

Forage de données

TP01: itemsets et règles d'association

Plan

- ▶ **Partie 01: (50%)**
 - ▶ Règles d'association avec Weka
- ▶ **Partie 02:(50%)**
 - ▶ Implémentation de l'algorithme Apriori pour la génération d'itemsets

Plan

- ▶ **Partie 01:**

- ▶ Règles d'association avec Weka

- ▶ **Partie 02:**

- ▶ Implémentation de l'algorithme Apriori pour la génération d'itemsets

Objectifs

- ▶ Approfondissement des connaissances sur la plateforme Weka
- ▶ Manipulation de jeux de données (*datasets*)
- ▶ *Workflow* pour la génération de règles d'association

Énoncé

- ▶ En utilisant Weka, réaliser une opération de forage de données sur un jeu de données intégré dans Weka
- ▶ Objectif de l'opération:
 - ▶ Détecter les habitudes d'achat des ménages sous la forme de règles d'association
- ▶ Étapes:
 1. Chargement de données:
 - ▶ Le jeu de données supermarket.arff sera utilisé
 - ▶ Le fichier se retrouve au niveau du répertoire data (data/supermarket.arff)
 2. Découverte de règles d'association en utilisant deux algorithmes
 1. Utiliser les algorithmes Apriori et FPGrowth avec différentes valeurs de leurs paramètres
 3. Analyse des résultats

Énoncé (suite)

- ▶ Une fois que les différentes étapes réalisées, produire un rapport détaillant le travail réalisé au niveau de chaque étape (avec captures d'écran)
- ▶ Revenir dans ce rapport sur le fonctionnement des deux algorithmes

Plan

- ▶ Partie 01:
 - ▶ Règles d'association avec Weka
- ▶ **Partie 02:**
 - ▶ Implémentation de l'algorithme Apriori pour la génération d'itemsets

Objectif

- ▶ Se familiariser davantage avec le fonctionnement interne de l'algorithme Apriori

Énoncé

► Implémentation de l'algorithme Apriori

- Donnez un pseudocode détaillé avec les structures de donnée d'un tel algorithme (ajoutez des commentaires)
- Implémentez ce pseudocode en utilisant Python, Ruby ou Java
 - Format des données:
 - Contenu: valeurs numériques
 - Fichier texte où chaque ligne est une transaction
 - Les items de chaque transaction sont séparés par un espace
 - Entrées: nFile, minSup
 - nFile: nom de fichier de données au format indiqué plus haut
 - minSup: support minimal (pourcentage minimal de présence d'un itemset dans la base pour qu'il soit considéré comme étant fréquent)
 - Sortie:
 - Tous les itemsets fréquents avec le support de chaque itemet

Énoncé (suite)

- ▶ Implémentation de l'algorithme Apriori
 - ▶ Un exemple de transactions à utiliser est donné dans la prochaine diapositive
 - ▶ Comparez les résultats obtenus avec votre algorithme avec ceux obtenus en utilisant la plateforme Weka
- ▶ Produisez un rapport détaillant votre implémentation et l'exécution de votre algorithme (avec captures d'écran)

Énoncé (suite): transactions

37I 20 6 485 23 65 3 7I 20 6 48

37I 20 6 485 7I 20 6 48 5 6

23 65 3 7I 20 6 48

37I 20 6 485 23 65 3 7I 20 6 48

23 65 3 6

37I 20 6 485 23 65 7I 20 6 48 6

7I 20 6 48 5

37I 20 6 485 23 65 3 7I 20 6 48 5

3 7I 20 6 48 5

37I 20 6 485 23 65 3 6

37I 20 6 485 23 65 3 5

37I 20 6 485 7I 20 6 48 5

23 65 3 7I 20 6 48 6

37I 20 6 485 23 65 3 7I 20 6 48

23 65 3 7I 20 6 48

37I 20 6 485 23 65 7I 20 6 48 5 6

3 7I 20 6 48 5

37I 20 6 485 23 65 3 7I 20 6 48 5 6

3 7I 20 6 48 5 6

37I 20 6 485 23 65 3 5



Directives

- ▶ Travail à effectuer en équipes de deux
- ▶ Fichiers à remettre:
 - ▶ Rapport word/libreoffice ou présentation pptx ou latex
 - ▶ Code source
- ▶ Date de remise:
 - ▶ 22 février 23h59
 - ▶ Il y aura éventuellement une séance pour la présentation des travaux (date à déterminer)