L3 IA – R pour l'Intelligence Artificielle – TP n°4

Nettoyage puis traitement d'un fichier de données avec le logiciel R

Construire un compte-rendu de TP complet.

Le fichier de données « assurance » comporte les variables suivantes :

Origine : origine de la voiture
1=américaine 2=japonaise
3=européenne

Taille : taille de la voiture
Type : type de voiture
1=petite
2=moyenne
3=grande
Type : type de voiture
1=familiale
2=sportive
3=utilitaire

■ Sinistre : l'assuré a déjà eu un sinistre 1=non 2=oui

NbSinistre : le nombre de sinistres durant les 12 derniers mois

Marital : statut matrimonial
1= célibataire avec des enfants

2=marié avec des enfants

3=célibataire 4=marié.

Enfant : nombre d'enfants

Sexe : sexe de l'assuré
1=homme
2=femme.

Age : age de l'assuré

Region : Densité de population ou habité l'assuré

1=forte densité (grande ville) 2=densité moyenne (petite ville)

3=faible densité (village)

On réalise les étapes suivantes avec le logiciel R.

- 1. Décoder et donner le bon type à chacune des variables.
- 2. Détecter les valeurs aberrantes des variables Enfant et Age afin de les supprimer.
- 3. Imputer les valeurs manquantes des différentes variables par la méthode de son choix.
- 4. Traiter les données nettoyées afin de sortir des résultats intéressants sous forme de tableaux et/ou graphiques à une ou deux variables.