

Algoritma Naïve bayes

Nugroho DS

Algoritma Data Mining

1. **Estimation** (Estimasi):

- Linear Regression, **Neural Network**, Support Vector Machine, etc

2. **Prediction/Forecasting** (Prediksi/Peramalan):

- Linear Regression, **Neural Network**, Support Vector Machine, etc

3. **Classification** (Klasifikasi):

- Naive Bayes, K-Nearest Neighbor, **C4.5**, ID3, CART, Linear Discriminant Analysis, etc

4. **Clustering** (Klastering):

- **K-Means**, K-Medoids, Self-Organizing Map (SOM), Fuzzy C-Means, etc

5. **Association** (Asosiasi):

- FP-Growth, **A Priori**, etc

Naïve Bayes

Algoritma *Naive Bayes* merupakan salah satu algoritma yang terdapat pada teknik klasifikasi.

Naive Bayes merupakan pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris *Thomas Bayes*, yaitu memprediksi peluang di masa depan berdasarkan pengalaman dimasa sebelumnya sehingga dikenal sebagai *Teorema Bayes*

- Probability of event H given evidence E:

$$Pr[H | E] = \frac{Pr[E | H]Pr[H]}{Pr[E]}$$

- Probabilitas hipotesis: $Pr[H]$
- Probabilitas dari kejadian sebelum bukti terlihat
- Probabilitas hipotesis berdasar kondisi E: $Pr[H | E]$

Thomas Bayes

Born: 1702 in London, England

Died: 1761 in Tunbridge Wells, Kent, England



Naïve bayes Klasifikasi

- perlu diketahui bahwa proses klasifikasi memerlukan sejumlah petunjuk untuk menentukan kelas apa yang cocok bagi sampel yang dianalisis tersebut. Karena itu, teorema *bayes* di atas disesuaikan sebagai berikut :

$$Pr[H | E] = \frac{Pr[E_1 | H]Pr[E_2 | H] \dots Pr[E_n | H]Pr[H]}{Pr[E]}$$

Probabilitas cuaca

Outlook	Temp	Humidity	Windy	Play
Sunny	Hot	High	False	No
Sunny	Hot	High	True	No
Overcast	Hot	High	False	Yes
Rainy	Mild	High	False	Yes
Rainy	Cool	Normal	False	Yes
Rainy	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Sunny	Mild	High	False	No
Sunny	Cool	Normal	False	Yes
Rainy	Mild	Normal	False	Yes
Sunny	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Rainy	Mild	High	True	No

Contoh Data Weather

Outlook	Temp.	Humidity	Windy	Play
Sunny	Cool	High	True	?

← *Evidence E*

$$Pr[yes \mid E] = Pr[Outlook = Sunny \mid yes]$$

$$\times Pr[Temperature = Cool \mid yes]$$

$$\times Pr[Humidity = High \mid yes]$$

$$\times Pr[Windy = True \mid yes]$$

$$\times \frac{Pr[yes]}{Pr[E]}$$

$$= \frac{\frac{2}{9} \times \frac{3}{9} \times \frac{3}{9} \times \frac{3}{9} \times \frac{9}{14}}{Pr[E]}$$

↑
*Probability of
class “yes”*

Outlook			Temperature			Humidity			Windy			Play	
Yes	No		Yes	No		Yes	No		Yes	No		Yes	No
Sunny	2	3	Hot	2	2	High	3	4	False	6	2	9	5
Overcast	4	0	Mild	4	2	Normal	6	1	True	3	3		
Rainy	3	2	Cool	3	1								
Sunny	2/9	3/5	Hot	2/9	2/5	High	3/9	4/5	False	6/9	2/5	9/14	5/14
Overcast	4/9	0/5	Mild	4/9	2/5	Normal	6/9	1/5	True	3/9	3/5		
Rainy	3/9	2/5	Cool	3/9	1/5								

A new day:

Outlook	Temp.	Humidity	Windy	Play
Sunny	Cool	High	True	?

Likelihood of the two classes

For "yes" = $2/9 \times 3/9 \times 3/9 \times 3/9 \times 9/14 = 0.0053$

For "no" = $3/5 \times 1/5 \times 4/5 \times 3/5 \times 5/14 = 0.0206$

<https://www.solidfiles.com/v/QkKygmmlYeP2L>