



PROGRAM STUDI
TEKNIK INFORMATIKA – S1
FAKULTAS ILMU KOMPUTER
UNIVERSITAS DIAN NUSWANTORO

MATA KULIAH
DATA MINING



<https://www.freepik.com/vectors/technology> Technology vector created by sentavio - www.freepik.com

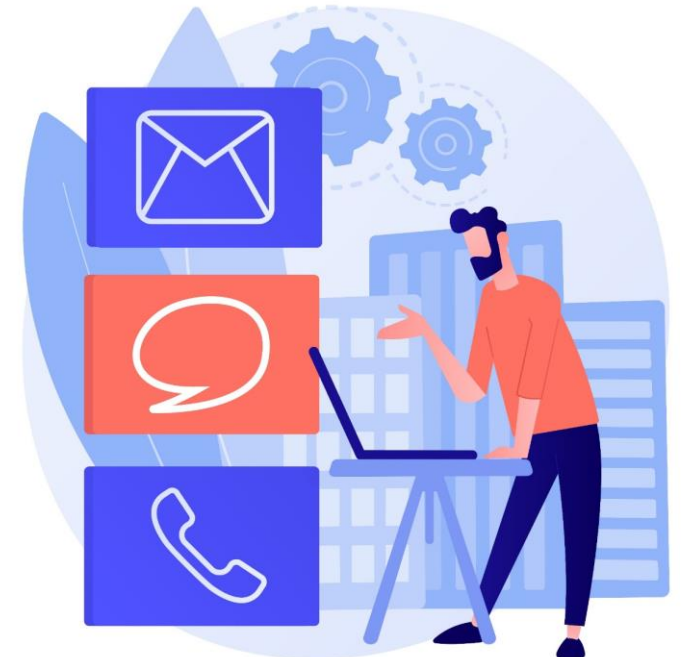
DATA MINING

“Asosiasi”

TIM PENGAMPU DOSEN DATA MINING
2023

Kontak Dosen

- Junta Zeniarja, M.Kom
- Email: junta@dsn.dinus.ac.id
- Youtube : <https://www.youtube.com/JuntaZeniarja>
- Scholar : <http://bit.do/JuntaScholar>



Asosiasi

- Dikenal dengan nama ***Market Basket Analysis*** atau ***Association Rule***.
- *Association Rule* merupakan hubungan (asosiasi) antara kombinasi beberapa item (barang, orang, produk, atau apapun yang diwakili oleh kata benda) yang sering muncul bersamaan.
- Banyak digunakan di toko-toko untuk mengatur penempatan barang atau mengontrol penjualan.
- Apriori adalah algoritma yang banyak digunakan.

Aturan (Rule)

- Apabila suatu hubungan sedemikian seringnya muncul di dalam data, maka hubungan ini kemudian dianggap sebagai suatu keharusan, atau “**aturan**” (**rule**) yang wajib dipenuhi.
- Di dalam notasi matematis, sebuah rule memiliki **antecedent** (hal yang mengawali) yang ditulis di kiri tanda panah, dan sebuah **consequent** (efek atau akibat) di sebelah kanan anak panah.
 - $\{ \text{antecedent} \} \Rightarrow \{ \text{consequent} \}$
- Nilai – nilai yang ada di dalam tanda kurung kurawal {} disebut **itemset**, bisa terdiri dari satu item atau lebih.
- Contoh: $\{ \text{roti} \} \Rightarrow \{ \text{mentega} \}$, rule tersebut menunjukkan bahwa bilamana roti dibeli, maka mentega juga (kemungkinan besar) akan dibeli (kemunculan roti mengakibatkan kemunculan mentega).

Support

- Support sebuah item (atau kumpulan item) adalah prosentase transaksi dimana item tersebut muncul.
- Support mengukur seberapa sering item tersebut muncul pada data transaksi.

Transaksi	Item
1	Soto, nasi putih, es teh
2	Perkedel, soto, kopi susu, bakso
3	Nasi putih, bakso, soto
4	Perkedel, kopi susu, es teh
5	Soto, nasi putih, es teh



Set	Support
Soto	80%
Nasi putih	60%
Es teh	60%
Perkedel	40%
Kopi susu	40%
Bakso	40%
	40%
	60%
	40%
	40%

$$\text{Support}(A \rightarrow B) = \frac{\text{Jumlah transaksi mengandung } (A \rightarrow B)}{\text{Total Transaksi}}$$

Confident

- Confident adalah rasio jumlah transaksi yang mengandung $X \cup Y$ terhadap jumlah transaksi yang mengandung X .
- Confidence mengukur seberapa kuat hubungan antar item pada setiap aturan asosiasi.

Transaksi	Item
1	Soto, nasi putih, es teh
2	Perkedel, soto, kopi susu, bakso
3	Nasi putih, bakso, soto
4	Perkedel, kopi susu, es teh
5	Soto, nasi putih, es teh



	Support (A)	
	80%	$60\% / 80\% = 75\%$
	80%	$40\% / 80\% = 50\%$
	40%	$40\% / 40\% = 100\%$
	60%	$20\% / 60\% = 33,3\%$
	80%	$40\% / 80\% = 50\%$
	40%	$20\% / 40\% = 50\%$

$$\text{Confidence}(A \rightarrow B) = \frac{\text{Support}(A \rightarrow B)}{\text{Support}(A)}$$

Pembentukan Aturan Asosiasi

Pembentukan Frequent Itemset:

1. Membuat daftar k -itemsets dan hitung nilai supportnya.
2. Gunakan k -itemsets yang memenuhi minimal support untuk menyusun $k + 1$ -itemsets.

Pembentukan Aturan Asosiasi:

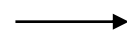
1. Buat aturan asosiasi untuk setiap k -itemsets dengan minimum support.
2. Hitung nilai confident untuk setiap aturan asosiasi.
3. Ekstrak aturan asosiasi yang memenuhi minimum confident.

Contoh

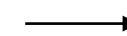
ID	Item Transaksi			
1	telur	beras		
2	telur	beras	minyak	topi
3	telur	beras	minyak	buku
4	buku	topi		
5	telur	beras	minyak	topi
6	telur	beras	minyak	
7	telur	minyak	buku	
8	beras	buku		

minsupp	5
min-conf	90%

L1	Support
telur	6
buku	4
beras	6
minyak	5
topi	3



L2	Support
telur-beras	5
telur-minyak	5
beras-minyak	4



L3	Support
telur-beras-minyak	4

Aturan Asosiasi	Confident
telur-beras	83.33%
beras-telur	83.33%
telur-minyak	83.33%
minyak-telur	100%



Implementasi Python (Apriori)

Pertemuan 13

Apriori di Python [Import Library dan Load Data]

```
# Langkah 1: Mengimpor library yang diperlukan
import numpy as np
import pandas as pd
from mlxtend.frequent_patterns import apriori, association_rules
```

```
# Langkah 2: Meload dan Mengeksplor data
# Memuat Data
data = pd.read_excel("Online_Retail.xlsx")
data.head()
```

Apriori di Python [Dataset Online Retail]

	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
0	536365	85123A	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	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom
2	536365	84406B	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	6	2010-12-01	3.39	17850.0	United Kingdom
4	536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	6	2010-12-01	3.39	17850.0	United Kingdom

	A	B	C	D	E	F	G	H
1	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
2	536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	6	01/12/2010 08.26	2.55	17850	United Kingdom
3	536365	71053	WHITE METAL LANTERN	6	01/12/2010 08.26	3.39	17850	United Kingdom
4	536365	84406B	CREAM CUPID HEARTS COAT HANGER	8	01/12/2010 08.26	2.75	17850	United Kingdom
5	536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	6	01/12/2010 08.26	3.39	17850	United Kingdom
6	536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	6	01/12/2010 08.26	3.39	17850	United Kingdom
7	536365	22752	SET 7 BABUSHKA NESTING BOXES	2	01/12/2010 08.26	7.65	17850	United Kingdom
8	536365	21730	GLASS STAR FROSTED T-LIGHT HOLDER	6	01/12/2010 08.26	4.25	17850	United Kingdom
9	536366	22633	HAND WARMER UNION JACK	6	01/12/2010 08.28	1.85	17850	United Kingdom
10	536366	22632	HAND WARMER RED POLKA DOT	6	01/12/2010 08.28	1.85	17850	United Kingdom
11	536367	84879	ASSORTED COLOUR BIRD ORNAMENT	32	01/12/2010 08.34	1.69	13047	United Kingdom
12	536367	22745	POPPY'S PLAYHOUSE BEDROOM	6	01/12/2010 08.34	2.1	13047	United Kingdom
13	536367	22748	POPPY'S PLAYHOUSE KITCHEN	6	01/12/2010 08.34	2.1	13047	United Kingdom
14	536367	22749	FELTCRAFT PRINCESS CHARLOTTE DOLL	8	01/12/2010 08.34	3.75	13047	United Kingdom
15	536367	22310	IVORY KNITTED MUG COSY	6	01/12/2010 08.34	1.65	13047	United Kingdom
16	536367	84969	BOX OF 6 ASSORTED COLOUR TEASPOONS	6	01/12/2010 08.34	4.25	13047	United Kingdom
17	536367	22623	BOX OF VINTAGE JIGSAW BLOCKS	3	01/12/2010 08.34	4.95	13047	United Kingdom
18	536367	22622	BOX OF VINTAGE ALPHABET BLOCKS	2	01/12/2010 08.34	9.95	13047	United Kingdom
19	536367	21754	HOME BUILDING BLOCK WORD	3	01/12/2010 08.34	5.95	13047	United Kingdom
20	536367	21755	LOVE BUILDING BLOCK WORD	3	01/12/2010 08.34	5.95	13047	United Kingdom

Apriori di Python [Explore Data]

```
# Mengeksplor kolom data  
data.columns
```

```
Index(['InvoiceNo', 'StockCode', 'Description', 'Quantity', 'InvoiceDate',  
      'UnitPrice', 'CustomerID', 'Country'],  
      dtype='object')
```

```
# Mengeksplor berbagai wilayah transaksi  
data.Country.unique()
```

```
array(['United Kingdom', 'France', 'Australia', 'Netherlands', 'Germany',  
      'Norway', 'EIRE', 'Switzerland', 'Spain', 'Poland', 'Portugal',  
      'Italy', 'Belgium', 'Lithuania', 'Japan', 'Iceland',  
      'Channel Islands', 'Denmark', 'Cyprus', 'Sweden', 'Austria',  
      'Israel', 'Finland', 'Bahrain', 'Greece', 'Hong Kong', 'Singapore',  
      'Lebanon', 'United Arab Emirates', 'Saudi Arabia',  
      'Czech Republic', 'Canada', 'Unspecified', 'Brazil', 'USA',  
      'European Community', 'Malta', 'RSA'], dtype=object)
```

Apriori di Python [Data Cleaning]

```
# Langkah 3: Membersihkan Data
# Menghapus ruang ekstra dalam deskripsi
data['Description'] = data['Description'].str.strip()

# Menghapus baris tanpa nomor faktur (InvoiceNo)
data.dropna(axis = 0, subset = ['InvoiceNo'], inplace = True)
data['InvoiceNo'] = data['InvoiceNo'].astype('str')

# Menghapus semua transaksi yang dilakukan secara kredit
data = data[~data['InvoiceNo'].str.contains('C')]
```

Apriori di Python [Data Split]

```
# Langkah 4: Memisahkan data sesuai dengan wilayah transaksi
# Transaksi dilakukan di Prancis
basket_France = (data[data['Country'] == "France"]
                 .groupby(['InvoiceNo', 'Description'])['Quantity']
                 .sum().unstack().reset_index().fillna(0)
                 .set_index('InvoiceNo'))

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

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

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


Apriori di Python [Pengkodean Data dengan Hot Encoding]

```
# Langkah 5: Pengkodean Data dengan "Hot Encoding"
# Mendefinisikan fungsi "Hot Encoding" untuk membuat data sesuai untuk library yang bersangkutan
def hot_encode(x):
    if(x<= 0):
        return 0
    if(x>= 1):
        return 1
```

```
# Mengkodekan dataset
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
```


Apriori di Python [Membangun Model untuk France]

```
# Langkah 6: Membangun model dan menganalisis hasilnya
# 1) France
# Membangun model
frq_items = apriori(basket_France, min_support = 0.05, use_colnames = True)

# Mengumpulkan aturan yang disimpulkan dalam dataframe
rules = association_rules(frq_items, metric="lift", min_threshold = 1)
rules = rules.sort_values(['confidence', 'lift'], ascending =[False, False])
print(rules.head())
```

	antecedents	... conviction
44	(JUMBO BAG WOODLAND ANIMALS)	... inf
258	(RED TOADSTOOL LED NIGHT LIGHT, PLASTERS IN TI...	... inf
271	(RED TOADSTOOL LED NIGHT LIGHT, PLASTERS IN TI...	... inf
300	(SET/20 RED RETROSPOT PAPER NAPKINS, SET/6 RED...	... 34.897959
301	(SET/20 RED RETROSPOT PAPER NAPKINS, SET/6 RED...	... 34.489796

[5 rows x 9 columns]

Apriori di Python [Membangun Model untuk United Kingdom]

```
# 2) United Kingdom
frq_items = apriori(basket_UK, min_support = 0.01, 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	...	conviction
117	(BEADED CRYSTAL HEART PINK ON STICK)	...	39.637371
2020	(SUKI SHOULDER BAG, JAM MAKING SET PRINTED)	...	26.096206
2296	(HERB MARKER THYME, HERB MARKER MINT)	...	21.947227
2301	(HERB MARKER ROSEMARY, HERB MARKER PARSLEY)	...	20.444951
2302	(HERB MARKER THYME, HERB MARKER PARSLEY)	...	20.443842

```
[5 rows x 9 columns]
```

Apriori di Python [Membangun Model untuk Portugal]

```
# Portugal
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())
```

	antecedents	...	conviction
1170	(SET 12 COLOUR PENCILS SPACEBOY)	...	inf
1171	(SET 12 COLOUR PENCILS DOLLY GIRL)	...	inf
1172	(SET OF 4 KNICK KNACK TINS LONDON)	...	inf
1173	(SET 12 COLOUR PENCILS DOLLY GIRL)	...	inf
1174	(SET 12 COLOUR PENCILS DOLLY GIRL)	...	inf

```
[5 rows x 9 columns]
```

Apriori di Python [Membangun Model untuk Sweden]

```
# Sweden
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())
```

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

```
[5 rows x 9 columns]
```

Latihan Soal (Kuis)

- Tentukan pola asosiasi yang terbentuk dari transaksi dibawah ini:

T1	Roti, selai, mentega
T2	Roti, mentega
T3	Roti, susu, mentega
T4	Cokelat, roti, susu, mentega
T5	Cokelat, susu

minsupp	30%
min-conf	60%

Referensi

1. Kusrini, Taufiq Emha, Algoritma Data Mining, *Penerbit Andi*, 2009.
2. Ian H. Witten, Frank Eibe, Mark A. Hall, Data mining: Practical Machine Learning Tools and Techniques 4th Edition, *Elsevier*, 2017.
3. Dios Kurniawan, Pengenalan Machine Learning dengan Python, Elex Media Komputindo, 2020.
4. <https://www.geeksforgeeks.org/implementing-apriori-algorithm-in-python/>
5. Sumber gambar: www.freepik.com.



THANKS

ANY QUESTIONS?

