

TP03 Fouille de Données

Guettouche Islem

Master 1 Génie de l'informatique logicielle (G.I.L)

20 Mai 2017

Table des matières

Introduction :3

Question13

Question 23

Question 33

Question 43

Question 73

Question 83

Question 94

Question 104

Question 114

Question 124

Question 134

Introduction :

Le système Weka est développé à l'Université de Waikato, Nouvelle Zélande. Il permet de prétraiter des données, de les analyser à l'aide d'une méthode de data mining et d'afficher le modèle résultant et ses performances. Weka est entièrement développé en Java. Il est diffusé sous licence publique GNU. Les ressources nécessaires à l'installation et à l'utilisation du système sont disponibles à l'adresse suivante <http://www.cs.waikato.ac.nz/ml/weka/>. Dans Weka, chaque méthode de transformation, de sélection d'attributs, d'apprentissage, de clustering ou de découverte d'associations est implémentée par une classe Java. La documentation des classes est accessible à partir du fichier packages.html. Weka traite des données au format ARFF (Attribute Relation Format File) ou CSV (Comma Separated Values).

Question 1

On ne peut pas appliquer l'algorithme.

Question 2

On ne peut pas appliquer l'algorithme.

Question 3

On ne peut pas appliquer l'algorithme.

Question 4

On ne peut pas appliquer l'algorithme.

Question 7

play=yes (64.29% [9/14])

humidity <= 80 (85.71% [6/7])

| temperature > 65 (100% [5/5])

windy = FALSE (75% [6/8])

| temperature <= 83 (85.71% [6/7])

| | humidity <= 86 (100% [5/5])

| humidity <= 86 (83.33% [5/6])

Question 8

play=yes (64.29% [9/14])

humidity = low-medium (85.71% [6/7])

windy = FALSE (75% [6/8])

Question 9

On ne peut pas parce que le **fichier bank-data.csv** contient des **données numériques**.

Question 10

bank-data5-2.csv.arff.

Question 11

bank-data5-3a.csv.arff

bank-data5-3b.csv.arff

Question 12

Ça marche pas rien ne s'affiche.

Question 13

- ✚ 0,400446 0,020000 */ age = '(-inf-34.333333]' ==> income = '(-inf-24386.173333]'
- ✚ 0,336064 0,121667 */ income = '(-inf-24386.173333]' ==> age = '(-inf-34.333333]'
- ✚ 0,320998 0,045000 */ income = '(-inf-24386.173333]' ==> save_act = NO or age = '(-inf-34.333333]'
- ✚ 0,303847 0,013333 */ age = '(-inf-34.333333]' ==> income = '(-inf-24386.173333]' or children = '(2.25-inf)'
- ✚ 0,302589 0,010000 */ pep = NO and age = '(-inf-34.333333]' ==> income = '(-inf-24386.173333]'
- ✚ 0,299144 0,003333 */ income = '(43758.136667-inf)' ==> age = '(50.666667-inf)'
- ✚ 0,298962 0,018333 */ age = '(-inf-34.333333]' ==> region = SUBURBAN or income = '(-inf-24386.173333]'
- ✚ 0,289126 0,010000 */ age = '(-inf-34.333333]' ==> income = '(-inf-24386.173333]' or current_act = NO
- ✚ 0,288708 0,035000 */ children = '(0.75-1.5]' ==> pep = YES
- ✚ 0,285261 0,016667 */ age = '(-inf-34.333333]' ==> region = RURAL or income = '(-inf-24386.173333]'