

Examen Fouille de Données - Session 1 M2GIL et M1ITA

Notes personnelles manuscrites autorisées

14 janvier 2020

Répondez de manière claire aux questions. Le barème est donné à titre indicatif et n'est en aucun cas définitif. Merci de remettre une copie propre et lisible (1pt).

Nous disposons des données suivantes relatives à des patients porteurs de lentilles de contact au format .arff.

```
@relation lentilles
@attribute prescription {myope, hypermetrope}
@attribute astigmatisme {no, yes}
@attribute production-larmes {reduite, normale}
@attribute couleur-yeux {marron, noir, vert, bleu}
@attribute genre {M, F}
@attribute lentilles {yes, no}
@data
myope,no,reduite,bleu,M,no
myope,no,reduite,noir,M,no
myope,no,normale,noir,F,no
myope,yes,reduite,marron,F,no
myope,yes,reduite,marron,M,no
myope,yes,normale,vert,F,yes
myope,yes,normale,vert,M,yes
hypermetrope,no,reduite,noir,M,no
hypermetrope,no,normale,vert,F,yes
hypermetrope,no,normale,vert,F,yes
hypermetrope,no,normale,vert,M,yes
hypermetrope,no,normale,bleu,M,yes
```

Nous souhaitons étudier les associations entre les différentes valeurs des attributs y compris celles de la classe.

1 Apriori (7 pts)

1. Trouver un moyen pour représenter la base sous forme de transactions.
2. Déroulez l'algorithme Apriori (les différentes itérations) pour rechercher les motifs fréquents ayant un support supérieur ou égal à 30%.

3. Dessinez le treillis associé que vous obtenez.
4. Donnez la liste des règles d'association obtenues avec une confiance de 50%.
5. Donnez une interprétation générale des règles obtenues.
6. Pour chacune des règles obtenues calculez leur lift. Est-ce que cela confirme votre interprétation des règles obtenues ?

2 Close (8 pts)

On souhaite maintenant extraire des règles d'association à partir de l'ensemble des patients en utilisant l'algorithme Close.

1. Déroulez l'algorithme Close (les différentes itérations) et donnez la liste des règles d'association exactes et approximatives avec un support minimum de 50% et une confiance minimale de 50%.
2. Dessinez le treillis associé.
3. Donnez une interprétation générale des règles obtenues.
4. Pour chacune des règles obtenues calculez leur lift. Est-ce que cela confirme votre interprétation des règles obtenues ?
5. Quelles remarques faites-vous par rapport aux règles extraites avec Apriori ?
6. Si vous étiez un commercial, est-ce que les résultats obtenus par la fouille (algos Apriori et Close) vous permettraient de lancer une campagne publicitaire ? Si oui, laquelle ? Si non, pourquoi ?

3 Questions générales (4 pts)

Pour chacune des questions ci-dessous il est attendu pour chaque réponse, une explication claire, argumentée et rédigée sous forme de texte (entre 3 et 5 lignes par réponse).

1. Quelles sont les différentes étapes de la fouille de données ?
2. Décrivez le principe de l'apprentissage supervisé. Citez au moins une méthode appartenant à ce type d'apprentissage et donnez-en un exemple.
3. Quels sont les avantages et inconvénients des arbres de décision ?