



## Projet IDEX Formations Innovantes INFOSHS Initiation au machine learning avec R

<b>Présentation</b>	Formation compacte intense d'utilisation de R pour une initiation aux techniques de classification sur des données linguistique. Initiation au langage de programmation R à partir d'un script, application à un package spécifique conçu pour la classification automatique. La formation contient des cours et des travaux pratiques en alternance au cours de la journée. Des liens vers des ressources complémentaires pour approfondissement seront proposés.
<b>Mode de validation</b>	Une attestation de présence sera délivrée à chaque étudiant ayant suivi l'ensemble de la formation. Un certificat de réussite sera délivré aux étudiants ayant passé avec succès le QCM et l'épreuve de validation.
<b>Programme détaillé</b>	<ul style="list-style-type: none"> <li>• initiation à la logique des packages sous R. Présentation du package Rweka: classer des données. modéliser des données par classification supervisée et non-supervisée. Initiation aux classifieurs linéaires (Flach2012)</li> <li>• 9h-12h le package RWeka (niveau I) <ul style="list-style-type: none"> <li>◦ comparer des corpus de corpus de taille différentes</li> <li>◦ notion de <i>lazy algorithm</i> : exemples d'analyse des k plus proches voisins.</li> </ul> </li> <li>• 14h-18h autres packages <ul style="list-style-type: none"> <li>◦ initiation aux techniques de classification</li> <li>◦ importation de jeux de données/résultats de concordances propres</li> <li>◦ Les arbres d'inférences conditionnelles (packages {rpart} et {party}).</li> <li>◦ initiation à la critique de modèles</li> </ul> </li> </ul>
<b>Objectifs pédagogiques</b>	Amener en peu de temps à savoir utiliser des classifieurs à partir du package {RWeka}. Apprendre à s'appropriier un package de R à partir de sa documentation. S'initier aux techniques de classification sur des données linguistiques (Baayen 2008, chapitre 5).
<b>Formateur(s)</b>	<b>Nicolas Ballier</b> est professeur de linguistique anglaise à Paris Diderot et pratique R depuis plus de trois ans. Il a notamment co-édité <i>Automatic Treatment and Analysis of Learner Corpora</i> chez Benjamins.
<b>Pré-requis</b>	Savoir se servir de R et de son environnement RStudio.
<b>Public visé</b>	Étudiants en Master 1, 2. Doctorants de 1ère année - Disciplines : langues, linguistique, littérature. S'adresse plus spécifiquement à des étudiants concernés par l'outillage informatique des langues peu dotées.
<b>Effectif</b>	maximum 20 personnes



## Informations pratiques

<b>Email (contact pédagogique, pour toutes questions sur le contenu de la formation)</b>	nicolas.ballier@univ-paris-diderot.fr
<b>Date(s)</b>	mai 2016 - de 9h à 18h
<b>Lieu</b>	Université Paris Diderot
<b>Durée de la formation</b>	Une journée
<b>Nombre de jours validés par le Cfdip</b>	
<b>Lien vers l'inscription (obligatoire)</b>	