

Penambahan Data

# ***Politeknik Caltex Riau***

Apriori

Dosen/PLP :

**Dini Nurmalasari, S.T.,M.T**

**Asmarini, S.Tr.Kom**

Identitas :

Nama Lengkap : Nanda Habibie Erwin

NIM : 2257301100

Kelas : 2 SI A

**PRODI SISTEM INFORMASI  
JURUSAN TEKNOLOGI INFORMASI**

```
import numpy as np
import pandas as pd
from mlxtend.frequent_patterns import apriori, association_rules
```

Analisa:

1. Codingan ini mengimport library yg diperlukan untuk data mining model apriori

```
data = pd.read_excel("/content/drive/MyDrive/Data Mining/P-11/Online Retail.xlsx")
data.head()
```

/usr/local/lib/python3.10/dist-packages/ipykernel/ipkernel.py:283: DeprecationWarning: `should\_run\_async` will not call `transform\_cell` au and should\_run\_async(code)

	InvoiceNo	StockCode	lower	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
0	536365	85123A	white hanging heart t-light holder	WHITE HANGING HEART T-LIGHT HOLDER	6	2010-12-01 08:26:00	2.55	17850.0	United Kingdom
1	536365	71053	white metal lantern	WHITE METAL LANTERN	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom
2	536365	84406B	cream cupid hearts coat hanger	CREAM CUPID HEARTS COAT HANGER	8	2010-12-01 08:26:00	2.75	17850.0	United Kingdom
3	536365	84029G	knitted union flag hot water bottle	KNITTED UNION FLAG HOT WATER BOTTLE	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom
4	536365	84029E	red woolly hottie white heart.	RED WOOLLY HOTTIE WHITE HEART.	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom

Analisa:

1. Codingan ini menampilkan isi dari file excel Online Retail

```
data['Description'] = data['Description'].str.strip()
data.dropna(axis=0, subset=['InvoiceNo'], inplace=True)
data['InvoiceNo'] = data['InvoiceNo'].astype('str')
data = data[~data['InvoiceNo'].str.contains('C')]
```

Analisa:

2. Codingan ini membersihkan dan mempersiapkan dataframe dengan menghapus spasi ekstra, baris dengan nilai kosong pada kolom 'InvoiceNo' dan mengubah 'InvoiceNo' ke tipe data string

```
basket_France = (data[data['Country']=="France"]
                  .groupby(['InvoiceNo', 'Description'])['Quantity']
                  .sum().unstack().reset_index().fillna(0)
                  .set_index('InvoiceNo'))

/usr/local/lib/python3.10/dist-packages/ipykernel/ipkernel.py:283: DeprecationWarning: `should_run_async` is deprecated and should be removed in a future version of IPython.
and should_run_async(code)

basket_UK = (data[data['Country']=="United Kingdom"]
             .groupby(['InvoiceNo', 'Description'])['Quantity']
             .sum().unstack().reset_index().fillna(0)
             .set_index('InvoiceNo'))

/usr/local/lib/python3.10/dist-packages/ipykernel/ipkernel.py:283: DeprecationWarning: `should_run_async` is deprecated and should be removed in a future version of IPython.
and should_run_async(code)

basket_Por = (data[data['Country']=="Portugal"]
             .groupby(['InvoiceNo', 'Description'])['Quantity']
             .sum().unstack().reset_index().fillna(0)
             .set_index('InvoiceNo'))

/usr/local/lib/python3.10/dist-packages/ipykernel/ipkernel.py:283: DeprecationWarning: `should_run_async` is deprecated and should be removed in a future version of IPython.
and should_run_async(code)

basket_Sweden = (data[data['Country']=="Sweden"]
                 .groupby(['InvoiceNo', 'Description'])['Quantity']
                 .sum().unstack().reset_index().fillna(0)
                 .set_index('InvoiceNo'))
```

Analisa:

3. Codingan ini mengambil data transaksi untuk pelanggan yang berasal dari France, UK, Portugal, dan Sweden, lalu mengelompokkannya berdasarkan InvoiceNo dan description. Kemudian, kuantitas setiap barang dalam satu faktur dijumlahkan. Data ini

kemudian diubah menjadi format yang lebih lebar dengan mengubah description menjadi kolom-kolom baru. Setelah itu, indeks dari dataframe direset agar nomor faktur menjadi kolom biasa, dan nilai yang kosong diisi dengan 0 kemudian nomor faktur dijadikan indeks dataframe

```
def hot_encode(x):  
    if(x<=0):  
        return 0  
    if(x>=1):  
        return 1
```

Analisa:

4. Codingan ini menginisialisasi fungsi hot\_encode untuk mengubah value menjadi biner

```
basket_encoded = basket_France.applymap(hot_encode)  
basket_France = basket_encoded  
  
basket_encoded = basket_UK.applymap(hot_encode)  
basket_UK = basket_encoded  
  
basket_encoded = basket_Por.applymap(hot_encode)  
basket_Por = basket_encoded  
  
basket_encoded = basket_Sweden.applymap(hot_encode)  
basket_Sweden = basket_encoded
```

Analisa:

5. Codingan ini mengubah value yg telah ada menjadi value biner dengan value 0 atau 1

```
frq_items = apriori(basket_France, min_support = 0.05, use_colnames=True)

rules = association_rules(frq_items, metric="lift", min_threshold=1)
rules = rules.sort_values(['confidence', 'lift'], ascending=[False, False])
print(rules.head())
```

	antecedents \
45	(JUMBO BAG WOODLAND ANIMALS)
260	(PLASTERS IN TIN CIRCUS PARADE, RED TOADSTOOL ...
272	(PLASTERS IN TIN WOODLAND ANIMALS, RED TOADSTO...
301	(SET/20 RED RETROSPOT PAPER NAPKINS, SET/6 RED...
300	(SET/20 RED RETROSPOT PAPER NAPKINS, SET/6 RED...

  

	consequents	antecedent support	consequent support \
45	(POSTAGE)	0.076531	0.765306
260	(POSTAGE)	0.051020	0.765306
272	(POSTAGE)	0.053571	0.765306
301	(SET/6 RED SPOTTY PAPER PLATES)	0.102041	0.127551
300	(SET/6 RED SPOTTY PAPER CUPS)	0.102041	0.137755

  

	support	confidence	lift	leverage	conviction	zhangs_metric
45	0.076531	1.000	1.306667	0.017961	inf	0.254144
260	0.051020	1.000	1.306667	0.011974	inf	0.247312
272	0.053571	1.000	1.306667	0.012573	inf	0.247978
301	0.099490	0.975	7.644000	0.086474	34.897959	0.967949
300	0.099490	0.975	7.077778	0.085433	34.489796	0.956294

Analisa:

- Codingan ini mengekstrak aturan asosiasi dari order pelanggan di France yg sudah dijadikan biner untuk memberikan insight tentang kombinasi produk yang sering dibeli bersama oleh pelanggan di Prancis, serta confidence dan lift dari kombinasi tersebut

```
frq_items = apriori(basket_UK, min_support = 0.05, use_colnames=True)

rules = association_rules(frq_items, metric="lift", min_threshold=1)
rules=rules.sort_values(['confidence', 'lift'], ascending=[False, False])
#print(rules.head())
rules.head(10)
```

Analisa:

7. Codingan ini mengekstrak aturan asosiasi dari order pelanggan di United Kingdom yg sudah dijadikan biner untuk memberikan insight tentang kombinasi produk yang sering dibeli bersama oleh pelanggan di Prancis, serta confidence dan lift dari kombinasi tersebut. Pada codingan ini tidak muncul outputnya karena kurangnya korelasi belanja di United Kingdom

```
frq_items = apriori(basket_Por, min_support = 0.05, use_colnames=True)

rules = association_rules(frq_items, metric="lift", min_threshold=1)
rules=rules.sort_values(['confidence', 'lift'], ascending=[False, False])
print(rules.head())
```

/usr/local/lib/python3.10/dist-packages/ipykernel/ipkernel.py:283: DeprecationWarning: and should\_run\_async(code)

/usr/local/lib/python3.10/dist-packages/mlxtend/frequent\_patterns/fpcommon.py:110: warnings.warn(

	antecedents	consequents	\
1170	(SET 12 COLOUR PENCILS SPACEBOY)	(SET 12 COLOUR PENCILS DOLLY GIRL)	
1171	(SET 12 COLOUR PENCILS DOLLY GIRL)	(SET 12 COLOUR PENCILS SPACEBOY)	
1172	(SET OF 4 KNICK KNACK TINS LONDON)	(SET 12 COLOUR PENCILS DOLLY GIRL)	
1173	(SET 12 COLOUR PENCILS DOLLY GIRL)	(SET OF 4 KNICK KNACK TINS LONDON)	
1174	(SET OF 4 KNICK KNACK TINS POPPIES)	(SET 12 COLOUR PENCILS DOLLY GIRL)	

  

	antecedent support	consequent support	support	confidence	lift	\
1170	0.051724	0.051724	0.051724	1.0	19.333333	
1171	0.051724	0.051724	0.051724	1.0	19.333333	
1172	0.051724	0.051724	0.051724	1.0	19.333333	
1173	0.051724	0.051724	0.051724	1.0	19.333333	
1174	0.051724	0.051724	0.051724	1.0	19.333333	

  

	leverage	conviction	zhangs_metric
1170	0.049049	inf	1.0
1171	0.049049	inf	1.0
1172	0.049049	inf	1.0
1173	0.049049	inf	1.0
1174	0.049049	inf	1.0

Analisa:

- Codingan ini mengekstrak aturan asosiasi dari order pelanggan di Portugal yg sudah dijadikan biner untuk memberikan insight tentang kombinasi produk yang sering dibeli bersama oleh pelanggan di Prancis, serta confidence dan lift dari kombinasi tersebut

```

frq_items = apriori(basket_sweden, min_support = 0.05, use_colnames=True)

rules = association_rules(frq_items, metric="lift", min_threshold=1)
rules=rules.sort_values(['confidence', 'lift'], ascending=[False,False])
print(rules.head())

```

/usr/local/lib/python3.10/dist-packages/ipykernel/ipkernel.py:283: DeprecationWarning: and should\_run\_async(code)

/usr/local/lib/python3.10/dist-packages/mlxtend/frequent\_patterns/fpcommon.py:11: warnings.warn(

	antecedents	consequents
0	(12 PENCILS SMALL TUBE SKULL)	(PACK OF 72 SKULL CAKE CASES)
1	(PACK OF 72 SKULL CAKE CASES)	(12 PENCILS SMALL TUBE SKULL)
4	(ASSORTED BOTTLE TOP MAGNETS)	(36 DOILIES DOLLY GIRL)
5	(36 DOILIES DOLLY GIRL)	(ASSORTED BOTTLE TOP MAGNETS)
180	(CHILDRENS CUTLERY CIRCUS PARADE)	(CHILDRENS CUTLERY DOLLY GIRL)

  

	antecedent	support	consequent	support	support	confidence	lift
0		0.055556		0.055556	0.055556	1.0	18.0
1		0.055556		0.055556	0.055556	1.0	18.0
4		0.055556		0.055556	0.055556	1.0	18.0
5		0.055556		0.055556	0.055556	1.0	18.0
180		0.055556		0.055556	0.055556	1.0	18.0

  

	leverage	conviction	zhangs_metric
0	0.052469	inf	1.0
1	0.052469	inf	1.0
4	0.052469	inf	1.0
5	0.052469	inf	1.0
180	0.052469	inf	1.0

Analisa:

- Codingan ini mengekstrak aturan asosiasi dari order pelanggan di Sweden yg sudah dijadikan biner untuk memberikan insight tentang kombinasi produk yang sering dibeli bersama oleh pelanggan di Prancis, serta confidence dan lift dari kombinasi tersebut