



sec</>de

Security and Code Development

www.secode.id

LINEAR REGRESSION

Oleh: SEC</>DE

PENGERTIAN LINEAR REGRESSION

- ❑ Linear Regression adalah model statistik yang digunakan untuk membangun model hubungan antara dua variabel.
- ❑ Salah satu variabel ini disebut variabel prediktor yang nilainya didapatkan melalui percobaan.
- ❑ Variabel lain disebut variabel respon yang nilainya berasal dari variabel prediktor.
- ❑ Garis linear menunjukkan pola hubungan antara dua variabel misalnya variabel X dan Y.
- ❑ Garis linear sebenarnya hanya merupakan garis taksiran yang dipakai untuk mewakili pola sebaran data tersebut .

TUJUAN LINEAR REGRESSION

- ☐ Mencari dan menjelaskan korelasi antara kriterium dengan prediktor.
- ☐ Mengukur "seberapa kuat" atau "derajat kedekatan" suatu relasi yang terjadi antar variabel (signifikan).
- ☐ Mengetahui pola relasi dalam bentuk persamaan regresi.
- ☐ Menemukan sumbangan relatif prediktor.

RUMUS LINEAR REGRESSION SEDERHANA

Linear regression dapat dapat ditulis dengan rumus matematika sederhana, yaitu:

$$y = a + bx$$

$$a = \frac{\sum y - b \sum x}{n}$$

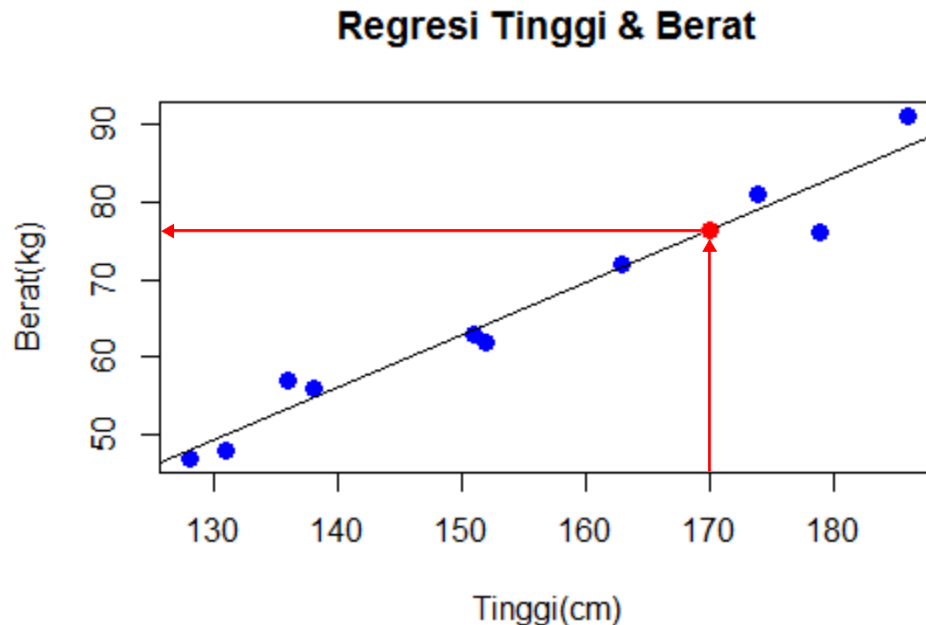
$$b = \frac{n \sum (xy) - (\sum x)(\sum y)}{n \sum x^2 - (\sum x)^2}$$

Keterangan:

- 1. y adalah variable respon.*
- 2. x adalah variable prediktor.*
- 3. a dan b adalah konstanta atau kofisien.*

CARA KERJA LINEAR REGRESSION

Secara sederhana cara kerja linear regression dapat dijelaskan dengan gambar di bawah ini.



Keterangan:

1. Titik-titik biru adalah data-data hasil pengamatan hubungan tinggi dan berat badan. Sebagai contoh untuk titik biru pertama dapat diketahui tinggi badan adalah 128 cm dengan berat badan adalah 47 kg.
2. Kemudian garis hitam adalah persamaan yang paling dapat mewakili hubungan relasi seluruh titik-titik biru. Persamaan tersebut dapat ditentukan dengan linear regression.
3. Sedangkan titik merah merupakan hasil prediksi data baru dari orang yang memiliki tinggi 170 cm. Dengan menarik garis dari titik 170,0 sampai menyentuh garis hitam, kemudian menarik garis ke arah sumbu y maka dapat diketahui informasi dari orang yang memiliki tinggi 170 cm adalah memiliki berat kira-kira 76 kg.



sec</>de

Security and Code Development

www.secode.id

STUDI KASUS LINEAR REGRESSION

Oleh: SEC</>DE

- ❑ Salah satu contoh kasus yang dapat diselesaikan dengan linear regression adalah prediksi berat badan seseorang ketika diketahui tinggi badannya.



Tabel Tinggi dan Berat Anak

No. (n)	Tinggi (x)	Berat (y)
1	151	63
2	174	81
3	138	56
4	186	91
5	128	47
6	136	57
7	179	76
8	163	72
9	152	62
10	131	48

Data disamping adalah data hasil pengamatan tinggi dan berat badan.

Dari data tersebut kita akan menghitung persamaan linear regression sederhananya.

Tentukan prediksi berat badan jika diketahui tinggi badan 170 dan 185 cm.

PERHITUNGAN MANUAL

Menghitung nilai a dan b

n	x	y	xy	x^2	y^2
1	151	63	9513	22801	3969
2	174	81	14094	30276	6561
3	138	56	7728	19044	3136
4	186	91	16926	34596	8281
5	128	47	6016	16384	2209
6	136	57	7752	18496	3249
7	179	76	13604	32041	5776
8	163	72	11736	26569	5184
9	152	62	9424	23104	3844
10	131	48	6288	17161	2304
Jumlah	1538	653	103081	240472	44513

Menghitung nilai a dan b

$$b = \frac{n \sum(xy) - (\sum x)(\sum y)}{n \sum x^2 - (\sum x)^2}$$

$$b = \frac{10(103081) - (1538)(653)}{10(240472) - (2364444)}$$

$$b = \frac{1030810 - 1004314}{2404720 - 2364444}$$

$$b = \frac{26496}{39276}$$

$$b = 0,67461$$

$$a = \frac{\sum y - b \sum x}{n}$$

$$a = \frac{653 - ((0,67461) * (1538))}{10}$$

$$a = \frac{-384,55}{10}$$

$$a = -38,4551$$

Persamaan Linear Regression-nya:

$$y = -38,4551 + 0,67461x$$

Menghitung prediksi tinggi badan 170 cm dan dan 185 cm

Persamaan Linear Regression-nya:

$$y = -38,4551 + 0,67461x$$

Prediksi berat badan dengan tinggi 170 cm:

$$y = -38,4551 + 0,67461(170)$$

$$y = 76,22869$$

Prediksi berat badan dengan tinggi 185 cm:

$$y = -38,4551 + 0,67461(185)$$

$$y = 86,34785$$

IMPLEMENTASI LINEAR REGRESSION PADA PEMROGRAMAN R

Oleh: SEC</>DE

Untuk melakukan linear regression pada lingkungan R maka perlu dilakukan langkah-langkah sederhana di bawah ini:

1. Temukan data nyata hasil pengamatan dari tinggi dan berat badan.
2. Buat model hubungan antar tinggi dan berat badan dengan menggunakan fungsi `lm()` pada lingkungan R.
3. Setelah mendapatkan koefisien dari model yang telah dibuat kemudian dapat dibuat persamaan matematikanya.
4. Dapatkan rangkuman (summary) dari model untuk mengetahui rata-rata error pada proses prediksi, atau sering disebut sebagai residual.
5. Gunakan fungsi `predict()` untuk memprediksi berat badan dari data yang baru

Implementasi langkah-langkah di atas pada lingkungan R dijelaskan dibawah ini.

Data nyata hasil pengamatan untuk tinggi dan berat dapat dilihat pada kode di bawah ini.

```
1 # input data tinggi dan berat badan
2 tinggi = c(151, 174, 138, 186, 128, 136, 179, 163, 152, 131)
3 berat = c(63, 81, 56, 91, 47, 57, 76, 72, 62, 48)
4 tinggi
5 berat
6
```

Variable tinggi adalah variable prediktor sedangkan variable berat adalah variabel respon. Selanjutnya akan dibuat model dari data tersebut dengan menggunakan fungsi lm(). Fungsi lm() mempunyai sintaks seperti berikut:

```
7 # sintaks fungsi lm()
8 # lm(formula, data)
9
```

Dalam penggunaannya untuk membuat model hubungan antara tinggi dan berat badan dapat dilakukan dengan cara seperti berikut ini.

```
10 # membuat model hubungan antara tinggi dan berat badan
11 relasi <- lm(berat~tinggi)
12
```

PRAKTIKUM

Setelah model dibuat maka dapat diketahui koefisien dari model dengan menggunakan memanggil variable relasi.

```
> # mengetahui koefisien dari model
> relasi

Call:
lm(formula = berat ~ tinggi)

Coefficients:
(Intercept)      tinggi
   -38.4551       0.6746

> |
```

Dengan didapatkannya koefisien tersebut maka dapat dibuat persamaan sebagai berikut:

```
15
16 # persamaan linear regression
17 # berat = 0.6746x -38455
18 # berat = 0.6746tinggi -38455
19 y <- -384551 + 0.6746
20
```


Sedangkan untuk mengetahui rangkuman (summary) dari model dapat digunakan fungsi `summary()` atau `summary.lm()`, berikut ini digunakan fungsi `summary.lm()`.

```
> # mengetahui rangkuman (summary) dari model
> summary.lm(relasi)

Call:
lm(formula = berat ~ tinggi)

Residuals:
    Min       1Q   Median       3Q      Max
-6.3002 -1.6629  0.0412  1.8944  3.9775

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -38.45509    8.04901  -4.778  0.00139 **
tinggi       0.67461    0.05191  12.997 1.16e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.253 on 8 degrees of freedom
Multiple R-squared:  0.9548,    Adjusted R-squared:  0.9491
F-statistic: 168.9 on 1 and 8 DF,  p-value: 1.164e-06

> |
```

PRAKTIKUM

Setelah model dibuat maka dapat dipergunakan untuk melakukan prediksi dengan fungsi `predict()`. Sebagai contoh jika ingin mengetahui berat seseorang yang memiliki tinggi 170, maka proses prediksi dapat dilakukan dengan cara di bawah ini.

```
24 # melakukan prediksi dengan fungsi predict()
25 # jika ingin mengetahui berat seseorang yang memiliki tinggi 170,
26 # maka proses prediksi dapat dilakukan dengan cara di bawah ini.
27 data_baru <- data.frame(tinggi = 170)
28 prediksi_data_baru <- predict(relasi, data_baru)
```

Hasil prediksi berat orang dengan tinggi 170 dapat dilihat dengan memanggil variabel `prediksi_data_baru`. Di bawah ini diketahui beratnya adalah 76.22869.

```
> prediksi_data_baru
      1
76.22869
> |
```

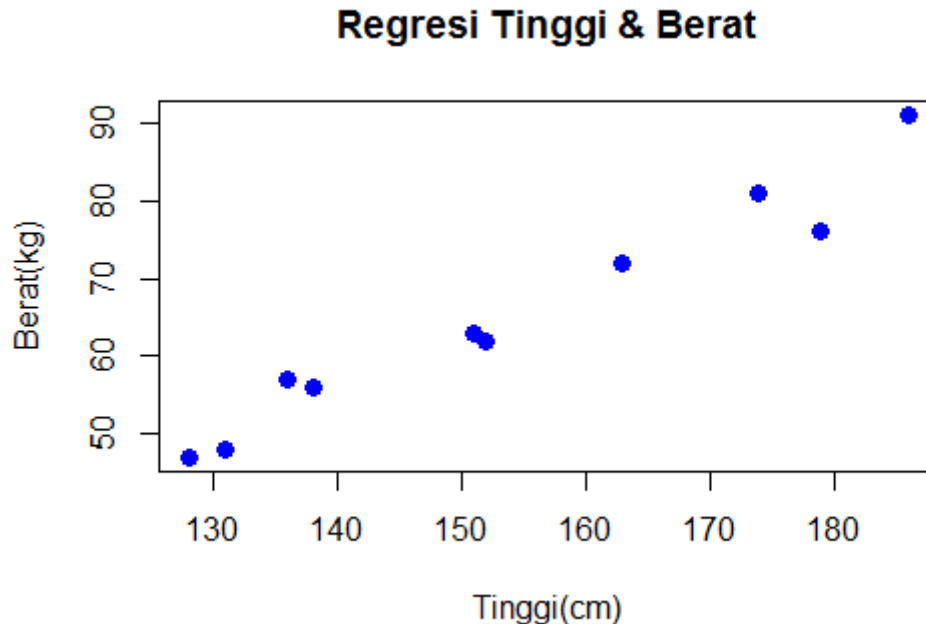
PRAKTIKUM (VISUALISASI)

Untuk membuat visualisasi model dan prediksi yang dibuat di atas maka dibuat kode berikut ini.

```
33
34 # membuat visualisasi model dan prediksi (tinggi 170)
35 plot(tinggi, berat, col = "blue", main = "Regresi Tinggi & Berat", cex = 1.3,
36      pch = 16, xlab = "Tinggi(cm)", ylab = "Berat(kg)")
37
38 abline(lm(berat~tinggi))
39 |
40 points(170, prediksi_data_baru, col = "red", cex = 1.3, pch = 16)
41
42 # memprediksi tinggi 185 dan memvisualisasikan data
43 data_baru <- data.frame(tinggi = 185)
44 prediksi_data_baru <- predict(relasi, data_baru)
45 prediksi_data_baru
46 points(185, prediksi_data_baru, col = "green", cex = 1.3, pch = 16)
47
```

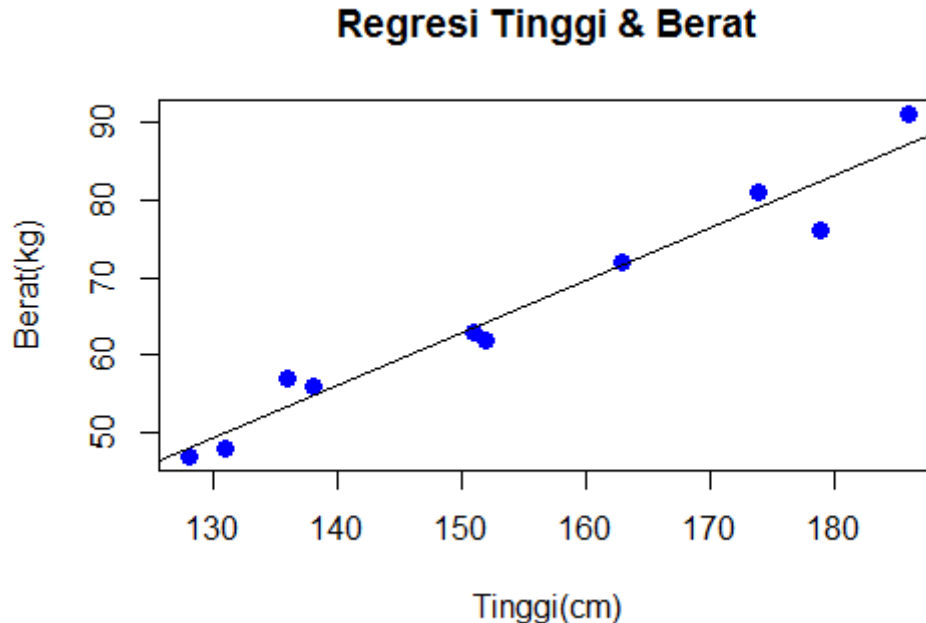
PRAKTIKUM (VISUALISASI)

Baris pertama dari kode di atas akan melakukan plot data tinggi dan berat. Hasil dari kode baris pertama adalah seperti gambar di bawah ini.



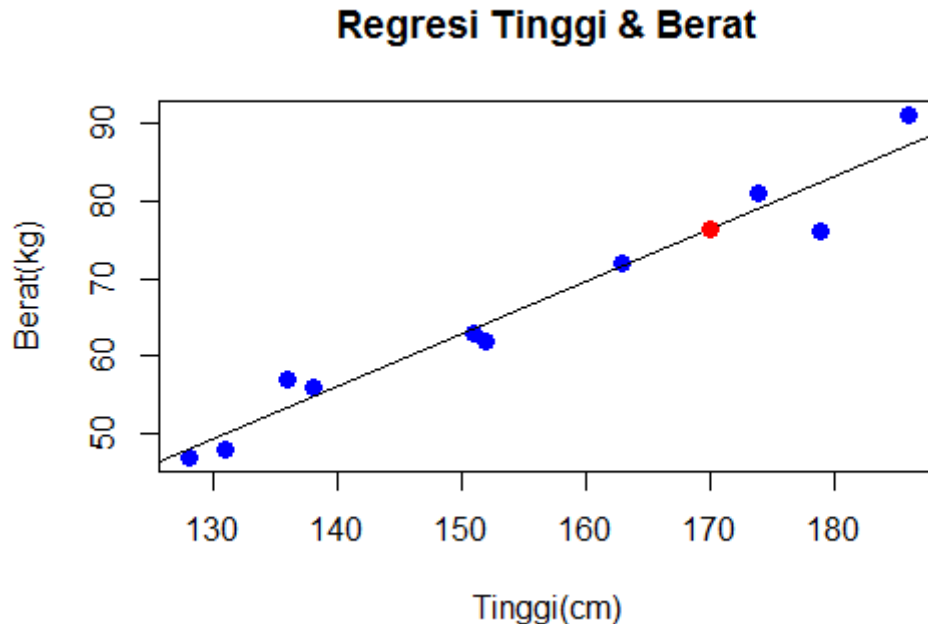
PRAKTIKUM (VISUALISASI)

Sedangkan baris kedua bertujuan untuk menggambar garis linear hasil dari model yang dibuat dari fungsi `lm()`. Garis tersebut mempunyai persamaan: $y = 0.6746x - 38.4551$.



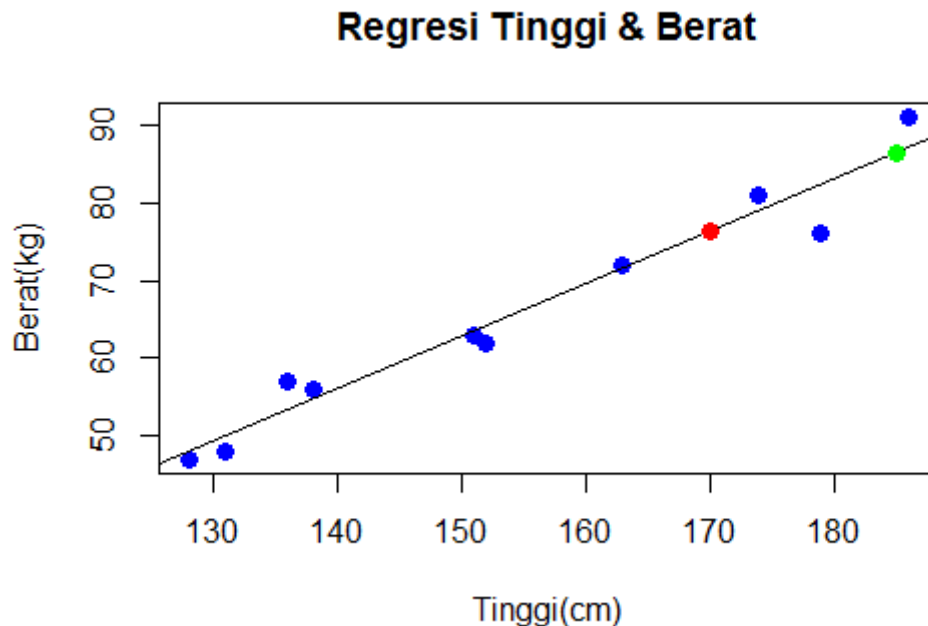
PRAKTIKUM (VISUALISASI)

Baris terakhir berfungsi untuk menunjukkan titik memprediksi data baru, dimana diketahui tinggi seseorang adalah 170 cm. Dari gambar dapat dilihat titik merah yang berada pada garis model linear regression.



PRAKTIKUM (VISUALISASI)

Sedangkan pada gambar di bawah ini adalah prediksi data baru untuk tinggi 185 cm yang hasilnya bisa dilihat pada titik warna hijau.





sec</>de

Security and Code Development

www.secode.id

IMPLEMENTASI LINEAR REGRESSION PADA RAPID MINER

Oleh: SEC</>DE

Menggunakan operator Linear Regression pada Rapid Miner

The screenshot displays the RapidMiner Studio Free 9.0.002 interface. The main workspace shows a process diagram with the following operators: 'Retrieve Tinggi Bera...' (input), 'Linear Regression' (model), and 'Apply Model' (evaluation). The 'Linear Regression' operator is highlighted with a green checkmark, indicating it is selected. The 'Apply Model' operator is also highlighted with a green checkmark. The 'Retrieve Tinggi Bera...' operator is highlighted with a yellow warning icon. The 'Process' tab is active, showing the workflow. The 'Parameters' panel on the right shows the 'logverbosity' set to 'init' and the 'logfile' field. The 'Help' panel on the right shows the 'Process' section, which is the root operator of every process. The 'Operators' panel on the left shows a list of operators, including 'Apply Forecast', 'Apply Threshold', and 'Apply Model'. The 'Data Editor' panel at the bottom shows a table with columns for 'Case sensitive' and 'Drag&Drop an Example Set from the repository or click 'Load Example Set' or 'Create new Example Set' to start.

Hasil Persamaan Linear Regression dengan Rapid Miner

<new process*> - RapidMiner Studio Free 9.0.002 @ Samidi-PC

File Edit Process View Connections Cloud Settings Extensions Help

Views: Design Results Turbo Prep Auto Model

Find data, operators...etc All Studio

Result History

LinearRegression (Linear Regression) ExampleSet (Apply Model)

Attribute	Coefficient	Std. Error	Std. Coefficient	Tolerance	t-Stat	p-Value	Code
l _a Tinggi	0.675	0.052	0.977	1	12.997	0.000	****
(Intercept)	-38.455	8.049	?	?	-4.778	0.001	***

Repository

Import Data

- Samples
- DB
- Data Mining Buy Computer (Samidi)
 - Data (Samidi)
 - Proses (Samidi)
 - Data Mining Buy Computer (Samidi - v1, 12/28)
 - Data Mining Buy Computer Decision Tr
 - Kelulusan Mahasiswa (Samidi - v1, 12/28)
 - Kelulusan Mahasiswa Training (Samidi - v1, 12/28)
 - Tinggi Berat Badan Testing (Samidi - v1, 12/28)
 - Tinggi Berat Badan Training (Samidi - v1, 12/28)
- Local Repository (Samidi)
- Cloud Repository (disconnected)

Hasil Persamaan Linear Regression dengan Rapid Miner

The screenshot displays the RapidMiner Studio Free 9.0.002 interface. The main workspace shows the results of a Linear Regression model. The equation displayed is $0.675 * i_{\text{Tinggi}} - 38.455$. The interface includes a menu bar (File, Edit, Process, View, Connections, Cloud, Settings, Extensions, Help), a toolbar with icons for file operations and execution, and a 'Views' section with tabs for Design, Results (selected), Turbo Prep, and Auto Model. A search bar and 'All Studio' dropdown are also present. The 'Result History' pane on the left shows the 'LinearRegression (Linear Regression)' model. The 'Repository' pane on the right lists various data sources and processes, including 'Samples', 'DB', 'Data Mining Buy Computer (Samidi)', 'Proses (Samidi)', 'Data Mining Buy Computer (Samidi - v1)', 'Data Mining Buy Computer Decision Tree (Samidi - v1)', 'Kelulusan Mahasiswa (Samidi - v1, 12/20)', 'Kelulusan Mahasiswa Testing (Samidi - v1)', 'Kelulusan Mahasiswa Training (Samidi - v1)', 'Tinggi Berat Badan Testing (Samidi - v1)', 'Tinggi Berat Badan Training (Samidi - v1)', 'Local Repository (Samidi)', and 'Cloud Repository (disconnected)'.

LinearRegression

Data

$$0.675 * i_{\text{Tinggi}} - 38.455$$

Description

Annotations

Repository

- Import Data
- Samples
- DB
- Data Mining Buy Computer (Samidi)
 - Data (Samidi)
 - Proses (Samidi)
 - Data Mining Buy Computer (Samidi - v1)
 - Data Mining Buy Computer Decision Tree (Samidi - v1)
 - Kelulusan Mahasiswa (Samidi - v1, 12/20)
 - Kelulusan Mahasiswa Testing (Samidi - v1)
 - Kelulusan Mahasiswa Training (Samidi - v1)
 - Tinggi Berat Badan Testing (Samidi - v1)
 - Tinggi Berat Badan Training (Samidi - v1)
- Local Repository (Samidi)
- Cloud Repository (disconnected)

Hasil Prediksi Linear Regression dengan Rapid Miner

<new process*> - RapidMiner Studio Free 9.0.002 @ Samidi-PC

File Edit Process View Connections Cloud Settings Extensions Help

Views: Design Results Turbo Prep Auto Model

Find data, operators...etc All Studio

Result History LinearRegression (Linear Regression) ExampleSet (Apply Model)

ExampleSet (2 examples, 2 special attributes, 1 regular attribute)

Filter (2 / 2 examples): all

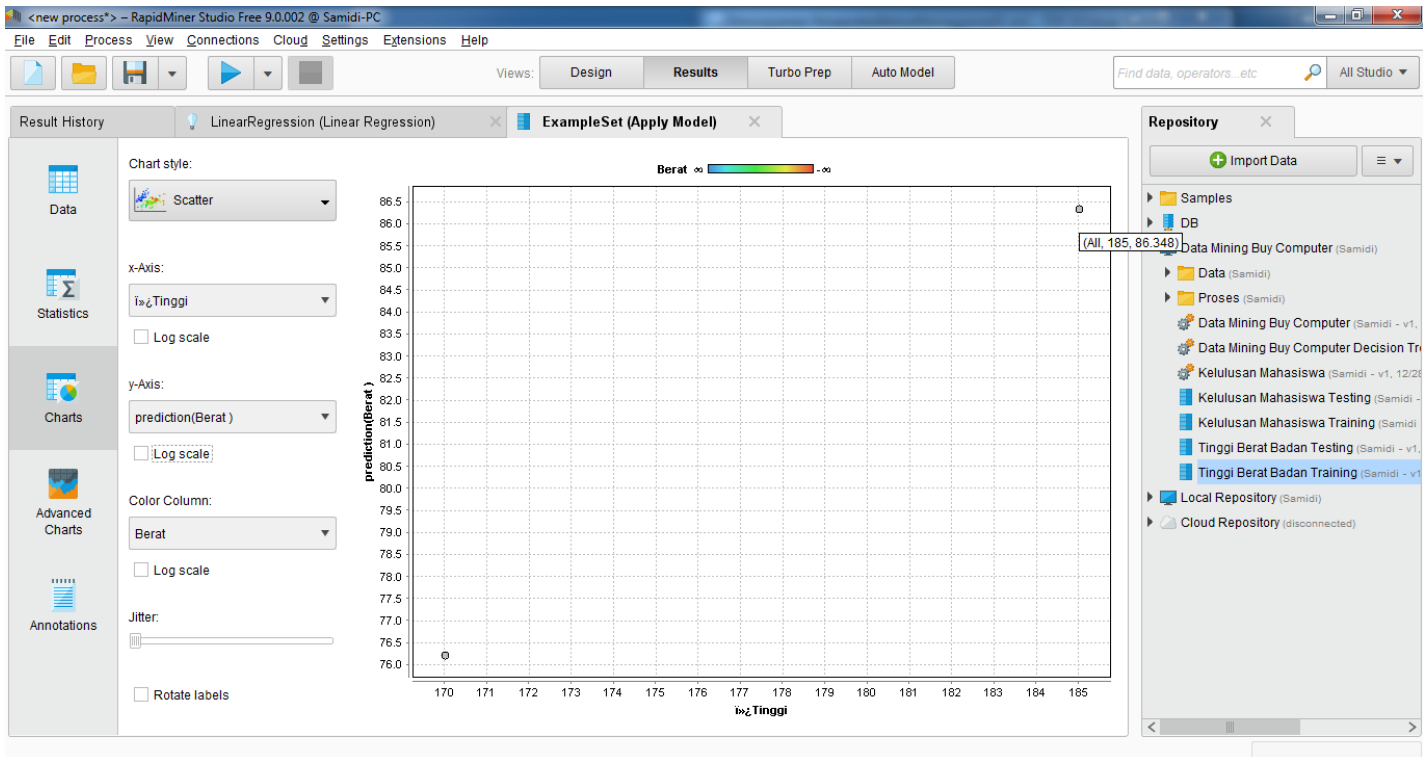
Row No.	Berat	prediction(B...	IszTinggi
1	?	76.229	170
2	?	86.348	185

Repository

Import Data

- Samples
- DB
- Data Mining Buy Computer (Samidi)
 - Data (Samidi)
 - Proses (Samidi)
 - Data Mining Buy Computer (Samidi - v1, 12/2/2019)
 - Data Mining Buy Computer Decision Tree (Samidi - v1, 12/2/2019)
 - Kelulusan Mahasiswa (Samidi - v1, 12/2/2019)
 - Kelulusan Mahasiswa Testing (Samidi - v1, 12/2/2019)
 - Kelulusan Mahasiswa Training (Samidi - v1, 12/2/2019)
 - Tinggi Berat Badan Testing (Samidi - v1, 12/2/2019)
 - Tinggi Berat Badan Training (Samidi - v1, 12/2/2019)
- Local Repository (Samidi)
- Cloud Repository (disconnected)

Plot Linear Regression dengan Rapid Miner





sec</>de

Security and Code Development

www.secode.id

Thank You