



Minggu ke-8

## Mid-Test

### Topik: Classification for Imputation

Ali Ridho Barakbah

Knowledge Engineering Research Group  
Department of Information and Computer Engineering  
Electronic Engineering Polytechnic Institute of Surabaya



Electronic Engineering  
Polytechnic Institute of Surabaya

Ali Ridho Barakbah

Knowledge Engineering  
(knoWing) Research Group

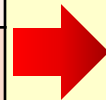


# Penerapan Klasifikasi untuk Imputasi

---

- Kita dapat menerapkan teknik klasifikasi untuk melakukan imputasi (pengisian nilai missing values dari atribut)
- Cara ini relatif lebih baik dibandingkan dengan imputasi (pengisian missing values) dengan rata-rata per class. Tapi cara ini cocok dipakai bila tidak banyak atribut yang mempunyai missing values.
- Langkah-Langkah:
  - Atribut yang mempunyai missing values diposisikan dengan sebagai class (class aslinya dijadikan atribut)
  - Data-data yang mempunyai missing values pada atribut tersebut, dijadikan sebagai sebagai test data. Data-data yang tidak ada missing values dijadikan sebagai train data.
  - Lakukan klasifikasi test data terhadap train data. Hasil klasifikasi dijadikan sebagai nilai untuk mengisi yang missing value.
  - Setelah klasifikasi selesai, semua nilai yang missing sudah terisi, gabungkan test data ke train data.
  - Kembalikan class sebagai atribut asalnya.

	a1	a2	a3	class
data1	4	5	7	1
data2	7	7	5	1
data3	4	4	4	2
data4	2		3	1
data5	6	1	5	2
data6	9		8	2
data7	7	3	3	2
data8	8	8	5	1
data9	4	6	8	2
data10	6	3	9	1



	a1	a3	class	a2
data1	4	7	1	5
data2	7	5	1	7
data3	4	4	2	4
data4	2	3	1	
data5	6	5	2	1
data6	9	8	2	
data7	7	3	2	3
data8	8	5	1	8
data9	4	8	2	6
data10	6	9	1	3

class



	a1	a3	class	a2
data1	4	7	1	5
data2	7	5	1	7
data3	4	4	2	4
data5	6	5	2	1
data7	7	3	2	3
data8	8	5	1	8
data9	4	8	2	6
data10	6	9	1	3



	a1	a3	class	a2
data4	2	3	1	
data6	9	8	2	

train data class

	a1	a3	class	a2
data1	4	7	1	5
data2	7	5	1	7
data3	4	4	2	4
data5	6	5	2	1
data7	7	3	2	3
data8	8	5	1	8
data9	4	8	2	6
data10	6	9	1	3

Klasifikasi

test data class

	a1	a3	class	a2
data4	2	3	1	
data6	9	8	2	

test data class

	a1	a3	class	a2
data4	2	3	1	c1
data6	9	8	2	c2

	a1	a3	class	a2
data1	4	7	1	5
data2	7	5	1	7
data3	4	4	2	4
data4	2	3	1	c1
data5	6	5	2	1
data6	9	8	2	c2
data7	7	3	2	3
data8	8	5	1	8
data9	4	8	2	6
data10	6	9	1	3

	a1	a2	a3	class
data1	4	5	7	1
data2	7	7	5	1
data3	4	4	4	2
data4	2	c1	3	1
data5	6	1	5	2
data6	9	c2	8	2
data7	7	3	3	2
data8	8	8	5	1
data9	4	6	8	2
data10	6	3	9	1

# UTS - #Assignment

## Klasifikasi untuk Imputasi

1. **dataset**  $\leftarrow$  titanic.csv
2. **data**  $\leftarrow$  ambil dataset kolom fitur (Sex, Age, Pclass, Fare, Survived).
3. **train\_data**  $\leftarrow$  ambil fitur (Sex, Pclass, Fare, Survived) pada **data** yang Age $\neq$ null
4. **train\_label**  $\leftarrow$  ambil fitur (Age) pada **data** yang Age $\neq$ null
5. **test\_data**  $\leftarrow$  ambil fitur (Sex, Pclass, Fare, Survived) pada **data** yang Age=null
6. **train\_data**  $\leftarrow$  lakukan normalisasi pada **train\_data** dengan Min-Max 0-1 (catat nilai min dan max setiap atribut)
7. **test\_data**  $\leftarrow$  lakukan normalisasi pada **test\_data** dengan Min-Max 0-1 (dengan nilai min dan max setiap atribut pada Langkah 6)
8. **class\_result**  $\leftarrow$  Lakukan klasifikasi **test\_data** terhadap **train\_data** dengan 3-NN (output mepakai class pada **train\_label**)
9. **data** (Age)  $\leftarrow$  lakukan pengisian missing values pada **data** yang Age=null dengan nilai **class\_result**
10. **test\_dataset**  $\leftarrow$  titanic\_test.csv
11. **train\_data**  $\leftarrow$  ambil fitur (Sex, Age, Pclass, Fare) dari **data**
12. **train\_label**  $\leftarrow$  ambil fitur (Survived) dari **data**
13. **test\_data**  $\leftarrow$  ambil **test\_dataset** kolom fitur (Sex, Age, Pclass, Fare). Hilangkan baris data yang terdapat missing values
14. **test\_label**  $\leftarrow$  titanic\_testlabel.csv (urutan sesuai test\_data no.13)
15. **train\_data**  $\leftarrow$  lakukan normalisasi pada **train\_data** dengan Min-Max 0-1 (catat nilai min dan max setiap atribut)
16. **test\_data**  $\leftarrow$  lakukan normalisasi pada **test\_data** dengan Min-Max 0-1 (dengan nilai min dan max setiap atribut pada Langkah 15)
17. **class\_result**  $\leftarrow$  Lakukan klasifikasi **test\_data** terhadap **train\_data** dengan 3-NN (output mepakai class pada **train\_label**)
18. **error**  $\leftarrow$  Bandingkan hasil klasifikasi **class\_result** dengan **test\_label**. Jika tidak sama berarti error. Hitunglah jumlah error dari seluruh **class\_result**
19. **error\_ratio**  $\leftarrow$  **error** dibagi jumlah **test\_data**, dikali 100 (%)

# Pengumpulan Tugas

---

- Buatlah coding dengan Bahasa pemrograman/tools apapun untuk semua assignment
- Buatlah laporan dalam slide ppt. Laporan terdiri dari screenshot coding dan hasil running untuk setiap assignment.
- Simpan laporan dalam file pdf dengan format penamaan:  
DM\_UTS\_NRP\_namadepan.pdf
- Upload file tersebut di alamat ini:  
<http://ridho.lecturer.pens.ac.id/lecture.php>  
(Pilih Mata Kuliah : Data Mining)
- Deadline upload: Rabu, 19 Mei 2021