

Nama : Anindita Khusnul Oktavia  
Program : Introduction to Python for Data Science

## SUMMARY

### Classification 2

#### Naive Bayes Classifier

Naive Bayes adalah teknik klasifikasi statistik berdasarkan Bayes Theorem. Ini adalah salah satu supervised learning algorithms yang paling sederhana. Naive Bayes classifier adalah algoritme yang cepat, akurat, dan andal. Naive Bayes classifier memiliki akurasi dan kecepatan tinggi pada kumpulan data besar.

Naive Bayes classifier mengasumsikan bahwa efek fitur tertentu dalam kelas tidak bergantung pada fitur lainnya. Misalnya, pemohon pinjaman diloloskan atau tidak tergantung pada pendapatannya, pinjaman sebelumnya dan riwayat transaksi, usia, dan lokasi. Meskipun fitur ini saling bergantung, fitur ini masih dianggap terpisah. Asumsi ini menyederhanakan komputasi, dan itulah mengapa dianggap naive. Asumsi ini disebut class conditional independence.

$$P(h|D) = \frac{P(D|h)P(h)}{P(D)}$$

- $P(h)$ : the probability of hypothesis  $h$  being true (regardless of the data). This is known as the **prior** probability of  $h$ .
- $P(D)$ : the probability of the data (regardless of the hypothesis). This is known as the **prior** probability.
- $P(h|D)$ : the probability of hypothesis  $h$  given the data  $D$ . This is known as **posterior** probability.
- $P(D|h)$ : the probability of data  $d$  given that the hypothesis  $h$  was true. This is known as **posterior** probability.

#### How Naive Bayes classifier works?

Naive Bayes classifier menghitung probabilitas suatu peristiwa dalam langkah-langkah berikut:

- Step 1: Calculate the prior probability for given class labels
- Step 2: Find Likelihood probability with each attribute for each class
- Step 3: Put these value in Bayes Formula and calculate posterior probability.
- Step 4: See which class has a higher probability, given the input belongs to the higher probability class.

Untuk menyederhanakan perhitungan probabilitas prior dan posterior kita dapat menggunakan dua tables frequency dan likelihood tables. Kedua tabel ini akan membantu kita menghitung probabilitas prior dan posterior. Frequency table berisi kemunculan label untuk semua fitur. Ada dua tabel kemungkinan. Likelihood Table 1 menunjukkan probabilitas sebelumnya dari label dan Likelihood Table 2 menunjukkan probabilitas posterior.

Whether	Play
Sunny	No
Sunny	No
Overcast	Yes
Rainy	Yes
Rainy	Yes
Rainy	No
Overcast	Yes
Sunny	No
Sunny	Yes
Rainy	Yes
Sunny	Yes
Overcast	Yes
Overcast	Yes
Rainy	No

  

Frequency Table		
Whether	No	Yes
Overcast		4
Sunny	2	3
Rainy	3	2
Total	5	9

  

Likelihood Table 1				
Whether	No	Yes		
Overcast		4	=4/14	0.29
Sunny	2	3	=5/14	0.36
Rainy	3	2	=5/14	0.36
Total	5	9		
	=5/14	=9/14		
	0.36	0.64		

  

Likelihood Table 2				
Whether	No	Yes	Posterior Probability for No	Posterior Probability for Yes
Overcast		4	0/5=0	4/9=0.44
Sunny	2	3	2/5=0.4	3/9=0.33
Rainy	3	2	3/5=0.6	2/9=0.22
Total	5	9		

Untuk implementasinya ada pada file .ipynb