



Université des Antilles
UFR des Sciences Exactes et Naturelles
Master Info

TP3
Intelligence Artificielle et Fouille de Données

Exercice I - Recherche de motifs fréquents

On souhaite étudier les motifs fréquents sur des données climatiques. Nous étudions pour cela un jeu de données qui donne la praticabilité d'un terrain selon plusieurs attributs météorologiques.

www.erickstattner.com/terrain.csv

Chaque ligne du jeu de données fourni se compose de 5 variables:

1. Le temps (sunny, overcast, rainy)
2. La température (valeur numérique)
3. Le taux d'humidité (valeur numérique)
4. La présence de vent (valeur booléen)
5. La praticabilité du terrain (valeur booléen)

On souhaite extraire de ce jeu de données des schémas récurrents "intéressants". Vous décidez pour cela d'utiliser un algorithme de recherche d'itemsets fréquents.

1. Dans un tel jeu de données, quel serait pour vous des itemsets fréquents intéressants ?
2. En utilisant l'algorithme "*FPGrowth_itemsets_with_strings*" de l'outil *smpfvu* en cours, extraire et commenter les itemsets fréquents obtenus avec un seuil de support de 0.4.
3. Modifier le jeu de données de façon à ce que
 - a. la température devienne une valeur nominale indiquant "chaude", "moyenne" ou "fraîche".
 - b. l'humidité devienne une valeur nominale indiquant "élevée" ou "normale".
4. Relancer la recherche des itemsets fréquents avec un seuil de support de 0.4 et commenter les résultats.