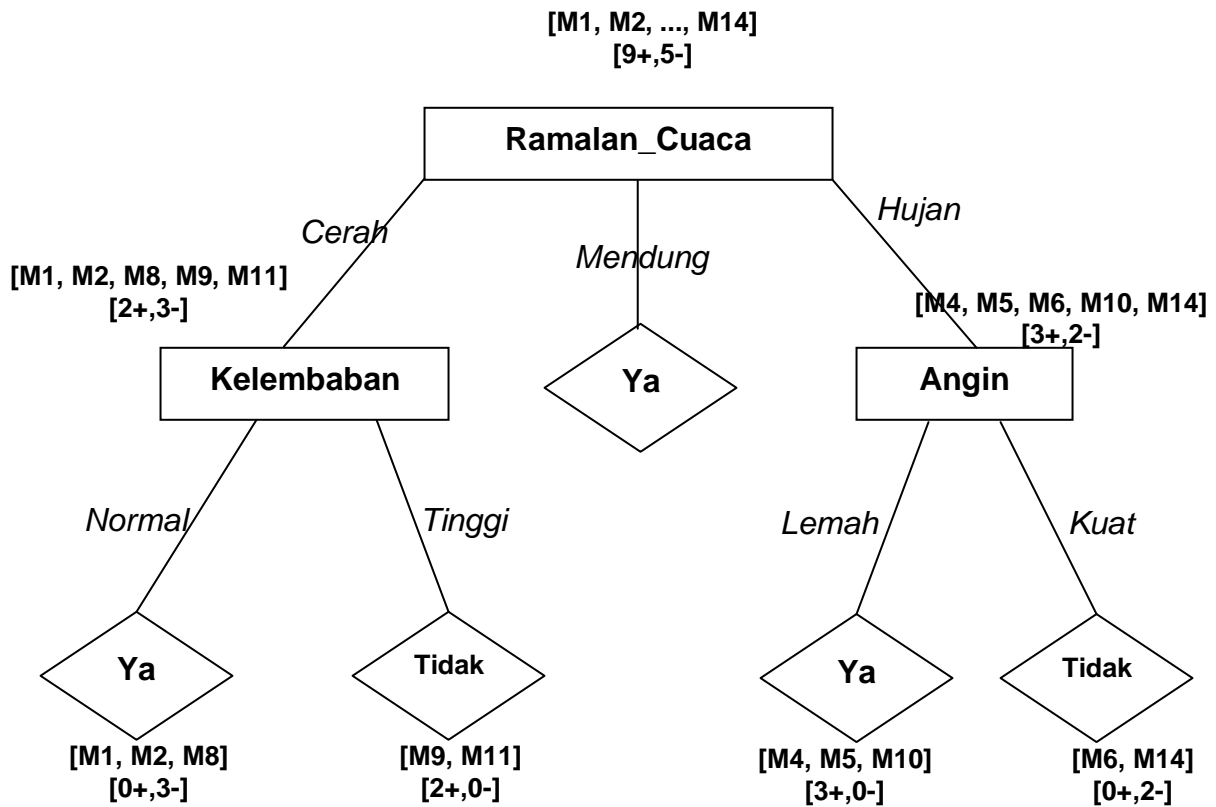
$S_{Kuat} = [0+, 2-]$

$Gain(S_{Hujan}, Angin) = Entropy(S_{Hujan}) - (3/5)Entropy(S_{Lemah}) - (2/5)Entropy(S_{Kuat})$

$= 0.97075 - (3/5)0.00000 - (2/5)0.00000$

$= \underline{\underline{0.97075}}$

Atribut *Angin* menyediakan prediksi terbaik pada level ini.



Algoritma :

**If Ramalan\_Cuaca = Cerah AND Kelembaban = Tinggi THEN  
Bermain\_Tenis = Tidak**

**If Ramalan\_Cuaca = Cerah AND Kelembaban = Normal THEN  
Bermain\_Tenis = Ya**

**If Ramalan\_Cuaca = Mendung THEN Bermain\_Tenis = Ya**

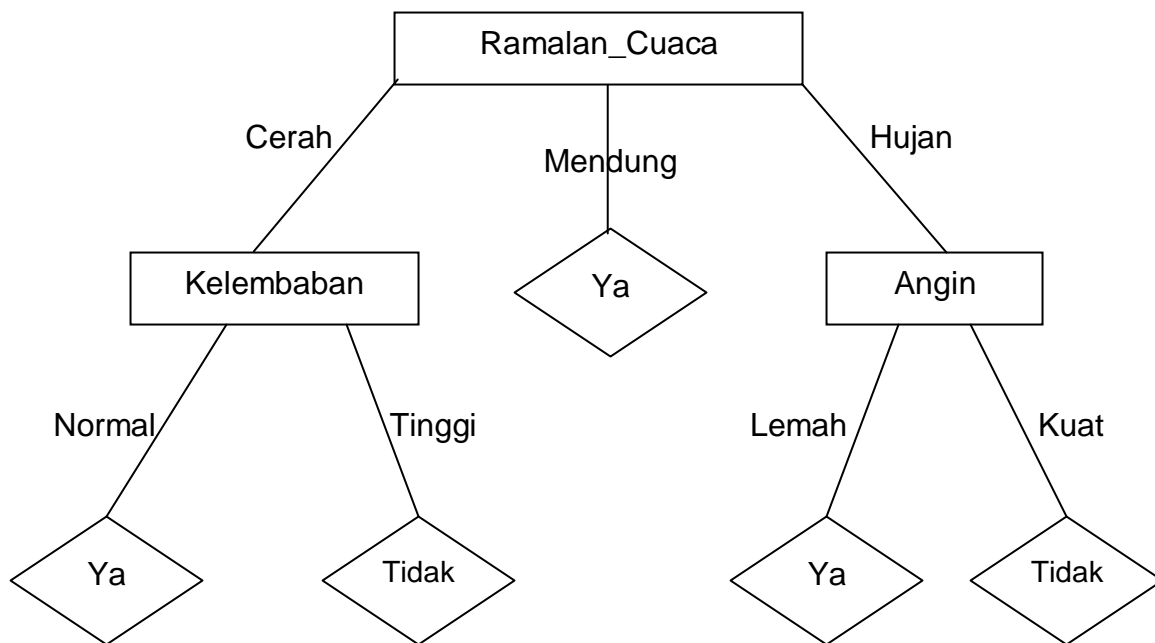
**If Ramalan\_Cuaca = Hujan AND Angin = Kuat THEN  
Bermain\_Tenis = Tidak**

**If Ramalan\_Cuaca = Hujan AND Angin = Lemah THEN  
Bermain\_Tenis = Ya**

Ramalan_Cuaca	Suhu	Kelembaban	Angin	Bermain_Tenis
Cerah	Panas	Tinggi	Kuat	Tidak
Cerah	Panas	Tinggi	Lemah	Tidak
Cerah	Panas	Normal	Kuat	Ya
Cerah	Panas	Normal	Lemah	Ya
Cerah	Sejuk	Tinggi	Kuat	Tidak
Cerah	Sejuk	Tinggi	Lemah	Tidak
Cerah	Sejuk	Normal	Kuat	Ya
Cerah	Sejuk	Normal	Lemah	Ya
Cerah	Dingin	Tinggi	Kuat	Tidak
Cerah	Dingin	Tinggi	Lemah	Tidak
Cerah	Dingin	Normal	Kuat	Ya
Cerah	Dingin	Normal	Lemah	Ya
Mendung	Panas	Tinggi	Kuat	Ya
Mendung	Panas	Tinggi	Lemah	Ya
Mendung	Panas	Normal	Kuat	Ya
Mendung	Panas	Normal	Lemah	Ya
Mendung	Sejuk	Tinggi	Kuat	Ya
Mendung	Sejuk	Tinggi	Lemah	Ya
Mendung	Sejuk	Normal	Kuat	Ya
Mendung	Sejuk	Normal	Lemah	Ya
Mendung	Dingin	Tinggi	Kuat	Ya
Mendung	Dingin	Tinggi	Lemah	Ya
Mendung	Dingin	Normal	Kuat	Ya
Mendung	Dingin	Normal	Lemah	Ya
Hujan	Sejuk	Tinggi	Kuat	Tidak
Hujan	Sejuk	Tinggi	Lemah	Ya
Hujan	Sejuk	Normal	Kuat	Tidak
Hujan	Sejuk	Normal	Lemah	Ya
Hujan	Dingin	Tinggi	Kuat	Tidak
Hujan	Dingin	Tinggi	Lemah	Ya
Hujan	Dingin	Normal	Kuat	Tidak
Hujan	Dingin	Normal	Lemah	Ya



Flowchart :



#### A. KEUNTUNGAN DAN KERUGIAN ID3

Keuntungan :

- Dapat membuat aturan prediksi yang mudah dimengerti.
- Membangun pohon keputusan dengan cepat.
- Membangun pohon keputusan yang pendek.
- Hanya membutuhkan beberapa tes atribut hingga semua data diklasifikasikan.

Kerugian :

- Jika contoh yang diteliti terlalu kecil / sederhana mungkin membuat data *over-classified*
- Hanya satu atribut yang dapat dites dalam satu waktu untuk membuat keputusan.
- Mengelompokkan data yang berkelanjutan mungkin terhitung mahal, sebanyak pohon yang harus dibuat untuk melihat dimana menghentikan proses kelanjutannya.

Example  
output :

```
D:\UKSW~1\TR938D~1\AI\PROGRAM\AI.EXE
Selamat Datang di Penentuan Bermain Tennis!!!

Apakah ramalan cuaca hari ini?
1. Cerah
2. Mendung
3. Hujan
Jawaban Anda : 1

Bagaimana suhunya?
1. Panas
2. Sejuk
3. Dingin
Jawaban Anda : 1

Bagaimana keadaan kelembabannya?
1. Tinggi
2. Normal
Jawaban Anda : 1

Bagaimana keadaan anginnya?
1. Kuat
2. Lemah
Jawaban Anda : 1

Kesimpulan : Permainan tenis DITIADAKAN!

Anda ingin mengulang? (y/t)
< ☐
```

## Referensi

Manongga, Danny (2005). *Teori&Aplikasi Iterative Dichotomizer Three dalam Pembelajaran Mesin*. Fakultas Teknologi Informasi UKSW, Salatiga.

<http://www.hansmichael.com/download/diktatid3.pdf>