Association Rule

Knowledge Discovery

Hanif Izzudin Rahman

Tugas 1

Assignment

1. **dataset** ← transaction.csv, and show it
2. **data** ← get data from **dataset** for country="Portugal"
3. **transaksi** ← get StockCode from **data** for each transaction (1 code on InvoiceNo = 1 transaction), and show it
4. Find association rule on **transaksi** with minimum support=0.2 and minimum confidence=0.7, and show it

➔Menampilkan Dataset “Transaction”

```
1  import pandas as pd
2  import csv
3
4  contacts = []
5
6  with open('D:/KULIAH/MASTER/S2 PENS/KULIAH/SEMESTER 2/Sistem Temu PengetahuanP Ali Ridho /P15_Asosiatif Rule/transaction.csv
7      csv_reader = csv.reader(csv_file, delimiter=",")
8      for row in csv_reader:
9          contacts.append(row)
10
11 labels = contacts.pop(0)
12
13 print(f'{labels[0]}\t{labels[1]}\t\t{labels[2]}\t{labels[3]}\t\t{labels[4]}\t{labels[5]}')
14 print("-"*34)
15
16 for data in contacts:
17     print(f'{data[0]}\t{data[1]}\t{data[2]}\t{data[3]}\t{data[4]}\t{data[5]}')
```

InvoiceNo	StockCode	Qty	InvoiceDate	CustomerID	Country
537626	22725	830	12/7/2010 14:57	12347	Iceland
537626	22729	948	12/7/2010 14:57	12347	Iceland
537626	22195	695	12/7/2010 14:57	12347	Iceland
542237	22725	636	1/26/2011 14:30	12347	Iceland
542237	22729	536	1/26/2011 14:30	12347	Iceland
542237	47559	919	1/26/2011 14:30	12347	Iceland
542237	21154	803	1/26/2011 14:30	12347	Iceland
542237	21035	532	1/26/2011 14:30	12347	Iceland
549222	23076	383	4/7/2011 10:43	12347	Iceland
549222	21791	389	4/7/2011 10:43	12347	Iceland
549222	22550	500	4/7/2011 10:43	12347	Iceland
549222	22432	875	4/7/2011 10:43	12347	Iceland
549222	22195	434	4/7/2011 10:43	12347	Iceland
549222	21975	735	4/7/2011 10:43	12347	Iceland
556201	23171	135	6/9/2011 13:01	12347	Iceland
556201	23172	974	6/9/2011 13:01	12347	Iceland
556201	23175	82	6/9/2011 13:01	12347	Iceland
573511	47559	922	10/31/2011 12:25	12347	Iceland
573511	21791	296	10/31/2011 12:25	12347	Iceland
573511	22992	412	10/31/2011 12:25	12347	Iceland
573511	22561	715	10/31/2011 12:25	12347	Iceland
573511	22621	566	10/31/2011 12:25	12347	Iceland
573511	22725	850	10/31/2011 12:25	12347	Iceland
573511	23308	492	10/31/2011 12:25	12347	Iceland
573511	22195	377	10/31/2011 12:25	12347	Iceland
539318	21981	957	12/16/2010 19:09	12348	Finland
539318	84988	802	12/16/2010 19:09	12348	Finland
539318	22952	732	12/16/2010 19:09	12348	Finland
539318	22952	112	12/16/2010 19:09	12348	Finland
541998	21980	737	1/25/2011 10:42	12348	Finland
548955	23077	779	4/5/2011 10:47	12348	Finland
548955	23076	922	4/5/2011 10:47	12348	Finland
548955	22437	513	4/5/2011 10:47	12348	Finland

➔Mendapatkan Data Dari Country = Portugal

```python
import pandas as pd
import csv

contacts = []

with open('D:/KULIAH/MASTER/S2 PENS/KULIAH/SEMESTER 2/Sistem Temu PengetahuanP Ali Ridho /P15_Asosiatif Rule/transaction.csv
    csv_reader = csv.reader(csv_file, delimiter=",")
    for row in csv_reader:
        contacts.append(row)

labels = contacts.pop(0)

print(f'{labels[0]}\t{labels[1]}\t\t{labels[2]}\t{labels[3]}\t\t{labels[4]}\t{labels[5]}')
print("-"*34)

for data in contacts:
    data5=f'{data[5]}'
    #print(f'{data[0]}\t{data[1]}\t{data[2]}\t{data[3]}\t{data[4]}\t{data[5]}')
    if (data5 == "Portugal"):
        print(f'{data[0]}\t{data[1]}\t{data[2]}\t{data[3]}\t{data[4]}\t{data[5]}')
```

InvoiceNo	StockCode	Qty	InvoiceDate	CustomerID	Country
541430	22195	649	1/18/2011 9:50	12356	Portugal
541430	22435	460	1/18/2011 9:50	12356	Portugal
541430	84378	304	1/18/2011 9:50	12356	Portugal
541430	22646	896	1/18/2011 9:50	12356	Portugal
541430	84987	157	1/18/2011 9:50	12356	Portugal
541430	84380	208	1/18/2011 9:50	12356	Portugal
549435	21124	444	4/8/2011 12:33	12356	Portugal
549435	22957	615	4/8/2011 12:33	12356	Portugal
549435	84378	744	4/8/2011 12:33	12356	Portugal
567929	20723	707	9/22/2011 17:31	12425	Portugal
567929	21154	546	9/22/2011 17:31	12425	Portugal
567929	22215	119	9/22/2011 17:31	12425	Portugal
567929	22329	946	9/22/2011 17:31	12425	Portugal
567929	22628	160	9/22/2011 17:31	12425	Portugal
567929	23077	476	9/22/2011 17:31	12425	Portugal
567929	23076	721	9/22/2011 17:31	12425	Portugal
553017	21980	435	5/12/2011 19:01	12757	Portugal
574099	22487	919	11/3/2011 10:05	12808	Portugal
574099	22619	586	11/3/2011 10:05	12808	Portugal
556829	22667	424	6/15/2011 10:51	12809	Portugal
556829	23306	252	6/15/2011 10:51	12809	Portugal
556829	23308	445	6/15/2011 10:51	12809	Portugal
556829	21218	544	6/15/2011 10:51	12809	Portugal
545191	22859	513	2/28/2011 15:39	12811	Portugal
545191	22957	178	2/28/2011 15:39	12811	Portugal
545191	21929	102	2/28/2011 15:39	12811	Portugal
545191	21928	441	2/28/2011 15:39	12811	Portugal
545191	22411	198	2/28/2011 15:39	12811	Portugal
547444	84279	363	3/23/2011 10:55	12811	Portugal
547444	84279	785	3/23/2011 10:55	12811	Portugal
547444	21164	543	3/23/2011 10:55	12811	Portugal
547444	22745	292	3/23/2011 10:55	12811	Portugal
547444	22747	133	3/23/2011 10:55	12811	Portugal
547444	22746	835	3/23/2011 10:55	12811	Portugal
547444	23176	878	3/23/2011 10:55	12811	Portugal
547444	22968	637	3/23/2011 10:55	12811	Portugal

➔Mencari Association Rule Dari Transaksi

```
1  import pandas as pd
2  from mlxtend.frequent_patterns import apriori
3  from mlxtend.frequent_patterns import association_rules
4
5  dataset = pd.read_csv ('D:/KULIAH/MASTER/S2 PENS/KULIAH/SEMESTER Sistem Temu Pengetahuan P Ali Ridho /P15_Asosiatif Rule/tra
6  transaksi = dataset.groupby(['InvoiceNo StockCode '])['Qty'].
7
8  transaksi = transaksi.unstack().reset_index().fillna(0).set_index('InvoiceNo')
9  transaksi[transaksi >0]=1
10
11 print('Tabel Transaksi \n', transaksi)
12
13 frequent_itemsets = apriori(transaksi,min_support=0.2,use_colnames=True)
14 rules=association_rules(frequent_itemsets,metric="confidence",min_threshold = 0.7)
15
16 print('\nAssociation Rules:\n', rules[['antecedents', 'consequents', 'confidence']])
```

Tabel Transaksi:

StockCode	16161	20713	20718	20723	20977	20981	21035	21124	21154	\
InvoiceNo										
537246	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
537818	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
537915	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
538311	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
539353	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
540519	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	
540546	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	
541430	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
542147	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
544495	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	
545191	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
545937	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
547444	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
547897	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	

StockCode	21164	...	47599	48184	82484	84077	84279	84378	84380	84509	\
InvoiceNo		...									
537246	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
537818	0.0	...	0.0	0.0	0.0	0.0	0.0	1.0	1.0	0.0	
537915	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
538311	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
539353	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
540519	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
540546	0.0	...	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	
541430	0.0	...	0.0	0.0	0.0	0.0	0.0	1.0	1.0	0.0	
542147	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	
544495	0.0	...	0.0	1.0	0.0	0.0	0.0	0.0	0.0	1.0	
545191	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
545937	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
547444	1.0	...	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	
547897	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
548470	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	

Association Rules:

	antecedents	consequents	confidence
0	(21928)	(21929)	0.900000
1	(21929)	(21928)	1.000000
2	(21928)	(22411)	0.900000
3	(22411)	(21928)	0.818182
4	(21929)	(22411)	1.000000
5	(22411)	(21929)	0.818182
6	(21928, 21929)	(22411)	1.000000
7	(21928, 22411)	(21929)	1.000000
8	(21929, 22411)	(21928)	1.000000
9	(21928)	(21929, 22411)	0.900000
10	(21929)	(21928, 22411)	1.000000
11	(22411)	(21928, 21929)	0.818182

TUGAS KNOWLEDGE
DISCOVERY
Latihan 2

T1	{roti, selai, mentega}
T2	{roti, mentega}
T3	{roti, susu, mentega}
T4	{cokelat, roti, susu, mentega}
T5	{cokelat, susu}

- Suatu supermarket mempunyai jumlah transaksi seperti dalam tabel
- Carilah association rule dari data tersebut dengan cara menghitung support dan confidence
- Pakailah metode apriori dengan minimum support = 0,3 dan confidence = 0,8

Jawab:

		Itemset	SP
T1	{roti, selai, mentega}	{roti}	0,8
T2	{roti, mentega}	{selai}	0,2
T3	{roti, susu, mentega}	{mentega}	0,8
T4	{cokelat, roti, susu, mentega}	{susu}	0,6
T5	{cokelat, susu}	{cokelat}	0,4

Itemset	SP
{roti, mentega, susu}	0,4

Itemset	SP
{roti, mentega}	0,8
{roti, susu}	0,4
{roti, cokelat}	0,2
{mentega, susu}	0,4
{mentega, cokelat}	0,2
{susu, cokelat}	0,4

Maka, mencari asosiasi pada semua himpunan bagian:

{R} → {M,S}	Keyakinannya adalah	0,4 / 0,8 = 50%
{M} → {R,S}	Keyakinannya adalah	0,4 / 0,8 = 50%
{S} → {R,M}	—"—	0,4 / 0,6 = 67%
{M,S} → {R}	—"—	0,4 / 0,4 = 100%
{R,S} → {M}	—"—	0,4 / 0,4 = 100%
{R,M} → {S}	—"—	0,4 / 0,8 = 100%

Association Rulenya adalah {M,S} → {R}, {R,S} → {M} & {R,M} → {S}

JOYKO 30 Lines, 6 mm

Tugas 2

ANALISA

- Berdasarkan hasil penerapan Association Rule dengan algoritma Aprioripada dataset ‘transaction’ menggunakan minimum support 0.2 dan minimum confidence 0.7 diperoleh12 rules hubungan antar indicator. Banyaknya jumlah rules yang dihasilkan mengakibatkan banyakanya pengetahuan mengenai pola hubungan antar indicator. Yang juga pengguna dapat lebih banyak memilih pemilihan rules keterkaitannya.
- Semakin besar nilai minimum support dan minimum confidence, maka semakin banyak pula association rules yang didapat dan sebaliknya