

# IA - Classification Challenge

Le but de ce challenge est de tester vos implémentations des  $K$  plus proches voisins sur un nouveau dataset anonymisé. Le challenge se déroule en 3 étapes :

- Un premier dataset nommé « data.csv » est disponible sur DVO. Celui-ci a une structure différente que le dataset Iris car il est constitué de 6 variables d'entrée et une variable qualitative en sortie. Il est composé de 5 classes. Vous êtes libre de le couper en deux ensembles train/test afin de l'entraîner, faire varier des paramètres de votre approche et de l'évaluer par la suite.
- vous recevrez un deuxième dataset nommé « preTest.csv » dans quelques jours. Celui-ci n'aura d'autre but que de permettre de tester votre modèle sur de nouvelles données et vous assurer que tout fonctionne. Vous serez libre d'en faire ce que vous voulez.
- Enfin, (et l'évaluation reposera principalement sur cette dernière partie), vous recevrez un troisième dataset nommé « finalTest.csv ». Celui-ci porte sur le même problème qu'à la première étape mais la variable à prédire n'est pas donnée (la dernière colonne). Vous devrez utiliser le modèle que vous avez élaboré afin de prédire la classe pour chaque exemple du fichier « finalTest.csv ». Vous devrez créer un fichier de sortie nommé de votre nom et de l'extension « txt ».

Ce fichier de sortie contiendra chaque prédiction (une prédiction par ligne : classA ou classB ... suivie d'un retour à la ligne). Le langage de programmation utilisé n'a pas d'importance, seul le fichier de sortie doit respecter **strictement** ces consignes car votre fichier sera analysé automatiquement. Un exemple de fichier vous sera fourni ainsi que des programmes permettant de tester votre sortie. Tout écart à ce format sera fortement pénalisé **(-5points)**.

Vous devrez déposer votre fichier de sortie sur DVO dans un dépôt spécifique. Dans un autre dépôt, vous devrez également rendre votre code (le **code exact** qui a permis de générer ce même fichier de sortie ainsi que d'un mini-rapport au format **pdf** expliquant les choix utilisés dans votre  $k$  plus proche voisins. Ce rapport est un tableau de bord et doit refléter vos tests, vos idées et vos choix qui vous ont amenés à la version déposée.

Le tout est à rendre au plus tard le **vendredi 19 avril**.

Bon Courage !