

# Data Mining - Latihan 1

Teknik Informatika  
Universitas Al-Azhar Indonesia

March 15, 2017

## Klasifikasi Bayesian

1. Sebuah taksi terlibat dalam tabrak lari semalam. Ada dua perusahaan taksi, Hijau dan Biru, yang beroperasi di kota tersebut. Anda diberikan data sebagai berikut:
  - (a) 85% dari taksi yang beroperasi di kota adalah taksi Hijau dan 15% adalah taksi Biru.
  - (b) Seorang saksi mengidentifikasi bahwa taksi yang terlibat dalam kecelakaan pada malam tersebut adalah taksi Biru. Pengadilan kemudian menguji saksi dengan mereka ulang kejadian pada malam kecelakaan terjadi. Hasilnya, saksi tersebut dapat dengan benar mengidentifikasi tiap-tiap operator taksi 80% dan melakukan kesalahan pada 20% sisanya.

Berapa peluangnya bahwa taksi yang terlibat dalam kecelakaan tersebut adalah taksi Biru? [1]

2. 100 dari 10.000 wanita berusia 40 tahun yang berpartisipasi dalam penapisan rutin mengidap kanker payudara. 80 dari 100 wanita dengan kanker payudara akan mendapatkan hasil tes *mammography* positif. 950 dari 9.900 tanpa kanker payudara ternyata juga akan mendapatkan hasil tes *mammography* positif. Jika ada 10.000 wanita yang menjalani proses penapisan rutin ini, berapa banyak wanita dengan hasil tes *mammography* positif ternyata memang mengidap kanker payudara? [2]

## Naïve Bayes

1. Anda diberikan data dengan format  $(x_1, x_2, x_3, y)$ , dengan  $x_1$ ,  $x_2$ , dan  $x_3$  adalah fitur, dan  $y$  adalah kelas. Setiap elemen tersebut bernilai 0 atau 1. Data latih yang Anda dapatkan adalah sebagai berikut:

$(1, 0, 1, 1), (0, 1, 0, 1), (0, 1, 1, 0), (0, 0, 0, 0), (1, 1, 1, 1), (0, 0, 0, 0), (1, 0, 1, 0)$

Apakah kemungkinan label yang paling besar dari data baru dengan  $x_1 = 1$ ,  $x_2 = 1$ , dan  $x_3 = 0$ ?

2. Terdapat 15 data yang diambil dari Iris dataset seperti pada Table 2. Asumsikan bahwa tiap atribut terdistribusi Gaussian. Apakah kemungkinan label untuk data baru berikut:
  - (a) petal\_length = 5.1, petal\_width = 2.4
  - (b) petal\_length = 4.2, petal\_width = 1.5
  - (c) petal\_length = 1.1, petal\_width = 0.1
  - (d) petal\_length = 4.0

## References

- [1] Tversky, A. and Kahneman, D., 1982. Evidential impact of base rates. In *Judgement under uncertainty: Heuristics and biases*. Cambridge University Press.
- [2] Tan, P.N., 2006. *Introduction to Data Mining*. Pearson Education India.

Table 1: Dataset Iris		
petal_length	petal_width	species
1.7	0.4	setosa
1.4	0.3	setosa
1.5	0.2	setosa
1.4	0.2	setosa
1.5	0.1	setosa
4.5	1.3	versicolor
4.7	1.6	versicolor
3.3	1.0	versicolor
4.6	1.3	versicolor
3.9	1.4	versicolor
6.6	2.1	virginica
4.5	1.7	virginica
6.3	1.8	virginica
5.8	1.8	virginica
6.1	2.5	virginica