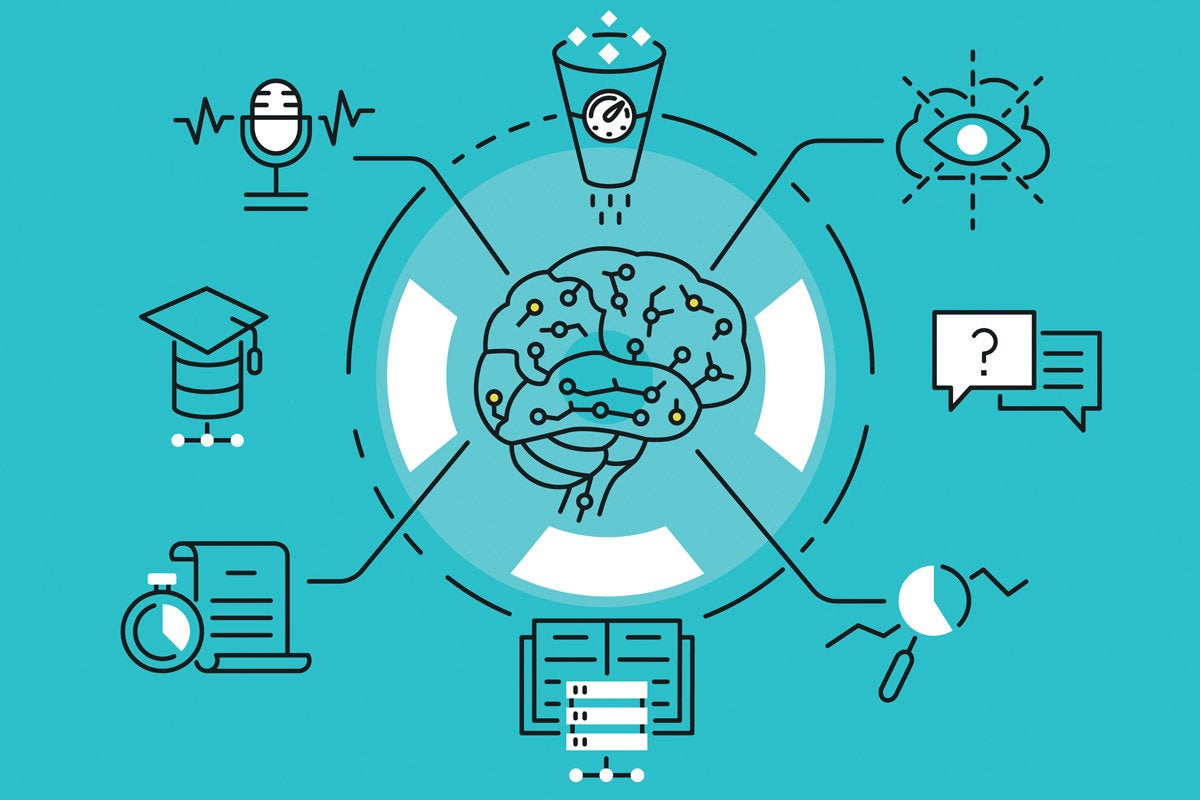
**Université Abdelmalek Essaâdi**

**Faculté des Sciences et Techniques-Tanger**

**Département Génie Informatique**

[**Machine Learning (Master SIBD)**](https://classroom.google.com/c/Njc0NzE2MzQxNjA5)

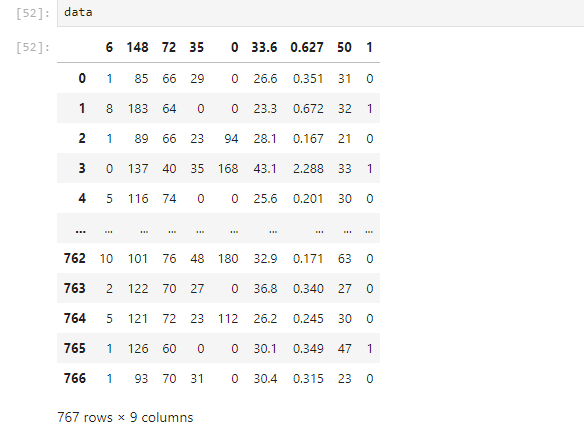
**Atelier 2 « «Classification»»**

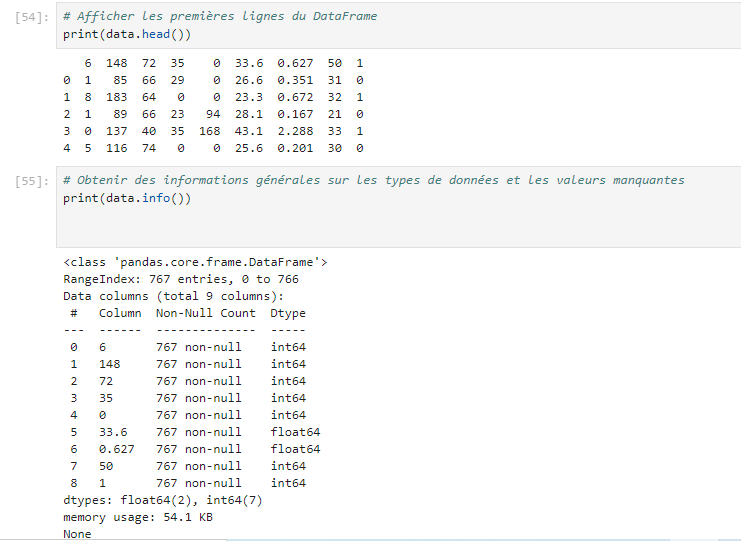
Réalisé Par : Encadre Par :

Yossra safi chetouan Pr . EL AACHAk LOTFI

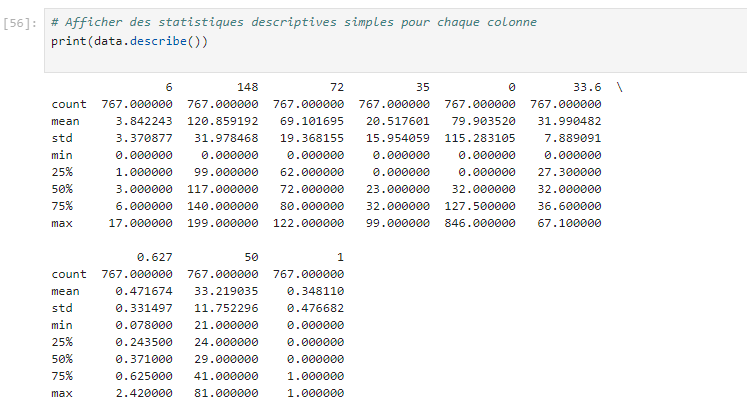
Partie 1 (Data Visualisation et Feature Selection et Normalisation):

1. En utilisant pandas essayer d’explorer les données du Data set.

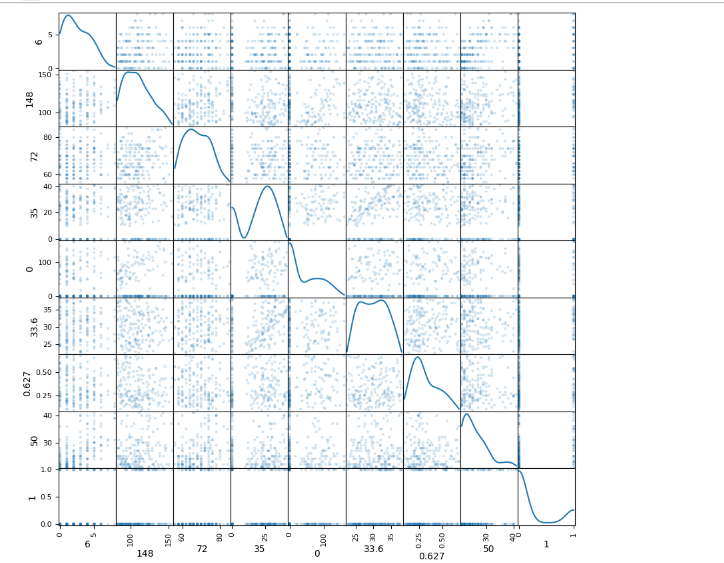




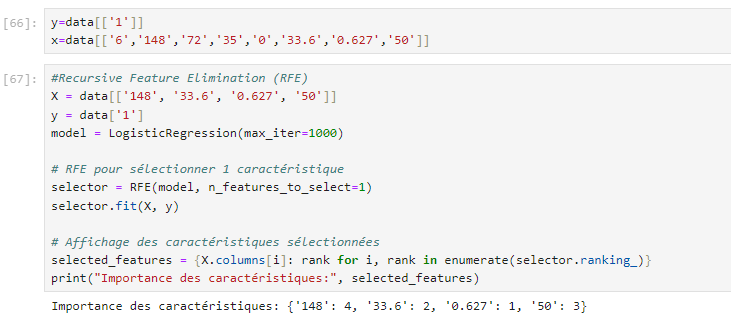
1. Afficher le résumer statistique du Data Sets avec une interprétation des résultats obtenues.

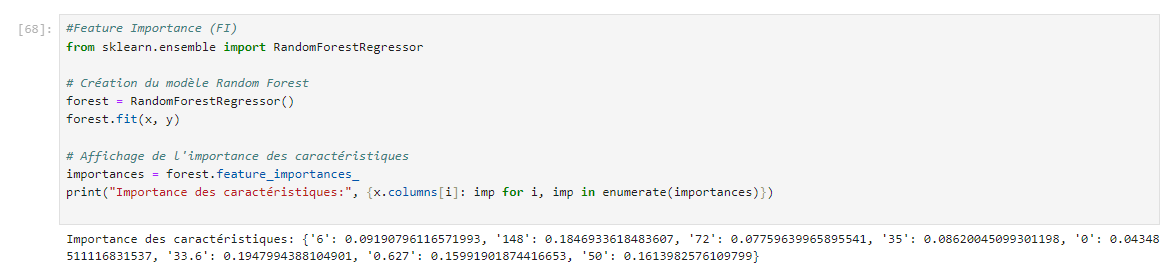


Afficher les nuages des points du data set selon les propriétés « Features » en utilisant matplotlib et pandas « scatter\_matrix ».

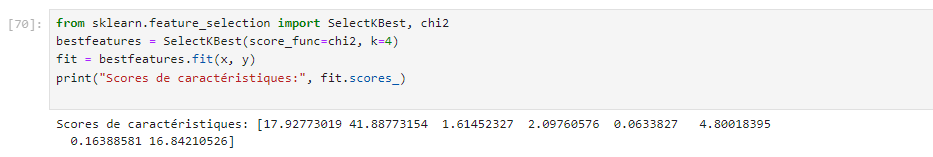


1. Appliquer les 4 méthodes de Features selection « Univariate Selection, PCA, Recursive Feature Elimination et Feature Importance ».

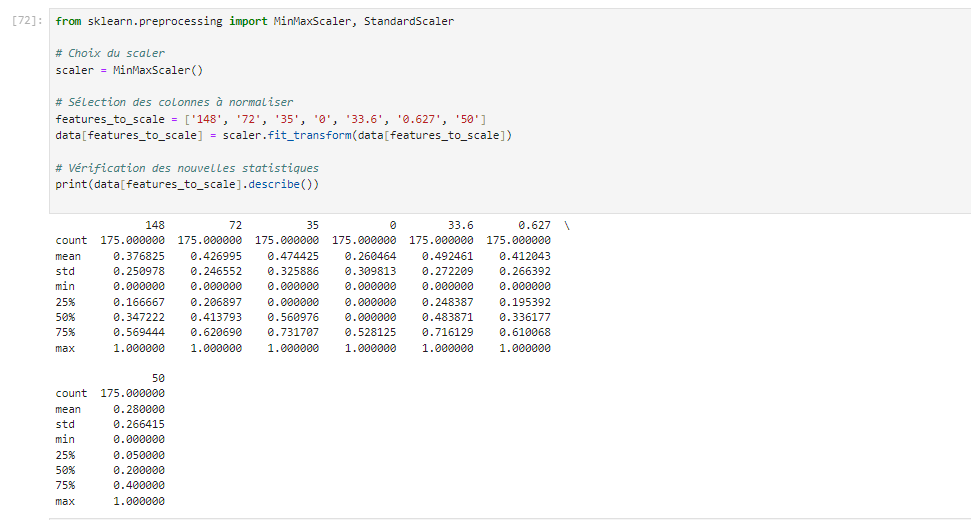


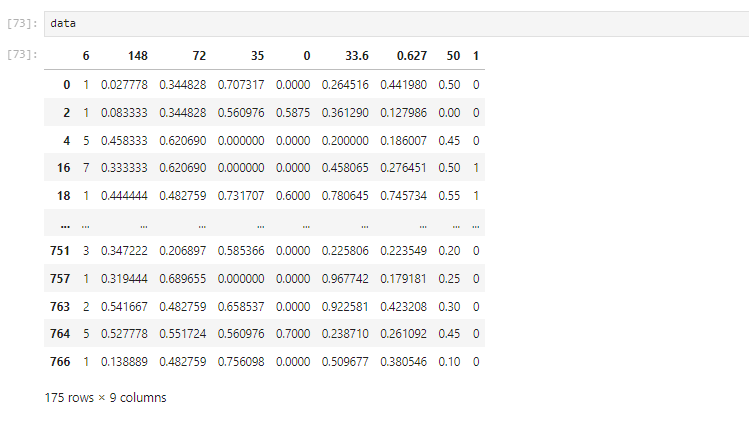






5. Normaliser les données des attributs qui nécessitent une normalisation.





Partie 2 (Classification choix de algorithme adéquat ):

1. En utilisant l’API sklearn entraîner les modèles en utilisant ces algorithmes « KNN, Decision Tree, ANN,Naive Bayes, SVM selon les kernels suivants : Linear, polynomial et guassain».



1. Sauvegarder les 5 modèles



1. Évaluer les modèles en utilisant ces métriques:

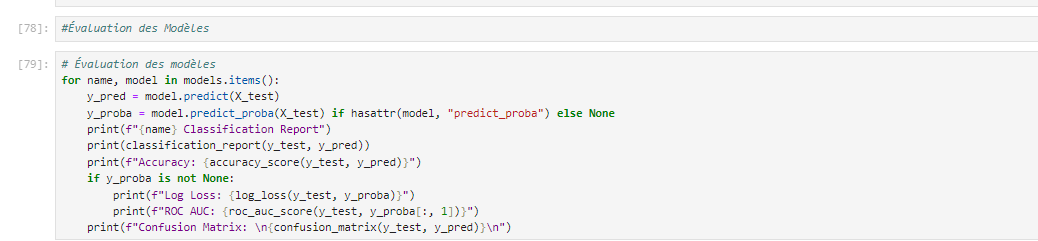
Classification Accuracy.

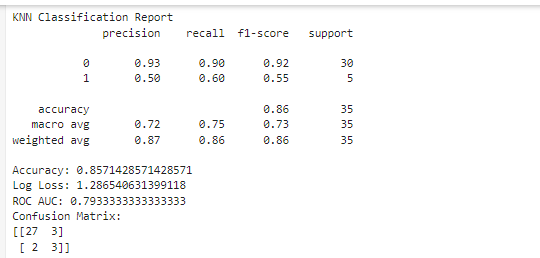
Logarithmic Loss.

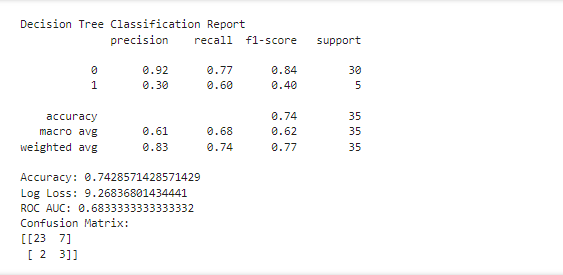
Area Under ROC Curve.

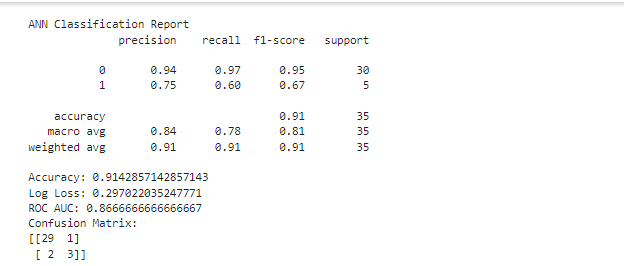
Confusion Matrix.

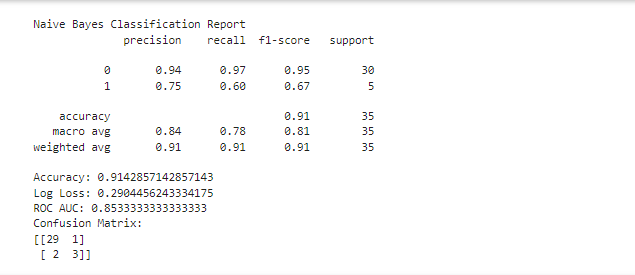
Classification Report.

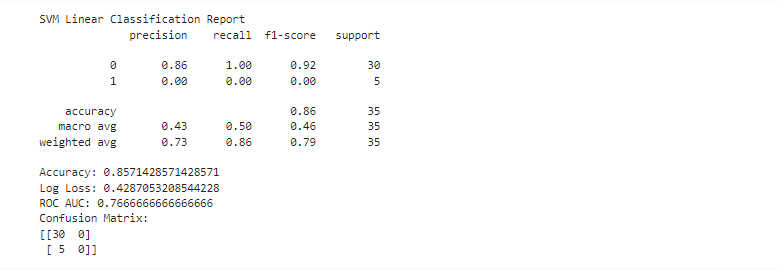


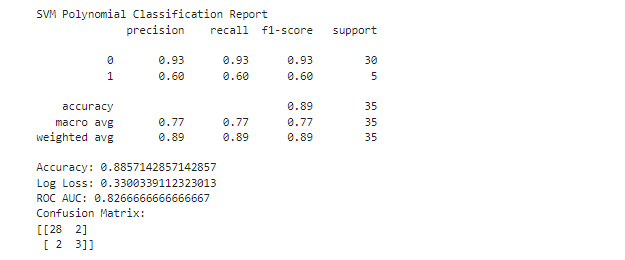


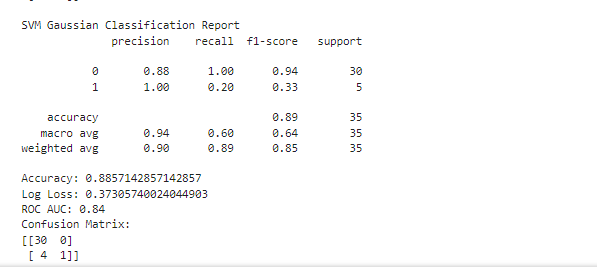




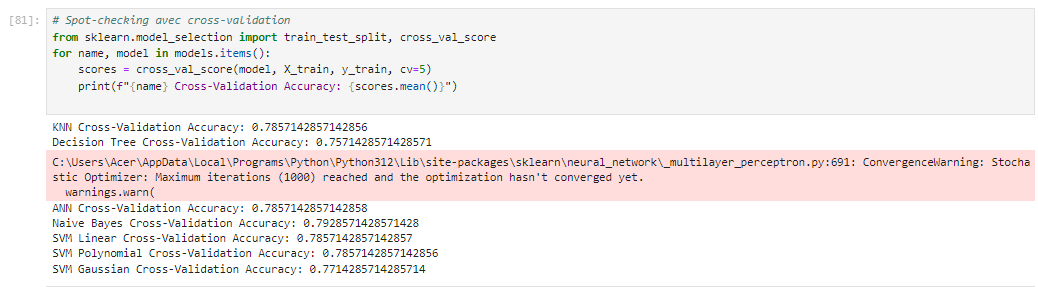








1. Comparer la performance des 8 algorithmes en utilisant la technique Spot-checking.



1. Appliquer cette fois les trois techniques d’ensemble learning « bagging , stacking et boosting »

