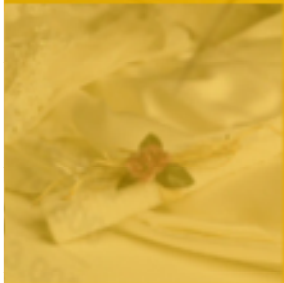


Algoritma Klasifikasi : Decision Tree

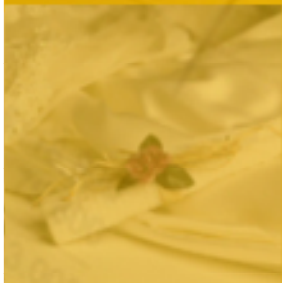
Tessy Badriyah, SKom. MT.



PENS - ITS

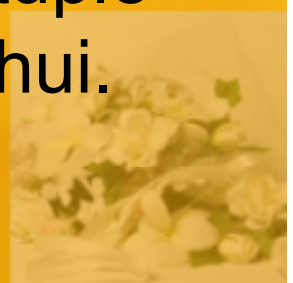
KLASIFIKASI

- Hal-hal yang berhubungan dengan klasifikasi adalah :
 - Meramalkan kategori label kelas (nominal atau terpisah)
 - Menggolongkan data (membangun suatu model) yang didasarkan pada pelatihan menetapkan dan nilai-nilai (label kelas) di (dalam) suatu penggolongan atribut dan penggunaan [itu] di (dalam) penggolongan data baru
- Sedangkan aplikasi umum untuk Klasifikasi adalah :
 - Persetujuan kredit
 - Target marketing
 - Diagnosa medis
 - Analisis keefektifan tindakan



ALGORITMA KLASIFIKASI

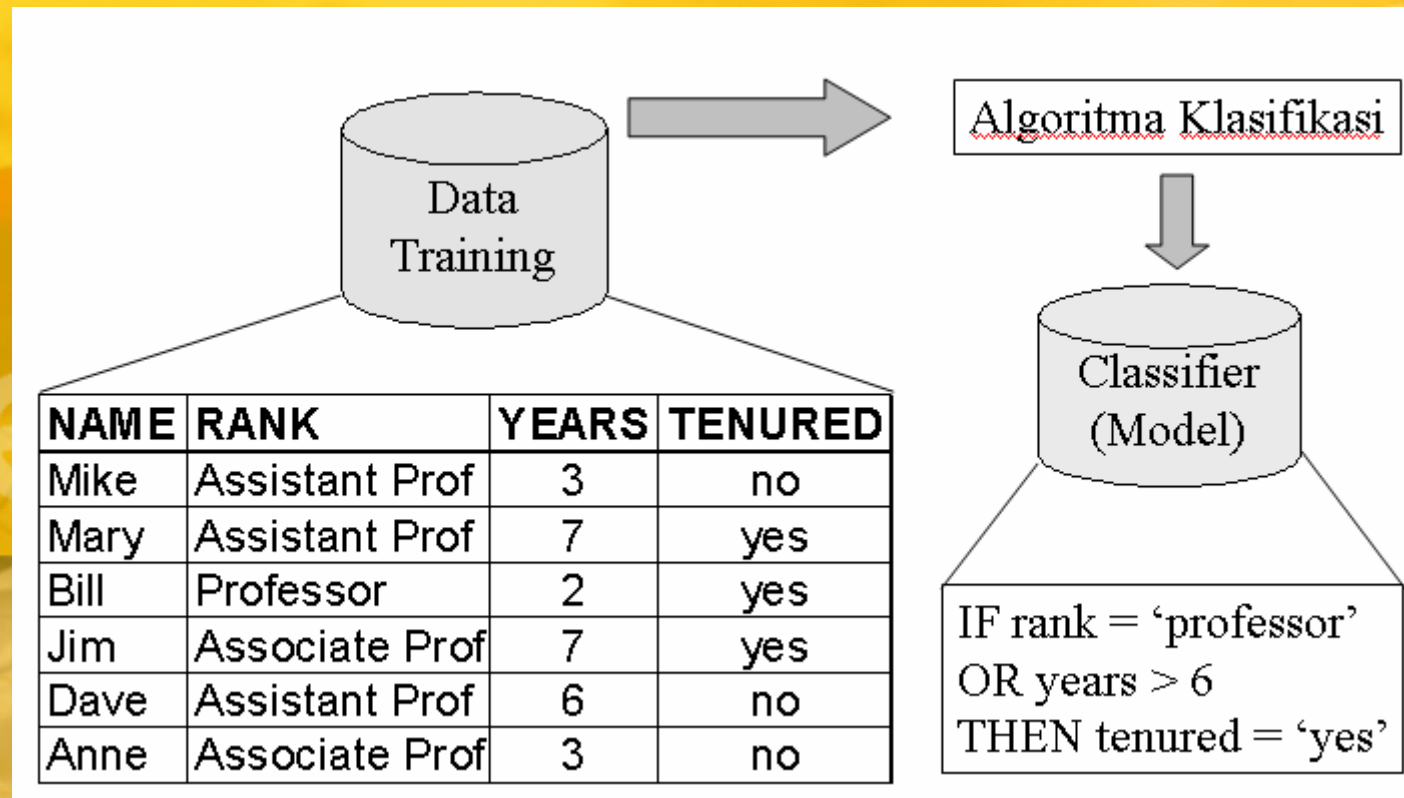
- Tahapan dalam algoritma Klasifikasi :
 - Konstruksi model: menguraikan suatu himpunan kelas yang ditentukan sebelumnya
 - Penggunaan model : setelah dibuat, model digunakan untuk mengklasifikasikan tuple data yang label kelasnya tidak diketahui.



Konstruksi model

- Setiap tuple/sample dimisalkan masuk kedalam suatu kelas yang didefinisikan sebelumnya, seperti yang ditetapkan melalui label atribut kelas.
- Himpunan dari tuple yang digunakan untuk konstruksi model dinamai himpunan pelatihan.
- Model disajikan sebagai kaidah klasifikasi, pohon keputusan, atau rumus matematika

Konstruksi model



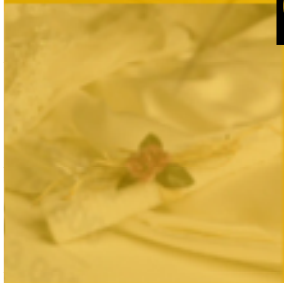
Contoh Algoritma Klasifikasi

- Algoritma klasifikasi yang sudah umum digunakan antara lain :
 - *Decision tree.*
 - *Bayesian Network.*
 - *Adaptive Bayesian network.*
 - *Naive Bayes dan sebagainya.*



DECISION TREE

- Decision tree adalah salah satu metode klasifikasi yang paling populer karena mudah diinterpretasikan oleh manusia. Decision tree digunakan untuk pengenalan pola dan termasuk dalam pengenalan pola secara statistik.



Struktur dari Decision Tree

- Decision tree dibentuk dari 3 tipe dari simpul: simpul root, simpul perantara, dan simpul leaf.
 - Simpul leaf memuat suatu keputusan akhir atau kelas target untuk suatu pohon keputusan.
 - Simpul root adalah titik awal dari suatu decision tree.
 - Setiap simpul perantara berhubungan dengan suatu pertanyaan atau pengujian

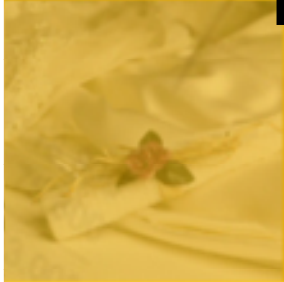


Algoritma Untuk Induksi Decision Tree

- Pohon dibangun dalam suatu metoda rekursif topdown divide and-conquer.
 - Seluruh contoh pelatihan dimulai dari simpul root, lalu dilakukan pengujian
 - Mencabang ke jalur yang benar berdasarkan hasil pengujian.
 - Apakah simpul leaf ditemukan? Jika yes, masukkan contoh ini ke kelas target, jika tidak kembali ke langkah 1.
- Atribut-atribut berada dalam suatu kategori (jika bernilai kontinu, nilai-nilai tersebut didiskritkan terlebih dahulu)
- Contoh-contoh dipartisi secara rekursif berdasarkan atribut terpilih
- Atribut-atribut uji dipilih berdasarkan heuristik atau pengukuran statistik (misal, information gain).

Contoh Pembentukan Tree (1)

- Untuk lebih memperjelas bagaimana decision tree menyusun dan menentukan node awal, berikut ini akan diberikan contoh secara detail. Contoh yang diberikan ini berhubungan dengan data kondisi cuaca untuk bermain golf.



Contoh Pembentukan Tree (2)

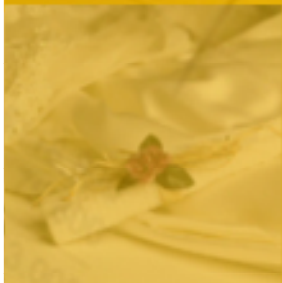
No	Outlook	Temperature	Humidity	Windy	Play/Don't play
1	sunny	hot	high	false	don't play
2	sunny	hot	high	true	don't play
3	overcast	hot	high	false	play
4	rain	mild	high	false	play
5	rain	cool	normal	false	play
6	rain	cool	normal	true	don't play
7	overcast	cool	normal	true	play
8	sunny	mild	high	false	don't play
9	sunny	cool	normal	false	play
10	rain	mild	normal	false	play
11	sunny	mild	normal	true	play
12	overcast	mild	high	true	play
13	overcast	hot	normal	false	play
14	rain	mild	high	true	don't play

PENS - ITS

Contoh Pembentukan Tree (3)

- Dari contoh data training tersebut, atribut kategori menentukan apakah seseorang perlu bermain (play) atau tidak (not play). Sedangkan atribut non-kategori yaitu :

ATRIBUT	Nilai yang mungkin
Outlook	sunny, overcast, rain
temperature	hot, mild, cool
Humidity	high, normal
Windy	true, false



Contoh Pembentukan Tree (4)

- Langkah menentukan Data yang terpilih menjadi Tree :
- Menentukan Node Terpilih
 - Untuk menentukan node terpilih, gunakan nilai entrophy dari setiap kriteria dengan data sample yang ditentukan. Node terpilih adalah kriteria dengan Entropy yang paling kecil.
- Keterangan :
 - Pernyataan Play (+)
 - Pernyataan Don't Play (-)



Prinsip Perhitungan nilai Log

- $\log_2(1) = 0$
- $\log_2(2) = 1$
- $\log_2(4) = 2$
- $\log_2(1/2) = -1$
- $\log_2(1/4) = -2$
- $(1/2)\log_2(1/2) = (1/2)(-1) = -1/2$

Rumus pencarian rata-rata entropy

$$\text{Average Entropy} = \sum_b \left(\frac{n_b}{n_1} \right) \times \left[\sum_c - \left(\frac{n_{bc}}{n_b} \right) \log_2 \left(\frac{n_{bc}}{n_b} \right) \right]$$

P_b = Probability an instance on a branch b is positive

$$= \frac{\text{number of positive instances on branch}}{\text{total number of instances on branch}} = \frac{n_{bc}}{n_b}$$

PENS - ITS

Contoh Pembentukan Tree (5)

Outlook

- Misal :
- B1:sunny terdapat 2 play(+) dan 3 don't play(-)
- B2:overcast terdapat 4 play(+)
- B3:rain terdapat 3 play(+) dan 2 don't play(-)
- average entropy untuk Outlook
- $= \frac{5}{14} [-\frac{3}{5} \log_2 (\frac{3}{5}) - \frac{2}{5} \log_2 (\frac{2}{5})] +$
 $\frac{4}{14} (-\log_2 1) + \frac{5}{14} [-\frac{3}{5} \log_2 (\frac{3}{5}) - \frac{2}{5} \log_2 (\frac{2}{5})]$
- $= 0.686 \sim 0.691$

Contoh Pembentukan Tree (6)

Temperature

- Misal:
- B1: hot terdapat 2 (+) dan 2 (-)
- B2: mild terdapat 4 (+) dan 2 (-)
- B3: cool terdapat 3(+) dan 1(-)
- average entropy untuk Temperature
- $= \frac{4}{14} [-\frac{2}{4} \log_2 (\frac{2}{4}) - \frac{2}{4} \log_2 (\frac{2}{4})] +$
 $\frac{6}{14} [-\frac{4}{6} \log_2 (\frac{4}{6}) - \frac{2}{6} \log_2 (\frac{2}{6})] +$
 $\frac{4}{14} [-\frac{3}{4} \log_2 (\frac{3}{4}) - \frac{1}{4} \log_2 (\frac{1}{4})]$
- = 0.82



Contoh Pembentukan Tree (7)

Humidity

- Misal :
- B1: high terdapat 3 (+) dan 4 (-)
- B2: normal terdapat 6 (+) dan 1 (-)
- average entropy untuk Humidity
- $= \frac{7}{14} [-\frac{3}{7} \log_2 (\frac{3}{7}) - \frac{4}{7} \log_2 (\frac{4}{7})] + \frac{7}{14} [-\frac{6}{7} \log_2 (\frac{6}{7}) - \frac{1}{7} \log_2 (\frac{1}{7})]$
- $= 0.785$

Contoh Pembentukan Tree (8)

Windy

- Misal :
- B1: false terdapat 6 (+) dan 2 (-)
- B2: true terdapat 3 (+) dan 3 (-)
- average entropy untuk Humidity
- = $\frac{8}{14} [-\frac{6}{8} \log_2 (\frac{6}{8}) - \frac{2}{8} \log_2 (\frac{2}{8})] + \frac{6}{14} [-\frac{3}{6} \log_2 (\frac{3}{6}) - \frac{3}{6} \log_2 (\frac{3}{6})]$
- = 0.8922

Contoh Pembentukan Tree (9)

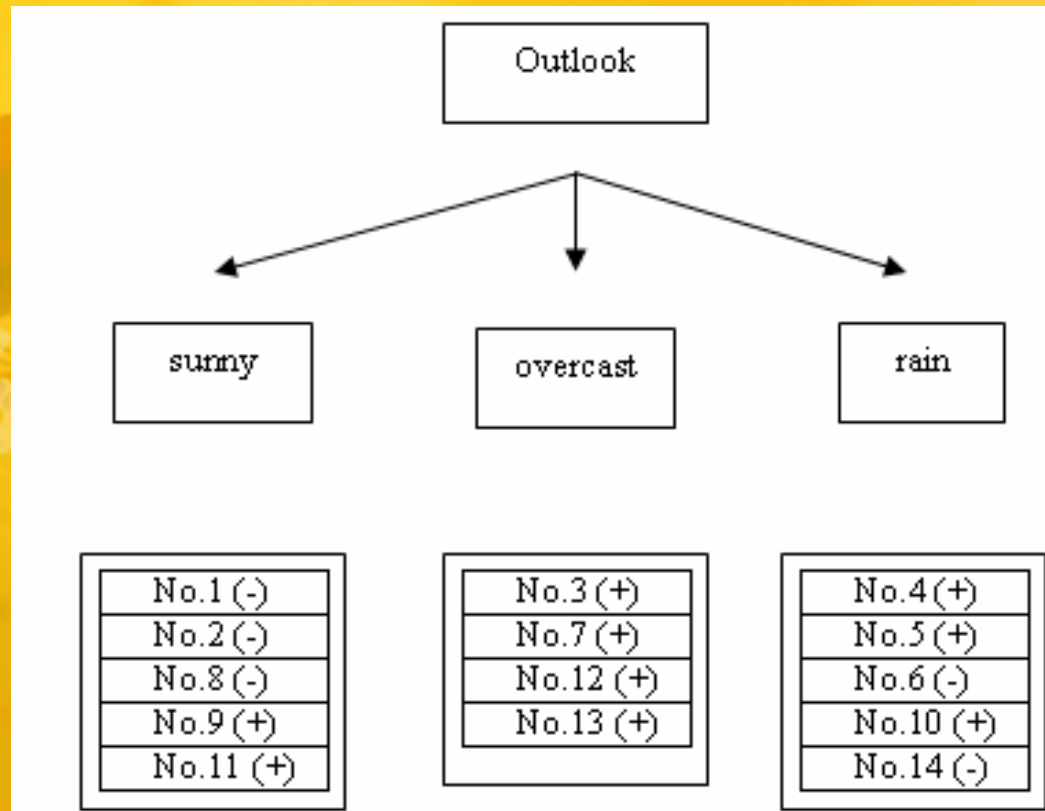
- Tabel 6.3 Hasil semua average entropy

Attribute	Average entropy
<i>Outlook</i>	0.686
<i>Temperature</i>	0.820
<i>Humidity</i>	0.785
<i>Windy</i>	0.8922

- Setelah menghitung masing-masing nilai entropi dari setiap atribut, didapatkan nilai entropi terkecil sebesar 0.686 yaitu atribut Outlook. Oleh karena itu atribut Outlook terpilih sebagai root.

Contoh Pembentukan Tree (10)

- Penyusunan Tree Awal



Contoh Pembentukan Tree (11)

- Leaf node berikutnya dapat dipilih pada bagian yang mempunyai nilai + dan - , pada contoh di atas ada dua yang mempunyai nilai + dan - yaitu Outlook=sunny dan Outlook=rain maka semuanya pasti mempunyai leaf node. Untuk melakukan leaf node lakukan satu persatu.

Contoh Pembentukan Tree (12)

- Data Training yang untuk Outlook=sunny

No	Temperature	Humidity	Windy	Play/Don't play
1	Hot	High	false	Don't play
2	Hot	High	true	Don't play
8	Mild	High	false	Don't play
9	Cool	Normal	false	Play
11	Mild	Normal	true	Play

Contoh Pembentukan Tree (13)

Temperature

- Misal:
- B1: hot terdapat 2 (-)
- B2: mild terdapat 1 (+) dan 1 (-)
- B3: cool terdapat 1 (+)
- average entropy untuk Temperature
- $= \frac{2}{5} [-\log_2 (1)] + \frac{2}{5} [-\frac{1}{2} \log_2 (1/2) - \frac{1}{2} \log_2 (1/2)] +$
- $\frac{1}{5} [-\log_2 (1)]$
- $= 0.4$

Contoh Pembentukan Tree (14)

Humidity

- Misal :
- B1: high terdapat 3 (-)
- B2: normal terdapat 2 (+)
- average entropy untuk Humidity
- $= \frac{3}{5} [-\log_2 (1)] + \frac{2}{5} [-\log_2 (1)]$
- $= 0$

Contoh Pembentukan Tree (15)

Windy

- Misal :
- B1: false terdapat 1 (+) dan 2 (-)
- B2: true terdapat 1 (+) dan 1 (-)
- average entropy untuk Humidity
- $= \frac{3}{5} [-\frac{2}{3} \log_2 (\frac{2}{3}) - \frac{1}{3} \log_2 (\frac{1}{3})] +$
- $\frac{2}{5} [-\frac{1}{2} \log_2 (\frac{1}{2}) - \frac{1}{2} \log_2 (\frac{1}{2})]$
- $= 0.317005$

Contoh Pembentukan Tree (16)

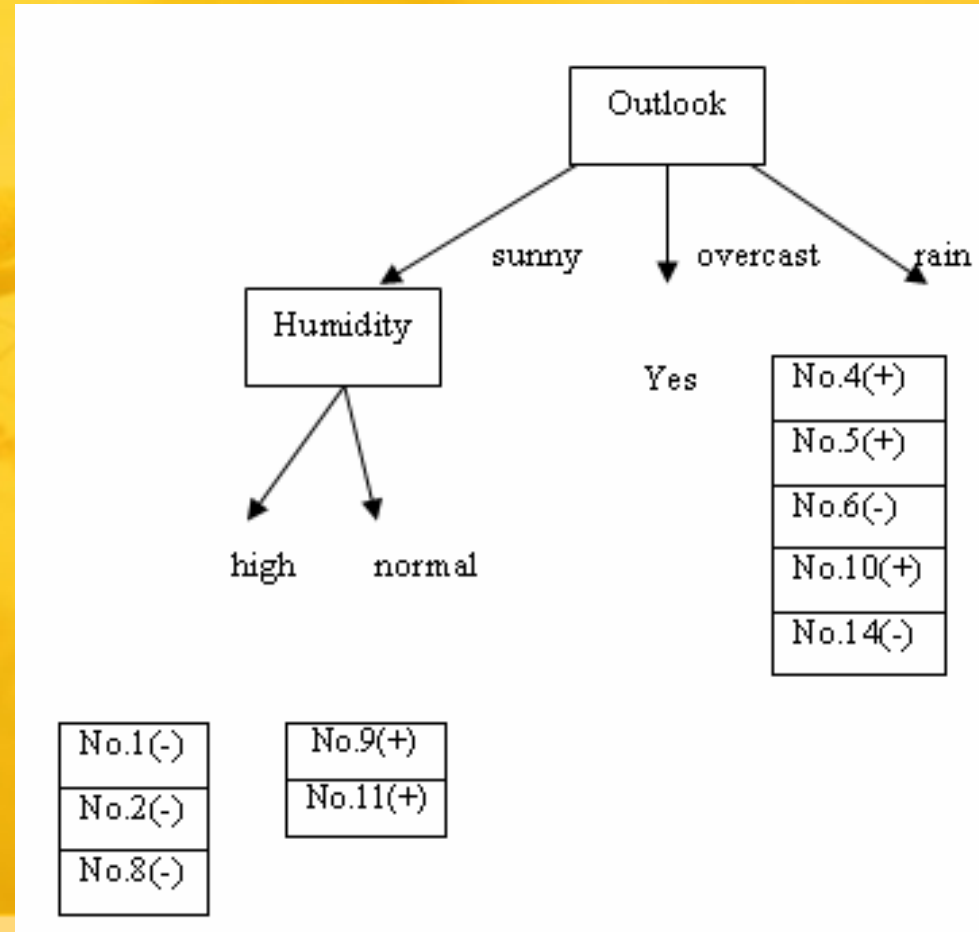
- Hasil average entropy untuk Outlook=sunny

Attribute	Average entropy
<i>Temperature</i>	0.4
<i>Humidity</i>	0.0
<i>Windy</i>	0.317005

- Atribut Humidity dipilih karena memiliki nilai entropi yang paling kecil yaitu 0.0

Contoh Pembentukan Tree (17)

- Menyusun tree lanjutan

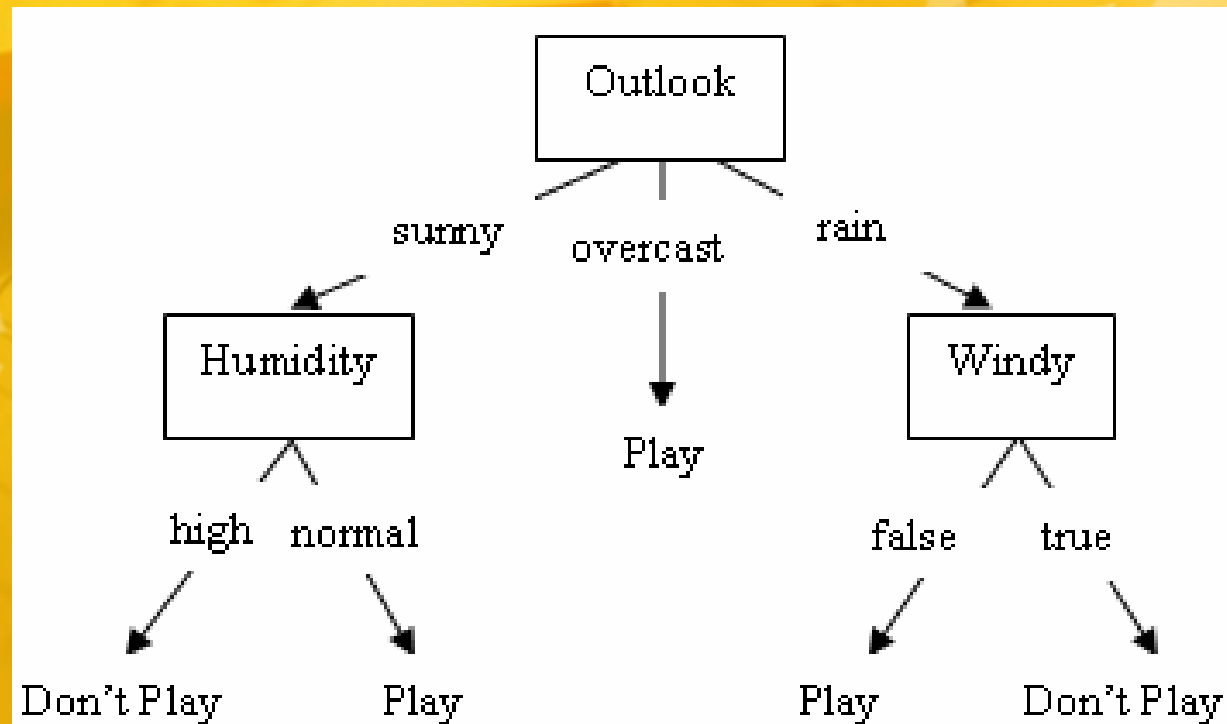


Contoh Pembentukan Tree (18)

- Pada tree tersebut, hanya outlook=rain yang mempunyai nilai + dan - .
- Untuk itu dilakukan perhitungan lagi agar semua menjadi leaf node.
- Caranya sama dengan cara diatas, yaitu dengan mencari nilai entropi terlebih dahulu

Contoh Pembentukan Tree (19)

- Hasil akhir dari Tree :



PENS - ITS