

EXAMEN S2

Exercice 1: Soit la matrice des transactions suivante:

	A	B	C	D	E
T1	0	1	1	1	0
T2	1	1	1	1	1
T3	1	0	0	1	0
T4	1	1	1	1	1
T5	0	1	1	1	0
T6	0	1	0	0	1

- 1) Donner les étapes d'extraction des itemsets fréquents fermés pour $\text{Minsup}=0.33$
- 2) Donner les règles d'association pour $\text{Minconf}=0.50$ et relatives aux 2-itemsets fréquents?

Exercice 2

Soient les données suivantes :

$A1=(1,9)$, $A2=(1,4)$, $A3=(7,3)$, $A4=(4,7)$, $A5=(6,4)$, $A6=(5,3)$, $A7=(0,1)$, $A8=(3,8)$.

- 1) Faire une classification de ces données en utilisant l'algorithme k-means pour $k=3$. On considère les classes de départ $C1=(A1,A2,A3)$, $C2=(A4,A5)$, $C3=(A6,A7,A8)$, On utilisera city-blocs pour mesurer la distance entre les données.

Exercice 3

- 1) Quelles sont les limites techniques du data mining ?
- 2) Parmi les approches d'extraction des itemsets étudiées, la quelle est la plus adaptée dans le cas d'un contexte dense et dans le cas d'un contexte épars. Justifiez.
- 3) Quel Si on considère, dans une approche d'extraction de règles d'association, que chaque item possède un poids représentant son importance, donnez la valeur de calcul du support?
- 4) Pour une classification supervisée, quant es ce qu'on dit qu'on a un phénomène de sur-apprentissage ?