

Daftar Kelompok :  
 1. Michael Peter Asher 41518120010  
 2. Firman Hidayat 41518120023

Training Set

No	Number of Legs	Number of Hands	Weight	Label
1	4	0	388	0
2	4	0	394	0
3	4	0	400	0
4	2	2	60	1
5	2	2	80	1
6	2	2	70	1

Testing Set

No	Number of Legs	Number of Hands	Weight	Label
1	4	0	80	1
2	2	2	110	1
3	4	0	350	0
4	2	2	80	1

Cow	0
Human	1

Kekurangan algoritma K Nearest Neighbour :

1. Perlu mengetahui Jumlah K (tetangga terdekat).
2. Berdasarkan perhitungan nilai jarak, tidak jelas perhitungan jarak mana yang sebaiknya digunakan dan atribut mana yang memberikan hasil terbaik.
3. Perlu komputasi yang tinggi karena perlu menghitung jarak dari nilai yang sedang di prediksi ke semua data yang ada di data training. data training yang terlalu banyak menyebabkan b
4. Rentan terhadap variabel yang non-informatif, kNN tidak dapat mengetahui variabel mana yang signifikan dalam klasifikasi dan mana yang tidak. Pada kasus ini, klasifikasi sebetulnya bisa dilakukan menggunakan variabel jumlah kaki dan jumlah tangan saja. Namun, kNN tidak dapat mengetahui bahwa variabel Bobot tidak informatif. Alhasil, instance nomor 1 kelas sapi terklasifikasi dengan salah, karena instance tersebut dekat dengan instance kelas manusia dalam dimensi Jumlah Kaki dan Jumlah Tangan. Andaikan kita hanya menggunakan variabel Jumlah kaki dan jumlah tangan dan membuang variabel Bobot, semua instance akan terklasifikasi dengan tepat. Pemilihan variabel sebelum menerapkan kNN dapat membantu menangani permasalahan di atas. Selain itu, kita juga bisa memberi bobot pada variabel dalam perhitungan jarak antar-instance. Variabel yang kita tahu noninformatif kita beri bobot yang kecil.
5. Rentan terhadap perbedaan rentang variabel  
 Dalam perhitungan jarak antar-instance, kNN menganggap semua variabel setara atau sama penting. Jika terdapat satu variabel yang memiliki rentang jauh lebih besar dibanding variabel-variabel lainnya, maka perhitungan jarak akan didominasi oleh tersebut. variabel, bobot, dengan rentan 0 sampai 1.000 dan variabel jumlah kaki dan tangan dengan rentan 0 sampai 4. Kuadrat selisih nilai variabel jumlah kaki dan tangan tidak akan lebih dari 10, sedangkan untuk variabel bobot kuadrat selisihnya bisa mencapai 1.000.000. Hal ini bisa mengecoh kNN sehingga kNN menganggap bobot tidak membawa pengaruh dalam perhitungan jarak karena rentangnya sangat besar dibanding rentang jumlah tangan dan kaki.  
 Untuk mengatasi perbedaan rentang, biasanya dilakukan preproses berupa standardisasi rentang semua variabel sebelum menerapkan algoritma kNN. Contohnya yaitu melalui operasi centre-scale atau operasi min-max.

Testing No 1

Distance	K=1	K=3	K=5	K=6
308			0	0
314			0	0
320				0
20.19900988		1	1	1
2.828427125	1	1	1	1
10.39230485		1	1	1

Testing No 3

Distance	K=1	K=3	K=5	K=6
38	0	0	0	0
44		0	0	0
50		0	0	0
290.0137928				1
270.0148144			1	1
280.0142853			1	1

Testing No 2

Distance	K=1	K=3	K=5	K=6
278.0144			0	0
284.0141			0	0
290.0138				0
50		1	1	1
30	1	1	1	1
40		1	1	1

Testing No 4

Distance	K=1	K=3	K=5	K=6
308.013			0	0
314.0127			0	0
320.0125				0
20		1	1	1
0	1	1	1	1
10		1	1	1

Saat variabel Bobot di buang pada kasus instance nomor 1 maka akan menghasilkan klasifikasi sapi, sehingga variabel bobot kurang informatif

Testing No 1 tanpa bobot

Distance	K=1	K=3	K=5	K=6
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
2.828427125			1	1
2.828427125			1	1
2.828427125			1	1

Berikut Penjelasan Langkah Testing pada Data Testing Nomor 3 dengan algoritma KNN

$$d_1 = \sqrt{((4-4))^2 + ((0-0))^2 + ((388-350))^2}$$

$$d_1 = \sqrt{(0+0+((388-350))^2)}$$

$$d_1 = \sqrt{((388-350))^2}$$

$$d_1 = 38$$

$$d_2 = \sqrt{((4-4))^2 + ((4-4))^2 + ((394-350))^2}$$

$$d_2 = \sqrt{(0+0+((394-350))^2)}$$

$$d_2 = \sqrt{((394-350))^2}$$

$$d_2 = 44$$

$$d_3 = \sqrt{((4-4))^2 + ((4-4))^2 + ((400-350))^2}$$

$$d_3 = \sqrt{(0+0+((400-350))^2)}$$

$$d_3 = \sqrt{((400-350))^2}$$

$$d_3 = 50$$

$$d_4 = \sqrt{((2-4))^2 + ((2-4))^2 + ((60-350))^2}$$

$$d_4 = \sqrt{((-2))^2 + ((-2))^2 + ((-290))^2}$$

$$d_4 = \sqrt{4+4+84100}$$

$$d_4 = \sqrt{84108}$$

$$d_5 = \sqrt{((2-4))^2 + ((2-4))^2 + ((80-350))^2}$$

$$d_5 = \sqrt{((-2))^2 + ((-2))^2 + ((-270))^2}$$

$$d_5 = \sqrt{4+4+72900}$$

$$d_5 = \sqrt{72908}$$

$$d_6 = \sqrt{((2-4))^2 + ((2-4))^2 + ((70-350))^2}$$

$$d_6 = \sqrt{((-2))^2 + ((-2))^2 + ((-280))^2}$$

$$d_6 = \sqrt{4+4+78400}$$