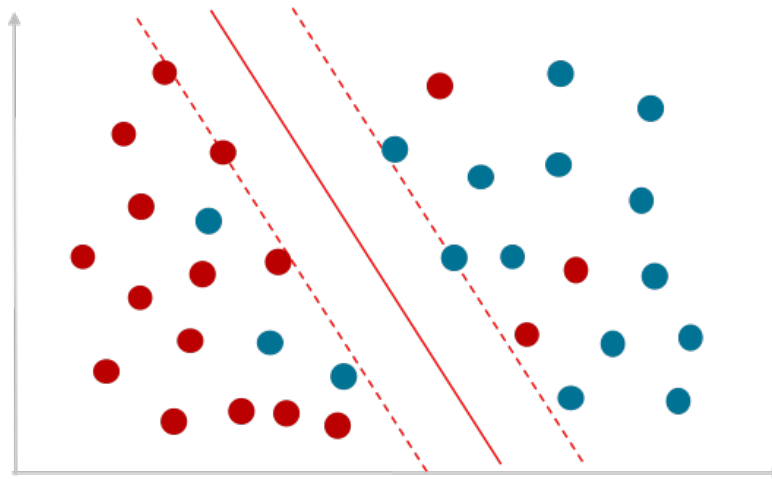


Atelier 2 «Classification»

Objective :

l'objectif principal de cet atelier est de pratiquer les concepts de la classification, en traitant les données d'une Data Sets, ainsi d'évaluer les algorithmes pour construire le modèle adéquat à notre problématique.



Outils : Python , Pandas, Sklearn, matplotlib.

Data Sets :

Diabète Data Set : <https://www.kaggle.com/kumargh/pimaindiansdiabetescsv>

Partie 1 (Data Visualisation et Feature Selection et Normalisation):

1. En utilisant pandas essayer d'explorer les données du Data set.
2. Afficher le résumer statistique du Data Sets avec une interprétation des résultats obtenues.
3. Afficher les nuages des points du data set selon les propriétés « Features » en utilisant matplotlib et pandas « scatter_matrix ».
4. Appliquer les 4 méthodes de Features selection « Univariate Selection, PCA, Recursive Feature Elimination et Feature Importance ».
5. Normaliser les données des attributs qui nécessitent une normalisation.



Université Abdelmalek Essaadi
Faculté des Sciences et techniques de Tanger
Département Génie Informatique
Cycle Ingénieur: LSI s4
Machine Learning
Pr . EL AACHAK LOTFI
2019/2020



Partie 2 (Classification choix de algorithme adéquat):

1. En utilisant l'API sklearn entraîner les modèles en utilisant ces algorithmes « KNN, Decision Tree, ANN, Naïve Bayes, SVM ».
 2. Sauvegarder les 5 models
 3. Évaluer les modèles en utilisant ces métriques:
 - Classification Accuracy.
 - Logarithmic Loss.
 - Area Under ROC Curve.
 - Confusion Matrix.
 - Classification Report.
- Interpréter le résultat de l'évaluation.
4. Comparer la performance des 5 algorithmes en utilisant la technique Spot-checking.
 5. Charger les 5 modèles puis Prédire les données du data set de test, en utilisant les 5 modèles.

Note : un rapport bien détailler doit être rédigé et imprimé.

Références :

<https://machinelearningmastery.com/compare-machine-learning-algorithms-python-scikit-learn/>

<https://machinelearningmastery.com/metrics-evaluate-machine-learning-algorithms-python/>

<https://machinelearningmastery.com/spot-check-classification-machine-learning-algorithms-python-scikit-learn/>