



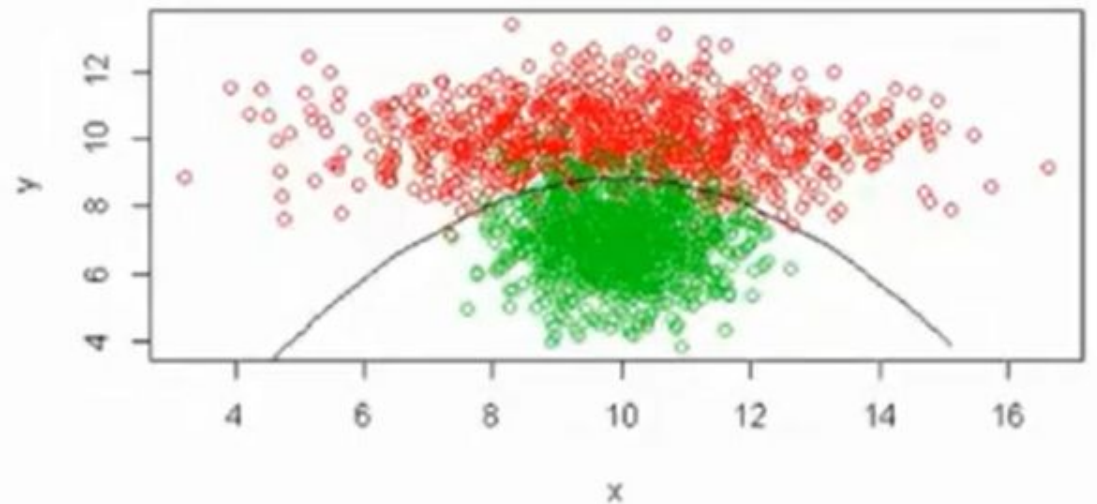
# Klasifikasi dengan Naïve Bayes

## KLASIFIKASI

- Klasifikasi adalah algoritma yang menggunakan data dengan **target/class/label** berupa nilai kategorikal (nominal).
- Contoh, apabila **target/class/label** adalah pendapatan, maka bisa digunakan nilai nominal (kategorikal) sbb: pendapatan besar, menengah, kecil.
- Contoh lain adalah rekomendasi contact lens, apakah menggunakan yang jenis **soft**, **hard** atau **none**.
- Algoritma klasifikasi yang biasa digunakan adalah: Naive Bayes, K-Nearest Neighbor, C4.5, ID3, CART, Linear Discriminant Analysis, etc.

## BAYESIAN CLASSIFICATION

- Bayesian Classification adalah pengklasifikasian statistik yang dapat digunakan untuk memprediksi *probabilitas* keanggotaan suatu class.
- Bayesian Classification didasarkan pada *teorema Bayes* yang memiliki kemampuan klasifikasi serupa *decision tree* dan *neural network*.
- Bayesian Classification terbukti memiliki *akurasi dan kecepatan* yang tinggi saat diaplikasikan ke dalam database dengan data yang besar.





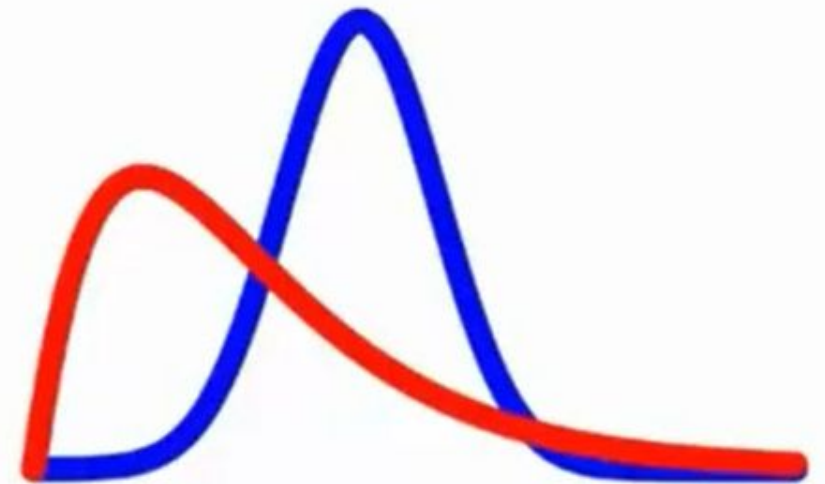
## RUMUS TEOREMA BAYES

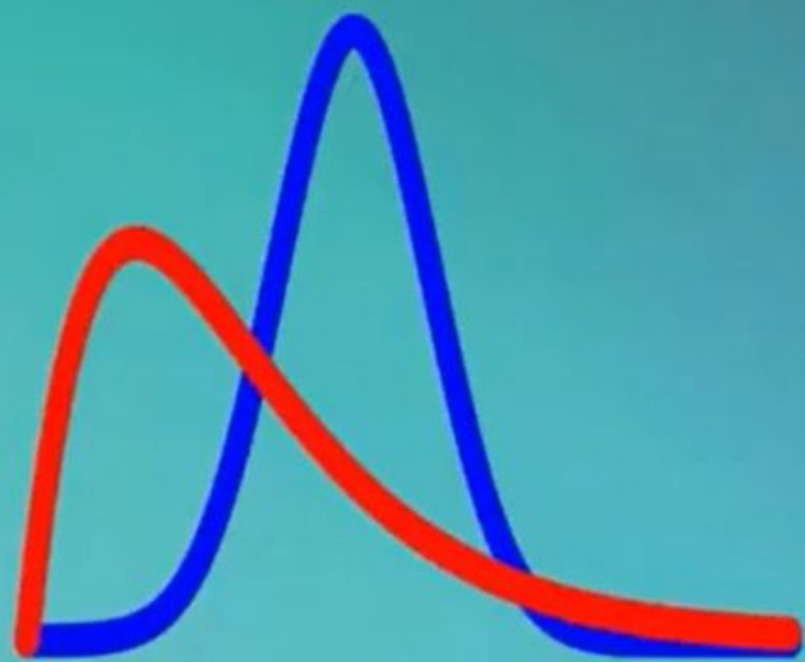
**Teorema Bayes** memiliki bentuk umum sbb :

$$P(H|X) = \frac{P(X|H) P(H)}{P(X)}$$

Keterangan :

- X : data dengan *class* yang belum diketahui
- H : hipotesis data X merupakan suatu *class* spesifik
- $P(H|X)$  : probabilitas hipotesis H berdasar kondisi X  
(*posteriori probability*)
- $P(H)$  : probabilitas hipotesis H (*prior probability*)
- $P(X|H)$  : probabilitas X berdasar kondisi hipotesis H
- $P(X)$  : probabilitas dari X





# Contoh Perhitungan

Klasifikasi menggunakan Naive Bayes

## DATA TRAINING

Terdapat dua class dari  
klasifikasi yang dibentuk  
yaitu :

C1 => buys\_computer = yes

C2 => buys\_computer = no

Nilai X yang belum  
diketahui Label / Kelas :

X =

(

age="<=30",  
income="Medium",  
student="Yes",  
credit\_rating="Fair";

)

$$P(C_i | \mathbf{X}) = P(C_i) \prod_{k=1}^N P(X_k = x_k | C_i)$$

Id	Age	Income	Student	Credit_rating	Class: buys_computer
1	<=30	High	No	Fair	No
2	<=30	High	No	Excellent	No
3	31..40	High	No	Fair	Yes
4	>40	Medium	No	Fair	Yes
5	>40	Low	Yes	Fair	Yes
6	>40	Low	Yes	Excellent	No
7	31..40	Low	Yes	Excellent	Yes
8	<=30	Medium	No	Fair	No
9	<=30	Low	Yes	Fair	Yes
10	>40	Medium	Yes	Fair	Yes
11	<=30	Medium	Yes	Excellent	Yes
12	31..40	Medium	No	Excellent	Yes
13	31..40	High	Yes	Fair	Yes
14	>40	Medium	No	Excellent	No



## PENYELESAIAN (1)

Dibutuhkan untuk memaksimalkan  $P(X|C_i) P(C_i)$  untuk  $i=1, 2$

- $P(C_i)$  merupakan prior probability untuk setiap class berdasar data contoh

$$P(\text{buys\_computer}=\text{"yes"}) = 9/14 = 0.643$$

$$P(\text{buys\_computer}=\text{"no"}) = 5/14 = 0.357$$

- Hitung  $P(X|C_i)$  untuk  $i=1, 2$

$$P(\text{age}=\text{"<=30"} | \text{buys\_computer}=\text{"yes"}) = 2/9 = 0.222$$

$$P(\text{age}=\text{"<=30"} | \text{buys\_computer}=\text{"no"}) = 3/5 = 0.600$$

$$P(\text{income}=\text{"medium"} | \text{buys\_computer}=\text{"yes"}) = 4/9 = 0.444$$

$$P(\text{income}=\text{"medium"} | \text{buys\_computer}=\text{"no"}) = 2/5 = 0.400$$

$$P(\text{student}=\text{"yes"} | \text{buys\_computer}=\text{"yes"}) = 6/9 = 0.667$$

$$P(\text{student}=\text{"yes"} | \text{buys\_computer}=\text{"no"}) = 1/5 = 0.200$$

$$P(\text{credit\_rating}=\text{"fair"} | \text{buys\_computer}=\text{"yes"}) = 6/9 = 0.667$$

$$P(\text{credit\_rating}=\text{"fair"} | \text{buys\_computer}=\text{"no"}) = 2/5 = 0.400$$

## PENYELESAIAN (2)

$$P(X | \text{buys\_computer} = \text{"yes"}) = 0.222 * 0.444 * 0.677 * 0.677 \\ = 0.044$$

$$P(X | \text{buys\_computer} = \text{"no"}) = 0.600 * 0.400 * 0.200 * 0.400 \\ = 0.019$$

- $P(X | \text{buys\_computer} = \text{"yes"}) P(\text{buys\_computer} = \text{"yes"})$   
 $= 0.044 * 0.643 = \mathbf{0.028}$
- $P(X | \text{buys\_computer} = \text{"no"}) P(\text{buys\_computer} = \text{"no"})$   
 $= 0.019 * 0.357 = 0.007$

Kesimpulan: ***buys\_computer = "yes"***



***Selesai***

---

