



## A black and white line drawing of a miner. The miner is wearing a hard hat and a long-sleeved shirt, and is bent over, using a pickaxe to break rocks. A small cart filled with rocks is on a track next to the miner.



# Profil



❖ Nama : Junta Zeniarja, M.Kom

❖ Alamat : Semarang

❖ Kontak

- Phone : -
- E-mail : [junta@dsn.dinus.ac.id](mailto:junta@dsn.dinus.ac.id)
- Room : Ruang Dosen TI-S1 (H.2.3)

❖ Pendidikan

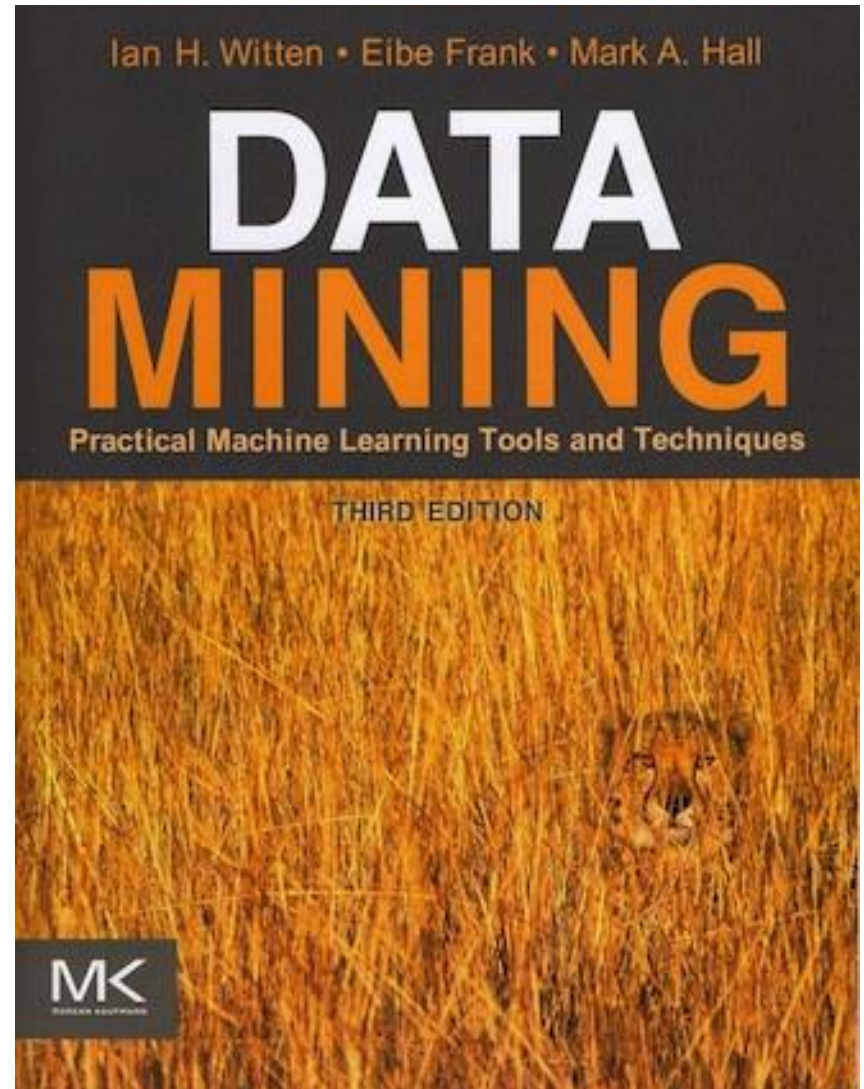
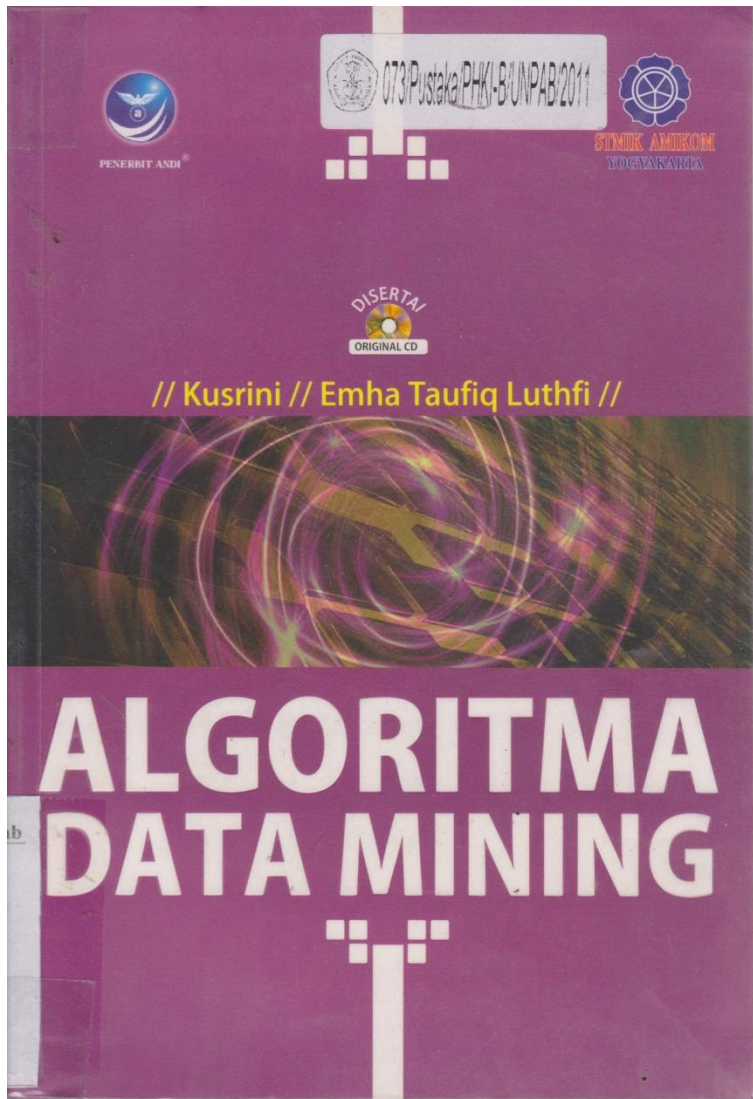
- S1 => TI – UDINUS
- S2 => TI – UDINUS
- S2 => Computer Science UTeM (Universiti Teknikal Malaysia Melaka)

❖ Konsultasi - Sharing

- 1:00 pm – 4:00 pm, Senin-Kamis.
- Appointment via phone or e-mail preferred



# Textbooks





## 1. Algoritma Data Mining

- Algoritma C4.5
- Algoritma Nearest Neighbor
- Bayesian Classification
- Algoritma Apriori
- Algoritma Fuzzy C Means



# **Algoritma Bayesian Classification**

# Introduction



- Bayesian Classification adalah pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu class.
- Bayesian Classification didasarkan pada teorema Bayes yang memiliki kemampuan klasifikasi serupa *decision tree* dan *neural network*.
- Bayesian Classification terbukti memiliki akurasi dan kecepatan yang tinggi saat diaplikasikan ke dalam database dengan data yang besar.

# Rumus Teorema Bayes



Teorema Bayes memiliki bentuk umum sbb :

$$P(H|X) = \frac{P(X|H) P(H)}{P(X)}$$

Keterangan :

X : data dengan *class* yang belum diketahui

H : hipotesis data X merupakan suatu *class* spesifik

P(H|X) : probabilitas hipotesis H berdasar kondisi X  
(*posteriori probability*)

P(H) : probabilitas hipotesis H (*prior probability*)

P(X|H) : probabilitas X berdasar kondisi hipotesis H

P(X) : probabilitas dari X



## ***Data training "All Electronics customer database"***

<b>Id</b>	<b>Age</b>	<b>Income</b>	<b>Student</b>	<b>Credit_rating</b>	<b>Class: buys_computer</b>
1	<=30	High	No	Fair	No
2	<=30	High	No	Excellent	No
3	31..40	High	No	Fair	Yes
4	>40	Medium	No	Fair	Yes
5	>40	Low	Yes	Fair	Yes
6	>40	Low	Yes	Excellent	No
7	31..40	Low	Yes	Excellent	Yes
8	<=30	Medium	No	Fair	No
9	<=30	Low	Yes	Fair	Yes
10	>40	Medium	Yes	Fair	Yes
11	<=30	Medium	Yes	Excellent	Yes
12	31..40	Medium	No	Excellent	Yes
13	31..40	High	Yes	Fair	Yes
14	>40	Medium	No	Excellent	No





- ❖ Terdapat dua class dari klasifikasi yang dibentuk yaitu :
  - C1 => buys\_computer = yes
  - C2 => buys\_computer = no
- ❖ Misal terdapat data X (belum diketahui *class*-nya).
- ❖ X = (age="≤30", income="Medium", student="Yes", credit\_rating="Fair")

# Penyelesaian (1)

- ❖ Dibutuhkan untuk memaksimalkan  $P(X|C_i)$   $P(C_i)$  untuk  $i=1, 2$ .
- ❖  $P(C_i)$  merupakan prior probability untuk setiap class berdasar data contoh:
  - $P(\text{buys\_computer}=\text{"yes"}) = 9/14 = 0.643$
  - $P(\text{buys\_computer}=\text{"no"}) = 5/14 = 0.357$
- ❖ Hitung  $P(X|C_i)$ , untuk  $i=1, 2$ .
  - $P(\text{age}=\text{"<=30"}|\text{buys\_computer}=\text{"yes"}) = 2/9 = 0.222$
  - $P(\text{age}=\text{"<=30"}|\text{buys\_computer}=\text{"no"}) = 3/5 = 0.600$
  - $P(\text{income}=\text{"medium"}|\text{buys\_computer}=\text{"yes"}) = 4/9 = 0.444$
  - $P(\text{income}=\text{"medium"}|\text{buys\_computer}=\text{"no"}) = 2/5 = 0.400$

## Penyelesaian (2)

- $P(\text{student}=\text{"yes"}|\text{buys\_computer}=\text{"yes"}) = 6/9 = 0.667$
- $P(\text{student}=\text{"yes"}|\text{buys\_computer}=\text{"no"}) = 1/5 = 0.200$
- $P(\text{credit\_rating}=\text{"fair"}|\text{buys\_computer}=\text{"yes"}) = 6/9 = 0.667$
- $P(\text{credit\_rating}=\text{"fair"}|\text{buys\_computer}=\text{"no"}) = 2/5 = 0.400$
- $$P(X|\text{buys\_computer}=\text{"yes"}) = 0.222 * 0.444 * 0.677 * 0.677$$
$$= 0.044$$
- $$P(X|\text{buys\_computer}=\text{"no"}) = 0.600 * 0.400 * 0.200 * 0.400$$
$$= 0.019$$
- $$P(X|\text{buys\_computer}=\text{"yes"}) P(\text{buys\_computer}=\text{"yes"})$$
$$= 0.044 * 0.643 = 0.028$$
- $$P(X|\text{buys\_computer}=\text{"no"}) P(\text{buys\_computer}=\text{"no"})$$
$$= 0.019 * 0.357 = 0.007$$

Kesimpulan: **buys\_computer = "yes"**

# Latihan



❖ Tentukan klas label dari X:

X = (Outlook=Rain, Temperature=Cool, Humidity=High, Wind=Weak)

Scenario	Outlook	Temperature	Humidity	Wind	PlayTennis
Day 1	Sunny	Hot	High	Weak	No
Day 2	Sunny	Hot	High	Strong	No
Day 3	Overcast	Hot	High	Weak	Yes
Day 4	Rain	Mild	High	Weak	Yes
Day 5	Rain	Cool	Normal	Weak	Yes
Day 6	Rain	Cool	Normal	Strong	No
Day 7	Overcast	Cool	Normal	Strong	Yes
Day 8	Sunny	Mild	High	Weak	No
Day 9	Sunny	Cool	Normal	Weak	Yes
Day 10	Rain	Mild	Normal	Weak	Yes
Day 11	Sunny	Mild	Normal	Strong	Yes
Day 12	Overcast	Mild	High	Strong	Yes
Day 13	Overcast	Hot	Normal	Weak	Yes
Day 14	Rain	Mild	High	Strong	No

# Referensi



1. Ian H. Witten, Frank Eibe, Mark A. Hall, Data mining: Practical Machine Learning Tools and Techniques 3rd Edition, *Elsevier*, 2011
2. Kusrini, Taufiq Emha, Algoritma Data Mining, *Penerbit Andi*, 2009

