

## Tugas 4

**Alvaro Austin - 2106752180**

1. Dengan minimum support = 0.6, jalankan algoritma secara naif untuk mendapatkan frequent itemset dengan memanfaatkan data transaksi diatas! Jelaskan setiap langkah yang perlu dilakukan.

- Langkah 1: Mengidentifikasi semua item yang terdapat pada semua transaksi

Jawab:

Dalam kasus ini, item uniknya adalah {Laptop, Mouse, Keyboard, Headphones, Webcam}.

- Langkah 2: Membuat kombinasi dari itemset:

Berikut adalah kombinasi yang kita dapat peroleh dari itemset diatas:

- {Laptop}
- {Mouse}
- {Keyboard}
- {Headphones}
- {Webcam}
- {Laptop, Mouse}
- {Laptop, Keyboard}
- {Laptop, Headphones}
- {Laptop, Webcam}
- {Mouse, Keyboard}
- {Mouse, Headphones}
- {Mouse, Webcam}
- {Keyboard, Headphones}
- {Keyboard, Webcam}
- {Laptop, Mouse, Keyboard}
- {Laptop, Mouse, Headphones}
- {Laptop, Mouse, Webcam}
- {Laptop, Keyboard, Headphones}
- {Laptop, Keyboard, Webcam}
- {Mouse, Keyboard, Headphones}
- {Mouse, Keyboard, Webcam}
- {Laptop, Mouse, Keyboard, Headphones}
- {Laptop, Mouse, Keyboard, Webcam}

- Langkah 3: Menghitung frekuensi dari itemset

- {Laptop}: 4
- {Mouse}: 4
- {Keyboard}: 5

- {Headphones}: 3
  - {Webcam}: 1
  - {Laptop, Mouse}: 3
  - {Laptop, Keyboard}: 4
  - {Laptop, Headphones}: 2
  - {Laptop, Webcam}: 1
  - {Mouse, Keyboard}: 4
  - {Mouse, Headphones}: 2
  - {Mouse, Webcam}: 1
  - {Keyboard, Headphones}: 3
  - {Keyboard, Webcam}: 1
  - {Laptop, Mouse, Keyboard}: 3
  - {Laptop, Mouse, Headphones}: 1
  - {Laptop, Mouse, Webcam}: 1
  - {Laptop, Keyboard, Headphones}: 2
  - {Laptop, Keyboard, Webcam}: 1
  - {Mouse, Keyboard, Headphones}: 2
  - {Mouse, Keyboard, Webcam}: 1
  - {Laptop, Mouse, Keyboard, Headphones}: 1
  - {Laptop, Mouse, Keyboard, Webcam}: 1
- Langkah 3: Menentukan support dari setiap itemset diatas:
- {Laptop}: 0.8
  - {Mouse}: 0.8
  - {Keyboard}: 1
  - {Headphones}: 0.6
  - {Webcam}: 0.2
  - {Laptop, Mouse}: 0.6
  - {Laptop, Keyboard}: 0.8
  - {Laptop, Headphones}: 0.4
  - {Laptop, Webcam}: 0.2
  - {Mouse, Keyboard}: 0.8
  - {Mouse, Headphones}: 0.4
  - {Mouse, Webcam}: 0.2
  - {Keyboard, Headphones}: 0.6
  - {Keyboard, Webcam}: 0.2
  - {Laptop, Mouse, Keyboard}: 0.6
  - {Laptop, Mouse, Headphones}: 0.2
  - {Laptop, Mouse, Webcam}: 0.2
  - {Laptop, Keyboard, Headphones}: 0.4
  - {Laptop, Keyboard, Webcam}: 0.2

- {Mouse, Keyboard, Headphones}: 0.4
    - {Mouse, Keyboard, Webcam}: 0.2
    - {Laptop, Mouse, Keyboard, Headphones}: 0.2
    - {Laptop, Mouse, Keyboard, Webcam}: 0.2
  - Langkah 4: Menentukan Frequent Itemset dengan minimum support 0.6
    - Untuk menemukan frequent Itemset dengan minimum support 0.6 artinya kita perlu menemukan itemset dengan support **setidaknya sebesar** 0.6.
    - Oleh karena itu kita dapat memperoleh frequent itemset yang memenuhi minimum support adalah:
      - {Laptop}
      - {Mouse}
      - {Keyboard}
      - {Headphones}
      - {Laptop, Mouse}
      - {Laptop, Keyboard}
      - {Mouse, Keyboard}
      - {Keyboard, Headphones}
      - {Laptop, Mouse, Keyboard}
2. Dengan minimum support = 0.6, jalankan algoritma Apriori dengan menuliskan  $F < K >$  sebagai frequent itemsets pada suatu  $K$  dan  $C < K+1 >$  sebagai kandidat untuk frequent itemsets selanjutnya dengan  $K$  adalah jumlah item dalam himpunan berdasarkan data transaksi diatas!
- $F_1 = \{\{Laptop\}, \{Mouse\}, \{Keyboard\}, \{Headphones\}\}$   
 $C_2 = \{\{Laptop, Mouse\}, \{Laptop, Keyboard\}, \{Laptop, Headphones\}, \{Mouse, Keyboard\}, \{Mouse, Headphones\}, \{Keyboard, Headphones\}\}$
  - **Hitung support di  $C_2$**   
 $C_2 = \{\{Laptop, Mouse\}, \{Laptop, Keyboard\}, \text{~~\{Laptop, Headphones\}}~~, \{Mouse, Keyboard\}, \text{~~\{Mouse, Headphones\}}~~, \{Keyboard, Headphones\}\}$   
 Untuk {Laptop, Headphones} dan {Mouse, Headphones} tidak dimasukkan karena hanya memiliki support sebesar 0,4

- $F_2 = \{\{\text{Laptop, Mouse}\}, \{\text{Laptop, Keyboard}\}, \{\text{Mouse, Keyboard}\}, \{\text{Keyboard, Headphones}\}\}$   
 $C_3 = \{\{\text{Laptop, Mouse, Keyboard}\}, \{\text{Laptop, Mouse, Headphones}\}, \{\text{Laptop, Keyboard, Headphones}\}, \{\text{Mouse, Keyboard, Headphones}\}\}$

- **Pruning dengan downward property  $C_3$**

$C_3 = \{\{\text{Laptop, Mouse, Keyboard}\}, \{\text{Laptop, Mouse, Headphones}\}, \{\text{Laptop, Keyboard, Headphones}\}, \{\text{Mouse, Keyboard, Headphones}\}\}$

Alasan:

- $\{\text{Laptop, Mouse, Headphones}\}$  tidak masuk karena subset  $\{\text{Laptop, Headphones}\}$  tidak ada di  $F_2$
- $\{\text{Laptop, Keyboard, Headphones}\}$  tidak masuk karena subset  $\{\text{Laptop, Headphones}\}$  tidak ada di  $F_2$
- $\{\text{Mouse, Keyboard, Headphones}\}$  tidak masuk karena subset  $\{\text{Mouse, Headphones}\}$  tidak ada di  $F_2$
- $F_3 = \{\{\text{Laptop, Mouse, Keyboard}\}\}$
- **Tidak ada  $C_4$**  karena tidak mungkin menghasilkan  $F_4$
- Menggabungkan semua hasil jadi  $F$ :  
 $F = \{\{\text{Laptop}\}, \{\text{Mouse}\}, \{\text{Keyboard}\}, \{\text{Headphones}\}, \{\text{Laptop, Mouse}\}, \{\text{Laptop, Keyboard}\}, \{\text{Mouse, Keyboard}\}, \{\text{Keyboard, Headphones}\}, \{\text{Laptop, Mouse, Keyboard}\}\}$

3. Dengan minimum confidence sebesar 0.5, temukan semua aturan asosiasi dari frequent sets yang telah ditemukan sebelumnya!

Saya akan menggunakan rules yang tertera pada slide, dimana:

$X \Rightarrow Y$ , apabila memenuhi:

- $\text{sup}(X \cup Y) \geq \text{minSup} (0.6)$
- $\text{conf}(X \Rightarrow Y) \geq \text{minConf} (0.5)$

Oleh karena itu, saya hanya akan mengevaluasi nilai  $X$  dan  $Y$  dari  $F$  saja.

Evaluasi:

- $X = \{\text{Laptop}\}, Y = \{\text{Mouse}\}, X \cup Y = \{\text{Laptop, Mouse}\}$   
 $\text{conf}(\{\text{Laptop}\} \Rightarrow \{\text{Mouse}\}) = \text{sup}(\{\text{Laptop, Mouse}\}) / \text{sup}(\{\text{Laptop}\}) = 0.6 / 0.8 = 0.75$
- $X = \{\text{Mouse}\}, Y = \{\text{Laptop}\}, X \cup Y = \{\text{Laptop, Mouse}\}$   
 $\text{conf}(\{\text{Mouse}\} \Rightarrow \{\text{Laptop}\}) = \text{sup}(\{\text{Laptop, Mouse}\}) / \text{sup}(\{\text{Mouse}\}) = 0.6 / 0.8 = 0.75$
- $X = \{\text{Laptop}\}, Y = \{\text{Keyboard}\}, X \cup Y = \{\text{Laptop, Keyboard}\}$

- $\text{conf}(\{\text{Laptop}\} \Rightarrow \{\text{Keyboard}\}) = \frac{\text{sup}(\{\text{Laptop}, \text{Keyboard}\})}{\text{sup}(\{\text{Laptop}\})} = 0.8/0.8 = 1$**
- $X = \{\text{Keyboard}\}, Y = \{\text{Laptop}\}, X \cup Y = \{\text{Laptop}, \text{Keyboard}\}$   
 **$\text{conf}(\{\text{Keyboard}\} \Rightarrow \{\text{Laptop}\}) = \frac{\text{sup}(\{\text{Laptop}, \text{Keyboard}\})}{\text{sup}(\{\text{Keyboard}\})} = 0.8/1 = 0.8$**
  - $X = \{\text{Mouse}\}, Y = \{\text{Keyboard}\}, X \cup Y = \{\text{Mouse}, \text{Keyboard}\}$   
 **$\text{conf}(\{\text{Mouse}\} \Rightarrow \{\text{Keyboard}\}) = \frac{\text{sup}(\{\text{Mouse}, \text{Keyboard}\})}{\text{sup}(\{\text{Mouse}\})} = 0.8/0.8 = 1$**
  - $X = \{\text{Keyboard}\}, Y = \{\text{Mouse}\}, X \cup Y = \{\text{Mouse}, \text{Keyboard}\}$   
 **$\text{conf}(\{\text{Keyboard}\} \Rightarrow \{\text{Mouse}\}) = \frac{\text{sup}(\{\text{Mouse}, \text{Keyboard}\})}{\text{sup}(\{\text{Keyboard}\})} = 0.8/1 = 0.8$**
  - $X = \{\text{Keyboard}\}, Y = \{\text{Headphones}\}, X \cup Y = \{\text{Keyboard}, \text{Headphones}\}$   
 **$\text{conf}(\{\text{Keyboard}\} \Rightarrow \{\text{Headphones}\}) = \frac{\text{sup}(\{\text{Keyboard}, \text{Headphones}\})}{\text{sup}(\{\text{Keyboard}\})} = 0.6/1 = 0.6$**
  - $X = \{\text{Headphones}\}, Y = \{\text{Keyboard}\}, X \cup Y = \{\text{Keyboard}, \text{Headphones}\}$   
 **$\text{conf}(\{\text{Headphones}\} \Rightarrow \{\text{Keyboard}\}) = \frac{\text{sup}(\{\text{Keyboard}, \text{Headphones}\})}{\text{sup}(\{\text{Headphones}\})} = 0.6/0.8 = 0.75$**
  - $X = \{\text{Laptop}, \text{Mouse}\}, Y = \{\text{Keyboard}\}, X \cup Y = \{\text{Laptop}, \text{Mouse}, \text{Keyboard}\}$   
 **$\text{conf}(\{\text{Laptop}, \text{Mouse}\} \Rightarrow \{\text{Keyboard}\}) = \frac{\text{sup}(\{\text{Laptop}, \text{Mouse}, \text{Keyboard}\})}{\text{sup}(\{\text{Laptop}, \text{Mouse}\})} = 0.6/0.6 = 1$**
  - $X = \{\text{Keyboard}\}, Y = \{\text{Laptop}, \text{Mouse}\}, X \cup Y = \{\text{Laptop}, \text{Mouse}, \text{Keyboard}\}$   
 **$\text{conf}(\text{Keyboard} \Rightarrow \{\text{Laptop}, \text{Mouse}\}) = \frac{\text{sup}(\{\text{Laptop}, \text{Mouse}, \text{Keyboard}\})}{\text{sup}(\{\text{Keyboard}\})} = 0.6/1 = 1$**
  - $X = \{\text{Laptop}, \text{Keyboard}\}, Y = \{\text{Mouse}\}, X \cup Y = \{\text{Laptop}, \text{Mouse}, \text{Keyboard}\}$   
 **$\text{conf}(\{\text{Laptop}, \text{Keyboard}\} \Rightarrow \{\text{Mouse}\}) = \frac{\text{sup}(\{\text{Laptop}, \text{Mouse}, \text{Keyboard}\})}{\text{sup}(\{\text{Laptop}, \text{Keyboard}\})} = 0.6/0.8 = 0.75$**
  - $X = \{\text{Mouse}\}, Y = \{\text{Laptop}, \text{Keyboard}\}, X \cup Y = \{\text{Laptop}, \text{Mouse}, \text{Keyboard}\}$   
 **$\text{conf}(\{\text{Mouse}\} \Rightarrow \{\text{Laptop}, \text{Keyboard}\}) = \frac{\text{sup}(\{\text{Laptop}, \text{Mouse}, \text{Keyboard}\})}{\text{sup}(\{\text{Mouse}\})} = 0.6/0.8 = 0.75$**
  - $X = \{\text{Mouse}, \text{Keyboard}\}, Y = \{\text{Laptop}\}, X \cup Y = \{\text{Laptop}, \text{Mouse}, \text{Keyboard}\}$   
 **$\text{conf}(\{\text{Mouse}, \text{Keyboard}\} \Rightarrow \{\text{Laptop}\}) = \frac{\text{sup}(\{\text{Laptop}, \text{Mouse}, \text{Keyboard}\})}{\text{sup}(\{\text{Mouse}, \text{Keyboard}\})} = 0.6/0.8 = 0.75$**
  - $X = \{\text{Laptop}\}, Y = \{\text{Mouse}, \text{Keyboard}\}, X \cup Y = \{\text{Laptop}, \text{Mouse}, \text{Keyboard}\}$   
 **$\text{conf}(\{\text{Laptop}\} \Rightarrow \{\text{Mouse}, \text{Keyboard}\}) = \frac{\text{sup}(\{\text{Laptop}, \text{Mouse}, \text{Keyboard}\})}{\text{sup}(\{\text{Laptop}\})} = 0.6/0.8 = 0.75$**

Sehingga semua aturan asosiasi yang ditemukan adalah:

- $\{\text{Laptop}\} \Rightarrow \{\text{Mouse}\}$
- $\{\text{Mouse}\} \Rightarrow \{\text{Laptop}\}$
- $\{\text{Laptop}\} \Rightarrow \{\text{Keyboard}\}$
- $\{\text{Keyboard}\} \Rightarrow \{\text{Laptop}\}$

- {Mouse} => {Keyboard}
  - {Keyboard} => {Mouse}
  - {Keyboard} => {Headphones}
  - {Headphones} => {Keyboard}
  - {Laptop, Mouse} => {Keyboard}
  - {Keyboard} => {Laptop, Mouse}
  - {Laptop, Keyboard} => {Mouse}
  - {Mouse} => {Laptop, Keyboard}
  - {Mouse, Keyboard} => {Laptop}
  - {Laptop} => {Mouse, Keyboard}
4. Untuk data transaksi yang lebih besar, pengerjaan dengan tangan akan sangat tidak efisien sehingga diperlukan untuk dapat mengimplementasikan algoritma untuk menemukan frequent itemsets, implementasikan algoritma untuk menghitung frequent itemsets tersebut tanpa menggunakan library module (make it from scratch)!

```
def prune(c, threshold):
    f = {}
    for item in c:
        if c[item] < threshold:
            continue
        f[item] = c[item]
    return f

inp = [
    {"Laptop", "Mouse", "Keyboard", "Headphones"},
    {"Laptop", "Mouse", "Keyboard"},
    {"Mouse", "Keyboard", "Headphones"},
    {"Laptop", "Keyboard", "Headphones"},
    {"Laptop", "Mouse", "Keyboard", "Webcam"}
]

# This one is also correct, feel free to try (answer is the same
# with Slide 12 of the association rules slide).
#
# Ref =
https://scele.cs.ui.ac.id/pluginfile.php/209218/mod\_resource/content/1/07\_assoc\_rules.pdf
# inp = [
```

```

#     {"Bread", "Coke", "Milk"},
#     {"Juice", "Bread"},
#     {"Juice", "Coke", "Diaper", "Milk"},
#     {"Juice", "Bread", "Diaper", "Milk"},
#     {"Coke", "Diaper", "Milk"}
# ]

min_support = 0.6

threshold = min_support * len(inp)

c1 = {}

for transaction in inp:
    for item in transaction:
        if (item in c1):
            c1[item] += 1
        else:
            c1[item] = 1

f1 = prune(c1, threshold)

c2 = {}
for i in f1:
    for j in f1:
        if i == j:
            continue

        if f"{{{i}, {j}}}" in c2 or f"{{{j}, {i}}}" in c2:
            continue

        c2[f"{{{i}, {j}}}" ] = 0
        for transaction in inp:
            if i in transaction and j in transaction:
                c2[f"{{{i}, {j}}}" ] += 1

f2 = prune(c2, threshold)

```

```

result = [set(key.strip("{}").split(', ')) for key in f1.keys()] +
[set(key.strip("{}").split(', ')) for key in f2.keys()]

F = [set(key.strip("{}").split(', ')) for key in f2.keys()]

k = 2

while F != []:
    C = {}
    # Generate Ck+1 by joining itemset-pairs in Fk;
    for i in range(len(F)):
        for j in range(i+1, len(F)):
            unions = F[i].union(F[j])
            if len(unions) == k + 1:
                C[frozenset(F[i]).union(frozenset(F[j]))] = 0

    for transaction in inp:
        for key in C:
            if key.issubset(transaction):
                C[key] += 1

    # Prune itemsets from Ck+1 that violate downward closure;
    F = prune(C, threshold)

    new_F = []

    for key, value in F.items():
        new_F.append(set(key))

    # Determine Fk+1 by support counting on Ck+1;

    result += [ set(key) for key in F.keys()]

    F = new_F

    k += 1

print(result)

```



```
PS D:\DataMining> py apriori.py  
[('Laptop', 'Mouse'), ('Keyboard', 'Headphones'), ('Laptop', 'Mouse'), ('Laptop', 'Keyboard'), ('Mouse', 'Keyboard'), ('Keyboard', 'Headphones'), ('Laptop', 'Mouse', 'Keyboard')]
```

5. Menurut Anda, apa kelebihan dan kekurangan dari pemanfaatan aturan asosiasi sebagai sistem rekomendasi jika dibandingkan dengan kebanyakan machine learning based algorithm untuk sistem rekomendasi e.g matrix factorization algorithm?

Kelebihan utama dari aturan asosiasi terletak pada kemampuannya untuk mengidentifikasi hubungan yang sederhana antara item berdasarkan sejarah pembelian atau interaksi pengguna. Aturan asosiasi dapat memberikan rekomendasi berdasarkan pola-pola transaksi yang muncul secara konsisten. Tidak hanya itu, aturan asosiasi juga memiliki kompleksitas yang cukup rendah dibandingkan algoritma lain dalam sistem rekomendasi.

Namun, kekurangan utama dari aturan asosiasi adalah ketidakmampuannya mengatasi permasalahan *cold start*, di mana sistem sulit memberikan rekomendasi untuk item atau pengguna baru yang belum memiliki riwayat interaksi. Selain itu, aturan asosiasi cenderung menghasilkan rekomendasi yang bersifat terbatas pada pola-pola yang sudah terlihat, dan sulit untuk menangkap preferensi kompleks atau tren yang lebih halus. Di sisi lain, algoritma machine learning berbasis matriks faktorisasi mampu menangani *cold start* dan dapat menggambarkan preferensi pengguna secara lebih dinamis dengan memperhitungkan faktor-faktor yang tersembunyi. Akan tetapi, algoritma ini memiliki komputasi yang cukup berat sehingga tidak begitu fleksibel.

Kelebihan aturan asosiasi:

- **Simpel dan intuitif** yang menyebabkan mudah dalam mengimplementasikan algoritma ini.
- Komputasi yang tidak terlalu berat menyebabkan **waktu pelatihan yang lebih cepat**.
- Aturan asosiasi dapat **berskala dengan baik terhadap jumlah data**. Ini membuatnya cocok untuk dataset yang sangat besar, karena tidak memerlukan komputasi yang rumit seperti beberapa algoritma pembelajaran mesin.

Kekurangan aturan asosiasi:

- Sulit dalam menangani permasalahan seperti *cold start* (tidak riwayat transaksi)
- Aturan asosiasi umumnya tidak mampu menangani data berkelanjutan atau variabel numerik secara langsung. Sehingga lebih baik digunakan pada **data kategori**.