



Data Mining



Junta Zeniarja, M.Kom, M.CS
junta@dsn.dinus.ac.id

Profil



❖ Nama : Junta Zeniarja, M.Kom

❖ Alamat : Semarang

❖ Kontak

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

❖ Pendidikan

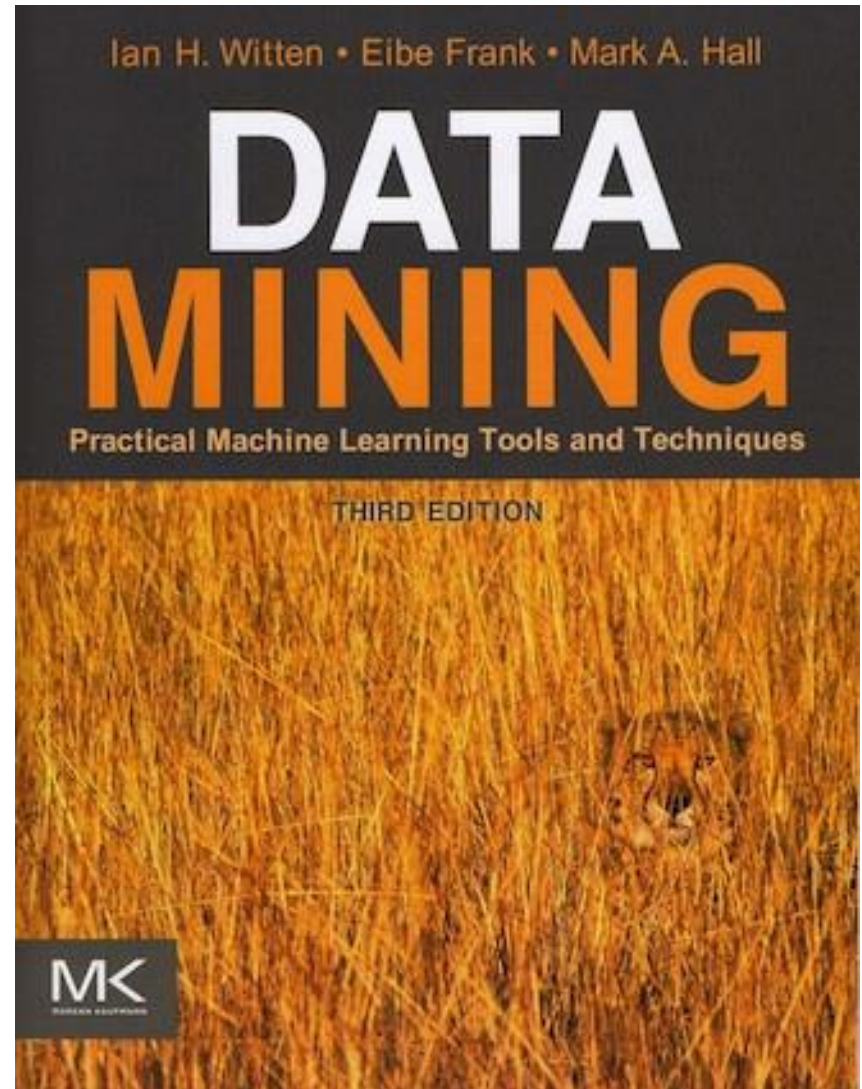
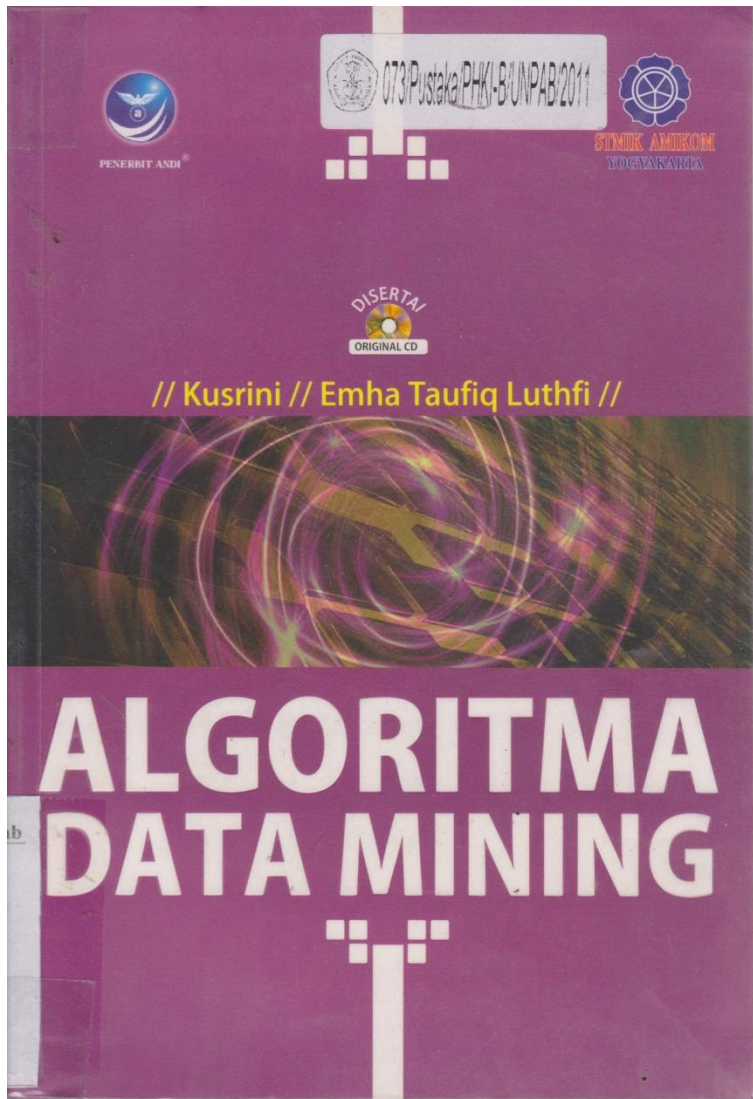
- S1 => TI – UDINUS
- S2 => TI – UDINUS
- S2 => Computer Science UTeM (Universiti Teknikal Malaysia Melaka)

❖ 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**
- Algoritma Apriori
- Algoritma Fuzzy C Means
- Bayesian Classification



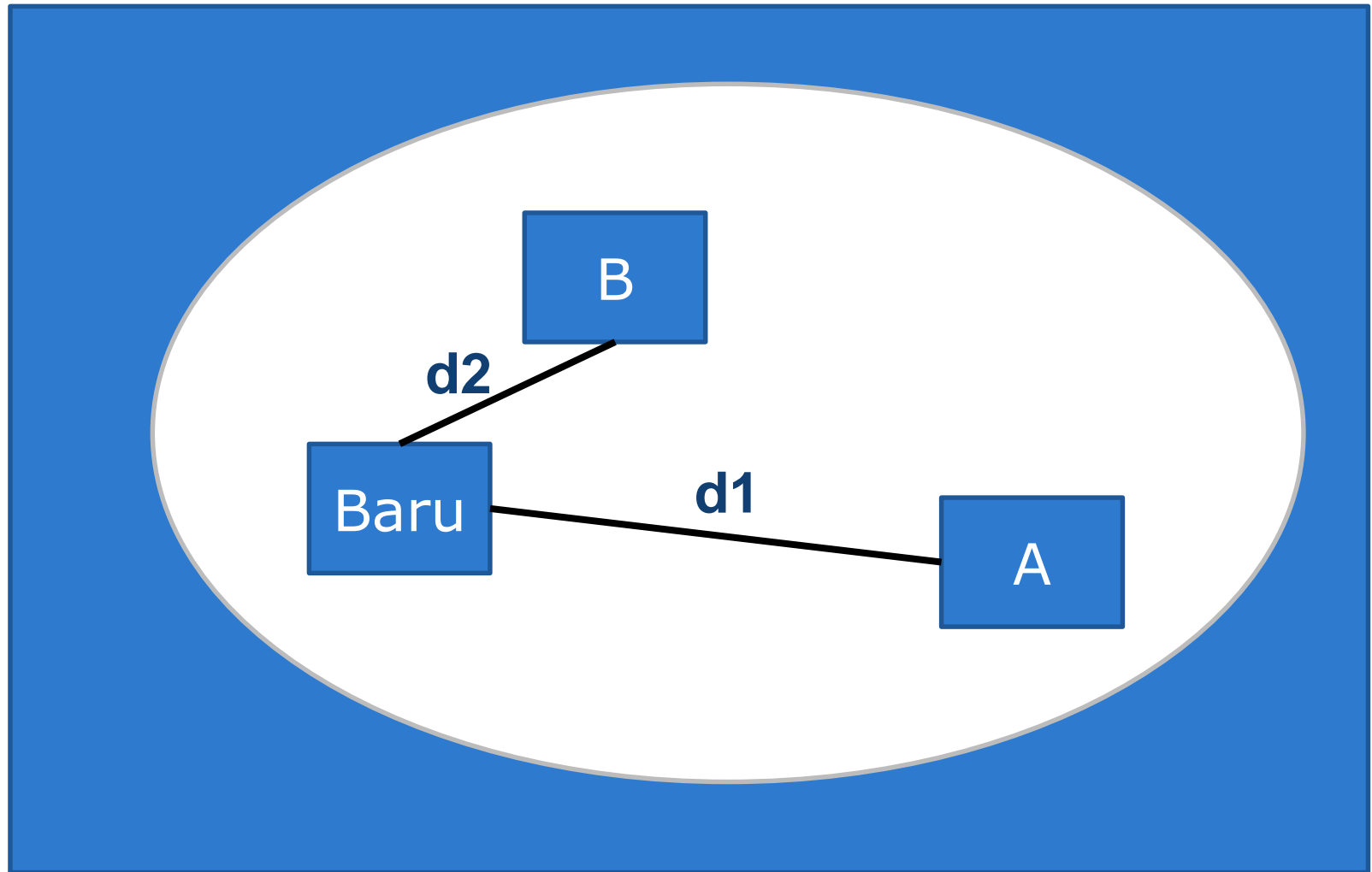
Algoritma Nearest Neighbor

Introduction



- Nearest Neighbor merupakan pendekatan untuk mencari kasus lama, yaitu berdasarkan pada pencocokan bobot dari sejumlah fitur yang ada.
- Misal untuk mencari solusi terhadap pasien baru dengan menggunakan solusi dari pasien terdahulu.
- Dengan menghitung kedekatan kasus pasien baru dengan semua kasus pasien lama.
- Kasus pasien lama dengan kedekatan terbesar lah yang akan diambil solusinya untuk digunakan pada kasus pasien baru.

Ilustrasi Kedekatan Kasus



Rumus Perhitungan Kedekatan



$$\text{similarity}(T, S) = \frac{\sum_{i=1}^n f(T_i, S_i) * W_i}{W_i}$$

Keterangan :

T : kasus baru

S : kasus yang ada dalam penyimpanan

n : jumlah atribut dalam tiap kasus

i : atribut individu antara 1 s.d. n

f : fungsi *similarity* atribut *i* antara kasus T dan S

w : bobot yang diberikan pada atribut ke-*i*

Nilai Kedekatan



- Kedekatan biasanya berada pada nilai antara 0 s.d. 1.
- Nilai 0 artinya kedua kasus mutlak tidak mirip.
- Nilai 1 artinya kedua kasus mirip dengan mutlak.

Analisa Kasus 1



- Kasus kemungkinan seorang nasabah bank akan bermasalah pembayarannya atau tidak.

No	Jenis Kelamin	Pendidikan	Agama	Bermasalah
1	L	S1	Islam	Ya
2	P	SMA	Kristen	Tidak
3	L	SMA	Islam	Ya

- Atribut Bermasalah merupakan atribut tujuan.
- Bobot antara satu atribut dengan atribut yang lain pada atribut bukan tujuan dapat didefinisikan dengan nilai berbeda.

Atribut	Bobot
Jenis Kelamin	0.5
Pendidikan	1
Agama	0.75

Kedekatan Nilai Atribut



Tabel 1. Kedekatan Nilai Atribut Jenis Kelamin

Nilai 1	Nilai 2	Kedekatan
L	L	1
P	P	1
L	P	0.5
P	L	0.5

Tabel 3. Kedekatan Nilai Atribut Agama

Nilai 1	Nilai 2	Kedekatan
Islam	Islam	1
Kristen	Kristen	1
Islam	Kristen	0.75
Kristen	Islam	0.75

Tabel 2. Kedekatan Nilai Atribut Pendidikan

Nilai 1	Nilai 2	Kedekatan
S1	S1	1
SMA	SMA	1
S1	SMA	0.4
SMA	S1	0.4

Contoh Kasus



- ❖ Misalkan ada kasus nasabah baru dengan nilai atribut berikut :
 - Jenis Kelamin : L
 - Pendidikan : SMA
 - Agama : Kristen

- ❖ Untuk memprediksi apakah nasabah tersebut akan bermasalah atau tidak dapat dilakukan langkah – langkah selanjutnya.

Langkah 1 (1)



1. Menghitung kedekatan kasus baru dengan kasus nomor 1.

Diketahui :

- a. Kedekatan nilai atribut Jenis Kelamin : 1
- b. Bobot atribut Jenis Kelamin : 0.5
- c. Kedekatan nilai atribut Pendidikan : 0.4
- d. Bobot atribut Pendidikan : 1
- e. Kedekatan nilai atribut Agama : 0.75
- f. Bobot atribut Agama : 0.75

Langkah 1 (2)



- Perhitungan :

$$\text{Jarak} = \frac{(a*b)+(c*d)+(e*f)}{b+d+f}$$

$$\text{Jarak} = \frac{(1*0.5)+(0.4*1)+(0.75*0.75)}{0.5+1+0.75}$$

$$\text{Jarak} = \frac{1.4625}{2.25}$$

$$\text{Jarak} = 0.65$$

Langkah 2 (1)



2. Menghitung kedekatan kasus baru dengan kasus nomor 2.

Diketahui :

- a. Kedekatan nilai atribut Jenis Kelamin : 0.5
- b. Bobot atribut Jenis Kelamin : 0.5
- c. Kedekatan nilai atribut Pendidikan : 1
- d. Bobot atribut Pendidikan : 1
- e. Kedekatan nilai atribut Agama : 0.75
- f. Bobot atribut Agama : 0.75

Langkah 2 (2)



- Perhitungan :

$$\text{Jarak} = \frac{(a*b)+(c*d)+(e*f)}{b+d+f}$$

$$\text{Jarak} = \frac{(0.5*0.5)+(1*1)+(0.75*0.75)}{0.5+1+0.75}$$

$$\text{Jarak} = \frac{1.8125}{2.25}$$

$$\text{Jarak} = 0.8$$

Langkah 3 (1)



3. Menghitung kedekatan kasus baru dengan kasus nomor 3.

Diketahui :

- a. Kedekatan nilai atribut Jenis Kelamin : 1
- b. Bobot atribut Jenis Kelamin : 0.5
- c. Kedekatan nilai atribut Pendidikan : 1
- d. Bobot atribut Pendidikan : 1
- e. Kedekatan nilai atribut Agama : 0.75
- f. Bobot atribut Agama : 0.75

Langkah 3 (2)



- Perhitungan :

$$\text{Jarak} = \frac{(a*b)+(c*d)+(e*f)}{b+d+f}$$

$$\text{Jarak} = \frac{(1*0.5)+(1*1)+(0.75*0.75)}{0.5+1+0.75}$$

$$\text{Jarak} = \frac{2.0625}{2.25}$$

$$\text{Jarak} = 0.9$$

Langkah 4 dan 5



4. Memilih kasus dengan kedekatan terdekat. Dari langkah 1, 2, 3 dapat diketahui bahwa nilai tertinggi adalah kasus 3. Berarti kasus yang terdekat dengan kasus baru adalah kasus 3.
5. Menggunakan klasifikasi dari kasus dengan kedekatan terdekat. Berdasarkan hasil pada langkah 4, maka klasifikasi dari kasus 3 yang akan digunakan untuk memprediksi kasus baru. Yaitu kemungkinan nasabah baru akan ***Bermasalah.***



❖ Tentukan prediksi menggunakan Nearest Neighbor jika ada kasus nasabah baru dengan nilai atribut berikut:

- Jenis Kelamin : P
- Pendidikan : S1
- Agama : Islam

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

