CONTOH KASUS
NAÏVE BAYES CLASSIFIER
MACHINE LEARNING

Disusun Oleh:

Dr. Jasman Pardede, S.Si., M.T.

TEKNIK INFORMATIKA
FAKULTAS TEKNIK INDUSTRI
INSTITUT TEKNOLOGI NASIONAL BANDUNG

Algoritma Naive Bayes merupakan sebuah metoda klasifikasi menggunakan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes. Algoritma Naïve Bayes adalah salah satu algoritma pembelajaran induktif yang paling efektif dan efisien untuk pembelajaran mesin dan *data mining*. Algoritma Naïve Bayes merupakan algoritma yang populer dalam aplikasi pembelajaran mesin karena kesederhanaan algoritmanya. Algoritma Naïve Bayes memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya, sehingga dikenal sebagai Teorema Bayes.

Algoritma Naïve Bayes menggunakan asumsi bahwa setiap atribut memiliki hubungan yang saling bebas yaitu tidak ada ketergantungan antara satu atribut dengan atribut lainnya. Sebagai contoh, buah dapat dianggap sebagai buah **appel,** jika buah tersebut memiliki atribut warna merah, bentuknya bulat, dan memiliki diameter sekitar 8 cm. Pada Naïve Bayes menggunakan asumsi bahwa tidak ada hubungan antara warna, bentuk, dan diameter dalam menentukan buah appel. Walaupun pada realitanya bahwa asumsi "hubungan yang saling bebas" atau *independensi* tersebut sangat jarang terjadi.

Ciri utama dari klasifikasi Naïve Bayes ini adalah asumsi yang sangat kuat (naïf) akan *independensi* dari masing-masing atribut atau kondisi/kejadian. Setiap atribut pada algoritma Naïve Bayes berkontribusi terhadap keputusan akhir secara setara dan *independen* atau bebas dari atribut lainnya. Sehingga algoritma Naïve Bayes efisiensi secara komputasi dan cocok untuk berbagai domain. Walaupun asumsi *independensi* tersebut dilanggar, tetapi kinerja klasifikasi dengan Naïve Bayes cukup tinggi. Hal ini telah dibuktikan pada berbagai hasil penelitian empiris. Pada penelitian Xhemali, dkk. (2009) menyatakan bahwa klasifikasi Naïve Bayes memiliki kinerja akurasi yang lebih baik dibandingkan dengan algoritma klasifikasi lainnya.

Algoritma Naïve Bayes memiliki kelebihan, diantaranya: mudah dipahami, mudah diimplementasikan pada suatu bahasa pemrograman tertentu, dapat digunakan untuk data kuantitatif maupun kualitatif, perhitungannya cepat dan efisien, tidak memerlukan jumlah data yang banyak, tidak memerlukan data *training* (pelatihan) yang banyak, dapat digunakan untuk klasifikasi *biner* atau *multi-class*, dapat mengabaikan data yang hilang dalam perhitungan, dan lain-lain.

Sedangkan kekurangan algoritma Naïve Bayes diantaranya: memerlukan pengetahuan awal atau pengetahuan masa lalu dalam membuat suatu keputusan, tidak cocok digunakan untuk kasus yang memiliki korelasi antara satu atribut dengan atribut lainnya, probabilitas prediksi akan bernilai nol jika probabilitas kondisional bernilai nol, dan lain-lain.

Rumus umum teorema Bayes adalah sebagai berikut:

$$P(\boldsymbol{c}|\boldsymbol{x}) = \frac{P(\boldsymbol{x}|\boldsymbol{c})\,P(\boldsymbol{c})}{P(\boldsymbol{x})}$$

Dimana:

x : data dengan class yang belum diketahui

c : hipotesis data merupakan suatu class yang spesifik

P(***x***|***c***) : probabilistik hipotesis ***x*** berdasarkan kondisi pada hipotesis ***c***

P(***c***) : probabilistik hipotesis ***c*** (prior probabilistik)

P(***c***|***x***) : probabilistik hipotesis ***c*** berdasarkan kondisi pada hipotesis ***x*** (posteriori probabilistik)

Contoh kasus 1:

Diberikan data seseorang berolah raga seperti yang dinyatakan pada Tabel 1. Tentukanlah apakah seseorang akan berolah raga jika diketahui *rain*, *mild*, *high*, dan *weak*?

Tabel 1. Data berolah raga sesorang

Day	Outlook	Temperature	Humidity	Wind	Play Tennis
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

Day	Outlook	Temperature	Humidity	Wind	Play Tennis
D14	Rain	Mild	High	Strong	No

Penyelesaian:

Berdasarkan Tabel 1 diperoleh bahwa banyaknya *sunny* adalah 5, yaitu: D1, D2, D8, D9, dan D11. Jumlah orang yang berolah raga dengan syarat *sunny* adalah 2, yaitu: D9 dan D11. Jumlah orang yang tidak berolah raga ketika *sunny* adalah 3, yaitu: D1, D2, dan D8. Dengan cara yang sama diperoleh tabel prekuensi orang berolah raga atau tidak seperti yang dinyatakan pada Tabel 2.

Tabel 2. Frekuensi seseorang bermain tennis

Tabel Frekuensi		**Play Tennis**	
		Yes	**No**
Outlook	Sunny	2	3
	Overcast	4	0
	Rain	3	2

Sehingga tabel likelihood dari tabel frekuensi seseorang bermain tennis adalah seperti yang dinyatakan pada Tabel 3.

Tabel 3. Likelihood ***outlook*** seseorang bermain tennis

Tabel Likelihood		**Play Tennis**		**Persentasi**
		Yes	**No**	
Outlook	Sunny	2/9	3/5	5/14
	Overcast	4/9	0/5	4/14
	Rain	**3/9**	**2/5**	5/14
		9/14	**5/14**	

Tabel 4. Likelihood ***temperature*** seseorang bermain tennis

Tabel Likelihood		**Play Tennis**		**Persentasi**
		Yes	**No**	
Temperature	Hot	2/9	2/5	4/14
	Mild	**4/9**	**2/5**	6/14
	Cool	2/9	2/5	4/14
		8/14	6/14	

Tabel 5. Likelihood ***humidity*** seseorang bermain tennis

Tabel Likelihood		**Play Tennis**		**Persentasi**
		Yes	**No**	
Humidity	High	**3/9**	**4/5**	7/14
	Normal	6/9	1/5	7/14

Tabel Likelihood	Play Tennis		Persentasi
	Yes	No	
	9/14	5/14	

Tabel 6. Likelihood ***wind*** seseorang bermain tennis

Tabel Likelihood		Play Tennis		Persentasi
		Yes	No	
Wind	Weak	**6/9**	**2/5**	8/14
	Strong	2/9	4/5	6/14
		8/14	6/14	

Berdasarkan Tabel 3 sampai dengan Tabel 6 diperoleh bahwa:

a. peluang seseorang bermain tennis adalah P(*yes*) = 9/14
b. peluang seseorang tidak bermain tennis adalah P(*no*) = 5/14
c. peluang *sunny* adalah P (*sunny*) = 5/14
d. peluang *overcast* adalah P(*overcast*) = 4/14
e. peluang *rain* adalah P(rain) = 5/14
f. peluang seseorang bermain tennis ketika *sunny* adalah P(*sunny*|*yes*) = 3/9
g. peluang seseorang tidak bermain tennis ketika *sunny* adalah P(*sunny*|*no*) = 2/5
h. peluang seseorang bermain tennis ketika *overcast* adalah P(*overcast*|*yes*) = 4/9
i. peluang seseorang tidak bermain tennis ketika *overcast* adalah P(*overcast*|*no*) = 0/5
j. peluang seseorang bermain tennis ketika *rain* adalah P(*rain*|*yes*) = 2/9
k. peluang seseorang tidak bermain tennis ketika *rain* adalah P(*rain*|*no*) = 3/5
l. peluang seseorang tidak bermain tennis ketika *rain* adalah P(*mild*|*yes*) = 4/9
m. peluang seseorang tidak bermain tennis ketika *rain* adalah P(*mild*|*no*) = 2/5
n. peluang seseorang tidak bermain tennis ketika *rain* adalah P(*high*|*yes*) = 3/9
o. peluang seseorang tidak bermain tennis ketika *rain* adalah P(*high*|*no*) = 4/5
p. peluang seseorang tidak bermain tennis ketika *rain* adalah P(*weak*|*yes*) = 6/9
q. peluang seseorang tidak bermain tennis ketika *rain* adalah P(*weak*|*no*) = 2/5

P(bermain tennis = *yes* | ***X***) = P(bermain_tennis=*yes*) * P(***outlook*** = *rain* | bermain_tennis= *yes*) *P (***temperatur*** = *mild* | bermain_tennis= *yes*) * P (***humidity*** = *high* | bermain_tennis= *yes*) * P (***wind*** = *weak* | bermain_tennis= *yes*)
= (9/14) * (3/9) * (4/9) * (3/9) * (6/9) = **0.02116402**

P(bermain tennis = *no* | ***X***) = P(bermain_tennis=*no*) * P(***outlook*** = *rain* | bermain_tennis= *no*) *P (***temperatur*** = *mild* | bermain_tennis= *no*) * P (***humidity*** = *high* | bermain_tennis= *no*) * P (***wind*** = *weak* | bermain_tennis= *no*)
= (5/14) * (3/5) * (2/5) * (4/5) * (2/5) = **0.02743**

Jadi, berdasarkan hasil perhitungan peluang bermain tennis di atas diperoleh bahwa hasil peluang (**ya**) bermain tennis = **0.01411** < peluang (**tidak**) bermain tennis = **0.02743** yaitu 1.41% < 2.74%, maka dapat disimpulkan bahwa seseorang bermain tennis dengan kondisi *outlook rain*, *temperature mild*, *humidity high*, dan *wind weak* TIDAK AKAN BERMAIN TENNIS.

Contoh kasus 2:

Tabel 7. Data sesorang melanggar rambu lalu lintas

Warna (X1)	Tipe (X2)	Asal (X3)	Tercuri (X4)
Merah	Sport	Domestik	Ya
Merah	Sport	Domestik	Tidak
Merah	Sport	Domestik	Ya
Kuning	SUV	Domestik	Tidak
Kuning	Sport	Import	Ya
Kuning	SUV	Import	Tidak
Kuning	SUV	Import	Ya
kuning	SUV	Domestik	Tidak
Merah	SUV	Import	Tidak
Merah	Sport	Import	Ya

Dari Tabel 7 di atas, data mobil yang melanggar rambu lalu lintas bisa dilihat dari atribut warna, tipe, dan asal. Misalkan kita ingin mengelompokkan mobil warna merah, tipe SUV, dan asal domestik. Tentukan probabilitas pelanggaran lalu lintas dan probabilitas tidak melanggar rambu lalu lintas, dan kemudian tentukan berapa persen mobil yang melanggar dan berapa persen mobil yang tidak melanggar, serta tentukan mobil dengan warna **merah**, tipe **SUV**, dan asal **domestik** tersebut melanggar lalu lintas atau tidak?
Penyelesaian:

Tabel 8. Frekuensi warna

		Tercuri	
		Ya	Tidak
Warna	Merah	3	2
	Kuning	2	3

Tabel 9. Likelihood warna

Likelihood		Tercuri		Persentasi
		Ya	Tidak	
Warna	Merah	3/5	2/5	5/10
	Kuning	2/5	3/5	5/10
		5/10	5/10	

Tabel 10. Frekuensi tipe

		Melanggar	
		Ya	Tidak
Tipe	Sport	4	1
	SUV	1	4

Tabel 11. Likelihood tipe

Likelihood		Melanggar		Persentasi
		Ya	Tidak	
Tipe	Sport	4/5	1/5	5/10
	SUV	1/5	4/5	5/10
		5/10	5/10	

Tabel 12. Frekuensi Asal

		Melanggar	
		Ya	Tidak
Asal	Domestik	2	3
	Import	3	2

Tabel 13. Likelihood Asal

Likelihood		Melanggar		Persentasi
		Ya	Tidak	
Asal	Domestik	2/5	3/5	5/10
	Import	3/5	2/5	5/10
		5/10	5/10	

Sehingga:

a. Peluang melanggar lalu lintas, P(ya) = 5/10 = 0.5
b. Peluang tidak melanggar lalu lintas, P(tidak) = 5/10 = 0.5
c. Peluang merah melanggar lalu lintas, P(merah| ya) = 3/5 = 0.6
d. Peluang SUV melanggar lalu lintas, P(SUV|ya) = 1/5 = 0.2
e. Peluang domestik melanggar lalu lintas, P(domestik|ya) = 2/5 = 0.4
f. Peluang merah tidak melanggar lalu lintas, P(merah| tidak) = 2/5 = 0.4
g. Peluang SUV tidak melanggar lalu lintas, P(SUV|tidak) = 4/5 = 0.8
h. Peluang domestik tidak melanggar lalu lintas, P(domestik|tidak) = 3/5 = 0.6
i. Peluang melanggar lalu lintas, P(melanggar=ya|X) = P(ya) * P(merah|ya) * P(SUV|ya) * P(domestik|ya) = 0.5 * 0.6 * 0.2 * 0.4 = 0.024 = 2.4%

j. Peluang tidak melanggar lalu lintas, P(melanggar =tidak|X) = P(tidak) * P(merah|tidak) * P(SUV| tidak) * P(domestik| tidak) = 0.5 * 0.6 * 0.8 * 0.6 = 0.144 = 14.4%.

Jadi, berdasarkan hasil perhitungan pelanggaran lalu lintas di atas dengan hasil pelanggaran (tidak) > pelanggaran (ya) yaitu 14.4% > 2.4% maka dapat disimpulkan mobil dengan warna merah, tipe SUV, dan asal domestik TIDAK MELANGGAR rambu lalu lintas.

Contoh kasus 3:

Dari hasil pengamatan sebelumnya, diperoleh bahwa penggunaan Listrik dipengaruhi oleh jumlah tanggungan keluarga, luas rumah, pendapatan per bulan, daya listrik yang digunakan, dan perlengkapan yang dimiliki. Adapun data hasil pengamatan yang dilakukan adalah sebagai berikut:

Tabel 14. Data penggunaan listrik

No	Jumlah Tanggungan Keluarga	Luas Rumah	Pendapatan/ Bulan	Daya Listrik	Perlengkapan Yang Dimiliki	Penggunaan Listrik
1	banyak	besar	besar	tinggi	sedang	***tinggi***
2	sedang	kecil	sedang	tinggi	tinggi	***tinggi***
3	sedang	standar	besar	rendah	tinggi	***sedang***
4	sedikit	standar	kecil	tinggi	tinggi	***tinggi***
5	sedang	besar	besar	tinggi	tinggi	***rendah***
6	sedikit	besar	besar	rendah	sedang	***tinggi***
7	sedang	besar	sedang	sedang	tinggi	***tinggi***
8	banyak	besar	besar	tinggi	tinggi	***sedang***
9	sedang	standar	besar	sedang	tinggi	***tinggi***
10	sedikit	standar	sedang	sedang	sedang	***tinggi***
11	sedikit	besar	kecil	tinggi	sedang	***sedang***
12	sedang	kecil	kecil	tinggi	tinggi	***sedang***
13	banyak	besar	besar	tinggi	sedang	***tinggi***
14	banyak	besar	besar	sedang	tinggi	***tinggi***
15	sedang	besar	besar	sedang	tinggi	***tinggi***
16	sedang	standar	besar	tinggi	tinggi	***tinggi***
17	banyak	standar	sedang	tinggi	tinggi	***sedang***
18	sedang	besar	besar	sedang	tinggi	***tinggi***
19	banyak	besar	sedang	tinggi	tinggi	***tinggi***

No	Jumlah Tanggungan Keluarga	Luas Rumah	Pendapatan/ Bulan	Daya Listrik	Perlengkapan Yang Dimiliki	Penggunaan Listrik
20	sedikit	besar	besar	sedang	rendah	***sedang***
21	sedang	standar	besar	sedang	tinggi	***tinggi***
22	banyak	standar	besar	tinggi	tinggi	***sedang***
23	banyak	besar	kecil	tinggi	tinggi	***tinggi***
24	banyak	besar	sedang	tinggi	tinggi	***tinggi***
25	sedang	besar	sedang	sedang	sedang	***sedang***
26	banyak	kecil	kecil	rendah	tinggi	***rendah***
27	sedang	standar	sedang	tinggi	sedang	***sedang***
28	banyak	besar	sedang	sedang	sedang	***tinggi***
29	sedang	standar	besar	tinggi	tinggi	***sedang***
30	banyak	kecil	sedang	tinggi	sedang	***tinggi***
31	banyak	besar	besar	sedang	rendah	***rendah***
32	sedikit	besar	sedang	rendah	tinggi	***tinggi***
33	sedang	besar	besar	tinggi	rendah	***rendah***
34	banyak	kecil	besar	sedang	tinggi	***sedang***
35	banyak	besar	besar	tinggi	rendah	***sedang***
36	sedang	standar	kecil	tinggi	tinggi	***sedang***
37	banyak	standar	besar	tinggi	rendah	***sedang***
38	banyak	besar	sedang	rendah	sedang	***rendah***
39	sedikit	besar	kecil	tinggi	sedang	***rendah***
40	banyak	besar	besar	tinggi	sedang	***tinggi***
41	banyak	standar	besar	tinggi	tinggi	***sedang***
42	sedikit	besar	besar	tinggi	tinggi	***sedang***
43	banyak	besar	besar	tinggi	tinggi	***tinggi***
44	sedikit	besar	sedang	sedang	tinggi	***tinggi***
45	banyak	kecil	kecil	tinggi	tinggi	***tinggi***
46	banyak	standar	sedang	sedang	tinggi	***rendah***
47	banyak	kecil	sedang	tinggi	tinggi	***tinggi***
48	sedang	besar	sedang	sedang	tinggi	***sedang***
49	banyak	besar	sedang	tinggi	tinggi	***rendah***
50	sedang	besar	besar	tinggi	tinggi	***rendah***
51	banyak	kecil	kecil	tinggi	tinggi	***tinggi***
52	sedang	standar	kecil	tinggi	sedang	***tinggi***
53	sedang	kecil	sedang	rendah	tinggi	***sedang***

No	Jumlah Tanggungan Keluarga	Luas Rumah	Pendapatan/ Bulan	Daya Listrik	Perlengkapan Yang Dimiliki	Penggunaan Listrik
54	banyak	besar	sedang	tinggi	tinggi	***sedang***
55	banyak	standar	besar	sedang	tinggi	***tinggi***
56	banyak	kecil	sedang	tinggi	tinggi	***sedang***
57	sedang	besar	besar	rendah	tinggi	***tinggi***
58	banyak	besar	besar	tinggi	tinggi	***sedang***

Tentukanlah *Correctly* dan *incorrectly classified instance* dari data ***penggunaan listrik*** yang diberikan.

Penyelesaikan:

1. Menghitung Probabilitas class penggunaan listrik:

   Jumlah kelas penggunaan listrik adalah 3, yaitu: tinggi, sedang, dan rendah. Dengan jumlah masing-masing adalah: 28, 21, dan 9. Seperti yang dinyatakan pada **Tabel** 15.

   Probabilitas (Tinggi) = jumlah_tinggi/total_data $= \frac{\sum_{i=0}^{58} Peng\_listrik_{tinggi}}{n} \frac{28}{58} = 0.4828$

   Tabel 15. Probabilitas Penggunaan Listrik

Jumlah Kejadian 'Penggunaan Listrik'			Probabilitas		
Tinggi	Sedang	Rendah	Tinggi	Sedang	Rendah
28	21	9	0,4828	0,3621	0,1552

2. Menghitung Probabilitas bersyaratnya:

   Pada kasus ini, atribut bersayaratnya ada 5 (lima), yaitu:

   a. Jumlah tanggungan keluarga
   b. Luas tanah
   c. Pendapatan per bulan
   d. Daya listrik'
   e. Perlengkapan yang dimiliki.

   Perhitungan numerik terhadap peluang bersyarat Penggunaan Listrik dengan syarat Jumlah tanggungan adalah sebegai berikut:

   Pada atribut Jumlah tanggungan memiliki 3 kelas, yaitu: banyak, sedang, dan sedikit. Adapun jumlah pengguna listrik dengan syarat jumlah tanggungan adalah seperti yang dinyatakan pada **Tabel** 16.

Tabel 16. Probabilitas Jumlah Tanggungan

Jumlah Tanggungan	Jumlah Kejadian 'Pengguna Listrik'			Probabilitas		
	Sedang	Rendah	Tinggi	Sedang	Rendah	Tinggi
Banyak	10	5	14	0,4762	0,5556	0,5000
Sedang	8	3	9	0,3810	0,3333	0,3214
Sedikit	3	1	5	0,1429	0,1111	0,1786
	21	9	28			

Nilai 10 menyatakan bahwa jumlah pengguna listrik **sedang** dan jumlah tanggungannya **banyak** adalah 10 keluarga. Nilai 8 menyatakan bahwa jumlah pengguna listrik **sedang** dan jumlah tanggungannya **sedang** adalah 8 keluarga. Sedangkan untuk pengguna listrik **sedang** dengan jumlah tanggungan **sedikit** ada sebanyak 3 keluarga. Sehingga jumlah total pengguna listrik sedang adalah (10+8+3) = 21. Probabilitas pengguna listrik **sedang** dengan jumlah tanggungan **banyak** adalah 0.4762. Dengan cara yang sama dilakukan untuk probabilitas setiap kelas penggunaan listrik dengan syarat kelas jumlah tanggungan.

Dengan cara yang sama, untuk probabilitas bersyarat lainnya diperoleh seperti yang dinyatakan tabel berikut:

Probabilitas Luas Tanah

Luas Tanah	Jumlah Kejadian 'Pengguna Listrik'			Probabilitas		
	Sedang	Rendah	Tinggi	Sedang	Rendah	Tinggi
besar	9	7	16	0,428571429	0,777777778	0,571428571
kecil	4	1	5	0,19047619	0,111111111	0,178571429
standard	8	1	7	0,380952381	0,111111111	0,25
	21	9	28			

Probabilitas Pendapatan

Pendapatan	Jumlah Kejadian 'Pengguna Listrik'			Probabilitas		
	Sedang	Rendah	Tinggi	Sedang	Rendah	Tinggi
kecil	3	2	5	0,142857143	0,222222222	0,178571429
sedang	7	3	10	0,333333333	0,333333333	0,357142857
besar	11	4	13	0,523809524	0,444444444	0,464285714
	21	9	28			

Probabilitas Daya Listrik

Daya Listrik	Jumlah Kejadian 'Pengguna Listrik'			Probabilitas		
	Sedang	Rendah	Tinggi	Sedang	Rendah	Tinggi
rendah	2	2	3	0,095238095	0,222222222	0,107142857
sedang	4	2	10	0,19047619	0,222222222	0,357142857
tinggi	15	5	15	0,714285714	0,555555556	0,535714286
	21	9	28			

Probabilitas Perlengkapan

Perlengkapan	Jumlah Kejadian 'Pengguna Listrik'			Probabilitas		
	Sedang	Rendah	Tinggi	Sedang	Rendah	Tinggi
rendah	3	2	0	0,142857143	0,222222222	0,032258065
sedang	3	2	8	0,142857143	0,222222222	0,290322581
tinggi	15	5	20	0,714285714	0,555555556	0,677419355
	21	9	28			

3. Pengujian Metode Naïve Bayes terhadap penggunaan Listrik

   Hitung probabilitas penggunaan listrik dengan syarat dari setiap kategori kelas yang diberikan. Pada kasus pertama, jumlah tanggungan keluarga (x_1) = banyak, luas rumah (X_2) = besar, pendapatan per bulan (X_3) = besar, daya listrik yang digunakan (X_4) = tinggi, dan perlengkapan yang dimiliki (X_5) = sedang.

   Sehingga

   P(penggunaan=rendah | X) = (P(**jumlah tanggungan** = banyak | penggunaan=rendah) * P(**luas tanah** =besar | penggunaan=rendah) * P(**pendapatan** = besar |penggunaan = rendah) * P(**daya listrik** = tinggi | penggunaan = rendah) * P(perlengkapan = sedang | penggunaan = rendah)) * P(penggunaan = rendah) = 0,5556 *0,7778 *0,4444*0,5556*0,2222*0,15517241 =0,003679

   Dengan cara yang sama untuk

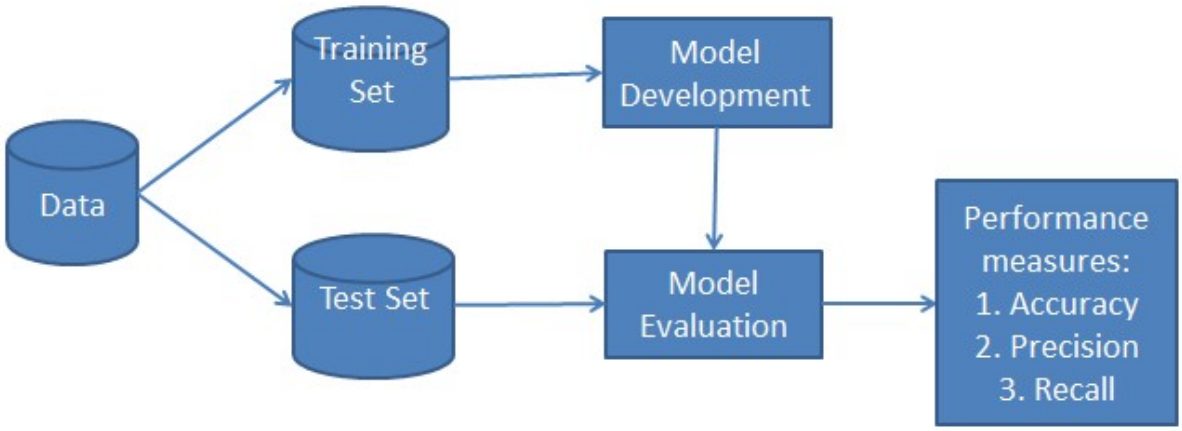
   - P(penggunaan=sedang|X) = 0,003950
   - P(penggunaan=tinggi|X) = 0,009960

   Dari probabilitas ketiga kelas peluang terbesar adalah 0.009960, sehingga prediksi sistem adalah pengguna tinggi.

No	Class	Input Kategorikal					Probabilitas_Pengguna			Probabilitas			Prob_max	Predict System	Kinerja
		X_1	X_2	X_3	X_4	X_5	rendah	sedang	tinggi	rendah	sedang	tinggi			
1	*tinggi*	banyak	besar	besar	tinggi	sedang	0,15517241	0,36206897	0,48275862	0,003679	0,003950	0,009960	0,009960	tinggi	1
2	*tinggi*	sedang	kecil	sedang	tinggi	tinggi	0,15517241	0,36206897	0,48275862	0,000591	0,004468	0,003591	0,004468	sedang	0
3	*sedang*	sedang	standar	besar	rendah	tinggi	0,15517241	0,36206897	0,48275862	0,000315	0,001872	0,001307	0,001872	sedang	1
4	*tinggi*	sedikit	standar	kecil	tinggi	tinggi	0,15517241	0,36206897	0,48275862	0,000131	0,001436	0,001397	0,001436	sedang	0
5	*rendah*	sedang	besar	besar	tinggi	tinggi	0,15517241	0,36206897	0,48275862	0,005519	0,015798	0,014940	0,015798	sedang	0
6	*tinggi*	sedikit	besar	besar	rendah	sedang	0,15517241	0,36206897	0,48275862	0,000294	0,000158	0,000711	0,000711	tinggi	1
7	*tinggi*	sedang	besar	sedang	sedang	tinggi	0,15517241	0,36206897	0,48275862	0,001656	0,002681	0,007662	0,007662	tinggi	1
8	*sedang*	banyak	besar	besar	tinggi	tinggi	0,15517241	0,36206897	0,48275862	0,009198	0,019748	0,023240	0,023240	tinggi	0
9	*tinggi*	sedang	standar	besar	sedang	tinggi	0,15517241	0,36206897	0,48275862	0,000315	0,003745	0,004358	0,004358	tinggi	1
10	*tinggi*	sedikit	standar	sedang	sedang	sedang	0,15517241	0,36206897	0,48275862	0,000032	0,000179	0,000798	0,000798	tinggi	1

No	Class	Input Kategorikal					Probabilitas_Pengguna			Probabilitas			Prob_max	Predict System	Kinerja
		X_1	X_2	X_3	X_4	X_5	rendah	sedang	tinggi	rendah	sedang	tinggi			
52	*tinggi*	sedang	standar	kecil	tinggi	sedang	0,15517241	0,36206897	0,48275862	0,000158	0,000766	0,001077	0,001077	tinggi	1
53	*sedang*	sedang	kecil	sedang	rendah	tinggi	0,15517241	0,36206897	0,48275862	0,000237	0,000596	0,000718	0,000718	tinggi	0
54	*sedang*	banyak	besar	sedang	tinggi	tinggi	0,15517241	0,36206897	0,48275862	0,006898	0,012567	0,017877	0,017877	tinggi	0
55	*tinggi*	banyak	standar	besar	sedang	tinggi	0,15517241	0,36206897	0,48275862	0,000526	0,004681	0,006778	0,006778	tinggi	1
56	*sedang*	banyak	kecil	sedang	tinggi	tinggi	0,15517241	0,36206897	0,48275862	0,000985	0,005585	0,005587	0,005587	tinggi	0
57	*tinggi*	sedang	besar	besar	rendah	tinggi	0,15517241	0,36206897	0,48275862	0,002207	0,002106	0,002988	0,002988	tinggi	1
58	*sedang*	banyak	besar	besar	tinggi	tinggi	0,15517241	0,36206897	0,48275862	0,009198	0,019748	0,023240	0,023240	tinggi	0
															36
													Correctly Classified Instance		62,07 %
													Incorrectly Classified Instance		37,93 %

Berdasarkan hasil prediksi yang dilakukan diperoleh data jumlah yang benar diprediksi sebanyak 36, sedangkan yang tidak tepat sebanyak 22. Sehingga nilai *Correctly classified*-nya adalah 62.07%, sedangkan *incorrectly classified* nya adalah 37.93%.



https://www.datacamp.com/community/tutorials/categorical-data

https://pbpython.com/categorical-encoding.html