

Regresi

Machine Learning

Metode Pembelajaran dalam Machine Learning

1. Supervised Learning
2. Unsupervised learning
3. Semi Supervised Learning
4. Reinforcement Learning

Decision tree

Regresi Linear

K-Nearest Neighbour (K-NN)

Bayesian

Neural Network

Dan lain-lain

Pembelajaran *Supervised*

- Adalah salah satu metode pembelajaran dari algoritma *machine Learning*.
- Yang bertujuan untuk memprediksi sebuah model, guna men-
aproksimasi fungsi yang memetakan atribut data ke atribut target
melalui proses generalisasi pola dari contoh data sehingga dapat
dipergunakan untuk memprediksi atribut target dari sebuah data
baru.

Aplikasi *Supervised*

1. Regresi :

- Proses pembelajaran yang bertujuan untuk memprediksi nilai variabel target t dari data baru x

2. Klasifikasi :

- Proses pembelajaran yang bertujuan untuk memprediksi kategori dari data baru x

Kedua proses pembelajaran di atas berdasarkan kepada himpunan S dan N buah data yang memiliki label : $S = \{(x_1, t_1), (x_2, t_2), \dots, (x_n, t_n)\}$

→ Setiap data $(x, t) \in S$ terdiri dari komponen variabel data x yang direpresentasikan dengan vector dari m buah fitur, (x_1, x_2, \dots, x_m) dimana x_i ($1 \leq i \leq m$) merupakan nilai fitur.

Sesuai dengan tujuan masing-masing:

- Variabel target pada masalah klasifikasi merupakan bilangan kategorikal atau $t \in K$, dimana K adalah himpunan label data
- Variabel target pada masalah regresi merupakan bilangan real atau $t \in R$

Data pada Machine Learning

- Data adalah salah satu komponen penting dari Machine Learning, selain Algoritma pembelajaran dan metode Learning
- Data diperlukan sebagai input bagi algoritma pembelajaran Machine Learning
- Data merupakan salah satu factor penentu dari kinerja model hasil training

contoh

- Mentraining sebuah model Machine Learning yang mampu mengenali gambar seekor kucing, maka sebagai input bagi algoritma pembelajaran diperlukan sejumlah besar gambar kucing.
- Gambar tersebut akan digunakan sebagai data training dalam pembelajaran konsep tentang “kucing” menggunakan algoritma pembelajaran *stochastic gradient descent*.
- Setelah proses training selesai, diharapkan model hasil training dapat mengenali gambar baru kucing dalam berbagai pose dan bentuk

- Data sebagai input bagi Algoritma pembelajaran Machine Learning harus dapat dibaca oleh komputer, dalam format bilangan, dan berupa tabel (tergantung bidang aplikasinya)
- Data training umumnya di representasikan dalam bentuk matriks dimana baris digunakan untuk entitas data, dan kolom sebagai variabel data (dalam berbagai satuan pengukuran).

Metode Transformasi Data

- Ketika data dipergunakan sebagai data training atau data testing dari proses pembelajaran Machine Learning , beberapa transformasi data diperlukan untuk mencapai kinerja yang baik

Beberapa metode transformasi data pada Machine Learning

- Diskretisasi
- Menangani nilai NaN
- Rescaling
- Menangani Variabel Kategorikal
- Menangani Variabel Multicollinearty

Diskretisasi

- Algoritma pada machine learning melakukan pekerjaan *ekstraksi knowledge* dari suatu database
- Sebagian besar algoritma hanya bisa diaplikasikan pada data numerik ataupun nominal, dan untuk atribut continuous dibutuhkan proses yang disebut dengan *diskretisasi* untuk merubah nilai atribut menjadi interval
- Diskretisasi adalah transformasi representasi data dari jenis data bilangan real menjadi data ordinal

Tujuan Diskretisasi

- Agar hasil prediksi model machine learning lebih mudah ditafsirkan

Misalnya:

- Prediksi intensitas hujan
 - Bagi orang umum, biasanya lebih sulit membedakan prediksi intensitas hujan sebesar 19.75 mm/hari dengan 20.25 mm/hari daripada membedakan kategori *hujan ringan* dan *hujan sedang*
- Dengan Diskretisasi:
 - Mentransformasikan variabel target yg semula merupakan data numerik menjadi data kategorikal
 - Masalah regresi dapat dirubah menjadi masalah klasifikasi

Diskretisasi

Misalnya :

- Data input dan prediksi intensitas curah hujan ditransformasi menjadi intensitas dengan kategori:
 - 0 (hujan ringan, intensitas hujan 0.5 – 20 mm/hari)
 - 1 (hujan sedang, intensitas hujan 20 – 50 mm/hari)
 - 2 (hujan lebat, intensitas hujan 50 – 100 mm/hari)
 - 3 (hujan sangat lebat, intensitas hujan 100 – 150 mm/hari)
 - 4 (hujan ekstrem, intensitas hujan > 150 mm/hari)

Menangani Nilai NaN

- NaN (Not a Number) : bukan angka
- Nilai bertipe data Numerik, mewakili nilai yang tidak ditentukan atau tidak terwakili
- Penggunaan NaN secara sistematis diperkenalkan dengan adanya standar titik mengambang IEEE 754 pada tahun 1985.

Menangani Nilai NaN - next

- Dalam bahasa pemrograman Python, nilai NaN merupakan jenis data yang khusus diberikan kepada yang tidak dapat diinterpretasikan,
- Beberapa operasi aritmetik blangan pecahan (*floating point*) yang membangkitkan nilai NaN antara lain:
 - Pembagian bilangan dengan bilangan nol.
 - Logaritma dari bilangan negative
 - Operasi aritmetik yang melibatkan biangan tidak terhngga (*infinity*)
 - Akar kuadrat dari bilangan negative adalah blangan imajiner yang tidak dapat direpresentasikan sebagai bilangan riil.
- Nilai NaN juga bisa digunakan untuk menandai data yan hilang (*missing data*)

Algoritma pembelajaran terhadap NaN

- Beberapa Algoritma pembelajaran machine learning yang toleransi terhadap jenis data NaN :
 - Decision tree
 - K-Nearest Neighbors
- Beberapa Algoritma pembelajaran machine learning yang tidak toleran terhadap jenis data NaN, seperti :
 - *random forest*
 - *support vector machine,*
- maka untuk penanganan khusus perlu dilakukan terhadap data jenis NaN sebelum memproses data training ataupun data testing.

Metode Menangani data NaN

- Beberapa metode yang dilakukan untuk menangani jenis data NaN adalah :
 - Menghapus entitas data yang berisi variabel bernilai NaN dari data training atau data testing
 - Menggantikan nilai NaN dengan sebuah nilai yang diprediksi menggunakan sebuah model atau besaran statistic misalnya: median, mode, atau rata-rata variabel tersebut

RESCALLING

- Adalah transformasi data yang bertujuan agar nilai variabel data memiliki skala atau kisaran (*range*) data yang sama.
- Rescaling sangat penting, mengingat banyak algoritma Machine Learning tidak bekerja secara maksimal, apabila variabel-variabel feature memiliki skala yang berbeda-beda.
- Contoh :
 - Algoritma Support Vector Machine (SVM) yang dalam penerapannya bekerja menggunakan fungsi jarak (*distance function*) untuk melakukan optimasi, akan memiliki performa yang kurang baik jika datanya tidak di Rescalling terlebih dahulu.
 - Misal, apabila pada beberapa feature yang memiliki nilai dari 0 sampai dengan 1, tetapi ada satu feature yang memiliki nilai bervariasi dari 0 hingga 1000, maka hal tersebut akan menimbulkan permasalahan dalam analisa datanya.

Pentingnya Rescalling data :

- Meningkatkan kecepatan konvergensi algoritma pembelajaran seperti: *stochastic gradient descent*
- Menghindari dominasi variabel dengan rentang nilai besar terhadap variabel dengan rentang nilai kecil didalam perhitungan fungsi jarak antara dua data misalnya pada model :
 - Regresi linier, k-Nearest neighbors atau k-means
 - Algoritma pengurangan dimensi seperti principal component analysis (PCA)

Metode Rescaling yang umum digunakan dalam Machine Learning

1. Standarisasi

- Adalah transformasi nilai setiap variabel data sehingga data hasil transformasi memiliki nilai tengah 0 dan deviasi standar 1 menggunakan persamaan :

$$x'_i = \frac{x_i - \bar{x}}{s_x}$$

- Dimana: x'_i adalah data hasil normalisasi, x_i adalah data asli, \bar{x} adalah rata-rata data, dan s_x deviasi standar dari data x

2. Normalisasi

- Adalah transformasi nilai setiap variabel data sehingga data hasil transformasi berada pada interval [0,1] menggunakan persamaan :

$$x'_i = \frac{x_i - \min(x)}{\max(x) - \min(x)}$$

- Dimana : $\min(x)$ adalah nilai minimum dari variabel x dan $\max(x)$ adalah nilai maksimum dari variabel x.

Variabel Kategorikal

- Adalah variabel yang alternative nilainya terbatas, baik berupa bilangan ataupun tekstual.
- Variabel kategorikal terdiri dari :
 - Variabel nominal:
 - Variabel yang nilainya tidak memiliki urutan
 - contoh: jenis kelamin(laki-laki, perempuan), status perkawinan (menikah, belum menikah, janda/duda)
 - Variabel ordinal :
 - Variabel yang nilainya memiliki urutan
 - Contoh: persepsi pelanggan terhadap kualitas layanan sebuah perusahaan, biasanya dalam skala Linkert(sangat tidak puas, cukup puas, puas, dan sangat puas)

Metode untuk menangani variabel kategorikal pada Algoritma Machine Learning :

1. Integer encoding

- Metode untuk menangani variabel ordinal dengan cara setiap nilai kategori direpresentasikan dengan sebuah bilangan

2. One-hit encoding

- Metode untuk menangani variabel nominal dengan cara setiap nilai kategori direpresentasikan dengan sebuah vector bilangan, satu bilangan untuk setiap nilai kategori.
- Misal, kategori warna terdiri dari tiga nilai: merah, hijau, dan biru, maka variabel warna direpresentasikan dengan sebuah vector yang terdiri dari tiga elemen sebagai berikut :

Merah = (1 0 0)

Hijau = (0 1 0)

Biru = (0 0 1)

Variabel Multicollinearity

- Adalah variabel yang memiliki korelasi dengan variabel lain di dalam variabel data.
- Adanya multicollinearity diantara variabel data akan menurunkan kinerja model Machine Learning meskipun dengan kompleksitas model yang tinggi (Garg & Tai, 2013)

Penyebab Multicollinearity

- Multicollinearity disebabkan adanya korelasi atau hubungan yang kuat antara dua variabel bebas atau lebih.
- Penyebab lainnya adalah:
 - Penggunaan variabel dummy yang tidak akurat, seperti ada lebih dari 1 variabel dummy di dalam model
 - Adanya perhitungan sebuah variabel bebas yang didasarkan pada variabel bebas lainnya di dalam model. Misalnya, dalam sebuah model regresi terdapat variabel X_1 , X_2 serta perkalian antara X_1 dan X_2 ($X_1 * X_2$) serta kolinearitas antara X_1 dan $X_1 * X_2$ serta kolinearitas X_2 dan $X_1 * X_2$
 - Adanya pengulangan variabel bebas di dalam sebuah model yang dikembangkan

Dampak dari Multicollinearity :

- Koefisien partial, misalnya pada regresi tidak terukur secara presisi yang dapat menyebabkan nilai standar errornya besar.
- Perubahan pada satu variabel dapat menyebabkan perubahan besar pada nilai koefisien variabel lainnya.
- Nilai *confidence* interval bisa sangat lebar, sehingga akan menjadi sulit untuk menolak hipotesis nol pada sebuah penelitian.

Salah satu cara menangani variabel yang bersifat multicollinearity adalah dengan metode *dimensionality reduction* seperti : Algoritma PCA (*Principal Component Analysis*)

mempersiapkan data dengan Python :

- Di dalam Library Python ada beberapa fungsi, untuk melakukan rescalling, seperti : *MinMaxScaler*, *RobustScaler*, *StandardScaler*. Dan *Normalizer*
- Contoh:
 - data karyawan dengan berbagai tipe data pada variabelnya.

Tabel 5.1. Contoh Data Konsumen dengan berbagai tipe data

Kota	Usia	Gaji	Promosi
Salatiga	45	66029	Ya
Solo	34	83088	Tidak
Klaten	38	81363	Ya
Klaten		93940	Ya
Solo	42	91738	Ya
Salatiga	29	96273	Tidak
Solo	30		Tidak
Klaten	28	61111	Ya
Klaten	34	67938	Ya
Salatiga	39	57189	Ya

Data pada tabel di atas berisikan variabel kota (kota asal pegawai) yang bertipe *string*, data usia, dan gaji yang bertipe *numeric* serta variabel promosi jabatan berupa data kategorikal (berisi “ya” dan “tidak”).

.. Pada data tersebut akan dilakukan *preprocessing* dengan langkah-langkah sebagai berikut.

1) Import library yang diperlukan, yaitu numpy dan pandas.

```
# Mengimpor library yang diperlukan
import numpy as np
import pandas as pd
```

2) Memanggil dataset.

Data pada tabel 5.1 yang telah disimpan dalam file bertipe CSV, selanjutnya dipanggil dengan perintah sebagai berikut.

```
# Memanggil dataset
filename = 'folderlokasi/DataKaryawan.csv'
dataset = pd.read_csv(filename)
X = dataset.iloc[:, :-1].values
y = dataset.iloc[:, 3].values

# Print out dataset
print(dataset)
```

Perintah print akan mencetak isi data dengan output seperti berikut

	Kota	Usia	Gaji	Promosi
0	Salatiga	45.0	66029.0	Ya
1	Solo	34.0	83088.0	Tidak
2	Klaten	38.0	81363.0	Ya
3	Klaten	NaN	93940.0	Ya
4	Solo	42.0	91738.0	Ya
5	Salatiga	29.0	98273.0	Tidak
6	Solo	30.0	NaN	Tidak
7	Klaten	28.0	61111.0	Ya
8	Klaten	34.0	67938.0	Ya
9	Salatiga	39.0	57189.0	Ya

3) Memproses data yang hilang (*missing value*).

Pada dataset terdapat sejumlah data yang *missing*, yang ditunjukkan dengan value NaN. Data yang hilang tersebut akan ditangani dengan menggunakan library *SimpleImputer* sebagai berikut.

```
# Memproses data yang hilang (missing)
from sklearn.impute import SimpleImputer
imputer = SimpleImputer(missing_values= np.
nan, strategy = 'mean')
imputer = imputer.fit(X[:, 1:3])
X[:, 1:3] = imputer.transform(X[:, 1:3])
print(X)
```

Perintah print akan mencetak isi data dengan output seperti berikut.

```
[['Salatiga' 45.0 66029.0]
 ['Solo' 34.0 83088.0]
 ['Klaten' 38.0 81363.0]
 ['Klaten' 35.44444444444444 93940.0]
 ['Solo' 42.0 91738.0]
 ['Salatiga' 29.0 98273.0]
 ['Solo' 30.0 77852.11111111111]
 ['Klaten' 28.0 61111.0]
 ['Klaten' 34.0 67938.0]
 ['Salatiga' 39.0 57189.0]]
```

Terlihat bahwa data-data pada tiga variabel pertama yang sebelumnya kosong sudah terisi.

- 4) Melakukan encoding data kategorikal dan variabel independen.

```
from sklearn.preprocessing import LabelEncoder,
    OneHotEncoder
from sklearn.compose import ColumnTransformer
labelencoder_X = LabelEncoder()
X[:, 0] = labelencoder_X.fit_transform(X[:, 0])
transformer = ColumnTransformer(
    [(Provinsi', OneHotEncoder(), [0])],
    remainder='passthrough')
X=np.array(transformer.fit_transform(X),
    dtype=np.integer)
print(X)
```

Perintah print akan mencetak isi data dengan output seperti berikut.

```
array([[ 0,  1,  0, 45, 66029],
       [ 0,  0,  1, 34, 83088],
       [ 1,  0,  0, 38, 81363],
       [ 1,  0,  0, 35, 93940],
       [ 0,  0,  1, 42, 91738],
       [ 0,  1,  0, 29, 98273],
       [ 0,  0,  1, 30, 77852],
       [ 1,  0,  0, 28, 61111],
       [ 1,  0,  0, 34, 67938],
       [ 0,  1,  0, 39, 57189]])
```

5) Melakukan encoding pada dependent variabel atau variabel Promosi.

```
# Encoding the Dependent Variable  
labelencoder_y = LabelEncoder()  
y = labelencoder_y.fit_transform(y)  
print (y)
```

Outputnya adalah sebagai berikut.

```
[1 0 1 1 1 0 0 1 1 1]
```

Hasil tersebut menunjukkan bahwa pada variabel promosi juga sudah dilakukan encoding menjadi data numerikal dengan nilai 0 dan 1.

Pengantar Regresi:

- Dari Himpunan Data di dapat
 - Bagaimana kecenderngan data tersebut di masa depan?
 - Apakah sama dengan data-data saat ini atau masa lalu?
 - Atau justru berbeda?
- Maka dari himpunan data tersebut perlu dilakukan :
 - Prediksi (*Prediction*)
 - Peramalan (*Forecasting*)
 - Proyeksi (*Projection*)
- Regresi bisa digunakan untuk Prediksi dan Peramalan

Apa itu Regresi (*Regression*)

- Regresi adalah proses identifikasi relasi dan pengaruhnya pada nilai-nilai objek
- Tujuan regresi adalah untuk menemukan suatu fungsi yang memodelkan dengan meminimalkan selisih antara nilai prediksi dengan nilai sebenarnya (*galat*)
- Regresi juga sebagai alat ukur untuk menentukan tingkat perubahan suatu variabel terhadap variabel lainnya.

Regresi :

- Regresi Linier :
 - Regresi linear adalah alat statistik yang dipergunakan untuk mengetahui pengaruh antara satu atau beberapa variabel terhadap satu buah variabel.
 - regresi yang variabel bebasnya (variabel X) berpangkat paling tinggi satu.
- Regresi Non Linier
 - Regresi non linier ialah bentuk hubungan atau fungsi di mana variabel bebas X dan atau variabel tak bebas Y dapat berfungsi sebagai faktor atau variabel dengan pangkat tertentu



Pengantar Analisis Regresi

Analisis Regresi

digunakan untuk mempelajari dan mengukur **pengaruh statistik** yang terjadi antara **dua atau lebih variabel**

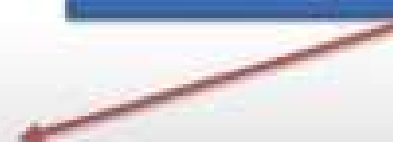
Analisis Regresi Sederhana

Analisis Regresi Berganda

1 variabel terikat (Y)

1 variabel bebas (X)

>1 variabel bebas (X)





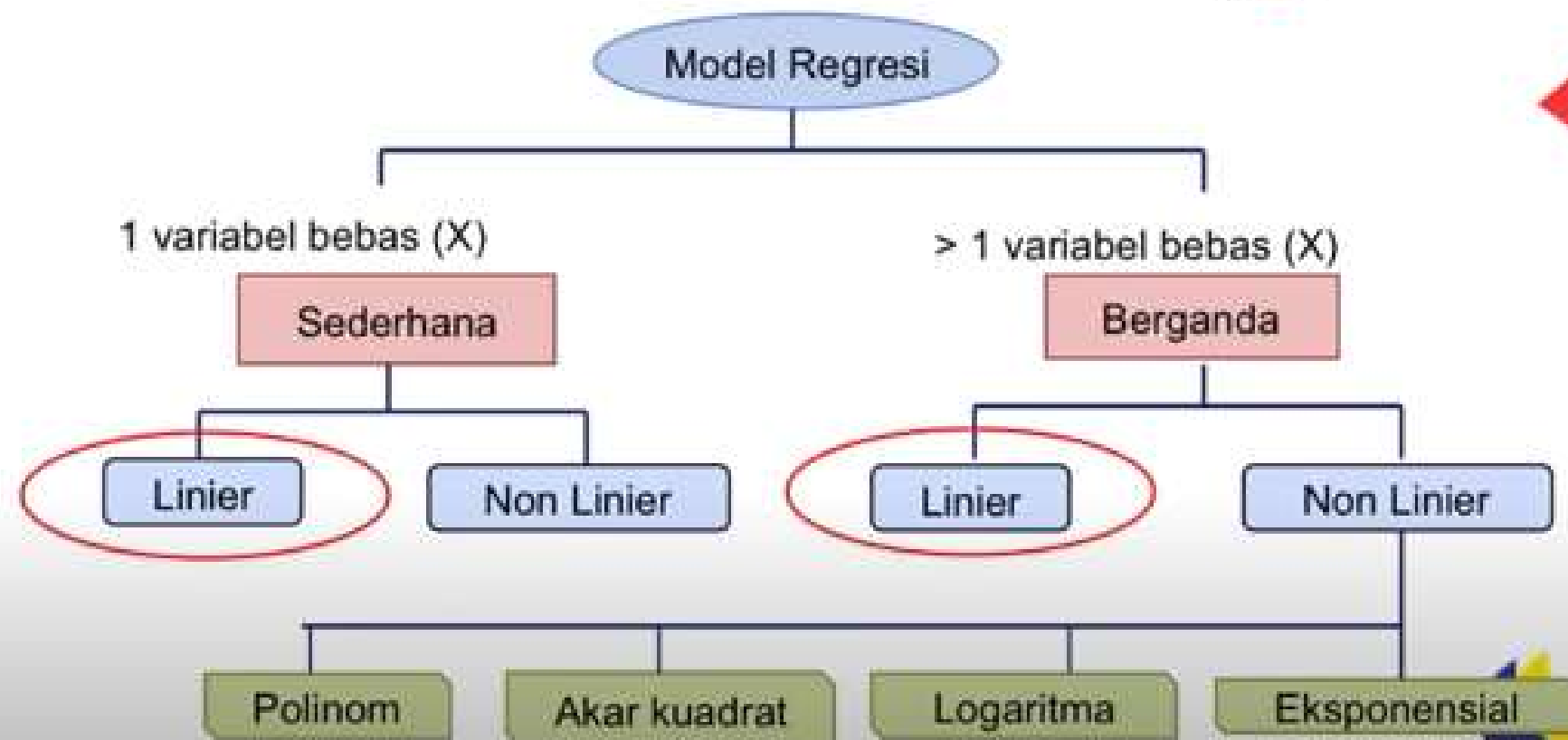
Perbedaan Regresi dan Korelasi

Korelasi



Regresi





Regresi Linier

- Regresi linear adalah alat statistik yang dipergunakan untuk mengetahui pengaruh antara satu atau beberapa variabel terhadap satu buah variabel.
- Regresi linier adalah regresi yang variabel bebasnya (variabel X) berpangkat paling tinggi satu.

Jenis Regresi Linier:

- Regresi Linier Sederhana
 - Mempunyai 1(satu) variabel terikat (*Dependen variable*) dan 1(satu) Variabel bebas (*Independen variable*)
- Regresi Linier Berganda
 - Mempunyai 1(satu) variabel terikat (*Dependen variable*) dan lebih dari 1(satu) Variabel bebas (*Independen variable*)

Regresi Linier Sederhana

- Regresi Linier sederhana adalah Mempunyai 1(satu) variabel terikat (*Dependen variable*) dan 1(satu) Variabel bebas (*Independen variable*)

Persamaan Regresi Linear dari Y terhadap X

Keterangan :

Y = variabel terikat

X = variabel bebas

a = intersep / konstanta

b = koefisien regresi / slop

Persamaan regresi linear di atas dpt pula dituliskan dlm bentuk

The diagram shows the equation $Y = a + bX$ with four labels and leader lines pointing to the corresponding parts: 'intersep' points to 'a', 'Koefisien regresi' points to 'b', 'Variabel bebas' points to 'X', and 'Variabel terikat' points to 'Y'.

$$Y = a + bX$$

$$Y = \left(\frac{\sum xy}{\sum x^2} \right) x$$



Ilustrasi Regresi Linier

Data Kontinu

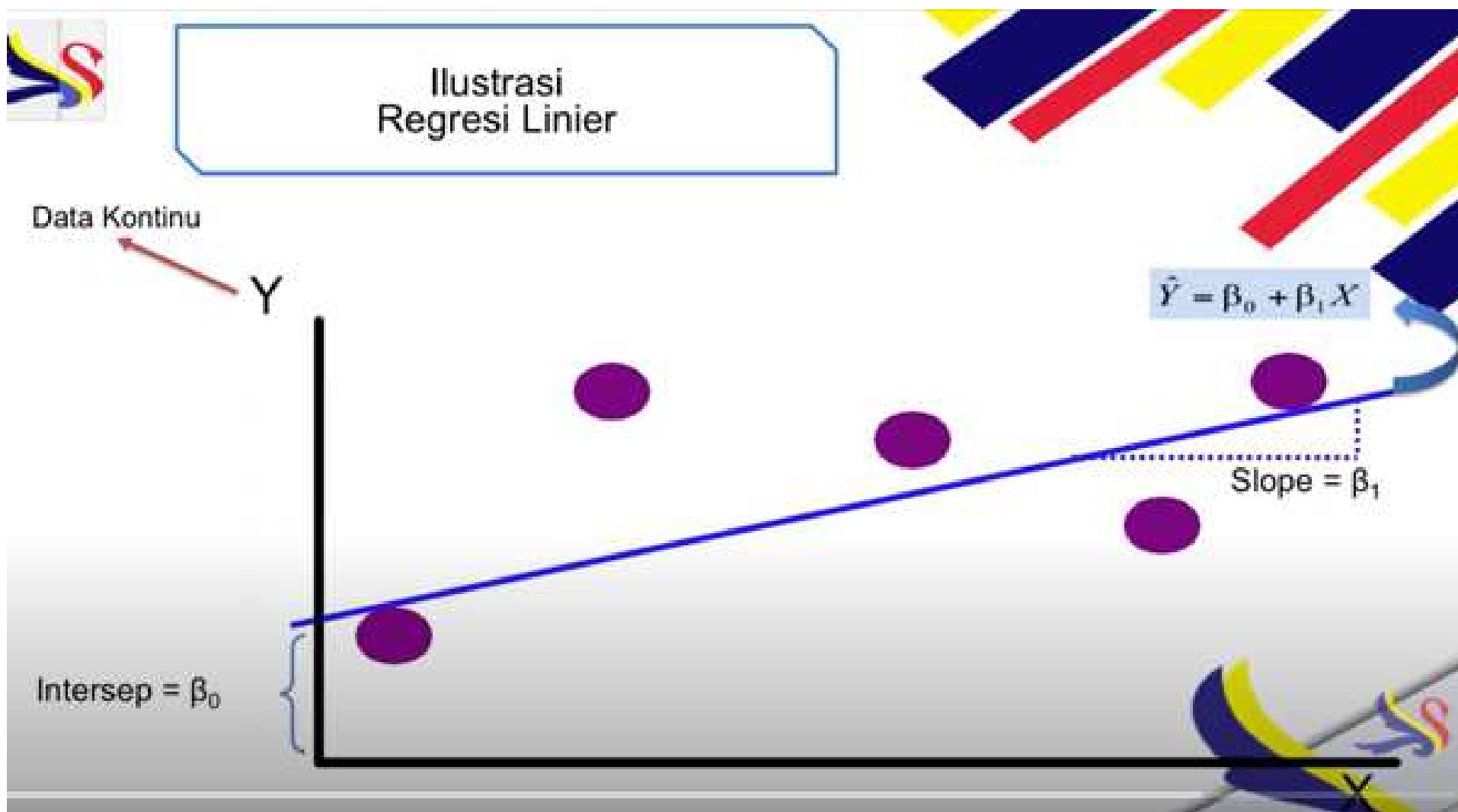
Y

$$\hat{Y} = \beta_0 + \beta_1 X$$

Slope = β_1

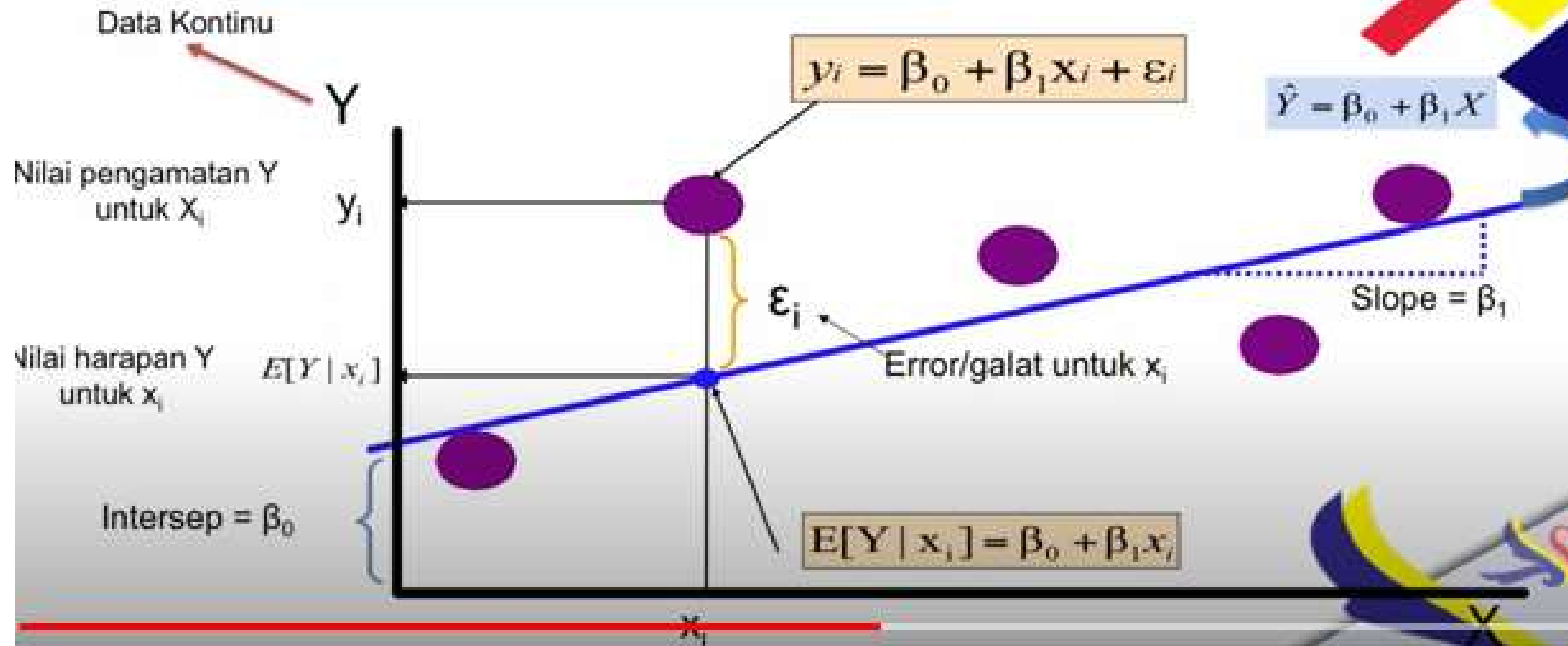
Intersep = β_0

X





Ilustrasi Regresi Linier



Mencari nilai a dan b

- Rumus 1

$$a = \frac{(\Sigma Y)(\Sigma X^2) - (\Sigma X)(\Sigma XY)}{(n)(\Sigma X^2) - (\Sigma X)^2}$$

$$b = \frac{(n)(\Sigma XY) - (\Sigma X)(\Sigma Y)}{(n)(\Sigma X^2) - (\Sigma X)^2}$$

- Pendekatan Matriks (rumus 2)

$$\begin{pmatrix} n & \Sigma X \\ \Sigma X & \Sigma X^2 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} \Sigma Y \\ \Sigma XY \end{pmatrix}$$

$$a = \frac{\det A_1}{\det A} \quad b = \frac{\det A_2}{\det A}$$

$$A = \begin{pmatrix} n & \Sigma X \\ \Sigma X & \Sigma X^2 \end{pmatrix} \quad A_1 = \begin{pmatrix} \Sigma Y & \Sigma X \\ \Sigma XY & \Sigma X^2 \end{pmatrix} \quad A_2 = \begin{pmatrix} \Sigma n & \Sigma Y \\ \Sigma X & \Sigma XY \end{pmatrix}$$

$$\det A = (n)(\Sigma X^2) - (\Sigma X)(\Sigma X)$$

$$\det A_1 = (\Sigma Y)(\Sigma X^2) - (\Sigma X)(\Sigma XY)$$

$$\det A_2 = (n)(\Sigma XY) - (\Sigma Y)(\Sigma X)$$

- Rumus 3

$$b = \frac{(n)(\Sigma XY) - (\Sigma X)(\Sigma Y)}{(n)(\Sigma X^2) - (\Sigma X)^2}$$

$$a = \bar{Y} - b.\bar{X}$$

Contoh Soal

- Berikut ini data mengenai pengalaman kerja dan penjualan
- X =pengalaman kerja (tahun)
- Y =omzet penjualan (ribuan)

X	2	3	2	5	6	1	4	1
Y	5	8	8	7	11	3	10	4

- Tentukan nilai a dan b (gunakan ketiga cara)!
- Buatkan persamaan regresinya!
- Berapa omzet penjualan dari seorang karyawan yg pengalaman kerjanya 3,5 tahun

Penyelesaian :

X	Y	X ²	Y ²	XY
2	5	4	25	10
3	8	9	64	24
2	8	4	64	16
5	7	25	49	35
6	11	36	121	66
1	3	1	9	3
4	10	16	100	40
1	4	1	16	4
24	56	96	448	198

$$\bar{X} = \frac{24}{8} = 3 \quad \bar{Y} = \frac{56}{8} = 7$$

Cara 1.

$$a = \frac{(56)(96) - (24)(198)}{(8)(96) - (24)^2}$$

$$a = \frac{5.376 - 4.752}{768 - 576} = 3,25$$

$$b = \frac{(8)(198) - (24)(56)}{(8)(96) - (24)^2}$$

$$b = \frac{1.584 - 1.344}{768 - 576} = 1,25$$

Cara 2.

$$\begin{pmatrix} 8 & 24 \\ 24 & 96 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 56 \\ 198 \end{pmatrix}$$

$$A = \begin{pmatrix} 8 & 24 \\ 24 & 96 \end{pmatrix} \quad A_1 = \begin{pmatrix} 56 & 24 \\ 198 & 96 \end{pmatrix} \quad A_2 = \begin{pmatrix} 8 & 56 \\ 24 & 198 \end{pmatrix}$$

$$\det A = (8)(96) - (24)(24) = 192$$

$$\det A_1 = (56)(96) - (24)(198) = 624$$

$$\det A_2 = (8)(198) - (56)(24) = 240$$

$$a = \frac{624}{192} = 3,25 \quad b = \frac{240}{192} = 1,25$$

Cara 3

$$b = \frac{(8)(198) - (24)(56)}{(8)(96) - (24)^2}$$

$$b = \frac{1.548 - 1.344}{768 - 576} = 1,25$$

$$a = 7 - 1,25(3)$$

$$a = 3,25$$

- a. Dari ketiga cara pengerjaan tersebut diperoleh nilai $a = 3,25$ dan nilai $b = 1,25$
- b. Persamaan regresi linearnya adalah $Y = 3,25 + 1,25X$
- c. Nilai duga Y , jika $X = 3,5$ adalah $Y = 3,25 + 1,25X$
 $Y = 3,25 + 1,25(3,5)$
 $= 7,625$

Koefisien Determinasi (R^2)

$$R^2 = \frac{((n)(\Sigma XY) - (\Sigma X)(\Sigma Y))^2}{(n(\Sigma X^2) - (\Sigma X)^2)(n(\Sigma Y^2) - (\Sigma Y)^2)}$$

$$R^2 = \frac{((8)(198) - (24)(56))^2}{(8(96) - (24)^2)(8(448) - (56)^2)}$$

$$R^2 = \frac{(1.584 - 1.344)^2}{(768 - 576)(3.584 - 3.136)}$$

$$R^2 = \frac{(240)^2}{(192)(448)} = \frac{57.600}{86.016} = 0,6696$$

Nilai determinasi (R^2) sebesar 0,6696, artinya sumbangan atau pengaruh pengalaman Kerja terhadap naik turunnya omzet penjualan adalah sebesar 66,96%. Sisanya 33,04% Disebabkan oleh faktor lain yang tidak dimasukkan dalam model.