Outlier Multi Dimension – Isolated Forest

Knowledge Discovery

Hanif Izzudin Rahman

➔ Data Frame

In [6]:

```
import pandas as pd
import numpy as np
from sklearn.ensemble import IsolationForest

data={'Data 1': [80, 60, 75, 73, 81, 66],
      'Data 2': [70, 84, 88, 90, 65, 60],
      'Data 3': [60, 40, 55, 58, 47, 49],
      'Data 4': [60, 65, 60, 70, 68, 72],
      'Data 5': [60, 35, 40, 20, 56, 57],
      'Data 6': [95, 98, 85, 87, 93, 95],
      'Data 7': [55, 70, 53, 64, 74, 77],
      'Data 8': [50, 53, 57, 63, 58, 40],
      'Data 9': [62, 64, 53, 58, 40, 45]}

df=pd.DataFrame(data, columns=['Data 1','Data 2', 'Data 3','D
df
```

Out[6]:

	Data 1	Data 2	Data 3	Data 4	Data 5	Data 6	Data 7	Data 8	Data 9
0	80	70	60	60	60	95	55	50	62
1	60	84	40	65	35	98	70	53	64
2	75	88	55	60	40	85	53	57	53
3	73	90	58	70	20	87	64	63	58
4	81	65	47	68	56	93	74	58	40
5	66	60	49	72	57	95	77	40	45

➔ Data – Transpose (6 Feature)

```
# Transponse
df=df.T
df
```

	0	1	2	3	4	5
Data 1	80	60	75	73	81	66
Data 2	70	84	88	90	65	60
Data 3	60	40	55	58	47	49
Data 4	60	65	60	70	68	72
Data 5	60	35	40	20	56	57
Data 6	95	98	85	87	93	95
Data 7	55	70	53	64	74	77
Data 8	50	53	57	63	58	40
Data 9	62	64	53	58	40	45

➔ Outlier Detection – Isolated Forest

```
1  clf = IsolationForest(contamination=0.3)
2  pred = clf.fit_predict(df)
3
4  df['Outlier']=pred.reshape(-1,1)
5
6  df
```

	0	1	2	3	4	5	Outlier
Data 1	80	60	75	73	81	66	1
Data 2	70	84	88	90	65	60	-1
Data 3	60	40	55	58	47	49	1
Data 4	60	65	60	70	68	72	1
Data 5	60	35	40	20	56	57	-1
Data 6	95	98	85	87	93	95	-1
Data 7	55	70	53	64	74	77	1
Data 8	50	53	57	63	58	40	1
Data 9	62	64	53	58	40	45	1