

Konsep Data Mining

Klasifikasi : Pohon Keputusan

Bertalya
Universitas Gunadarma
2009

Definisi Klasifikasi

- Proses untuk menyatakan suatu objek ke salah satu kategori yg sudah didefinisikan sebelumnya.
- Proses pembelajaran fungsi target (**model klasifikasi**) yg memetakan setiap sekumpulan atribut x (input) ke salah satu klas y yang didefinisikan sebelumnya.
 - Input : sekumpulan record (*training set*)
 - Setiap record terdiri atas sekumpulan atribut, salah satu atribut adalah klas.
 - Mencari model utk atribut klas sebagai fungsi dari nilai2 utk atribut yg lain.

Definisi Klasifikasi (lanj.)

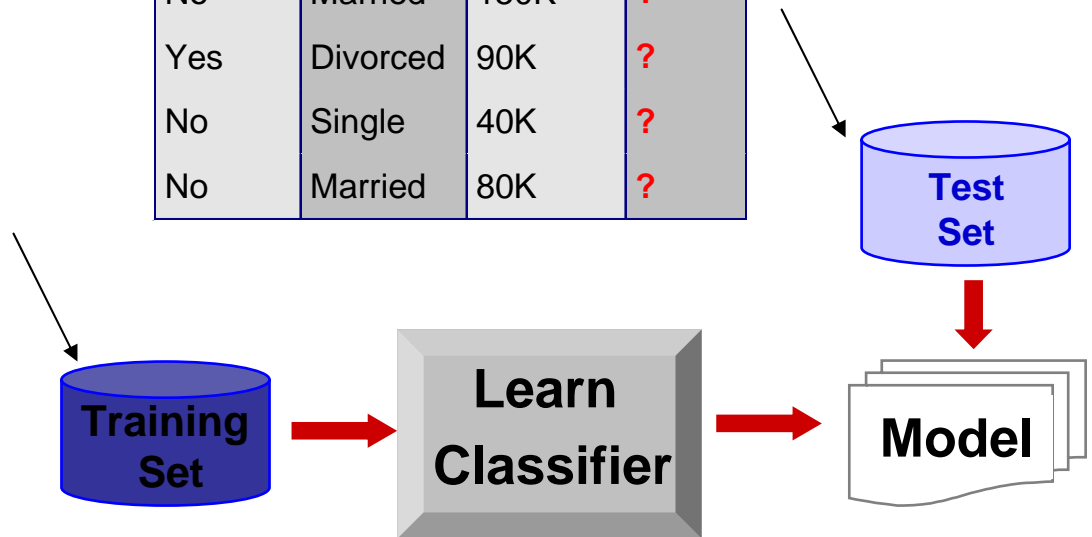
- Tujuannya adalah record2 yg sebelumnya tidak terlihat dinyatakan kelasnya seakurat mungkin.
- Model klasifikasi digunakan untuk
 - pemodelan deskriptif sebagai perangkat penggambaran utk membedakan objek2 dari klas berbeda.
 - Pemodelan prediktif digunakan utk memprediksi label klas utk record yg tidak diketahui atau tidak dikenal.

Contoh Klasifikasi

categorical
categorical
continuous
class

<i>Tid</i>	<i>Atr-1</i>	<i>Atr-2</i>	<i>Atr-3</i>	<i>Class</i>
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

<i>Atr-1</i>	<i>Atr-2</i>	<i>Atr-3</i>	<i>Class</i>
No	Single	75K	?
Yes	Married	50K	?
No	Married	150K	?
Yes	Divorced	90K	?
No	Single	40K	?
No	Married	80K	?



Classifier (Teknik Klasifikasi)

- Merupakan pendekatan sistematis utk membangun model klasifikasi dari sekumpulan data input.
- Contoh
 - *Decision tree* (pohon keputusan)
 - *Rule-based* (berbasis aturan)
 - *Neural network* (jaringan syaraf)
 - *Support Vector Machine* (SVM)
 - *Naïve Bayes*

Contoh Aplikasi 1

- Pendeteksi Kecurangan
 - Tujuan : memprediksi kecurangan pada transaksi kartu kredit
 - Pendekatan :
 - Gunakan transaksi kartu kredit & informasi mengenai pemegang rekening sebagai atribut2.
 - Kapan si pelanggan membeli, apa yg dibeli, seberapa sering pelanggan tsb membayar pada waktunya, dll
 - Tandai transaksi yg lama sebagai transaksi yg benar atau curang. Hal ini akan membentuk atribut klas.
 - Pelajari suatu model utk klas transaksi
 - Gunakan model ini utk mendeteksi kecurangan dgn mengamati transaksi kartu kredit pada suatu rekening.

Contoh Aplikasi 2

- Pemasaran secara langsung
 - Tujuan : mengurangi biaya pengiriman dengan menargetkan sekumpulan pelanggan yg kemungkinan membeli produk telpon selular baru
 - Pendekatan :
 - Gunakan data utk produk yg sama yg diperkenalkan sebelumnya.
 - Kita mengetahui pelanggan yg memutuskan membeli atau tidak. Hasil keputusan {beli, tidak beli} membentuk atribut klas.
 - Kumpulkan informasi mengenai pelanggan yg berhubungan dengan demografik, gaya hidup & interaksi dengan perusahaan.
 - Tipe bisnis, dimanan mereka tinggal, berapa banyak yg mereka belanjakan, dll
 - Gunakan informasi ini sebagai atribut input utk mempelajari model klasifikasi.

Pengukuran Kinerja Klasifikasi

- Akurasi = $\frac{\text{jumlah prediksi yang benar}}{\text{total jumlah prediksi}}$
- Rata2 kesalahan = $\frac{\text{jumlah prediksi yang salah}}{\text{total jumlah prediksi}}$

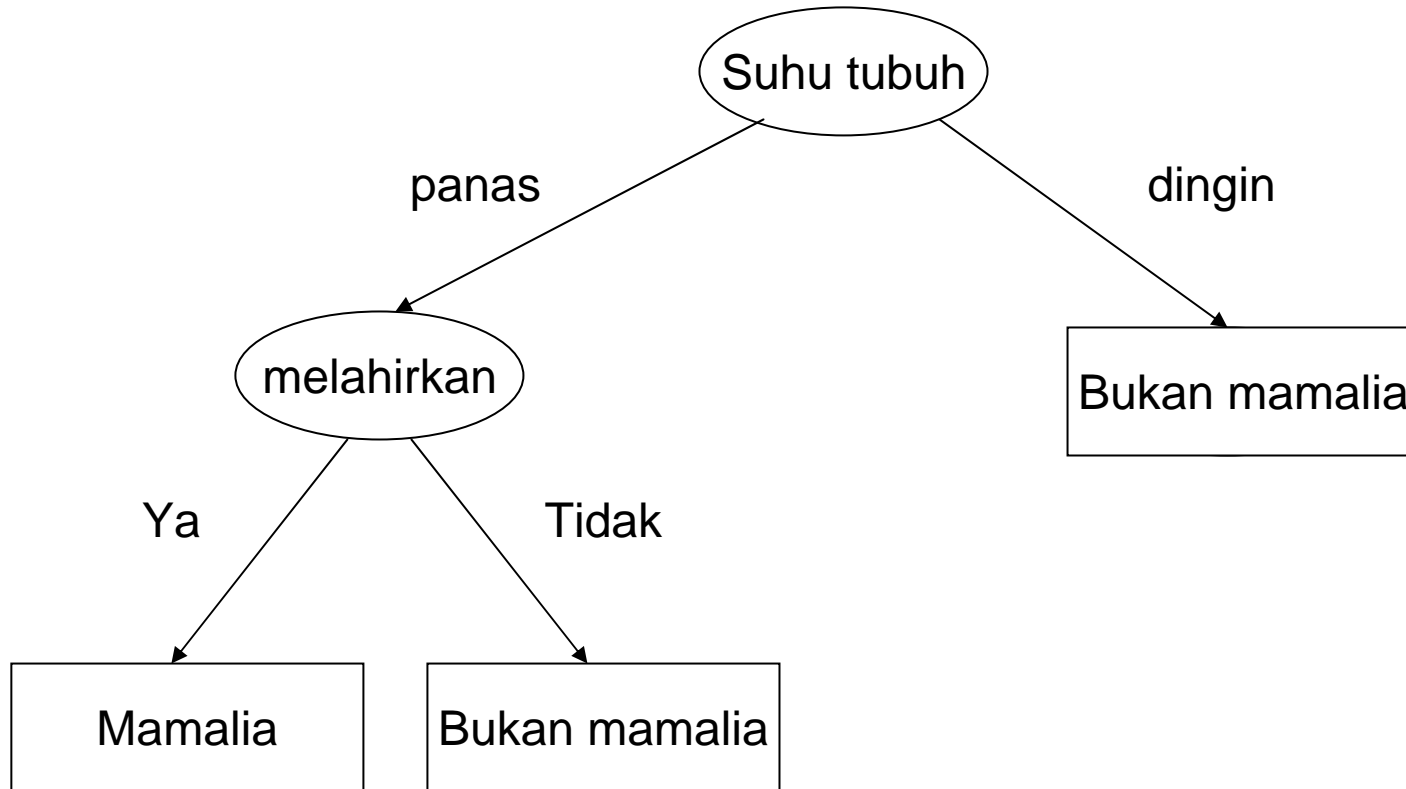
Pohon Keputusan (*Decision Tree*)

- Struktur hirarkis seperti pohon yg terdiri atas node & edge
 - *Root node*, node yg tidak mempunyai edge yg masuk & 0 atau banyak edge yg keluar
 - *Internal node*, node yg mempunyai satu edge yg masuk & dua atau lebih edge yg keluar
 - *Leaf atau terminal node*, node yg mempunyai satu edge yg masuk & tidak ada edge keluar

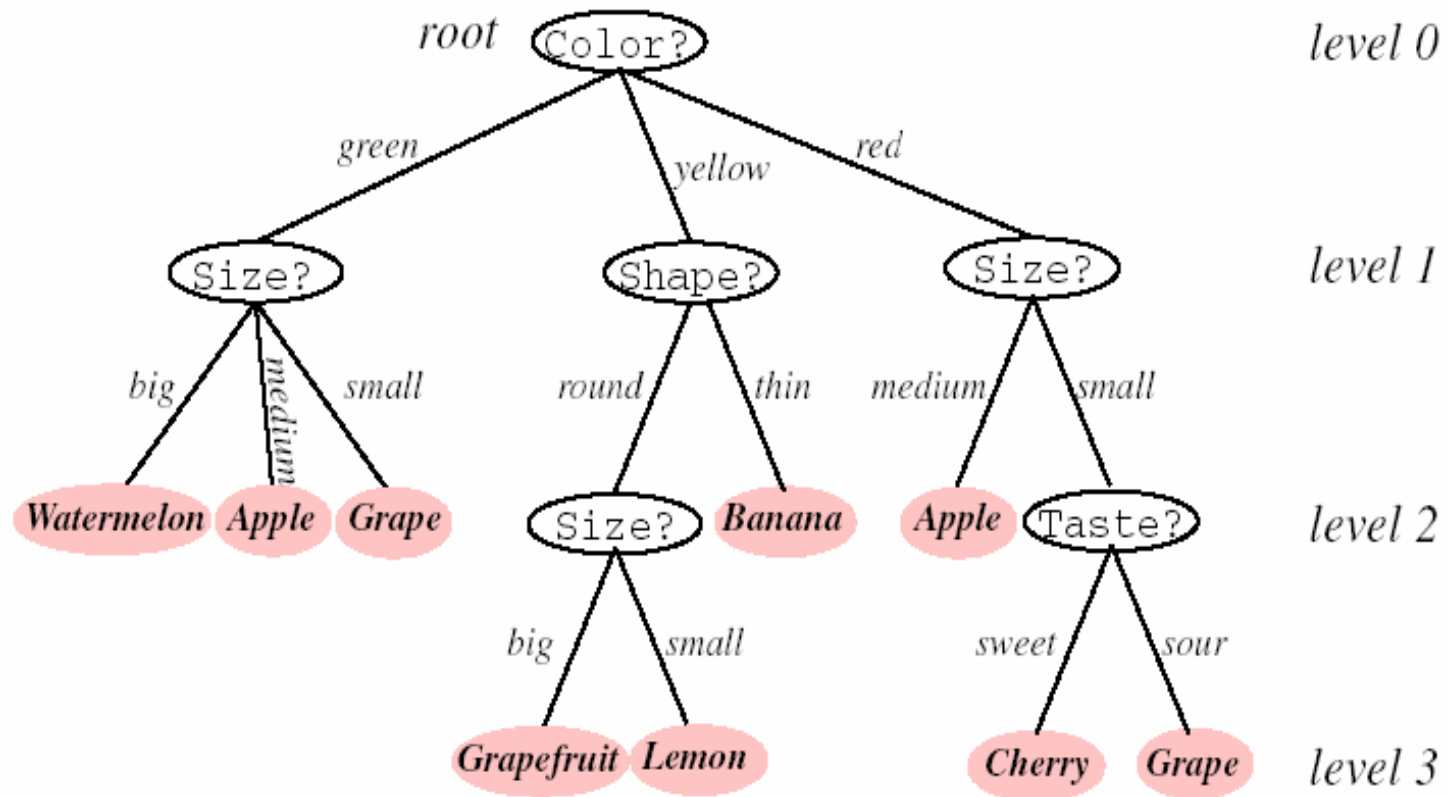
Leaf atau terminal node dinyatakan sebagai label klas.

Non-terminal node terdiri atas atribut kondisi tes utk membedakan record dgn karakteristik berbeda.

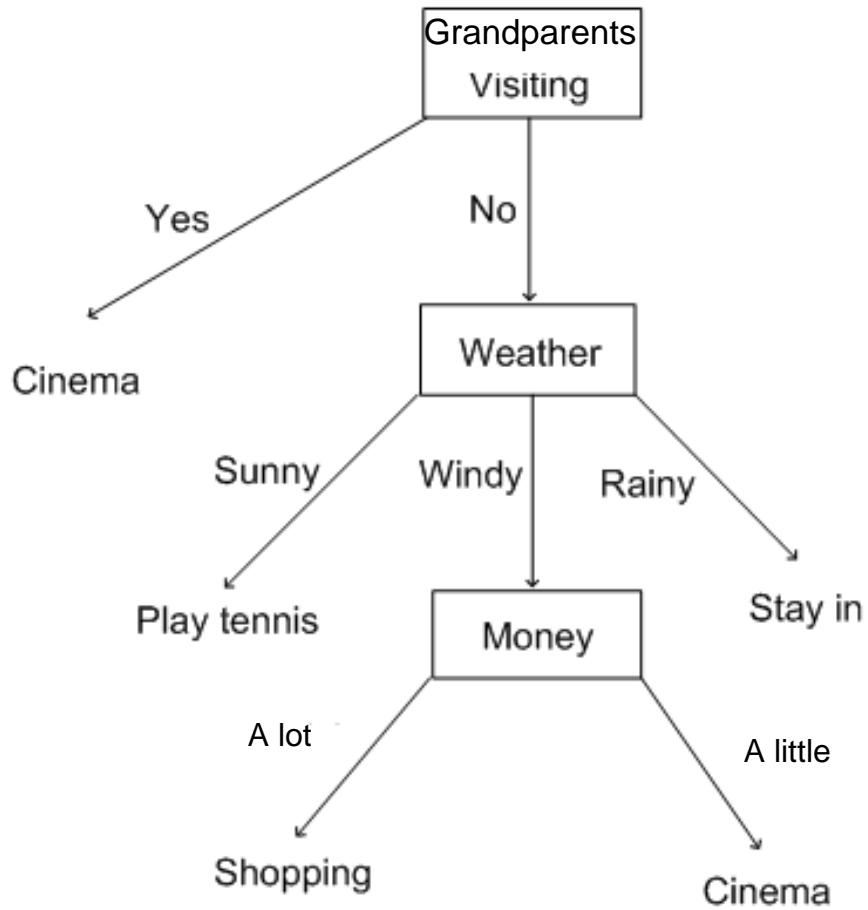
Contoh Pohon Keputusan : Klasifikasi Vertebrata



Classification and Regression Trees (CART)

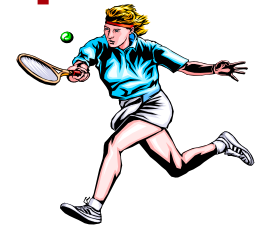


Contoh



Klasifikasi

DECISION TREE FOR THE CONCEPT

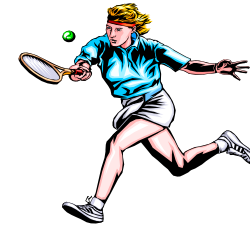


“Play Tennis”

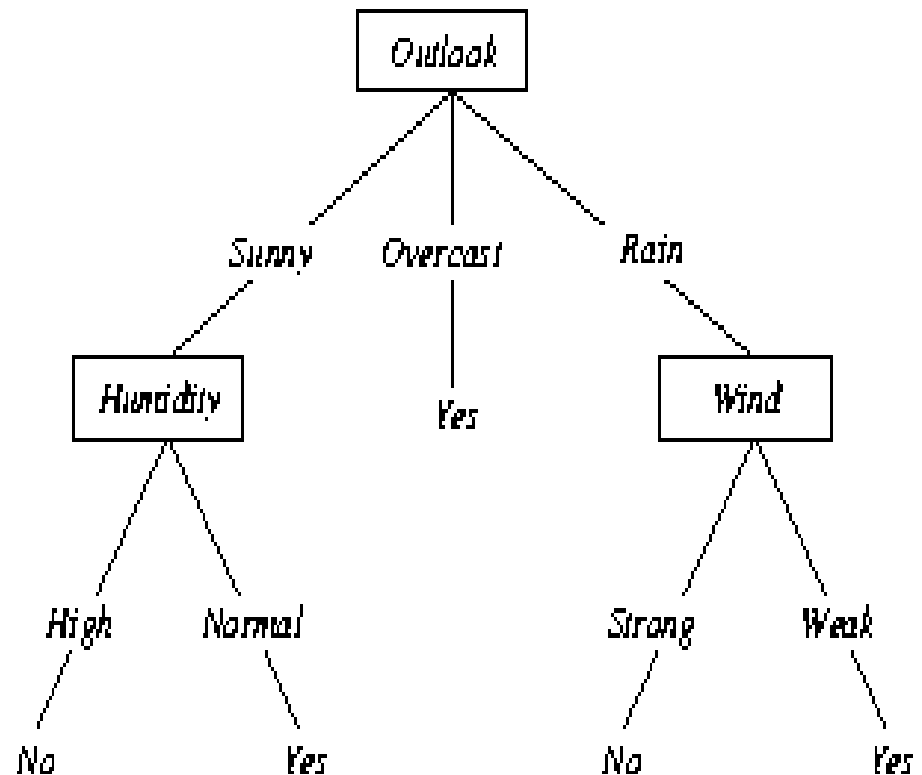
Day	Outlook	Temp	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

DECISION TREE FOR THE CONCEPT

“Play Tennis”



Day	Outlook	Temp	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No



Klasifikasi

[Mitchell, 1997] ¹⁴