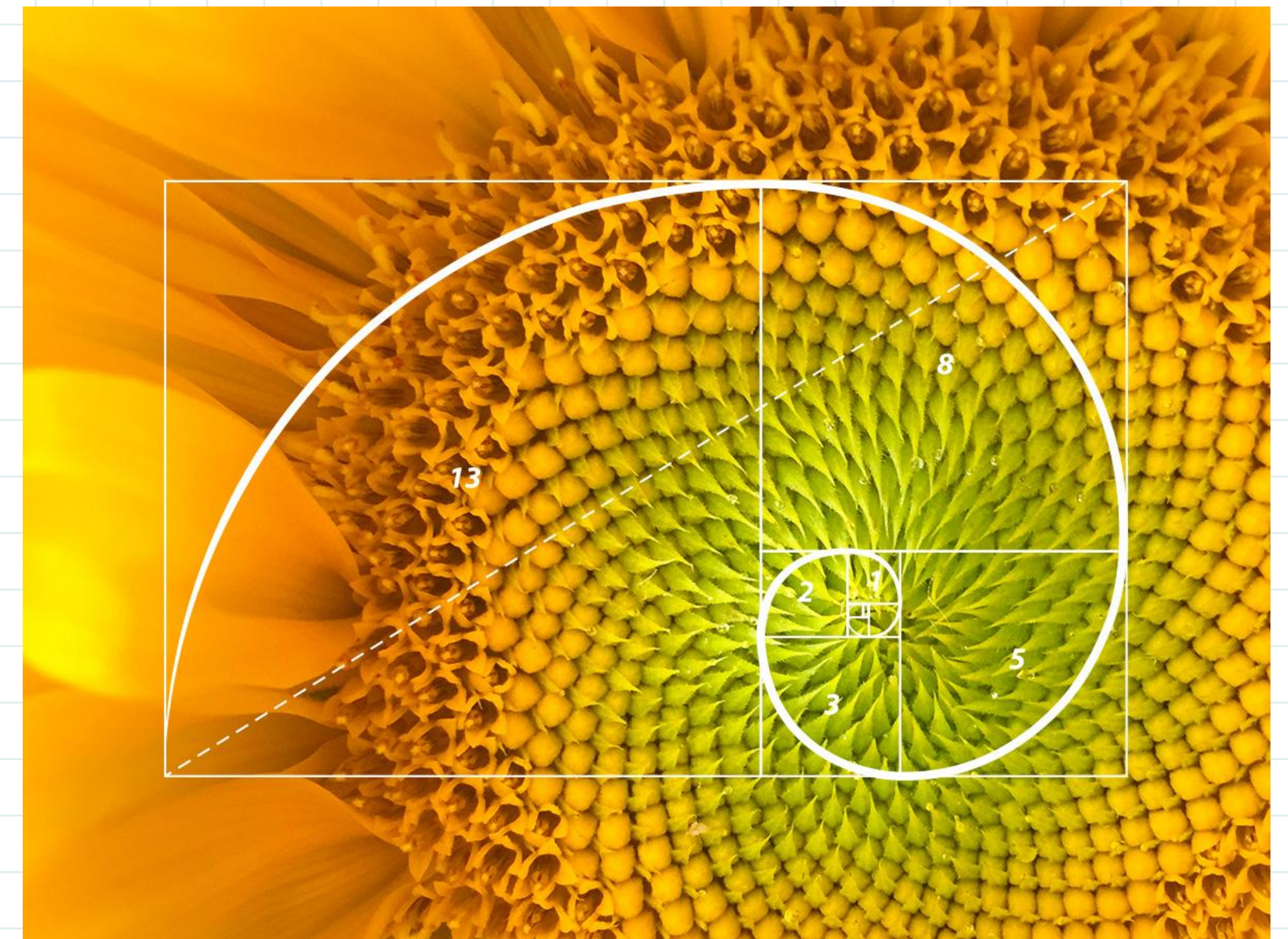


MACHINE LEARNING

Regresi Logistik

Dr. Sirojul Munir, S.Si., M.Kom.
rojulman@nurulfikri.ac.id

ARTIFICIAL INTELLIGENCE – INFORMATICS STTNF

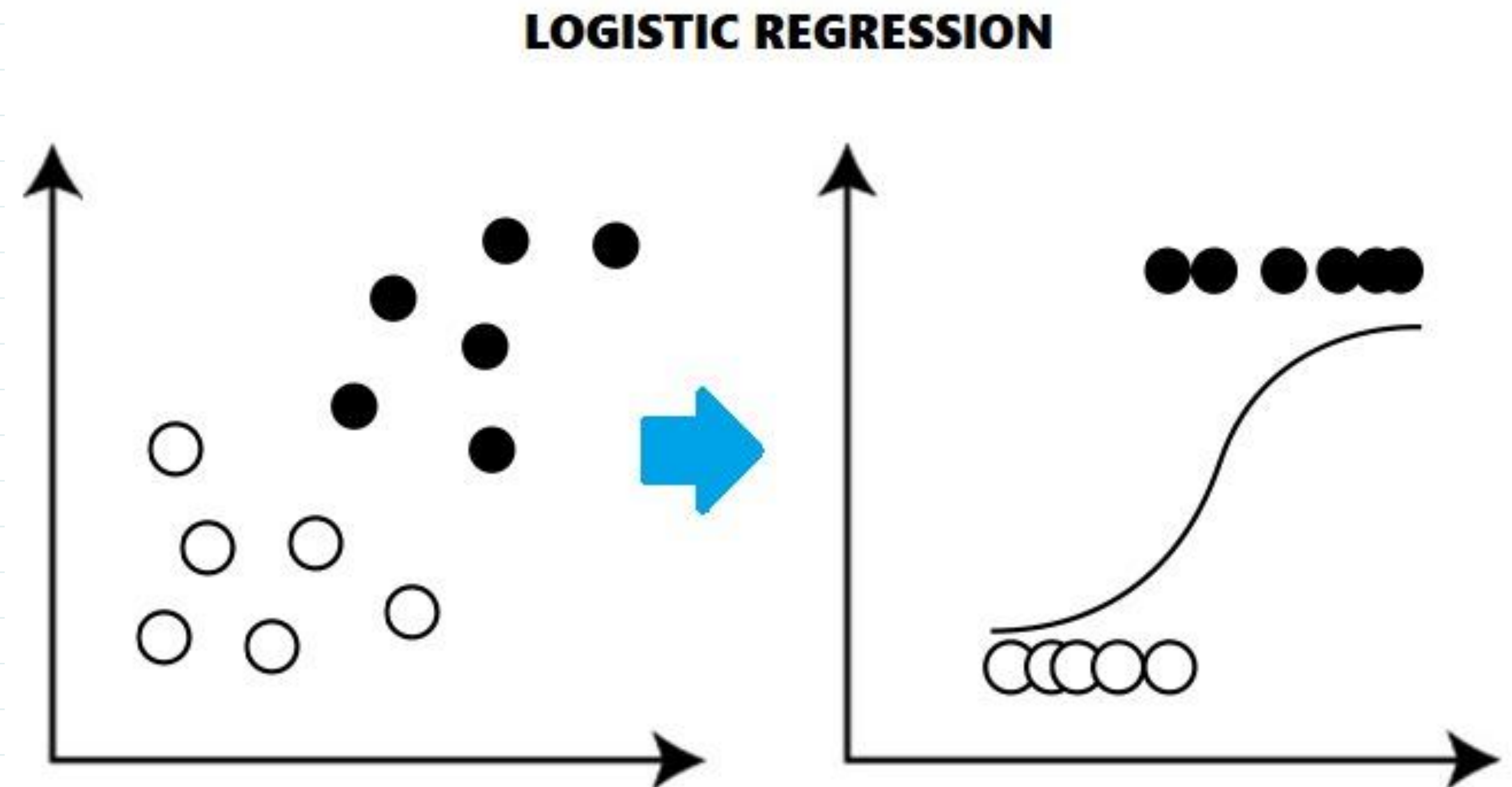


Daur ulang Project Data Science



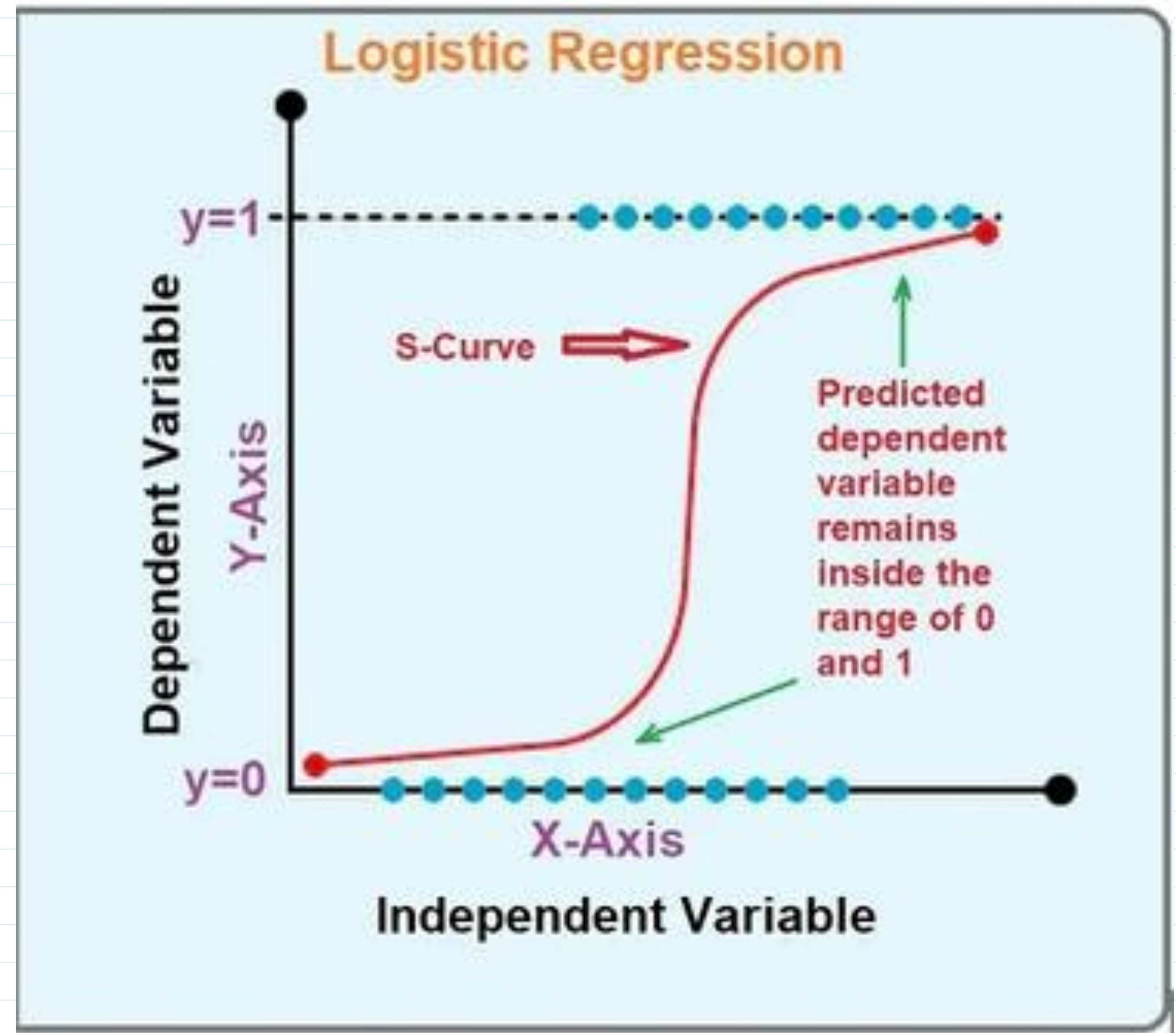
Logistic Regression

- ❑ Regresi: metode analisis data yang menggambarkan hubungan antara variabel respon dan satu atau lebih variabel prediktor
- ❑ Dalam kehidupan nyata sering ditemukan variable respon/target/dependen adalah bernilai diskrit yang mempunyai dua atau lebih kemungkinan hasil (kategori) .



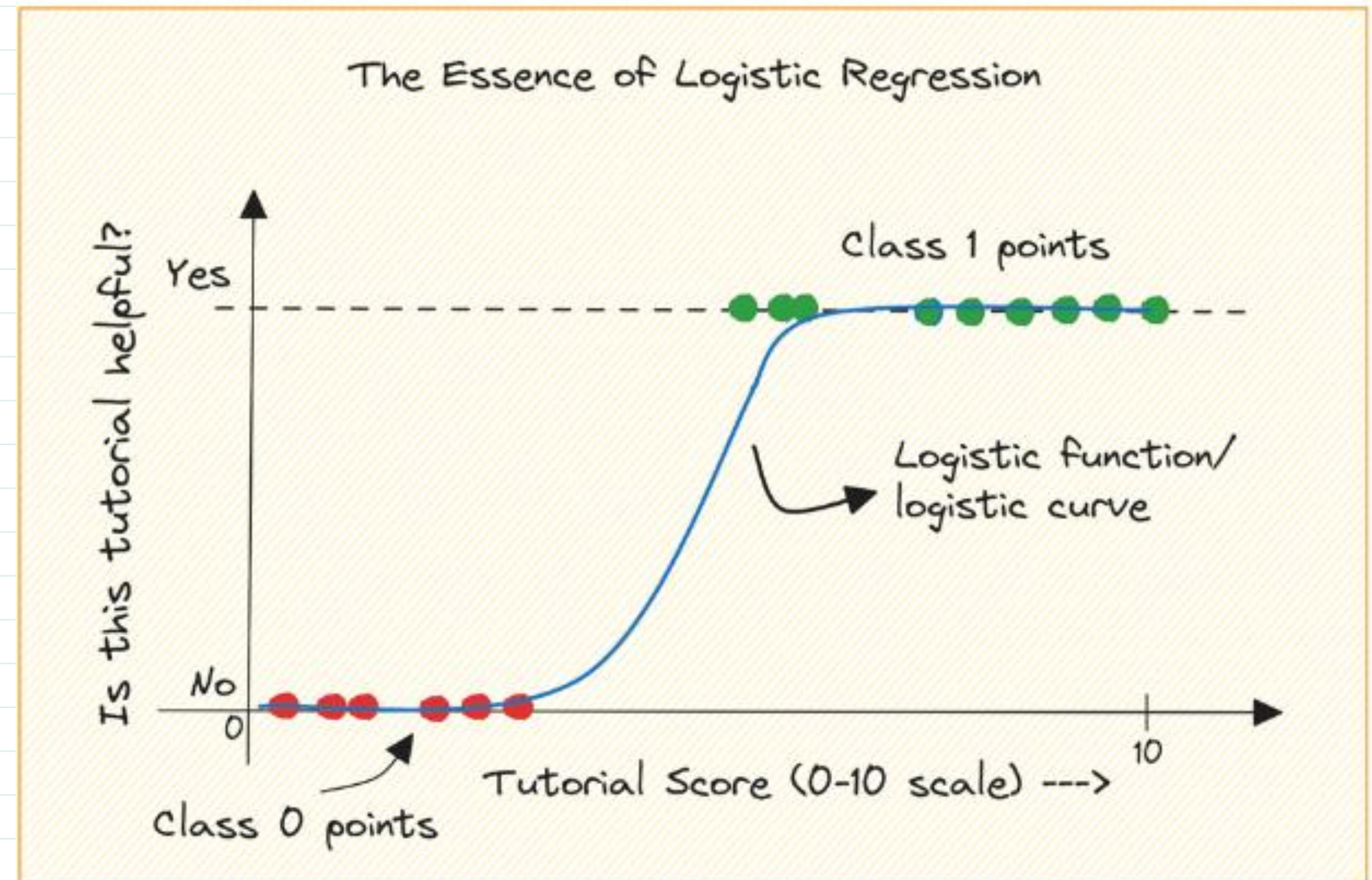
Logistic Regression

- ❑ Regresi: metode analisis data yang menggambarkan hubungan antara variabel respon dan satu atau lebih variabel prediktor
- ❑ Dalam kehidupan nyata sering ditemukan variable respon/target/dependen adalah bernilai diskrit yang mempunyai dua atau lebih kemungkinan hasil (kategori) .



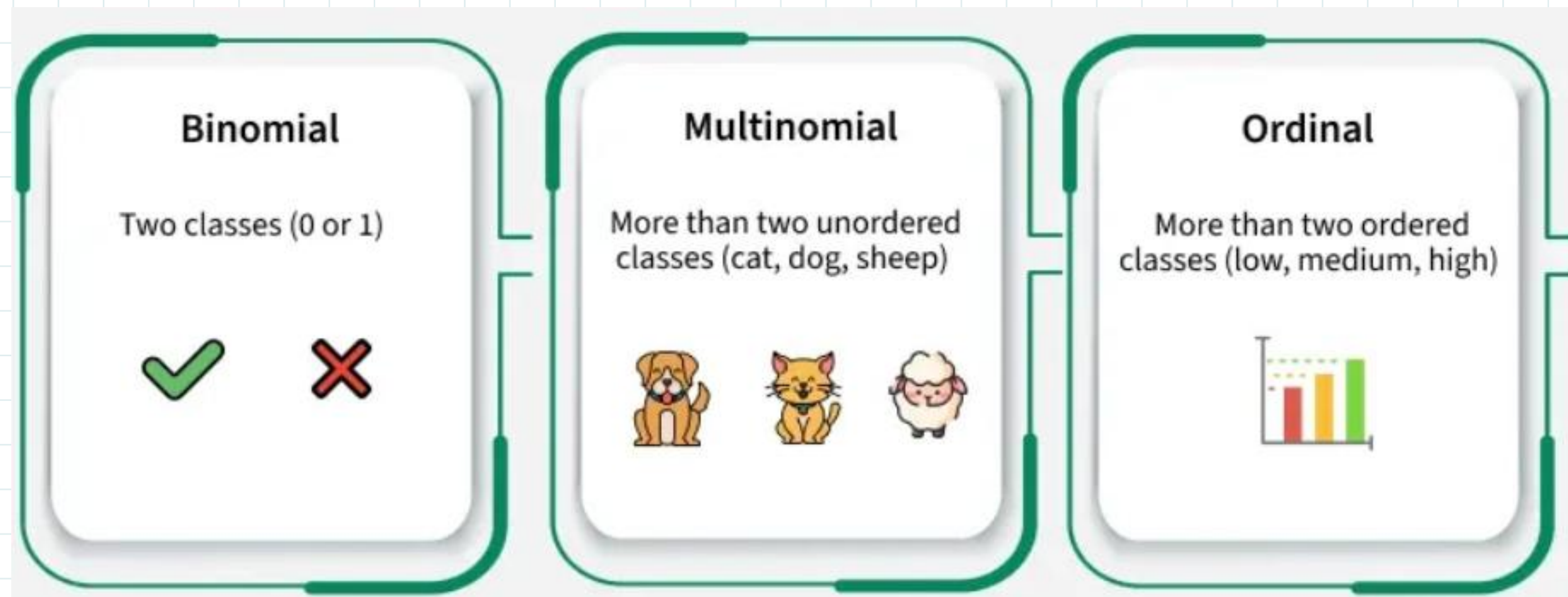
Logistic Regression

- ❑ Regresi: metode analisis data yang menggambarkan hubungan antara variabel respon dan satu atau lebih variabel prediktor
- ❑ Dalam kehidupan nyata sering ditemukan variable respon/target/dependen adalah bernilai diskrit yang mempunyai dua atau lebih kemungkinan hasil (kategori) .



Fungsi Logistik

- ❑ **Regresi Logistik:** Metode analisis yang dapat digunakan untuk menggambarkan hubungan antara variabel respon yang bersifat kategori dengan satu atau lebih variabel prediktor yang bersifat kategori
- ❑ **Logistik binomial:** variable respon terdiri dua kategori
- ❑ **Logistik multinomial:** variable respon bersifat nominal dan lebih dari dua kategori yang tidak terurut
- ❑ **Logistik Ordinal:** variable respon bersifat nominal dan lebih dari dua kategori yang tidak terurut



Persamaan - Regresi Logistik

□ **X – Matrix fitur (independent variable)**

$$X = \begin{bmatrix} x_{11} & \dots & x_{1m} \\ x_{21} & \dots & x_{2m} \\ \vdots & \ddots & \vdots \\ x_{n1} & \dots & x_{nm} \end{bmatrix}$$

□ **Y – Variable Dependent bernilai biner 0 atau 1**

$$Y = \begin{cases} 0 & \text{if Class 1} \\ 1 & \text{if Class 2} \end{cases}$$

□ **Z = Fungsi multi linear dengan inputan variable dependent X**

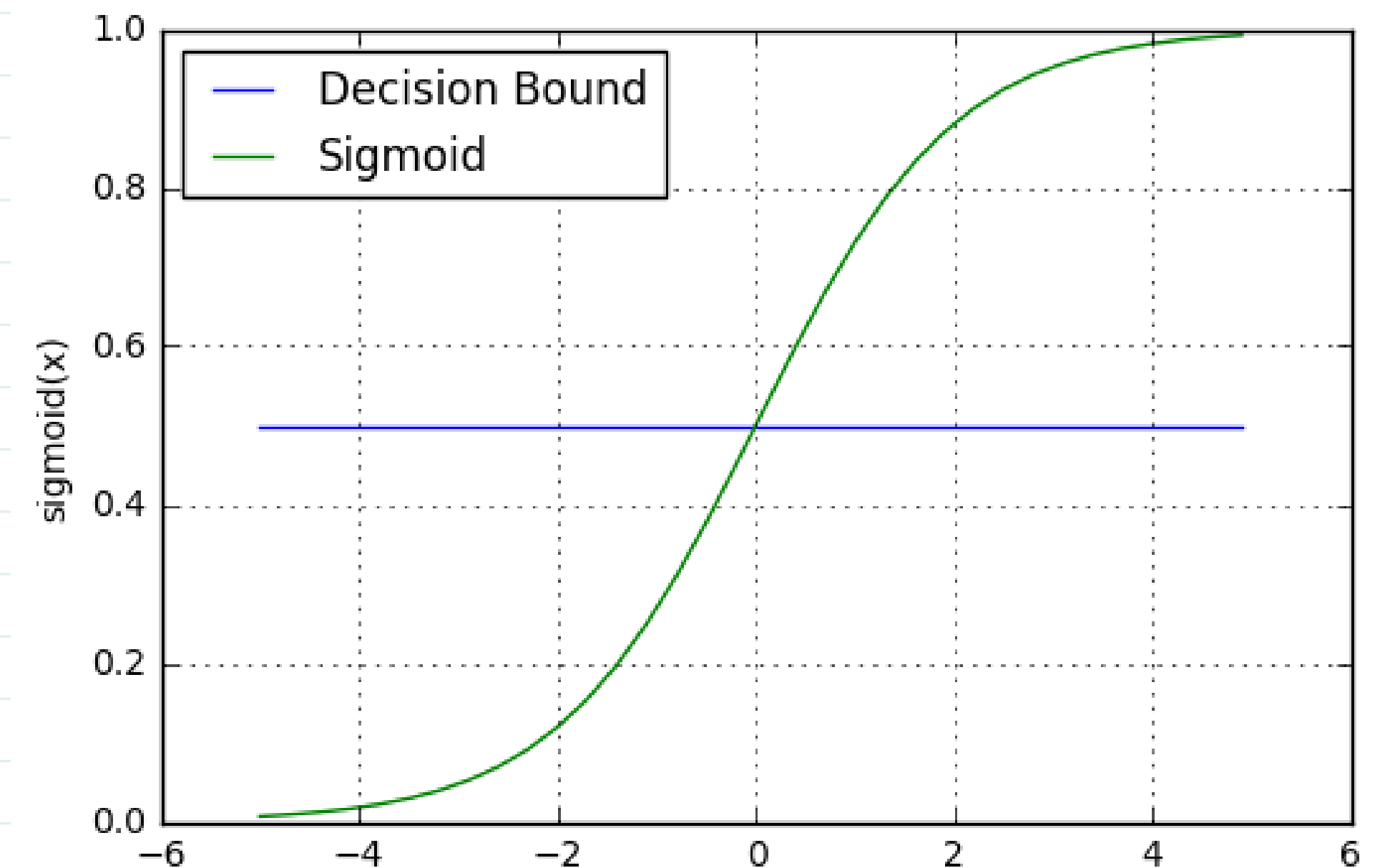
$$z = \left(\sum_{i=1}^n w_i x_i \right) + b$$

$$z = w \cdot X + b$$

$$w_i = [w_1, w_2, w_3, \dots, w_m]$$

Bobot koefisien dari bias / intecept

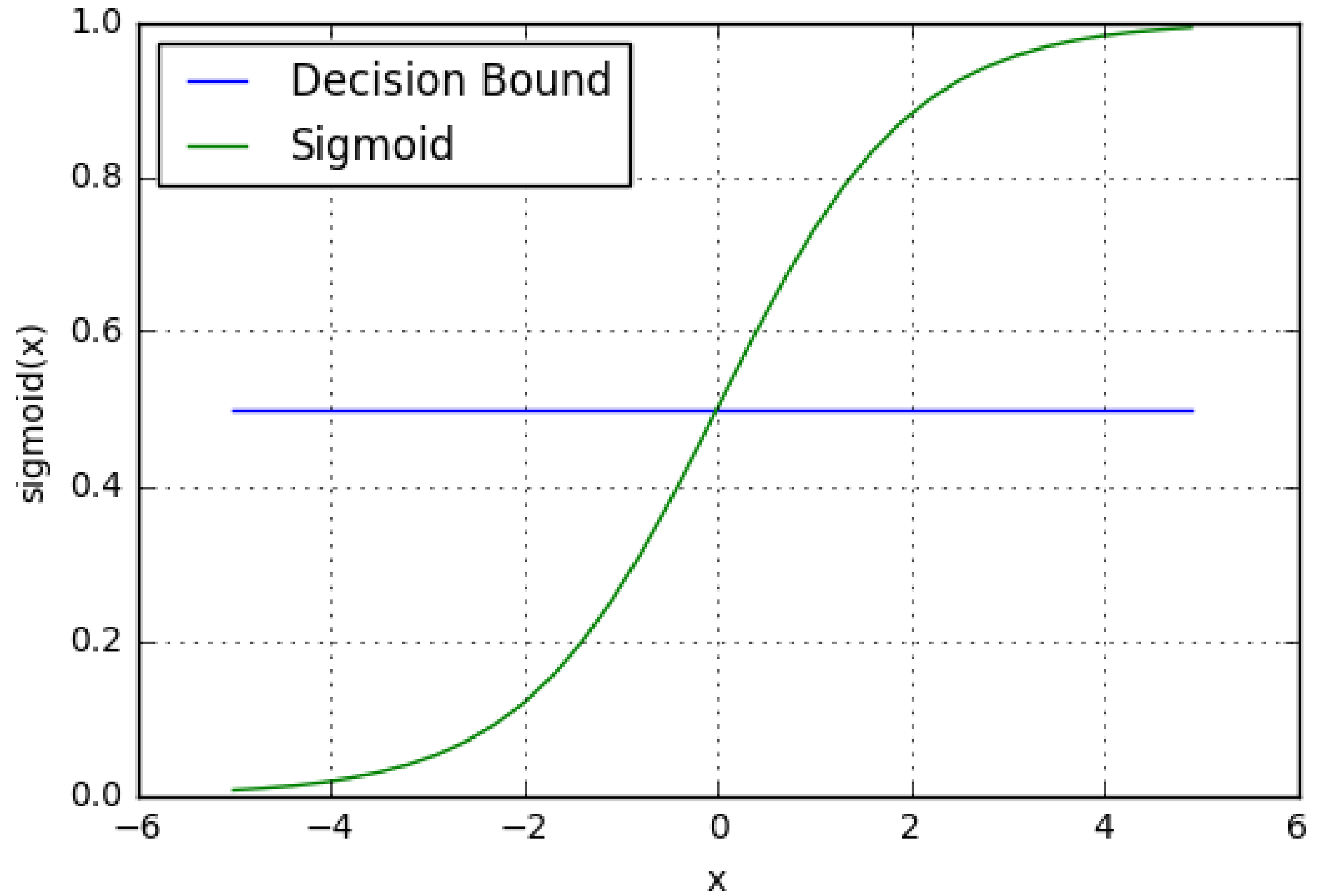
$$\sigma(z) = \frac{1}{1+e^{-z}}$$



- $\sigma(z)$ tends towards 1 as $z \rightarrow \infty$
- $\sigma(z)$ tends towards 0 as $z \rightarrow -\infty$
- $\sigma(z)$ is always bounded between 0 and 1

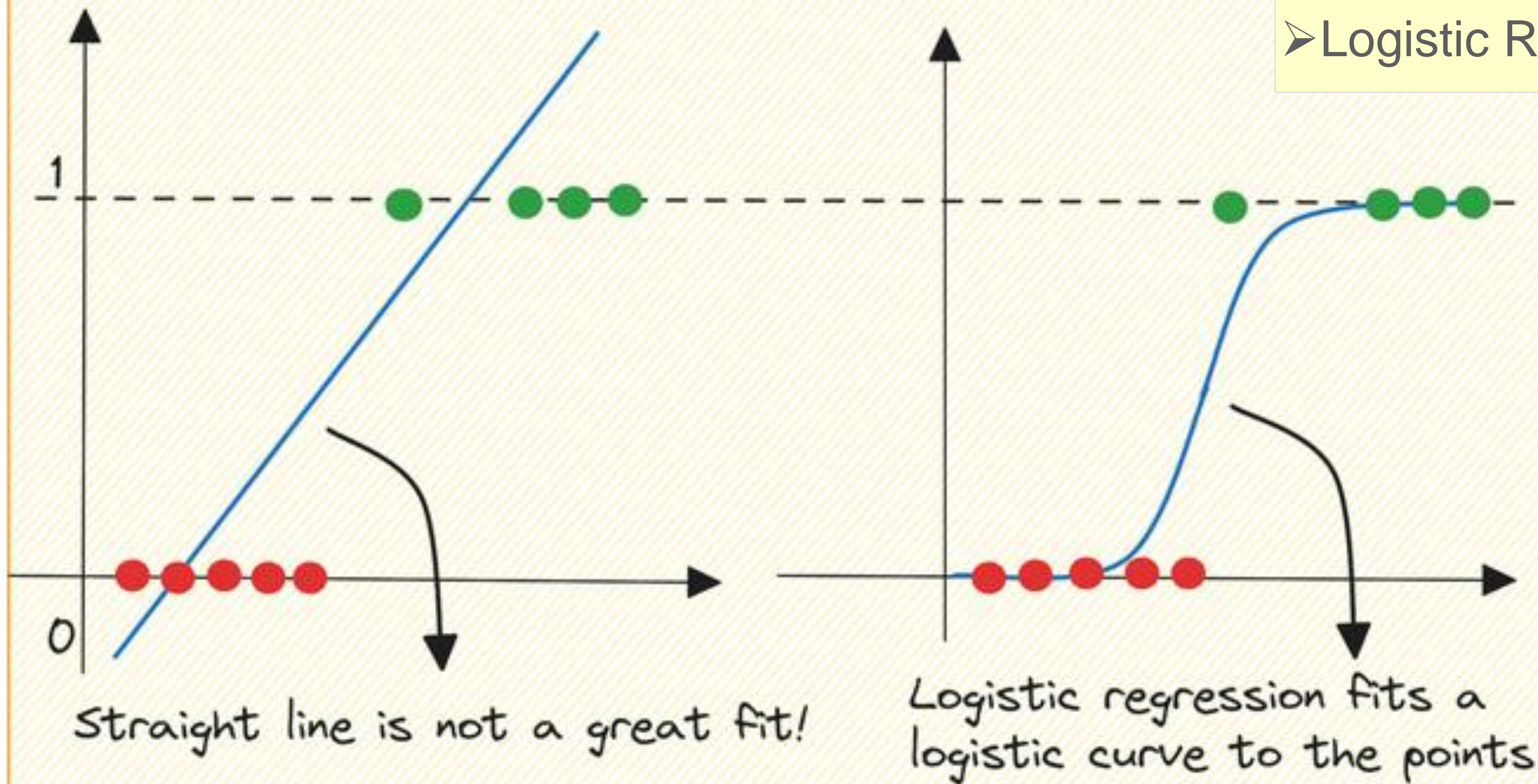
Regresi Logistik

- ❑ Fungsi sigmoid mengembalikan nilai probabilitas 0 atau 1
- ❑ Memetakan nilai probabilitas ke nilai diskrit 0 atau 1
- ❑ Ekpresi matematis:
 - $P \geq 0.5 \Rightarrow \text{class 1}$
 - $P < 0.5 \Rightarrow \text{class 0}$



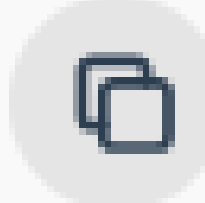
Linear vs. Logistic Regression

Linear vs. Logistic Regression



- Linear Regression: Prediksi Data Kontinu
- Logistic Regression: Prediksi Data Diskrit

Pustaka Program Logistic Regression - Binomial



```
from sklearn.datasets import load_breast_cancer
from sklearn.linear_model import LogisticRegression
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score

X, y = load_breast_cancer(return_X_y=True)

X_train, X_test, y_train, y_test = train_test_split(X, y,
test_size=0.20, random_state=23)

clf = LogisticRegression(max_iter=10000, random_state=0)
clf.fit(X_train, y_train)

acc = accuracy_score(y_test, clf.predict(X_test)) * 100
print(f"Logistic Regression model accuracy: {acc:.2f}%")
```


Pustaka Program Logistic Regression - Multinomial

```
from sklearn.model_selection import train_test_split
from sklearn import datasets, linear_model, metrics

digits = datasets.load_digits()

X = digits.data
y = digits.target

X_train, X_test, y_train, y_test = train_test_split(X, y,
test_size=0.4, random_state=1)

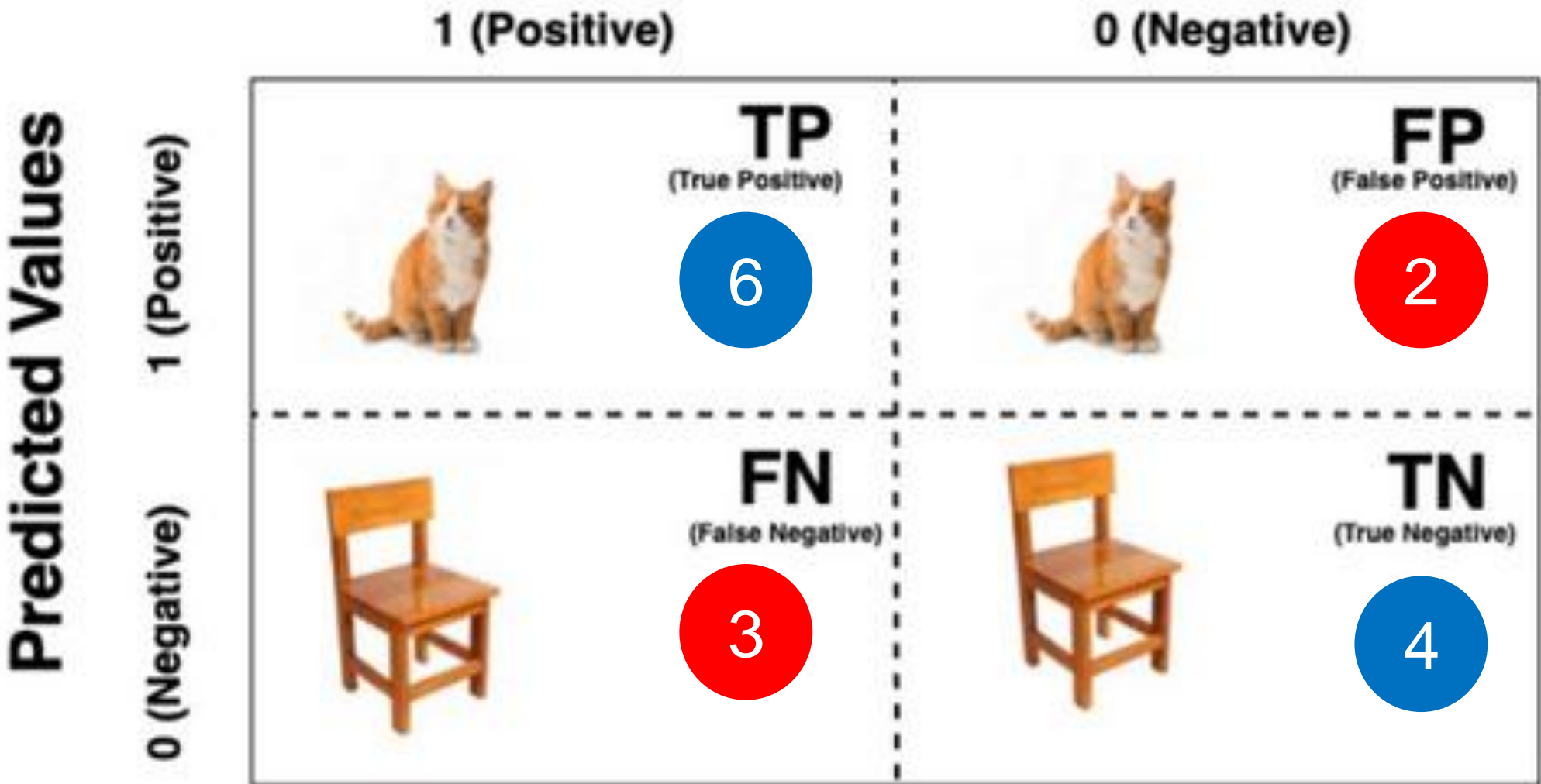
reg = linear_model.LogisticRegression(max_iter=10000,
random_state=0)
reg.fit(X_train, y_train)

y_pred = reg.predict(X_test)

print(f"Logistic Regression model accuracy:
{metrics.accuracy_score(y_test, y_pred) * 100:.2f}%")
```

Evaluasi Model Logistic Regression - Klasifikasi

	1 (Predicted)	0 (Predicted)
1 (Actual)	True Positive	False Negative
0 (Actual)	False Positive	True Negative



Data Ke:	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Prediksi	1	1	0	1	1	1	1	0	0	0	1	1	0	0	0
Aktual	0	1	1	1	1	1	0	0	0	1	1	1	1	0	0
Hasil	FP	TP	FN	TP	TP	TP	FP	TN	TN	FN	TP	TP	FN	TN	TN

Evaluasi Model Logistic Regression

1. **Accuracy:** Accuracy provides the proportion of correctly classified instances.

$$Accuracy = \frac{True\ Positives + True\ Negatives}{Total}$$

2. **Precision:** Precision focuses on the accuracy of positive predictions.

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives}$$

	1 (Predicted)	0 (Predicted)
1 (Actual)	True Positive	False Negative
0 (Actual)	False Positive	True Negative

Evaluasi Model Logistic Regression

3. Recall (Sensitivity or True Positive Rate): Recall measures the proportion of correctly predicted positive instances among all actual positive instances.

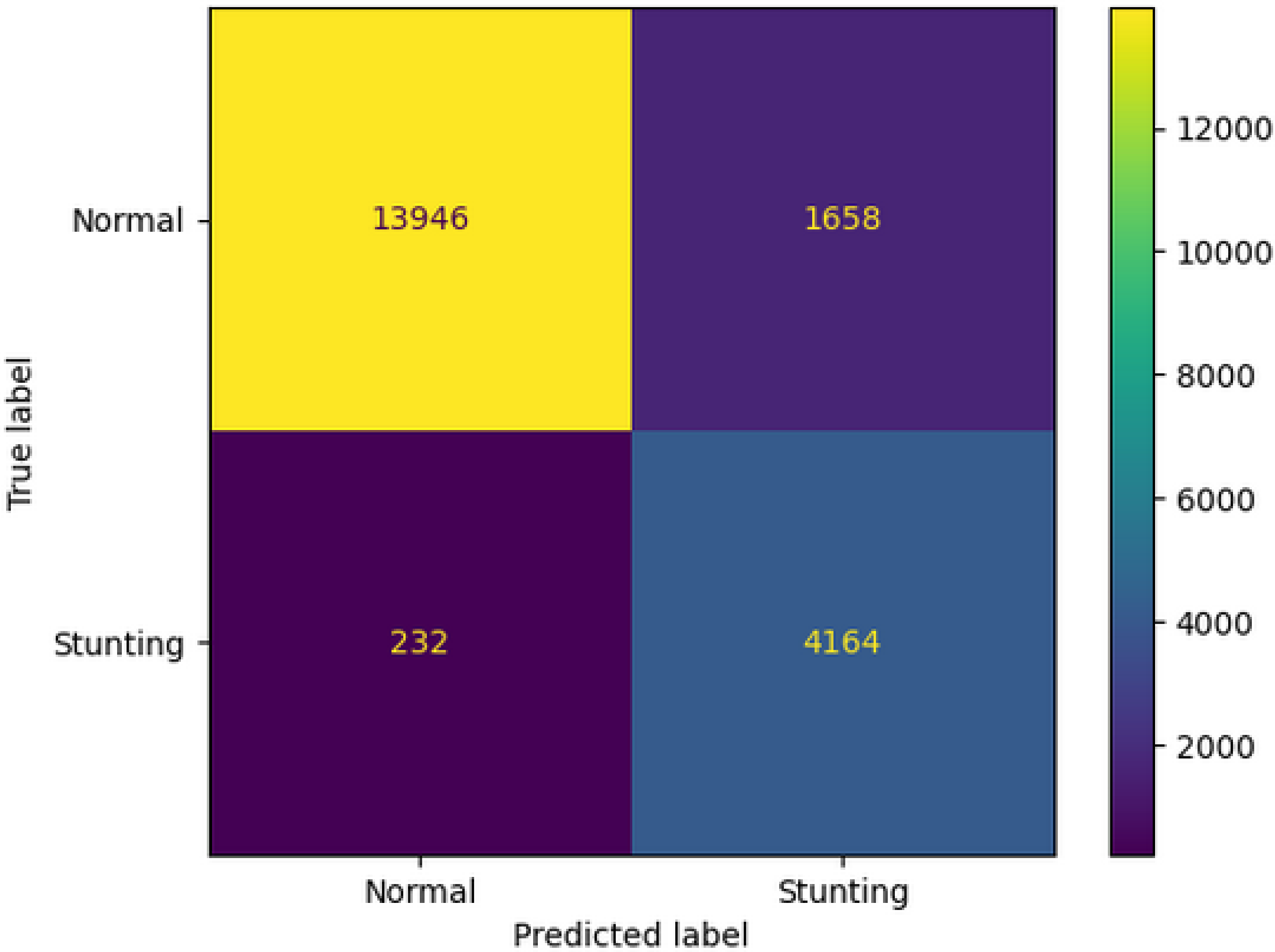
$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

4. F1 Score: F1 score is the harmonic mean of precision and recall.

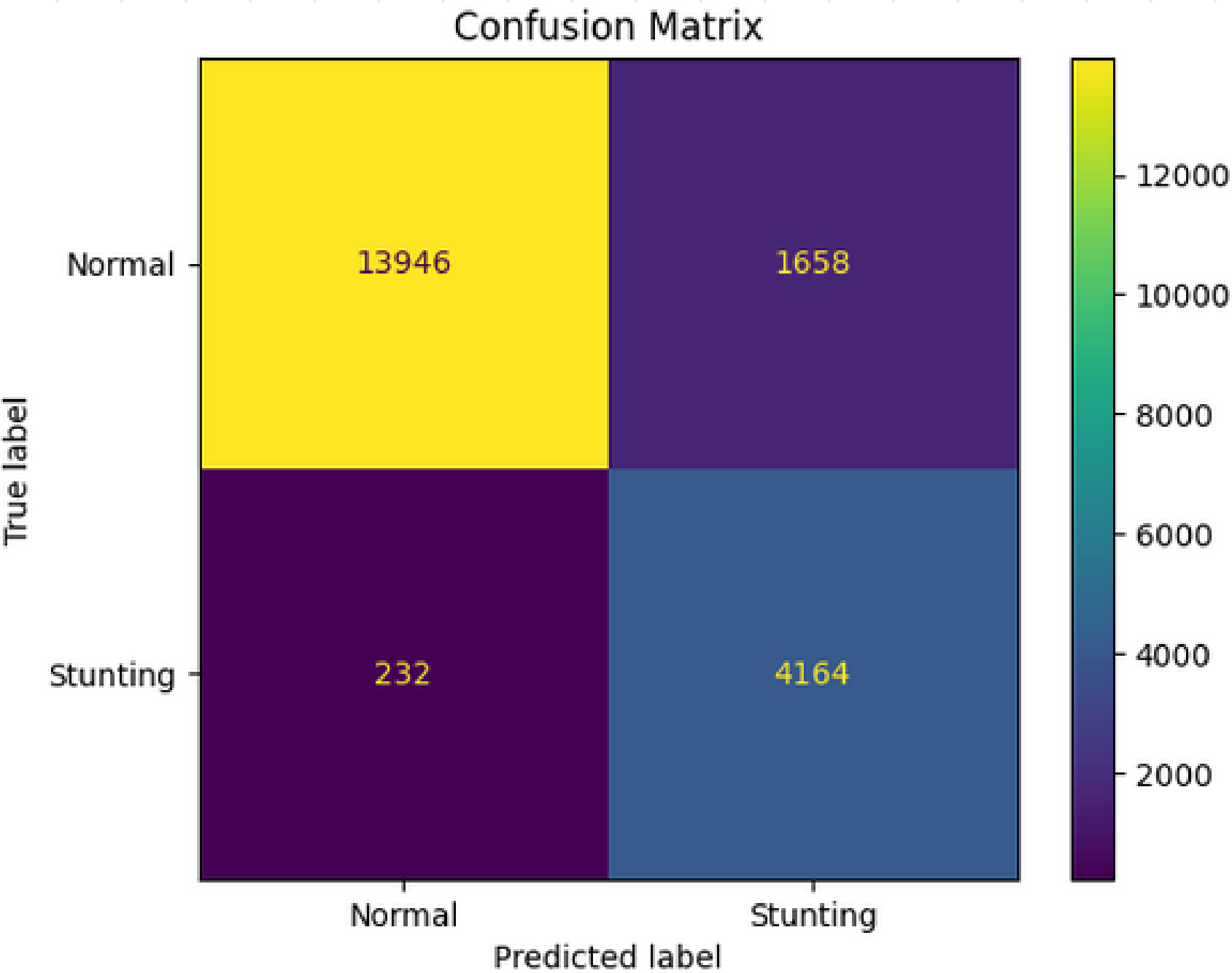
$$F1 \text{ Score} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$$

Evaluasi Model Logistic Regression

Confusion Matrix



Evaluasi Model Logistic Regression



		Actual Values	
		Positive (1)	Negative (0)
Predicted Values	Positive (1)	TP	FP
	Negative (0)	FN	TN

Kelas	Prediksi Benar	Prediksi Salah
Normal	13.946 benar (True Positive)	1.658 salah (False Positive)
Stunting	4.164 benar (True Negatif)	232 salah (False Negative)

Hitung Metrik Evaluasinya ?

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F1-score = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

Praktikum Case Study: Prediksi Berat Balita

Sampel Data:

	Jenis Kelamin	Umur (bulan)	Tinggi Badan (cm)	Berat Badan (kg)	Stunting	Wasting
0	Laki-laki	19	91.6	13.3	Tall	Risk of Overweight
1	Laki-laki	20	77.7	8.5	Stunted	Underweight
2	Laki-laki	10	79.0	10.3	Normal	Risk of Overweight
3	Perempuan	2	50.3	8.3	Severely Stunted	Risk of Overweight
4	Perempuan	5	56.4	10.9	Severely Stunted	Risk of Overweight

Prediksi balita normal atau stunting !

Praktikum Case Study: Prediksi Berat Balita

Sampel Data:

	Jenis Kelamin	Umur (bulan)	Tinggi Badan (cm)	Berat Badan (kg)	Stunting	Wasting	Stunting_bin	JK_bin
0	Laki-laki	19	91.6	13.3	Tall	Risk of Overweight	0	1
1	Laki-laki	20	77.7	8.5	Stunted	Underweight	1	1
2	Laki-laki	10	79.0	10.3	Normal	Risk of Overweight	0	1
3	Perempuan	2	50.3	8.3	Severely Stunted	Risk of Overweight	1	0
4	Perempuan	5	56.4	10.9	Severely Stunted	Risk of Overweight	1	0

Case Studi: Calon pembeli mobil

Dataset: <https://www.kaggle.com/code/ajieraflipamungkas/prediksi-calon-pembeli-regresi-logistik>

- **Usia:** Usia calon pembeli mobil
- **Status:** Status pernikahan calon pembeli (0=single, 1=menikah, 2=menikah mempunyai anak, 3=duda/janda)
- **Kelamin:** Jenis kelamin calon pembeli (0=pria, 1=wanita).
- **Memiliki Mobil:** Jumlah mobil yang dimiliki calon pembeli.
- **Penghasilan:** Penghasilan calon pembeli dalam setahun (jutaan).
- **Beli Mobil (Target):** Apakah calon pembeli mobil benar benar membeli mobil atau tidak (0=tidak membeli, 1=membeli mobil)



Praktikum Mandiri: Prediksi Pembelian Mobil

Sampel Data:

[1]:

	ID	Usia	Status	Kelamin	Memiliki_Mobil	Penghasilan	Beli_Mobil
0	1	32	1	0	0	240	1
1	2	49	2	1	1	100	0
2	3	52	1	0	2	250	1
3	4	26	2	1	1	130	0
4	5	45	3	0	2	237	1

Prediksi apakah pelanggan benar-benar akan membeli mobil!



Terima Kasih

<http://youtube.com/@rojulman>