**Bensalem Riad**

**Bousri Mohamed Charref eddin**

**SIQ4**

Rapport TP Feature Selection

Module ML

# Importation et préparation des données

Dans cette partie on procède à trois étapes :

1. Importation du dataset : lire de depuis le fichier CSV les donnes et créé le datafram en lui attachant le nom des colonnes récupérés depuis le ficher descriptif .names.
2. On prépare les données aux traitements :

* traiter les lignes contenants les informations manquantes ( ?) : ici on choisit de supprimer la ligne.
* Diviser les données pour l’entrainement et test du modèle. (20% pour le test)

# Application des approches de feature sélection

Après la première partie, les résultats d’un apprentissage initial sont les suivants :

C:\Users\Ryad\Desktop\accurancy.PNG

Afin d’amélioré ce résultat on applique deux approches pour la sélection d’attributs les plus expressifs.

## Wrapper feature selection

### La sélection d’attribue exhaustive

C:\Users\Ryad\Desktop\nlfnlnjffjfkdlfd.PNGCet algorithme procède en force brute en calculant pour toutes les combinaisons possible de sous-ensembles d’attributs la valeur d’une métrique qui va permettre d’ordonner ces ensembles et de choisir le meilleur. Cette métrique reflète la performance du modèle. Si par exemple, le classificateur est une régression logistique et que le jeu de données se compose de 4 attributs, l’algorithme évaluera les combinaisons d’attributs  (si min\_features = 1 et max\_features = 4).

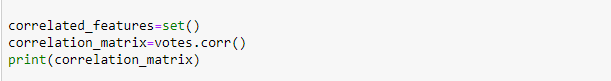
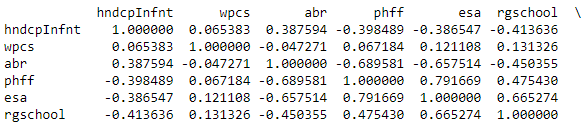
La métrique choisie est la précision. Cela donne les valeurs suivantes après l’apprentissage :

C:\Users\Ryad\Desktop\ljhdkjhgdhgkjdrhgkjehrijgdh.PNG

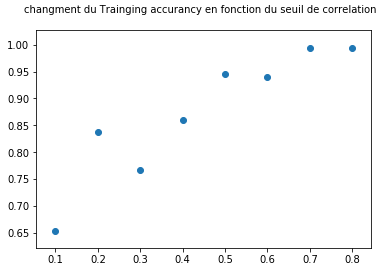
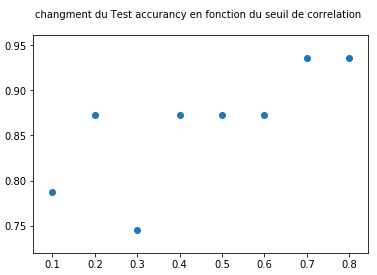
### Recursive feature selection

## Filter feature selection

### Eliminer les attributs corrélés

Cette méthode consiste premièrement à calculer la corrélation entre les colonnes deux a deux, on sauvegarde ces corrélations dans une matrice.

Après on choisit un seuil, pour le quel si la corrélation est supérieure, on élimine l’attribut(s) concerné(s).on change le seuil de corrélation on obtient plusieurs ensemble d’attributs.

Ci-dessous un graphe qui montre les changements de la précision en fonction des changements du seuil de corrélation (i.e. différents ensembles d’attributs).