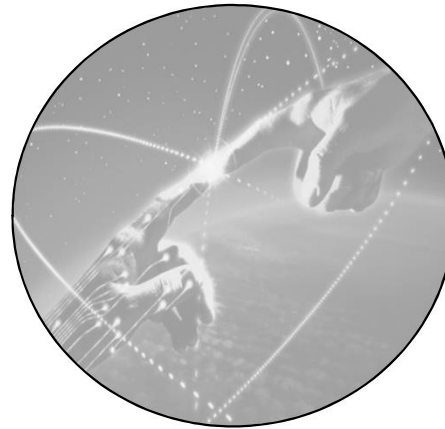


Metode Bayes



DATA MINING
(Model Klasifikasi)



PENDAHULUAN

- Teori keputusan Bayes adalah pendekatan statistik yang fundamental dalam pengenalan pola (*Pattern Recognition*).
- Pendekatan ini didasarkan pada kuantifikasi *trade-off* antara berbagai keputusan klasifikasi dengan menggunakan probabilitas dan resiko yang ditimbulkan dalam keputusan-keputusan tersebut.



Contoh Kasus

- Kita memiliki masalah yang bersifat hipotesis yaitu mendesain fungsi klasifikasi untuk memisahkan 2 jenis obyek ikan bandeng dan ikan kakap.
- Kedua obyek itu berjalan melalui *conveyor*.
- Seseorang bertugas mengamati obyek tersebut yang lewat *conveyor* secara acak.
- **Pengamat bertujuan memprediksi ikan apa yang akan lewat berikutnya.**



Penyelesaian

- Akan ada 2 kemungkinan yaitu: ikan bandeng atau ikan kakap.
- h_1 = ikan bandeng, h_2 = ikan kakap
- Jika jumlah ikan bandeng dan ikan kakap yang ditangkap sebelumnya adalah sama, maka peluang keduanya muncul di *conveyor* adalah sama besar.



Penyelesaian

- Selanjutnya *probabilitas apriori* , $P(h_1)$ dan $P(h_2)$ menyatakan peluang munculnya masing-masing ikan tersebut.
- Probabilitas prior ini menyatakan perkiraan kita akan jenis ikan apa yang muncul berikutnya sebelum benar-benar lewat di *conveyor*.
- Meskipun tidak diketahui dengan pasti tapi setidaknya bisa diestimasi dari data yang tersedia.



Penyelesaian

- Misalkan N = jumlah ikan total yang ditangkap sebelum masuk ke *conveyor*.
- N_1 = jumlah ikan bandeng, N_2 = jumlah ikan kakap, maka:

$$P(h_1) \approx \frac{N_1}{N} \qquad P(h_2) \approx \frac{N_2}{N}$$

- Prediksi adalah ikan bandeng (h_1) jika $P(h_1) > P(h_2)$, atau sebaliknya.



Permasalahan

- Bagaimana jika ikan yang harus kita perkirakan berjumlah banyak?
- Dalam banyak kasus, kita akan mengambil keputusan dengan informasi yang lebih banyak, tidak sekedar probabilitas prior saja.

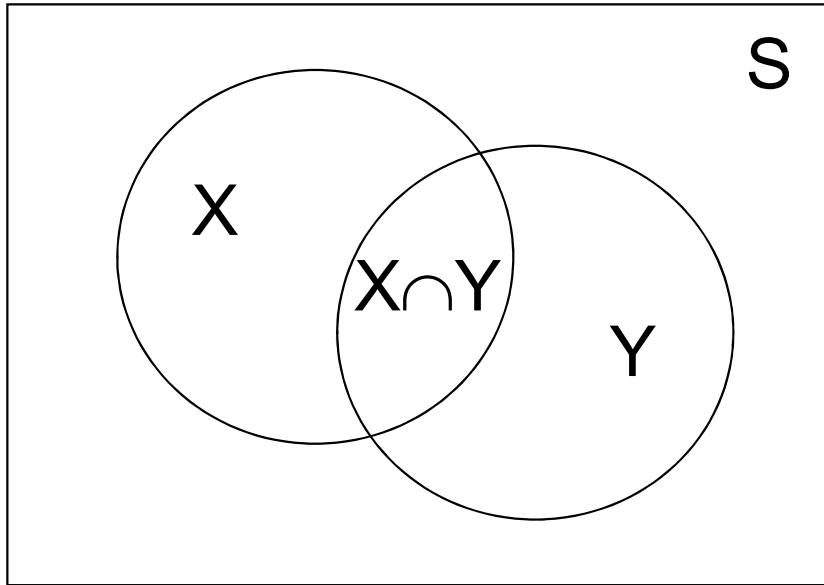


Permasalahan

- Kita bisa menggunakan informasi warna (x) untuk meningkatkan keakuratan prediksi.
- Dinyatakan dalam $P(x | h)$, menyatakan peluang muncul x jika diketahui h .
- Sehingga $P(x|h_1)$ dan $P(x|h_2)$ menyatakan perbedaan distribusi warna antara kedua ikan.



Probabilitas Bersyarat



$$P(X | Y) = \frac{P(X \cap Y)}{P(Y)}$$

Probabilitas X di dalam Y adalah probabilitas interseksi X dan Y dari probabilitas Y , atau dengan bahasa lain $P(X|Y)$ adalah prosentase banyaknya X di dalam Y



Probabilitas Bersyarat

- Dari kasus sebelumnya, misalkan diketahui probabilitas prior $P(h_j)$, dan probabilitas bersyarat $P(x|h_j)$, dan warna (x) , sehingga didapatkan *probabilitas posterior*:

$$P(h_j|x) = \frac{p(x|h_j)P(h_j)}{p(x)}$$

- $P(h_j|x)$ menyatakan probabilitas muncul h_j jika diketahui x .



Probabilitas Bersyarat

- Berdasarkan aturan Bayes, maka dapat ditetapkan sebagai berikut:
- Jika $P(h_1|x) < P(h_2|x)$, maka diklasifikasikan sebagai h_2 .
- Jika $P(h_1|x) > P(h_2|x)$, maka diklasifikasikan sebagai h_1 .



Probabilitas Bersyarat Dalam Data

#	Cuaca	Temperatur	Kecepatan Angin	Berolah-raga
1	Cerah	Normal	Pelan	Ya
2	Cerah	Normal	Pelan	Ya
3	Hujan	Tinggi	Pelan	Tidak
4	Cerah	Normal	Kencang	Ya
5	Hujan	Tinggi	Kencang	Tidak
6	Cerah	Normal	Pelan	Ya

Banyaknya data berolah-raga=ya adalah 4 dari 6 data maka dituliskan
 $P(\text{Olahraga}=\text{Ya}) = 4/6$

Banyaknya data cuaca=cerah dan berolah-raga=ya adalah 4 dari 6 data maka dituliskan

$$P(\text{cuaca}=\text{cerah dan Olahraga}=\text{Ya}) = 4/6$$

Jika cuaca cerah, orang
akankah berolah raga?

$$P(\text{cuaca} = \text{cerah} \mid \text{olahraga} = \text{ya}) = \frac{4/6}{4/6} = 1$$



Probabilitas Bersyarat Dalam Data

#	Cuaca	Temperatur	Berolahraga
1	cerah	normal	ya
2	cerah	tinggi	ya
3	hujan	tinggi	tidak
4	cerah	tinggi	tidak
5	hujan	normal	tidak
6	cerah	normal	ya

Banyaknya data berolah-raga=ya adalah 3 dari 6 data maka dituliskan
 $P(\text{Olahraga}=\text{Ya}) = 3/6$

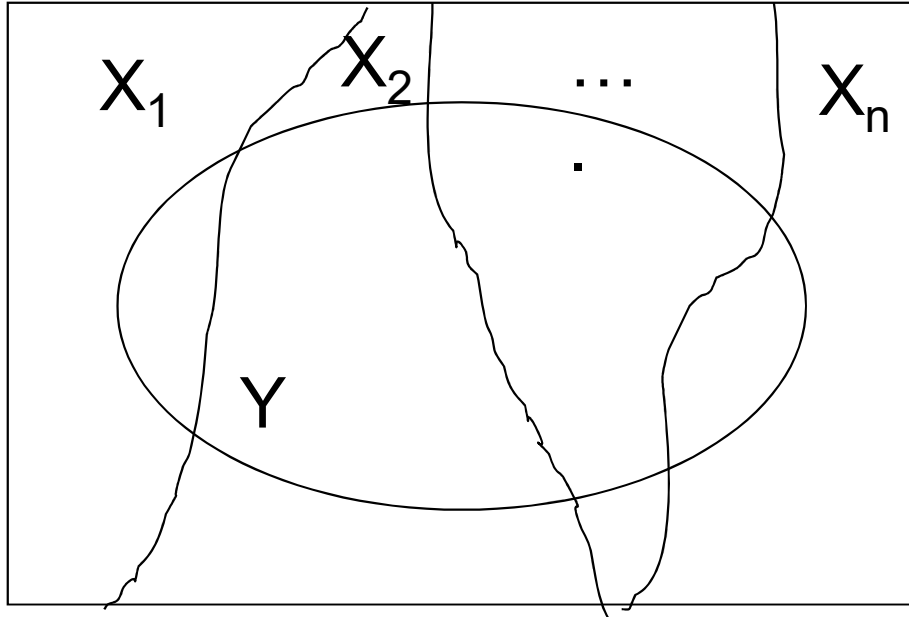
Banyaknya data cuaca=cerah, temperatur=normal dan berolah-raga=ya adalah 2 dari 6 data maka dituliskan

$$P(\text{cuaca}=\text{cerah}, \text{temperatur}=\text{normal}, \text{Olahraga}=\text{Ya}) = 2/6$$

$$P(\text{cuaca} = \text{cerah}, \text{temperatur} = \text{normal} \mid \text{olahraga} = \text{ya}) = \frac{2/6}{3/6} = \frac{2}{3}$$



Metode Bayes



$$P(X_k | Y) = \frac{P(Y | X_k)}{\sum_i P(Y | X_i)}$$

Keadaan Posterior (Probabilitas X_k di dalam Y) dapat dihitung dari keadaan prior (Probabilitas Y di dalam X_k dibagi dengan jumlah dari semua probabilitas Y di dalam semua X_i)



HMAP

HMAP (*Hypothesis Maximum a-Posteriori Probability*) menyatakan hipotesa yang diambil berdasarkan nilai probabilitas berdasarkan kondisi prior yang diketahui.

$$\begin{aligned} P(h | X) &= \frac{P(X | h) \cdot P(h)}{P(X)} \\ h_{\text{MAP}} &= \operatorname{argmax}_{h \in H} P(h | X) \\ &= \operatorname{argmax}_{h \in H} P(X | h) \cdot P(h) \end{aligned}$$

HMAP adalah model penyederhanaan dari metode bayes yang disebut dengan ***Naive Bayes***. HMAP inilah yang digunakan di dalam machine learning sebagai metode untuk mendapatkan hipotesis untuk suatu keputusan.



Contoh HMAP

Diketahui hasil survey yang dilakukan sebuah lembaga kesehatan menyatakan bahwa 90% penduduk di dunia menderita sakit paru-paru. Dari 90% penduduk yang sakit paru-paru ini 60% adalah perokok, dan dari penduduk yang tidak menderita sakit paru-paru 20% perokok.

Hipotesis: Bila diketahui seseorang merokok, apakah dia menderita sakit paru-paru?



Contoh HMAP

Fakta ini bisa didefinisikan dengan:

X=perokok, Y=sakit paru-paru.

Maka : $P(Y) = 0.9$

$$P(\neg Y) = 0.1$$

$$P(X|Y) = 0.6 \rightarrow P(\neg X|Y) = 0.4$$

$$P(X|\neg Y) = 0.2 \rightarrow P(\neg X|\neg Y) = 0.8$$

Dengan metode bayes dapat dihitung:

$$P(\{X\}|Y) = P(X|Y).P(Y) = (0.6) \cdot (0.9) = 0.54$$

$$P(\{X\}|\neg Y) = P(X|\neg Y).P(\neg Y) = (0.2) \cdot (0.1) = 0.02$$

Bila diketahui seseorang merokok, maka dia menderita sakit paru-paru karena $P(\{X\}|Y)$ lebih besar dari $P(\{X\}|\neg Y)$.

HMAP diartikan mencari probabilitas terbesar dari semua instance pada atribut target atau semua kemungkinan keputusan. Pada persoalan keputusan adalah sakit paru-paru atau tidak.



HMAP Dari Data Training

#	Cuaca	Temperatur	Kecepatan Angin	Berolah-raga
1	Cerah	Normal	Pelan	Ya
2	Cerah	Normal	Pelan	Ya
3	Hujan	Tinggi	Pelan	Tidak
4	Cerah	Normal	Kencang	Ya
5	Hujan	Tinggi	Kencang	Tidak
6	Cerah	Normal	Pelan	Ya

**Apakah bila cuaca
cerah dan
kecepatan angin
kencang, orang
akan berolahraga?**



HMAP Dari Data Training

#	Cuaca	Temperatur	Kecepatan Angin	Berolah-raga
1	Cerah	Normal	Pelan	Ya
2	Cerah	Normal	Pelan	Ya
3	Hujan	Tinggi	Pelan	Tidak
4	Cerah	Normal	Kencang	Ya
5	Hujan	Tinggi	Kencang	Tidak
6	Cerah	Normal	Pelan	Ya

Asumsi:

$Y = \text{berolahraga},$

$X_1 = \text{cuaca},$

$X_2 = \text{temperatur},$

$X_3 = \text{kecepatan angin}.$

Fakta menunjukkan:

$$P(Y=\text{ya}) = 4/6 \rightarrow P(Y=\text{tidak}) = 2/6$$



HMAP Dari Data Training

#	Cuaca	Temperatur	Kecepatan Angin	Berolah-raga
1	Cerah	Normal	Pelan	Ya
2	Cerah	Normal	Pelan	Ya
3	Hujan	Tinggi	Pelan	Tidak
4	Cerah	Normal	Kencang	Ya
5	Hujan	Tinggi	Kencang	Tidak
6	Cerah	Normal	Pelan	Ya

Apakah bila cuaca cerah dan kecepatan angin kencang, orang akan berolahraga?

Fakta: $P(X1=\text{cerah}|Y=\text{ya}) = 1$, $P(X1=\text{cerah}|Y=\text{tidak}) = 0$
 $P(X3=\text{kencang}|Y=\text{ya}) = 1/4$, $P(X3=\text{kencang}|Y=\text{tidak}) = 1/2$

HMAP dari keadaan ini dapat dihitung dengan:

$$\begin{aligned} P(X1=\text{cerah}, X3=\text{kencang} | Y=\text{ya}) &= \{ P(X1=\text{cerah}|Y=\text{ya}).P(X3=\text{kencang}|Y=\text{ya}) \} \cdot P(Y=\text{ya}) \\ &= \{ (1) \cdot (1/4) \} \cdot (4/6) = 1/6 \end{aligned}$$

$$\begin{aligned} P(X1=\text{cerah}, X3=\text{kencang} | Y=\text{tidak}) &= \{ P(X1=\text{cerah}|Y=\text{tidak}).P(X3=\text{kencang}|Y=\text{tidak}) \} \cdot P(Y=\text{tidak}) \\ &= \{ (0) \cdot (1/2) \} \cdot (2/6) = 0 \end{aligned}$$

KEPUTUSAN BEROLAHRAGA = YA



LATIHAN

#	Warna	Tipe	Asal	Tercuri ?
1	Merah	Sport	Domestik	Ya
2	Merah	Sport	Domestik	Tidak
3	Merah	Sport	Domestik	Ya
4	Kuning	Sport	Domestik	Tidak
5	Kuning	Sport	Import	Ya
6	Kuning	SUV	Import	Tidak
7	Kuning	SUV	Import	Ya
8	Kuning	SUV	Domestik	Tidak
9	Merah	SUV	Import	Tidak
10	Merah	Sport	Import	Ya

1. Apakah bila tipe SUV, asal import, kemungkinan besar akan tercuri?
2. Apakah bila warna merah, tipe sport, dan asal import kemungkinan akan tercuri?



Jawab Soal 1

Apakah bila tipe SUV, asal import, kemungkinan besar akan tercuri?

Asumsi:

$Y = \text{Tercuri},$
 $X_1 = \text{warna},$
 $X_2 = \text{tipe},$
 $X_3 = \text{asal}.$

Fakta menunjukkan:

$$P(Y=ya) = 5/10 \rightarrow P(Y=tidak) = 5/10$$

Fakta: $P(X_2=\text{SUV}|Y=ya) = 1/5$, $P(X_2=\text{SUV}|Y=tidak) = 3/5$
 $P(X_3=\text{import}|Y=ya) = 3/5$, $P(X_3=\text{import}|Y=tidak) = 2/5$

HMAP dari keadaan ini dapat dihitung dengan:

$$P(X_2=\text{SUV}, X_3=\text{import} | Y=ya)$$

$$= \{ P(X_2=\text{SUV}|Y=ya) \cdot P(X_3=\text{import}|Y=ya) \} \cdot P(Y=ya)$$

$$= \{ (1/5) \cdot (3/5) \} \cdot (5/10) = 15/250 = 3/50$$

$$P(X_2=\text{SUV}, X_3=\text{import} | Y=tidak)$$

$$= \{ P(X_2=\text{SUV}|Y=tidak) \cdot P(X_3=\text{import}|Y=tidak) \} \cdot P(Y=tidak)$$

$$= \{ (3/5) \cdot (2/5) \} \cdot (5/10) = 30/250 = 3/25$$

HIPOTESIS TERCURI = "TIDAK"



Jawab Soal 2

Apakah bila warna merah, tipe sport, dan asal import kemungkinan akan tercuri?

Asumsi:

$Y = \text{Tercuri},$
 $X_1 = \text{warna},$
 $X_2 = \text{tipe},$
 $X_3 = \text{asal}.$

Fakta menunjukkan:

$$P(Y=\text{ya}) = 5/10 \rightarrow P(Y=\text{tidak}) = 5/10$$

Fakta: $P(X_1=\text{Merah}|Y=\text{ya}) = 3/5$, $P(X_1=\text{Merah}|Y=\text{tidak}) = 2/5$
 $P(X_2=\text{Sport}|Y=\text{ya}) = 4/5$, $P(X_2=\text{Sport}|Y=\text{tidak}) = 2/5$
 $P(X_3=\text{import}|Y=\text{ya}) = 3/5$, $P(X_3=\text{import}|Y=\text{tidak}) = 2/5$

HMAP dari keadaan ini dapat dihitung dengan:

$P(X_1=\text{Merah}, X_2=\text{Sport}, X_3=\text{import} | Y=\text{ya})$

$$\begin{aligned} &= \{ P(X_1=\text{Merah}|Y=\text{ya}) \cdot P(X_2=\text{Sport}|Y=\text{ya}) \cdot P(X_3=\text{import}|Y=\text{ya}) \} \cdot P(Y=\text{ya}) \\ &= \{ (3/5) \cdot (4/5) \cdot (3/5) \} \cdot (5/10) = 180/1250 = 0,144 \end{aligned}$$

$P(X_1=\text{Merah}, X_2=\text{Sport}, X_3=\text{import} | Y=\text{tidak})$

$$\begin{aligned} &= \{ P(X_1=\text{Merah}|Y=\text{tidak}) \cdot P(X_2=\text{Sport}|Y=\text{tidak}) \cdot P(X_3=\text{import}|Y=\text{tidak}) \} \cdot P(Y=\text{tidak}) \\ &= \{ (2/5) \cdot (2/5) \cdot (2/5) \} \cdot (5/10) = 40/1250 = 0,032 \end{aligned}$$

HIPOTESIS TERCURI = “YA”