

# Data Mining

## Pertemuan 6

### Classification : Confusion Matrix

# Confussion Matrix

- Tata letak tabel khusus yang memungkinkan visualisasi kinerja suatu algoritma pada *supervised learning*
- Dalam *unsupervised learning* biasanya disebut *Matching Matrix*

# Tata letak Confusion Matrix

- Setiap baris matriks mewakili instance di kelas yang diprediksi sementara setiap kolom mewakili instance di kelas yang sebenarnya.

# Contoh

- Actual =  $[1, 1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0]$
- prediction =  $[0, 0, 0, 1, 1, 1, 1, 1, 0, 0, 0, 1, 1]$

|                 |   | Actual class |   |
|-----------------|---|--------------|---|
|                 |   | 1            | 0 |
| Predicted class | 1 | 5            | 2 |
|                 | 0 | 3            | 3 |

Benar = 8  
Salah = 5  
Akurasi =  $8 / 13$   
= 0.615

# Matrix

|                 |   | Actual class |           |
|-----------------|---|--------------|-----------|
|                 |   | P            | N         |
| Predicted class | P | <b>TP</b>    | FP        |
|                 | N | FN           | <b>TN</b> |

|                  |       | Actual class            |                         |
|------------------|-------|-------------------------|-------------------------|
|                  |       | 1                       | Not 1                   |
| Predict ed class | 1     | <b>5 True Positives</b> | 2 False Positives       |
|                  | Not 1 | 3 False Negatives       | <b>3 True Negatives</b> |

P = Positive; N = Negative;  
TP = True Positive; FP = False Positive;  
TN = True Negative; FN = False Negative.

# Multi Class

Untuk Class a

|   | a | b | c | d |
|---|---|---|---|---|
| a | 7 | 2 | 3 | 5 |
| b | 6 | 5 | 4 | 6 |
| c | 4 | 1 | 7 | 3 |
| d | 2 | 2 | 1 | 3 |

True Positive

False Positive

True Negative

False Negative

# Bingung?

- Ketika sumbu x dan y sama maka **True Positive**
- Sumbu x (kolom) pada dirinya (y) selain dirinya (x) adalah **False Positive**
- Sumbu y (baris) pada dirinya (x) selain dirinya (y) adalah **False Negative**
- Selain itu **True Negative**

# Precision

- Precision atau positive predictive value (PPV)

$$\text{PPV} = \frac{\text{TP}}{\text{TP} + \text{FP}} = 1 - \text{FDR}$$

- Kebalikannya : False discovery rate (FDR)

$$\text{FDR} = \frac{\text{FP}}{\text{FP} + \text{TP}} = 1 - \text{PPV}$$



# Recall

- Sensitivity, recall, hit rate, atau true positive rate (TPR)

$$\text{TPR} = \frac{\text{TP}}{\text{P}} = \frac{\text{TP}}{\text{TP} + \text{FN}} = 1 - \text{FNR}$$

- Kebalikannya : miss rate atau false negative rate (FNR)

$$\text{FNR} = \frac{\text{FN}}{\text{P}} = \frac{\text{FN}}{\text{FN} + \text{TP}} = 1 - \text{TPR}$$

# F1 Score

- Disebut juga harmonic mean dari precision dan recall

$$F_1 = 2 \cdot \frac{PPV \cdot TPR}{PPV + TPR} = \frac{2TP}{2TP + FP + FN}$$

# Akurasi berdasarkan Matrix

- Akurasi

$$ACC = \frac{TP + TN}{P + N} = \frac{TP + TN}{TP + TN + FP + FN}$$

- Balanced Accuracy

$$BA = \frac{TPR + TNR}{2}$$