

Data Mining

Technique de Data Mining

Partie N° : 2

Technique de Data Mining

■ Travail à réaliser

Prétraitement des données : normalisation

Programmer en Java la méthode de normalisation Min-Max.

$VALEUR_{(min, nouvelle)}$ & $VALEUR_{(max, nouvelle)}$:

. A définir par l'utilisateur .

. Pour les expérimentations : $VALEUR_{(min, nouvelle)} = 0$ & $VALEUR_{(max, nouvelle)} = 1$.

Programmer en Java la méthode de normalisation z-score.

Technique de Data Mining

■ Travail à réaliser

Prétraitement des données : discrétisation

Programmer en Java la méthode de discrétisation basée sur les Quantiles.

Le nombre de Quantiles Q :

. A définir par l'utilisateur.

. Pour les expérimentations : **Cas n°1** : $Q = 3$, **Cas n°2** : $Q = 4$, **Cas n°3** : $Q = 5$

Programmer en Java la méthode de discrétisation basée sur les intervalles égaux.

Le nombre d'intervalles Q :

. A définir par l'utilisateur.

. Pour les expérimentations : **Cas n°1** : $Q = 3$, **Cas n°2** : $Q = 4$, **Cas n°3** : $Q = 5$

Technique de Data Mining

■ Travail à réaliser

Extraction des itemsets fréquents

Programmer en Java l'algorithme Apriori.

Le support minimum *min_sup* :

- . A définir par l'utilisateur.
- . Pour les expérimentations : **Cas n°1** : $min_sup = 20\%$, **Cas n°2** : $min_sup = 30\%$

Programmer en Java l'algorithme Eclat.

Le support minimum *min_sup* :

- . A définir par l'utilisateur.
- . Pour les expérimentations : **Cas n°1** : $min_sup = 20\%$, **Cas n°2** : $min_sup = 30\%$

Technique de Data Mining

■ Travail à réaliser

Extraction des règles d'association

Programmer en Java l'algorithme d'extraction de règles d'association.

La confiance minimale min_conf :

. A définir par l'utilisateur.

. Pour les expérimentations : **Cas n°1** : $min_conf = 30\%$, **Cas n°2** : $min_conf = 50\%$

Technique de Data Mining

- **Travail à réaliser**

Extraction des règles de corrélation

Programmer en Java l'algorithme d'extraction de règles de corrélation.

Générer toutes les règles positivement corrélées

Technique de Data Mining

■ Travail à réaliser

Classification supervisée des instances du dataset

Programmer en Java l'algorithme de classification Naïve Bayésienne.

L'ensemble de test (T):

- . A définir par l'utilisateur.
- . Pour les expérimentations : **Cas n°1** : 20 instances de chaque classe
Cas n°2 : 25 instances de chaque classe

Technique de Data Mining

■ Travail à réaliser

Classification supervisée des instances du dataset

Programmer en Java l'algorithme de classification KNN.

L'ensemble de test (T) :

- . A définir par l'utilisateur.
- . Pour les expérimentations : **Cas n°1** : 20 instances de chaque classe
Cas n°2 : 25 instances de chaque classe

Le nombre K de voisins :

- . A définir par l'utilisateur.
- . Pour les expérimentations : **Cas n°1** : $K = 4$, **Cas n°2** : $K = 5$, **Cas n°3** : $K = 10$

Mesures de distance :

- . A définir par l'utilisateur (Euclidienne ou Manhattan).
- . Pour les expérimentations : **Cas n°1** : Euclidienne, **Cas n°2** : Manhattan

Technique de Data Mining

■ Travail à réaliser

Classification supervisée des instances du dataset

Calculer la matrice de confusion.

		Classes prédites		
		C_1	...	C_m
Classes réelles	C_1

	C_m

		C_i prédite	
		YES	NO
C_i réelle	YES	TP	FN
	NO	FP	TN

Technique de Data Mining

■ Travail à réaliser

Classification supervisée des instances du dataset

Comparer l'algorithme Naïve Bayésienne avec l'algorithme KNN.

ACCURACY pour chaque classe & ACCURACY moyenne :

$$ACCURACY = \frac{TP + TN}{TP + TN + FP + FN}$$

SENSITIVITY pour chaque classe & SENSITIVITY moyenne :

$$SENSITIVITY = \frac{TP}{TP + FN}$$

SPECIFICITY pour chaque classe & SPECIFICITY moyenne :

$$SPECIFICITY = \frac{TN}{FP + TN}$$

Technique de Data Mining

■ Travail à réaliser

Classification supervisée des instances du dataset

Comparer l'algorithme Naïve Bayésienne avec l'algorithme KNN.

PRECISION pour chaque classe & PRECISION moyenne :

$$PRECISION = \frac{TP}{TP + FP}$$

RAPPEL pour chaque classe & RAPPEL moyenne :

$$RAPPEL = \frac{TP}{TP + FN}$$

F-SCORE pour chaque classe & F-SCORE moyenne :

$$F - SCORE = \frac{2 * PRECISION * RAPPEL}{PRECISION + RAPPEL}$$