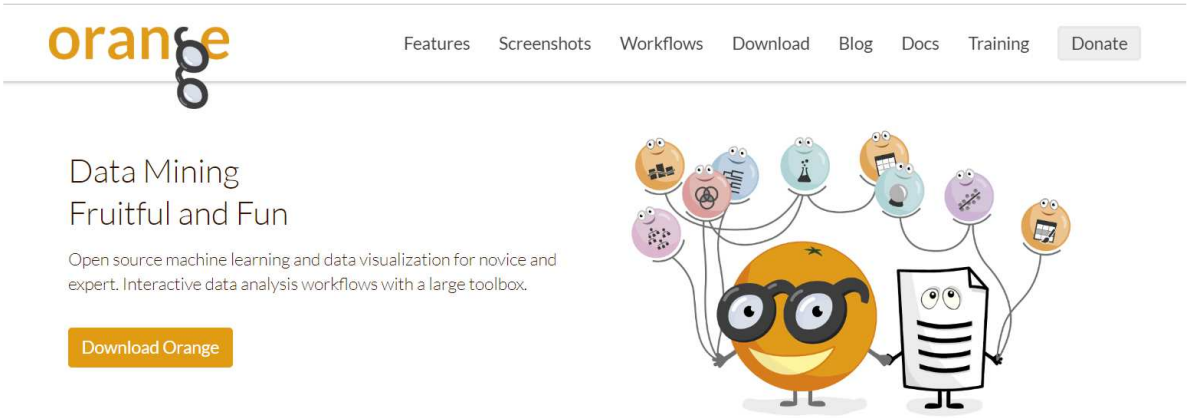
EISTI Ecole Internationale des Sciences du Traitement de l'Information	**SUJET DE PROJET S3 MI** **Analyse d'un jeu de données à l'aide de l'outil interactif Orange**
Rédigé par : Astrid Jourdan	A l'intention de : Elèves ING2 - MI
	Dernière modification : 10/10/2019

Comme vous avez pu le constater l'année dernière et cette année, extraire de l'information d'un jeu de données nécessite beaucoup de manipulations (tableur, script R, …). Orange est un logiciel libre d'exploration de données. Il propose des fonctionnalités de modélisation à travers une interface visuelle, une grande variété de modalités de visualisation et des affichages variés dynamiques[i].

Vous aurez donc à utiliser orange avec vos propres jeux de données. L'objectif est double. Il s'agit de fournir une analyse pertinente de vos données et de fournir votre démonstration des possibilités de Orange.
Voici quelques sites sur lesquels vous pouvez trouver vos jeux de données :

- Kaggle : https://www.kaggle.com/datasets
- UCI Machine learning repository : https://archive.ics.uci.edu/ml/index.php
- Open data world : https://data.world/datasets/open-data

Attention, le choix d'un jeu de données est une tâche chronophage. Par ailleurs, il est possible que votre jeu de données s'avère non pertinent au cours de votre étude. Je vous conseille donc d'en préparer plusieurs.



Vous aurez deux rendus à produire à la fin du semestre.

- Un rapport avec :
  - Une présentation du ou des jeux de données
  - Les problématiques relatives au jeu de données
  - Les moyens mis en œuvre pour répondre à ces problématiques
  - Une analyse critique des résultats
- Une démonstration de 20 min de votre utilisation de Orange

D'ici là, vous devez produire des rendus intermédiaires et vous organiser autour d'une méthode de gestion de projet à faire valider par votre enseignant en méthode Agile.

Quelques contraintes pédagogiques :

- Présenter des tâches de datamining supervisées **et** non supervisées
- Toutes les personnes du groupe doivent être en mesure d'utiliser Orange (au moins une tâche par personne)
- Intégrer votre propre script Python dans Orange
- Préciser le paramétrage utilisé pour les méthodes que vous mettrez en œuvre (métriques, nombre de clusters, initialisation, nombre de neurones,…)

[i] Attention, vous allez y prendre goût mais vos enseignements resteront illustrés avec R ☺