

## Fiche TP

*Intelligence artificielle, apprentissage automatique et big data*

*Licence "Agriculture de Précision" – L3*

Soit le fichier CSV « out2.csv » :

- 1) Charger ces données dans un dataframe
- 2) Afficher les colonnes
- 3) Pour chaque colonne, donner :
  - a. Nombre de lignes
  - b. Moyenne
  - c. Ecart type
  - d. Min
  - e. Le quantile à 25%
  - f. Le quantile à 50%
  - g. Le quantile à 75%
  - h. Max
- 4) Visualiser le nombre de données suivant les valeurs de chacune des colonnes : NDVI, ....
- 5) Visualiser toutes les paires d'indicateurs en coloriant les classes.  
Commentez les Figures les plus significatives.
- 6) Donner la cardinalité de chaque classe.
- 7) Donner les graphiques des quartiles de chaque colonne. Commenter !
- 8) Réduire les dimensions des données vers 2 dimensions, puis les visualiser. Que peut-on dire ?
- 9) On veut apprendre la classe de végétation suivant les colonnes données :
  - a. En utilisant les arbres de décision : quelle est la profondeur la plus précise ?
  - b. En utilisant SVM : quelle est le facteur de sur-apprentissage le plus précis ?
  - c. Quelle est la méthode ensembliste la plus précise parmi 'Random Forest', 'Bagging', 'AdaBoost' ?
- 10) On veut repérer le nombre de groupements le plus adéquat des données :
  - a. Comment le calculer ?
  - b. Visualiser les données en exploitant les données en 2D ?