

TD1 : Règles d'association

Exercice 1 :

Soit D la base de transactions suivante :

Id transactions	Items
100	pain beurre lait eau
200	chocolat beurre javel
300	pain chocolat beurre javel
400	chocolat lait

Appliquer l'algorithme à priori pour extraire les itemsets fréquents, sachant que le support_min = 2.

Sol :

Pain -> p, beurre -> b, lait -> l, eau -> e, chocolat -> c, javel -> j

C1 :

Itemsets	support
P	2
B	3
L	2
E	1
C	3
J	2

L1 :

P
B
L
C
J

C2 :

{P,B}	2
{P,L}	1
{P,c}	1
{P,J}	1
{b , L }	1
{b, c}	2
{b, J}	2
{l, c}	1
{l, j}	0
{J, c}	2

L2 :

{P,B}
{b, c}
{b, J}

{J, c}

C3 :

{b,c,j}	2
---------	---

Les itemsets fréquents :

{P},{B},{L},{C},{J},{P,B},{b, c},{b, J},{J, c},{b,c,j}

Exercice 2 :

Considérons seulement l'ensemble $\{A;B;C\}$; Trouver l'ensemble de règles qui permettent de prédire l'achat de deux produits, sachant que $conf_min = 60\%$

- 1 A D
- 2 A B C
- 3 A E
- 4 A B C D
- 5 B D

Sol :

R1 : A => BC ; conf(R1)= 2/4 => 50%

R2 : B =>AC ; conf(R2) = 2/3 => 66% accepté

R3 : C=>AB ; conf(R3)= 2/2 => 100% accepté

Exercice 3:

Soit la table de transactions suivante :

Id transaction	Items
T1	A, B, C, D
T2	A, B, C, D
T3	A, B, C
T4	B, C, D
T5	C, D

Sachant que $sup_min = 3$, chercher alors les *frequent itemsets*, les *closed itemsets* et les *maximal itemsets*.

Sol :

C1 :

itemsets	support
A	3
B	4
C	5
D	4

L1 :

itemsets
A
B
C
D

C2 :

itemsets	support
{A,B}	3
{A,C}	3
{A,D}	2
{B,C}	4
{B,D}	3
{C,D}	4

L2 :

itemsets
{A,B}
{A,C}
{B,C}
{B,D}
{C,D}

C3 :

{A,B,C} : 3

{B,C,D} : 3

L3 : {A,B,C} et {B,C,D}

Recherche des itemsets fermés et maximal :

{A} => n'est pas fermé et n'est pas maximal

{B} => n'est pas fermé et n'est pas maximal

{C} => fermé et n'est pas maximal

{D} => n'est pas fermé et n'est pas maximal

{A,B} => n'est pas fermé et n'est pas maximal

{A,C} => n'est pas fermé et n'est pas maximal

{B,C} => fermé mais n'est pas maximal

{B,D} => n'est pas fermé et n'est pas maximal

{C,D} => fermé mais n'est pas maximal

{A,B,C} => fermé et maximal

{B,C,D} => fermé et maximal

Exercice 4:

Soient les transactions suivantes ainsi que les produits achetés dans chacune.

Id transaction	Items
T1	HotDogs, Buns, Ketchup
T2	HotDogs, Buns
T3	HotDogs, Coke, Chips
T4	Chips, Coke
T5	Chips, Ketchup
T6	HotDogs, Coke, Chips

- 1) En utilisant l'algorithme à priori, trouver les itemsets candidats et les itemsets fréquents pour chaque scan de la base, sachant que *support_min* = 2.
- 2) Générer les règles d'association sachant que *conf_min* = 60%

Sol :

HotDogs :H, Buns: B, Ketchup:k, Coke: Co, Chips: Ch

C1:

Itemsets	Support
H	4
B	2
K	2
Co	3
Ch	4

L1:

Itemsets
H
B
K
Co
Ch

C2:

Itemsets	support
{H,B}	2
{H,k}	1
{H,Co}	2
{H,Ch}	2
{B,k}	1
{B,Co}	0
{B,Ch}	0
{k,Co}	0
{k,Ch}	1
{Co,Ch}	3

L2:

{H,B}
{H,Co}
{H,Ch}
{Co,Ch}

C3:

{H,Co,Ch}	2
-----------	---

Les itemsets frequents:

$\{H\}, \{B\}, \{Co\}, \{k\}, \{Ch\}, \{H, B\}, \{H, Co\}, \{H, Ch\}, \{Co, Ch\}, \{H, Co, Ch\}$

2) Pour $\{H, B\}$:

R1: $H \rightarrow B$ $conf(R1) = 2/4 \Rightarrow 50\%$

R2 : $B \rightarrow H$ $conf(R2) = 2/2 \Rightarrow 100\%$ Ok

Pour $\{H, Co\}$:

R3: $H \rightarrow Co$ $conf(R3) = 2/4 \Rightarrow 50\%$

R4: $Co \rightarrow H$ $conf(R4) = 2/3 \Rightarrow 66\%$ Ok

Pour $\{H, Ch\}$:

R5: $H \rightarrow Ch$ $conf(R5) = 2/4 \Rightarrow 50\%$

R6: $Ch \rightarrow H$ $conf(R6) = 2/4 \Rightarrow 50\%$

Pour $\{Co, Ch\}$:

R7 : $Co \rightarrow Ch$: $conf(R7) = 3/3 \Rightarrow 100\%$ Ok

R8: $Ch \rightarrow Co$ $conf(R8) = 3/4 \Rightarrow 75\%$ Ok

Pour $\{H, Co, Ch\}$:

R9: $H, Co \rightarrow Ch$, $conf(R9) = 2/2 \Rightarrow 100\%$

R91: $H \rightarrow Ch, Co$ $conf(R91) = 2/4 \Rightarrow 50\%$

R92: $Co \rightarrow Ch, H$ $conf(R92) = 2/3 \Rightarrow 66\%$

R10: $H, Ch \rightarrow Co$, $conf(R10) = 2/2 \Rightarrow 100\%$

R101: $Ch \rightarrow Co, H$, $conf(R101) = 2/4 \Rightarrow 50\%$

R11: $Co, Ch \rightarrow H$, $conf(R11) = 2/2 \Rightarrow 100\%$

Exercice 5:

ID transactions	Items
1	A, C, D
2	B, C, E
3	A, B, C, E
4	B, E
5	A, B, C, E

- 1) Extraire les closed itemsets , sachant que le support_min = 2.
- 2) D duire les r gles d'association

A : 3
 B : 4
 C : 4 fermé
 E : 4
 {A,B} : 2
 {A,C} : 3 fermé
 {A,E} : 2
 {B,C} : 3 fermé
 {B,E} : 4 fermé
 {C,E} : 3 fermé
 {A,B,C} : 2
 {A,B,E} : 2
 {A,C,E} : 2
 {B,C,E} : 2
 {A,B,C,E} : 2 fermé et maximal

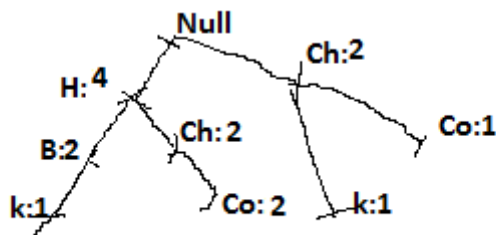
Exercice 6 :

Id transaction	Items
T1	HotDogs, Buns, Ketchup
T2	HotDogs, Buns
T3	HotDogs, Coke, Chips
T4	Chips, Coke
T5	Chips, Ketchup
T6	HotDogs, Coke, Chips

- 1) En utilisant la même base que l'exercice 4 construire le FP-tree.
- 2) Extraire les itemsets fréquents à partir de FP-tree.

Sol :

$L=[H,Ch,Co,B,K]$



Pour k :

Base des itemsets	Itemsets générés
<Ch> :1	-
<B,H> :1	-

Pour B :

Base des itemsets	Itemsets générés
-------------------	------------------

<H> :2	{H,B} :2
--------	----------

Pour Co :

Base des itemsets	Itemsets générés
<Ch> :1	{H,Ch,Co} :2
<H,Ch> :2	{H,Co} :2
	{Ch,Co} :3

Pour Ch :

Base des itemsets	Itemsets générés
<H> :2	{H,Ch} :2

Les itemsets fréquents :

{H}, {Ch}, {B}, {k}, {Co}, {H,B} , {H,Co} , {Ch,Co} , {H,Ch}, {H,Ch,Co}

Ex :

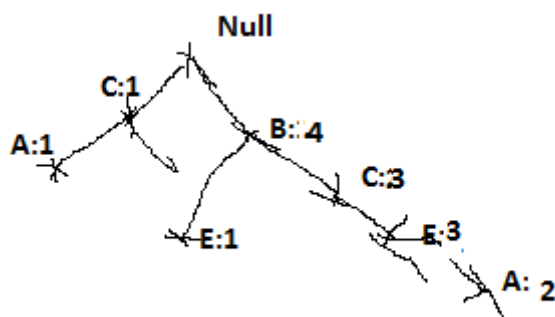
ID transactions	Items
1	A, C, D
2	B, C, E
3	A, B, C, E
4	B, E
5	A,B, C, E

- 1) Construire le FP-tree.
- 2) Extraire les itemsets fréquents à partir de FP-tree.

1)items fréquents :

A :3, B :4,C :4,E :4

L=[B,C,E,A]



Pour A :

Base	Itemsets générées
<E,C,B> :2	{A,E,C,B} :2
<C> :1	{A,E,C} :2

	{A,E,B} :2 {A,C,B} :2 {A,E} :2 {A,B} :2 {A,C} :3
--	---

Pour E :

Base	Itemsets générées
 :1	{B,C,E} :3
<B,C> :3	{B,E} :4
	{C,E} :3

Pour C :

Base	Itemsets générées
 :4	{B,C} :4