

# IFT870/BIN710 - Exercice : Réduction de dimensions et Clustering

Davy Ouedraogo (Dpt. Informatique, Fac. Sciences, Université de Sherbrooke)

Courriel: [wend.yam.donald.davy.ouedraogo@usherbrooke.ca](mailto:wend.yam.donald.davy.ouedraogo@usherbrooke.ca)

Dans le cadre de cet exercice, nous procéderons à l'importation du jeu de données prédéfini « iris ». Le lien de la documentation se trouve ci-après:

- [https://scikit-learn.org/stable/modules/generated/sklearn.datasets.load\\_iris.html#sklearn.datasets.load\\_iris](https://scikit-learn.org/stable/modules/generated/sklearn.datasets.load_iris.html#sklearn.datasets.load_iris)

---

## Question 1 : Jeu de données

- Charger le jeu de données à partir du dépôt scikit-learn.
- Charger le jeu de données à partir du fichier *iris\_data.csv*.
- fournir une description détaillée du jeu de données, incluant son format, les types des attributs ainsi qu'une explication de chacun des attributs.

---

## Question 2 : Prétraitement des données

- Le jeu de données nécessite-t-il un prétraitement (tel que le nettoyage, la transformation, etc.) avant l'application des techniques de forage de données ?
- Les données cibles sont-elles présentes ? Si oui, de quel type d'apprentissage s'agit-il (apprentissage supervisé ou non supervisé) ?
- Ainsi, opteriez-vous pour des techniques de classification ou des techniques de régression ? Justifiez votre choix et vos analyses.

---

## Question 3 : Clustering

3-1)

Appliquez une décomposition en deux composantes principales (2-CP) et trois composantes principales (3-CP) du jeu de données. Affichez une visualisation des données après transformation.

3-2)

Appliquez la méthode t-SNE sur le jeu de données. Affichez une visualisation des données après transformation.

3-3)

Que pouvez-vous conclure concernant les deux méthodes utilisées, à savoir PCA et t-SNE ?

3-4)

Instanciez les trois modèles suivants du module `sklearn.cluster`, entraînez-les sur votre jeu de données, puis effectuez des prédictions. Évaluez ensuite les résultats obtenus.

- `KMeans`
  - `AgglomerativeClustering`
  - `DBSCAN`
-