Introduction

La Data science est en fait un terme générique englobant toute une famille d'outils facilitant l'exploration et l'analyse des données pour des fins décisionnelles.

Les techniques mises en action lors de l'utilisation de cet instrument d'analyse et de prospection sont particulièrement efficaces pour extraire des informations significatives depuis de grandes quantités de données.

En dépit des méthodes classiques d'analyses statistiques, Cet instrument d'analyse est particulièrement adapté au traitement de grands volumes de données et avec l'augmentation de la capacité de stockage des supports informatiques, un maximum de renseignements sera capté, ordonnés et rangés « Comportement des acheteurs, caractéristiques des produits, historisation de la production », désormais plus rien n'échappe à la collecte.

Le travail présenté dans ce rapport rentre dans ce cadre et consiste en la réalisation d’une étude sur plusieurs sources d’informations afin de faire une comparaison entre ces différentes techniques d’analyse.

Le présent rapport est organisé de la manière suivante :

Partie 1 : Etude sur les données synthétiques

Dans cette partie nous ferons une étude sur les données synthétiques en passant par deux étapes à savoir :

* Exploratoire préliminaire : qui consiste en l’utilisation des méthodes classiques d’analyses statistiques.
* Classification supervisée : Dans cette étape nous appliquerons et comparerons les différentes approches de classifications supervisées .

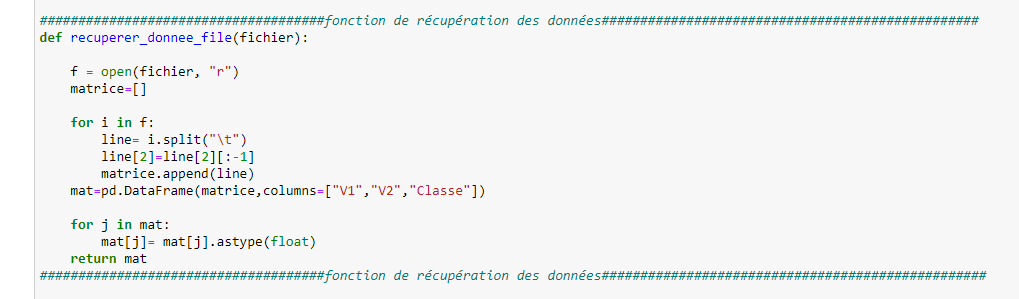
Partie 2 : Etude des données réels :

Dans cette partie nous allons faire une comparaison entre les approches de classifications supervisées sur les données «visa premiere, credit card fraude »

Enfin nous terminons par une conclusion générale.

Partie 1 : données synthétiques

Tout d’abord nous allons commencer par la récupération des données via une fonction.



Après avoir récupérer les données nous allons les normaliser

