

Home Work #3: Naïve Bayes

(Tugas dikerjakan sendiri dengan tulisan tangan dan dibuat softcopy pdf-nya yang masih mudah terbaca, tidak buram karena kurang cahaya atau terlihat kabur tulisannya.

Pengumpulan dilakukan secara online di situs kuliah.itb.ac.id paling lambat hari Kamis, 24 September pukul 23.55)

1. Diberikan tabel berikut dengan PlayTennis sebagai target output.

Day	<i>Outlook</i>	<i>Temperature</i>	<i>Humidity</i>	<i>Wind</i>	PlayTennis
D1	Rain	15	Normal	Weak	Yes
D2	Rain	16	Normal	Strong	No
D3	Overcast	18	Normal	Strong	Yes
D4	Sunny	25	High	Weak	No
D5	Sunny	17	Normal	Weak	Yes
D6	Rain	24	Normal	Strong	No
D7	Sunny	26	Normal	Strong	Yes
D8	Overcast	28	High	Strong	Yes
D9	Sunny	34	Normal	Weak	Yes
D10	Rain	27	High	Strong	No

Tabel 1. Data Pelatihan

Lengkapi nilai probability target output dan atribut untuk kelas tertentu dengan menggunakan *Maximum Likelihood Estimate* (MLE) dan tabel sebagai berikut:

PlayTennis	P(PlayTennis)
Yes	
No	

Outlook	P(Outlook PlayTennis=yes)
Rain	
Overcast	
Sunny	

Outlook	P(Outlook PlayTennis=no)
Rain	
Overcast	
Sunny	

Humidity	P(Humidity PlayTennis=yes)
Normal	
High	

Humidity	(Humidity PlayTennis=no)
Normal	
High	

Wind	P(Wind PlayTennis=yes)
Weak	
Strong	

Wind	P(Wind PlayTennis=no)
Weak	
Strong	

Jika atribut *Temperature* bernilai kontinyu dan untuk ini diasumsikan memiliki distribusi normal (Gaussian) $N = (\mu, \sigma)$. Tuliskan formula untuk menghitung probability atribut temperatur untuk kasus di mana μ dan σ independet terhadap kelas target (PlayTennis) dan untuk kasus di mana μ dan σ tidak independent terhadap kelas target.

2. Berikut diberikan daftar judul artile yang telah diberi label apakah menarik (Interesting) untuk di baca atau tidak. Judaul artikel telah di pra-proses sehingga siap untuk diproses lebih lanjut algoritma machine learning.

	Title of Article	Interesting?
1	oracle open sources java machine learning library	yes
2	the reason trump not trying save economy	no
3	it fed day time for the economy to get tuneup	no
4	auto ml bridging skill gaps with machine learning	yes
5	donald trump chance to peel off hispanic voters	no
6	definining ai machine learning deep learning martech	yes
7	ai macine learning tool can help separate covid related claim	yes
8	dbz taking machine learning for spin	yes
9	deepmind introduce machine learning model improve google maps	yes
10	we must use covid crisis reheape society economy	no

Tentukan kelas *interestingness* (yes/no) dari dua judul artikel berikut dengan Naïve Bayes :

1. system target ai adoption challenge machine learning builder
2. trump destroying economy though

Lakukan estimasi dengan *Maximum a Priory* (MAP) dengan asumsi bahwa semua kata (baik yang muncul dalam vocabulary atau tidak) memiliki kemunculan ‘halucinated’ sekali sebagai *prior knowledge* nya. Tunjukkan cara menghitungnya dimulai dengan formula yang dipakai, misal $P(\text{yes} | \text{kalimat}) = \dots\dots$, lalu isikan nilai probability sesuai dengan urutan dalam formula dalam bentuk pembilang/penyebut. Pecahan desimal hanya ditulis untuk hasil akhir.

Hint: hitung nilai probability yang hanya diperlukan untuk dapat mengklasifikasn kedua kalimat yang ditanyakan kelasnya di atas.