



**FAKULTAS ILMU KOMPUTER
UNIVERSITAS DIAN NUSWANTORO**

*MATA KULIAH
DATA MINING – SI S1*

Data Mining Multiple Linear Regression

TEAM PENYUSUN DATA MINING PRODI SISTEM INFORMASI

Daftar Pustaka

- Jiawei Han and Micheline Kamber, *Data Mining: Concepts and Techniques Third Edition*, Elsevier, 2012
- Ian H. Witten, Frank Eibe, Mark A. Hall, *Data mining: Practical Machine Learning Tools and Techniques 3rd Edition*, Elsevier, 2011
- Ethem Alpaydin, *Introduction to Machine Learning*, 3rd ed., MIT Press, 2014
- Florin Gorunescu, *Data Mining: Concepts, Models and Techniques*, Springer, 2011

Capaian Pembelajaran

- Memahami tentang Multiple Linear Regression



Multiple Linear Regression

Apa itu Multiple Linear Regression

- Multiple Linear Regression adalah analisis regresi yang menjelaskan hubungan antara peubah respon (variabel dependen) dengan faktor-faktor yang mempengaruhi lebih dari 1 prediktor (variabel independen). Regresi linier berganda hampir sama dengan regresi linier sederhana, hanya saja pada regresi linier berganda variabel bebasnya lebih dari satu variabel penduga.
- Tujuan analisis regresi linier berganda adalah untuk mencari bentuk hubungan (relasi) linier antara 1 variabel terikat y dan n variabel bebas x_1, x_2, \dots, x_n .

Model multiple linear regression

- Rumus: $y = a + b_1x_1 + b_2x_2 + \dots + b_nx_n$
- Dimana:
 - y : variabel tidak bebas (response/dependent variable)
 - a : konstanta
 - b : koefisien regresi
 - x : variabel bebas (predictor/independent variable)

Model multiple linear regression

- Rumus: $y = a + b_1x_1 + b_2x_2 + \dots + b_nx_n$
- Rumus $a = \bar{y} - b_1\bar{x}_1 - b_2\bar{x}_2$
- dimana
 - \bar{y} = rata-rata variabel tidak bebas (response/dependent variable)
 - a = konstanta
 - \bar{x}_1, \bar{x}_2 = rata-rata variabel bebas (predictor/independent variable)
 - b_1, b_2 = koefisien regresi

Model multiple linear regression: koefisien regresi

$$b_1 = \frac{(\sum x_2^2)(\sum x_1 y) - (\sum x_1 x_2)(\sum x_2 y)}{(\sum x_1^2)(\sum x_2^2) - (\sum x_1 x_2)^2}$$

$$b_2 = \frac{(\sum x_1^2)(\sum x_2 y) - (\sum x_1 x_2)(\sum x_1 y)}{(\sum x_1^2)(\sum x_2^2) - (\sum x_1 x_2)^2}$$

- dimana:
- y : variabel tidak bebas (response/dependent variable)
- $\sum x_1, \sum x_2$: jumlah variabel bebas (predictor/independent variable)
- b_1, b_2 : koefisien regresi

$$b_1 = \frac{(\sum x_2^2)(\sum x_1 y) - (\sum x_1 x_2)(\sum x_2 y)}{(\sum x_1^2)(\sum x_2^2) - (\sum x_1 x_2)^2}$$

Rumus

$$\sum x_i^2 = \sum x_i^2 - \frac{(\sum x_i)^2}{n}$$

Jika ada 2 fitur, maka

$$\sum x_1^2 = \sum x_1^2 - \frac{(\sum x_1)^2}{n}$$

$$\sum x_2^2 = \sum x_2^2 - \frac{(\sum x_2)^2}{n}$$

$$b_2 = \frac{(\sum x_1^2)(\sum x_2 y) - (\sum x_1 x_2)(\sum x_1 y)}{(\sum x_1^2)(\sum x_2^2) - (\sum x_1 x_2)^2}$$

Rumus

$$\sum x_i y = \sum x_i y - \frac{(\sum x_i)(\sum y)}{n}$$

Jika ada 2 fitur, maka

$$\sum x_1 y = \sum x_1 y - \frac{(\sum x_1)(\sum y)}{n}$$


$$\sum x_2 y = \sum x_2 y - \frac{(\sum x_2)(\sum y)}{n}$$

Contoh kasus

y	x_1	x_2
-3,7	3	8
3,5	4	5
2,5	5	7
11,5	6	3
5,7	2	1

Menghitung x_1^2 dan x_2^2

y	x_1	x_2
-3,7	3	8
3,5	4	5
2,5	5	7
11,5	6	3
5,7	2	1



y	x_1	x_2	$(x_1)^2$	$(x_2)^2$
-3,7	3	8	9	64
3,5	4	5	16	25
2,5	5	7	25	49
11,5	6	3	36	9
5,7	2	1	4	1

Menghitung x_1y dan x_2y

y	x_1	x_2	$(x_1)^2$	$(x_2)^2$		y	x_1	x_2	$(x_1)^2$	$(x_2)^2$	x_1y	x_2y
-3,7	3	8	9	64	→	-3,7	3	8	9	64	-11,1	-29,6
3,5	4	5	16	25		3,5	4	5	16	25	14	17,5
2,5	5	7	25	49		2,5	5	7	25	49	12,5	17,5
11,5	6	3	36	9		11,5	6	3	36	9	69	34,5
5,7	2	1	4	1		5,7	2	1	4	1	11,4	5,7

Menghitung $x_1 x_2$

y	x_1	x_2	$(x_1)^2$	$(x_2)^2$	$x_1 y$	$x_2 y$
-3,7	3	8	9	64	-11,1	-29,6
3,5	4	5	16	25	14	17,5
2,5	5	7	25	49	12,5	17,5
11,5	6	3	36	9	69	34,5
5,7	2	1	4	1	11,4	5,7



y	x_1	x_2	$(x_1)^2$	$(x_2)^2$	$x_1 y$	$x_2 y$	$x_1 x_2$
-3,7	3	8	9	64	-11,1	-29,6	24
3,5	4	5	16	25	14	17,5	20
2,5	5	7	25	49	12,5	17,5	35
11,5	6	3	36	9	69	34,5	18
5,7	2	1	4	1	11,4	5,7	2

Menghitung $\sum y$, $\sum x_1$, $\sum x_2$, $\sum (x_1)^2$, $\sum (x_2)^2$, $\sum x_1y$, $\sum x_2y$, $\sum x_1x_2$

y	x_1	x_2	$(x_1)^2$	$(x_2)^2$	x_1y	x_2y	x_1x_2
-3,7	3	8	9	64	-11,1	-29,6	24
3,5	4	5	16	25	14	17,5	20
2,5	5	7	25	49	12,5	17,5	35
11,5	6	3	36	9	69	34,5	18
5,7	2	1	4	1	11,4	5,7	2



y	x_1	x_2	$(x_1)^2$	$(x_2)^2$	x_1y	x_2y	x_1x_2	
-3,7	3	8	9	64	-11,1	-29,6	24	
3,5	4	5	16	25	14	17,5	20	
2,5	5	7	25	49	12,5	17,5	35	
11,5	6	3	36	9	69	34,5	18	
5,7	2	1	4	1	11,4	5,7	2	
Σ	19,5	20	24	90	148	95,8	45,6	99
	\downarrow	\downarrow	\downarrow	\downarrow	\downarrow	\downarrow	\downarrow	\downarrow
	Σy	Σx_1	Σx_2	Σx_1^2	Σx_2^2	Σx_1y	Σx_2y	Σx_1x_2

Mencari $\sum x_1^2$ dan $\sum x_2^2$

y	x_1	x_2	$(x_1)^2$	$(x_2)^2$	x_1y	x_2y	x_1x_2
-3,7	3	8	9	64	-11,1	-29,6	24
3,5	4	5	16	25	14	17,5	20
2,5	5	7	25	49	12,5	17,5	35
11,5	6	3	36	9	69	34,5	18
5,7	2	1	4	1	11,4	5,7	2
Σ	19,5	20	90	148	95,8	45,6	99

$\bullet \sum x_1^2 = \sum x_1^2 - \frac{(\sum x_1)^2}{n} = 90 - \frac{20^2}{5} = 10$
 $\bullet \sum x_2^2 = \sum x_2^2 - \frac{(\sum x_2)^2}{n} = 148 - \frac{24^2}{5} = 32,8$

Mencari $\sum x_1y$, $\sum x_2y$, $\sum x_1x_2$

y	x_1	x_2	$(x_1)^2$	$(x_2)^2$	x_1y	x_2y	x_1x_2	
-3,7	3	8	9	64	-11,1	-29,6	24	
3,5	4	5	16	25	14	17,5	20	
2,5	5	7	25	49	12,5	17,5	35	
11,5	6	3	36	9	69	34,5	18	
5,7	2	1	4	1	11,4	5,7	2	
Σ	19,5	20	24	90	148	95,8	45,6	99

$$\bullet \sum x_1y = \sum x_1y - \frac{(\sum x_1)(\sum y)}{n}$$

$$\sum x_1y = 95,8 - \frac{(20)(19,5)}{5} = 17,8$$

$$\bullet \sum x_2y = \sum x_2y - \frac{(\sum x_2)(\sum y)}{n}$$

$$\sum x_2y = 45,6 - \frac{(24)(19,5)}{5} = -48$$

$$\bullet \sum x_1x_2 = \sum x_1x_2 - \frac{(\sum x_1)(\sum x_2)}{n}$$

$$\sum x_1x_2 = 99 - \frac{(20)(24)}{5} = 3$$

Mencari b_1 dan b_2

- $\sum x_1^2 = \sum x_1^2 - \frac{(\sum x_1)^2}{n} = 90 - \frac{20^2}{5} = 10$
- $\sum x_2^2 = \sum x_2^2 - \frac{(\sum x_2)^2}{n} = 148 - \frac{24^2}{5} = 32,8$
- $\sum x_1 y = \sum x_1 y - \frac{(\sum x_1)(\sum y)}{n}$
 $\sum x_1 y = 95,8 - \frac{(20)(19,5)}{5} = 17,8$
- $\sum x_2 y = \sum x_2 y - \frac{(\sum x_2)(\sum y)}{n}$
 $\sum x_2 y = 45,6 - \frac{(24)(19,5)}{5} = -48$
- $\sum x_1 x_2 = \sum x_1 x_2 - \frac{(\sum x_1)(\sum x_2)}{n}$
 $\sum x_1 x_2 = 99 - \frac{(20)(24)}{5} = 3$



- $b_1 = \frac{(\sum x_2^2)(\sum x_1 y) - (\sum x_1 x_2)(\sum x_2 y)}{(\sum x_1^2)(\sum x_2^2) - (\sum x_1 x_2)^2}$
 $b_1 = \frac{(32,8)(17,8) - (3)(-48)}{(10)(32,8) - (3)^2}$
 $b_1 = \frac{727,84}{319} = 2,28$
- $b_2 = \frac{(\sum x_1^2)(\sum x_2 y) - (\sum x_1 x_2)(\sum x_1 y)}{(\sum x_1^2)(\sum x_2^2) - (\sum x_1 x_2)^2}$
 $b_2 = \frac{(10)(-48) - (3)(17,8)}{(10)(32,8) - (3)^2}$
 $b_2 = \frac{-533,4}{319} = -1,67$

Mencari a

$$\bullet \quad b_1 = \frac{(\sum x_2^2)(\sum x_1 y) - (\sum x_1 x_2)(\sum x_2 y)}{(\sum x_1^2)(\sum x_2^2) - (\sum x_1 x_2)^2}$$

$$b_1 = \frac{(32,8)(17,8) - (3)(-48)}{(10)(32,8) - (3)^2}$$

$$b_1 = \frac{727,84}{319} = 2,28$$

$$\bullet \quad b_2 = \frac{(\sum x_1^2)(\sum x_2 y) - (\sum x_1 x_2)(\sum x_1 y)}{(\sum x_1^2)(\sum x_2^2) - (\sum x_1 x_2)^2}$$

$$b_2 = \frac{(10)(-48) - (3)(17,8)}{(10)(32,8) - (3)^2}$$

$$b_2 = \frac{-533,4}{319} = -1,67$$



$$\bullet \quad a = \bar{y} - b_1 \bar{x}_1 - b_2 \bar{x}_2$$

$$a = \left(\frac{19,5}{5}\right) - \left[(2,28) \left(\frac{20}{5}\right)\right] - \left[(-1,67) \left(\frac{24}{5}\right)\right]$$

$$a = 2,796$$

Mencari y

$$\bullet \quad b_1 = \frac{(\sum x_2^2)(\sum x_1 y) - (\sum x_1 x_2)(\sum x_2 y)}{(\sum x_1^2)(\sum x_2^2) - (\sum x_1 x_2)^2}$$

$$b_1 = \frac{(32,8)(17,8) - (3)(-48)}{(10)(32,8) - (3)^2}$$

$$b_1 = \frac{727,84}{319} = 2,28$$

$$\bullet \quad b_2 = \frac{(\sum x_1^2)(\sum x_2 y) - (\sum x_1 x_2)(\sum x_1 y)}{(\sum x_1^2)(\sum x_2^2) - (\sum x_1 x_2)^2}$$

$$b_2 = \frac{(10)(-48) - (3)(17,8)}{(10)(32,8) - (3)^2}$$

$$b_2 = \frac{-533,4}{319} = -1,67$$

$$\bullet \quad a = \bar{y} - b_1 \bar{x}_1 - b_2 \bar{x}_2$$

$$a = \left(\frac{19,5}{5}\right) - \left[(2,28) \left(\frac{20}{5}\right)\right] - \left[(-1,67) \left(\frac{24}{5}\right)\right]$$

$$a = 2,796$$

$$\bullet \quad y = a + b_1 x_1 + b_2 x_2$$

$$y = 2,796 + 2,28x_1 + (-1,67x_2)$$

$$y = 2,796 + 2,28x_1 - 1,67x_2$$

→ intercept



THANKS

ANY QUESTIONS?