

Examen semestriel

Durée : 01H30

Modules "Fouille et extraction de données" & "Datamining"

Exercice 1 (12 points) :

On veut appliquer le modèle des "Règles d'association" à un problème de TextMining.

Le tableau suivant représente les mots-clés (les mots les plus importants) extraits à partir de 7 textes.

N° Texte	Mots clés
01	Finance, Marché, Budget, Economie
02	Ouverture, Finance, Economie
03	Ouverture, Assemblée, Handball, Sport
04	Directeur, Budget, Finance, Economie
05	Directeur, Assemblée, Handball, Sport
06	Ouverture, Marché, Economie
07	Ouverture, Assemblée, Directeur, Handball, Sport

A/ D'après-vous quel est l'objectif recherché de l'application des "Règles d'association" à ce problème ?.

B/ Sans faire de calcul, donnez une règle d'association du tableau dont la confiance est égale à 100%. Justifiez.

C/ Réécrivez le tableau précédent en gardant uniquement la première lettre de chaque mot-clé (pour simplifier la notation) . Appliquez l'algorithme a priori pour trouver toutes les règles d'association qui vérifient $minsup \geq 40\%$ et donnez leur confiance. Détaillez toutes les étapes.

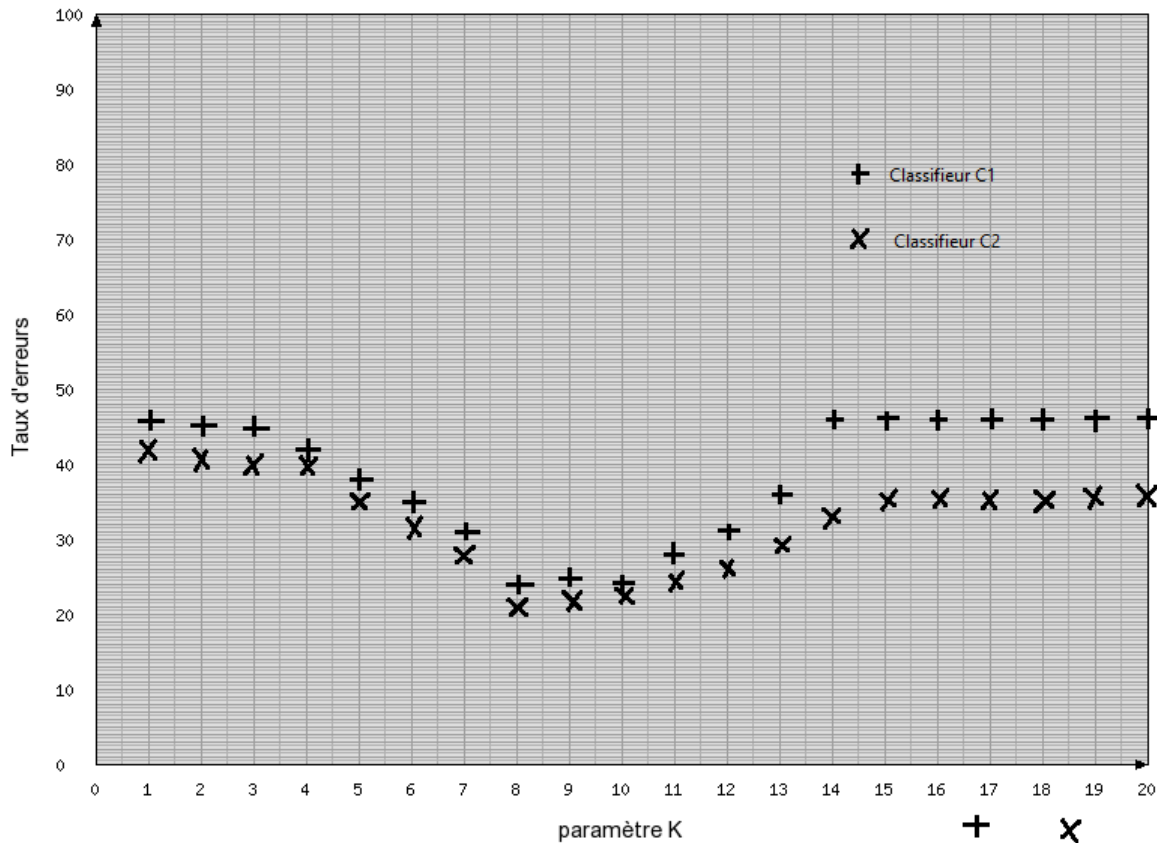
D/ On considère maintenant les données d'apprentissage d'un problème de classification en utilisant les "Règles d'association" (Remarquez la similitude avec le tableau précédent) :

N° Texte	Mots clés	Classe
01	Finance, Marché, Budget	Economie
02	Ouverture, Finance	Economie
03	Ouverture, Assemblée, Handball	Sport
04	Directeur, Budget, Finance	Economie
05	Directeur, Assemblée, Handball	Sport
06	Ouverture, Marché	Economie
07	Ouverture, Assemblée, Directeur, Handball	Sport

D'après-vous comment peut-on utiliser les résultats de la question C pour répondre à ce problème de classification : Etant donné un ou plusieurs mots-clés, on veut savoir à quelle classe ils renvoient ?.

Exercice 2 (8 points) :

Le graphique suivant représente les résultats d'une comparaison des performances de deux classifieurs C1 et C2. Les deux classifieurs sont basés sur le modèle KNN (K Nearest Neighbors), mais le premier utilise la règle du "vote majoritaire", alors que le second utilise la règle de "l'inverse de la distance au carré". L'évaluation a été faite sur les mêmes données (leur nombre est 15), et en utilisant la même distance.



A/ Expliquez comment a-t-on obtenu les résultats de l'évaluation de ces classifieurs (les étapes qui ont été suivies) ?.

B/ Dans les deux courbes (C1 et C2), il y'a une tendance commune caractéristique du modèle KNN. Expliquez-la brièvement.

C/ D'après le graphique, quel est le meilleur classifieur ? . Comment pouvez-vous l'expliquer ?

D/ Peut-on généraliser ce résultat ? Justifiez.