



# EXAMEN FOUILLE DE DONNÉE SUJET 2

## EXAMEN DATA MINING

### EXERCICE 1 : CLASSIFICATION PAR ARBRE DE DÉCISION 05PTS

Une banque dispose des informations suivantes sur un ensemble de ses clients :

Client	M	A	R	E	I
1	moyen	moyen	village	oui	oui
2	élevée	moyen	bourg	non	non
3	faible	âgé	bourg	non	non
4	faible	moyen	bourg	oui	oui
5	moyen	jeune	ville	oui	oui
6	élevée	âgé	ville	oui	non
7	moyen	âgé	ville	oui	non
8	faible	moyen	village	non	non

#### QUESTIONS :

- 1 Quelle est l'entropie de la population ?
- 2 Pour la construction de l'arbre de décision, utilisez-vous l'attribut numéro de client ? Pourquoi ?
- 3 Lors de la construction de l'arbre de décision, quel est l'attribut à tester à la racine de l'arbre ?
- 4 Construisez l'arbre de décision complet.
- 5 Quel est le taux d'erreur de cet arbre estimé sur l'ensemble des clients 1 à 8 ?
- 6 Donnez un intervalle de valeurs pour l'erreur réelle en utilisant une confiance de 90%.


#### ON SE DONNE LES 4 CLIENTS SUIVANTS :

Client	M	A	R	I	E
9	moyen	âgé	village	oui	non
10	élevée	jeune	ville	non	non
11	faible	âgé	village	non	non
12	moyen	moyen	bourg	oui	oui


7) Comment chacun de ces clients est-il classé avec l'arbre de décision que vous avez proposé dans la question 4?

Pour ces 4 clients, on apprend par ailleurs que les clients 9 et 10 gèrent leur compte par Internet, et que les clients 11 et 12 ne le font pas.





효과적으로 말소리를 키워줍니다. 건강 청력처럼 말소리와 주변음을 구분하는 신규 보청기 체험



더 알아보기 →

\*이 제품은 '의료기기'이며, '사용상의 주의사항'과 '사용방법'을 잘 읽고 사용하십시오. 하이닥원은 보청기 판매원입니다. 광고심의필 32023-110-01-0050 (유효기간 26.01.05)

8) Quel est le taux d'erreur estimé sur les clients 9, 10, 11 et 12? Combien y a-t-il de faux positifs et de faux négatifs?

## EXERCICE 2 : CLASSIFICATION AVEC LA TECHNIQUE DES K-MEANS 05PTS

Utilisez l'algorithme du K-means et la distance euclidienne pour regrouper les 8 exemples suivants en 3 clusters : A1(2,10), A2(2,5), A3(8,4), A4(5,8), A5(7,5), A6(6,4), A7(1,2), A8(4,9).

La matrice de distance basée sur la distance euclidienne est fournie ci-dessous :

	A1	A2	A3	A4	A5	A6	A7	A8
A1	0	$\sqrt{25}$	$\sqrt{36}$	$\sqrt{13}$	$\sqrt{50}$	$\sqrt{52}$	$\sqrt{65}$	$\sqrt{5}$
A2		0	$\sqrt{37}$	$\sqrt{18}$	$\sqrt{25}$	$\sqrt{17}$	$\sqrt{10}$	$\sqrt{20}$
A3			0	$\sqrt{25}$	$\sqrt{2}$	$\sqrt{2}$	$\sqrt{53}$	$\sqrt{41}$
A4				0	$\sqrt{13}$	$\sqrt{17}$	$\sqrt{52}$	$\sqrt{2}$
A5					0	$\sqrt{2}$	$\sqrt{45}$	$\sqrt{25}$
A6						0	$\sqrt{29}$	$\sqrt{29}$
A7							0	$\sqrt{58}$
A8								0


www.PandaCodeur.com

On considère comme centre de classe à l'initialisation les points A1, A4 et A7.


1) Déroulez une itération de l'algorithme de K-means pour ces données et cette initialisation et donnez :

- Les nouveaux clusters ;
- Les centres de chaque cluster ;
- Faites une représentation graphique montrant les points d'étude, les clusters et les centres de clusters.

2) Combien d'itérations supplémentaire sont-elles nécessaires pour converger ?



효과적으로 말소리를 키워줍니다. 건강 청력처럼 말소리와 주변음을 구분하는 신규 보청기 체험



[더 알아보기 →](#)

기드형보청기: "이 제품은 '의료기기'이며, '사용상의 주의사항'과 '사용방법'을 잘 읽고 사용하십시오.  
헤어닷컴은 보청기 판매원입니다. 광고심의필 2023-110-01-0050 (유효기간 26.01.05)

Dessinez les résultats de chaque itération nécessaire.

## EXERCICE 3 : RÉSEAUX DE NEURONES 05PTS

Soit A et B deux variables booléennes.

- 1 Concevoir un réseau de neurones à deux entrées permettant d'implémenter la fonction booléenne  $A \wedge B$ .
- 2 Concevoir un réseau de neurones à deux couches implémentant la fonction booléenne  $A \text{ XOR } B$ .
- 3 En déduire un réseau de neurones implémentant la fonction booléenne :  $\neg(A \text{ XOR } B)$  sachant que :  $\neg(A \text{ XOR } B) = \neg((A \wedge \neg B) \vee (\neg A \wedge B))$   
 $= \neg(A \wedge \neg B) \wedge \neg(\neg A \wedge B)$

## EXERCICE 4 : GENERATION DE RÈGLES D'ASSOCIATION 05PTS

Soit X la base de transactions contenant un ensemble de transactions décrivant des achats de produits dans l'ensemble {A, B, C, D, E}.

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

1. Montrer que l'union de deux itemsets fréquents n'est pas toujours fréquent ;
2. Générer les règles d'association avec  $\text{minConf} = 1$  pour les données de la base X ;
3. Considérons seulement l'ensemble {A, B, C} ; Trouver l'ensemble de règles qui permettent de prédire l'achat de deux produits tout en améliorant la prédiction par rapport à la mesure statistique induite par la base de données ;

4. Répéter la même question pour trouver l'ensemble de règles qui permettent de prédire l'achat d'un produit.

## CORRIGE EXAMEN FOUILLE DE DONNEE :

bientôt

# TÉLÉCHARGER L'EXERCICE SOUS FORME DE PDF



PANDACODEUR

PANDACODEUR

PANDACODEUR

Si vous avez trouvé les examens corrigés en fouille de donnée de Mr JoëlYk intéressants et utiles, pourquoi ne pas les partager avec d'autres personnes qui pourraient également en bénéficier ? Partagez ce lien sur les réseaux sociaux ou envoyez-le à vos amis et collègues. Vous pourriez aider quelqu'un à améliorer ses compétences en programmation ou à trouver des solutions à des problèmes complexes. N'oubliez pas que la connaissance doit être partagée pour grandir. Merci pour votre soutien et votre partage !

Contact WhatsApp : +237 652027193 | Réaliser Par [Joël\\_Yk](#)

[ENCORE PLUS D'EXAMEN](#)[ENCORE PLUS D'EXAMEN CORRIGE ICI !!!](#)

corrige ici !!!



☆☆☆☆☆ Aucune note. Soyez le premier à attribuer une note !