

# 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 exactement la même structure que le dataset Iris avec 4 variables d'entrée et une variable qualitative en sortie. La seule différence avec Iris est qu'il y a 10 classes au lieu de 3. 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 vous n'avez pas fait de sur-apprentissage. 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 vos noms (vos deux noms séparés par un underscore si en binômes) et de l'extension « txt ».

Ce fichier de sortie contiendra chaque prédiction (une prédiction par ligne : A ou B ... jusqu'à J 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. Tout écart à ce format sera sanctionné.

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 (quelques lignes suffisent, elles doivent juste présenter vos différents choix de paramètres). Même si vous avez déjà envoyé votre code du k plus proche voisins sur DVO, vous devrez le renvoyer à ce nouveau dépôt.

Le tout est à rendre au plus tard le **vendredi 22 mai**.

Bon Courage !