

# TRAVAIL À DOMICILE

## COURS DE GESTION D'INFOCENTRE

### L2 INFORMATIQUE DE GESTION / DEPT. DES MATHS & INFO / FAC. DES SCIENCES / UNIKIN

---

**NOTA :** *Le travail est à téléverser (codes + explications) dans votre compte GitHub. Pour ceux qui n'ont pas de compte GitHub, ils peuvent le créer au <https://github.com>.*

*Le lien de votre travail est à envoyer au [gradi.kamingu@unikin.ac.cd](mailto:gradi.kamingu@unikin.ac.cd) au plus tard le 31 août 2022 à 23h59.*

Soit le jeu de données « data\_vin.xlsx » contenant 1598 consommateurs du vin.

Pour chaque individu, les caractéristiques retenues sont :

- L'*identifiant de l'individu* ;
- L'*acidité fixe* (acide tartrique en  $g/dm^3$ ): la plupart des acides impliqués dans le vin sont fixes ou non volatils (ne s'évaporent pas facilement) ;
- L'*acidité volatile* (acide acétique en  $g/dm^3$ ): la quantité d'acide acétique dans le vin, qui, à des niveaux trop élevés, peut conduire à un goût de vinaigre désagréable ;
- L'*acide citrique* (en  $g/dm^3$ ) : présent en petite quantité, l'acide citrique peut apporter de la « fraîcheur » et de la saveur aux vins ;
- Le *sucré résiduel* (en  $g/dm^3$ ) : la quantité de sucre restant après l'arrêt de la fermentation, il est rare de trouver des vins avec moins de 1 gramme/litre et les vins avec plus de 45 grammes/litre sont considérés comme doux ;
- Les *chlorures* (chlorure de sodium en  $g/dm^3$ ) : la quantité de sel dans le vin ;
- Le *dioxyde de soufre libre* (en  $mg/dm^3$ ) : empêche la prolifération microbienne et l'oxydation du vin ;
- Le *dioxyde de soufre* ( $SO_2$ ) total (en  $mg/dm^3$ ) : à de faibles concentrations, le  $SO_2$  est pratiquement indétectable dans le vin, mais à des concentrations de  $SO_2$  libre supérieures à 50 ppm, le  $SO_2$  devient évident dans le nez et le goût du vin ;
- La *densité* ( $g/cm^3$ ) : la densité du vin est proche de celle de l'eau en fonction du pourcentage d'alcool et de la teneur en sucre ;

- Le *pH* : décrit le degré d'acidité ou de basicité d'un vin sur une échelle de 0 (très acide) à 14 (très basique) ; la plupart des vins se situent entre 3 et 4 sur l'échelle de pH ;
- les *sulfates* (sulfate de potassium  $g/dm^3$ ): un additif du vin qui peut contribuer aux niveaux de gaz sulfureux ( $SO_2$ ), qui agit comme antimicrobien et antioxydant ;
- D'*alcool* (% en volume): le pourcentage d'alcool du vin.

En utilisant le langage de programmation **R** ou **Python**,

- (1) Importez le jeu de données depuis votre « **Mes Documents/My Documents** », dans un dossier nommé « **dataset\_votreprenomVotrenomVotreprenom** ».
- (2) Afficher les caractéristiques statistiques de cet ensemble de données ;
- (3) À l'aide de l'algorithme *k*-means, segmentez l'ensemble de données en 10 classes ; chaque classe représentant la qualité du vin;
- (4) Déterminez la taille de chaque classe ;
- (5) Représentez graphiquement ses classes dans un graphique bidimensionnel (indice : vous pouvez utiliser l'analyse en composantes principales (ACP) pour la réduction de la dimensionnalité);
- (6) Déterminez les caractéristiques statistiques de chacune des classes ;
- (7) Séparez l'ensemble de données résultant (de la segmentation) en données d'apprentissage et en données de test ;
- (8) Utilisez les méthodes suivantes pour construire vos modèles prédictifs :
  - *La régression logistique ;*
  - *Le réseau de neurones ;*
  - *L'arbre de décision (avec la méthode CART) ;*
  - *L'arbre de décision (avec la méthode ID3) ;*
  - *La méthode de k plus proches voisins.*
- (9) À l'aide des données de test, discutez de l'efficacité de chacune des méthodes utilisées ;
- (10) Choisissez arbitrairement un vin quelconque, même un vin fictif (tout en fournissant ses caractéristiques), puis prédire sa qualité.

*Bon travail !*

*Asst. Gradi L. Kamingu, M.Sc.*

*Asst. Jeannot Tshiala Mutombo, B.Sc.*