

# Machine Learning 1 (méthodes classiques)

HAI817I

**Konstantin Todorov**

LIRMM

Konstantin.Todorov@lirmm.fr

<http://www.lirmm.fr/~todorov>



**Pascal Poncelet**

LIRMM

Pascal.Poncelet@lirmm.fr

<http://www.lirmm.fr/~poncelet>

# Les informations sont disponibles sous Moodle


- Nom : Machine Learning 1 (méthodes classiques)
- Code : HAI817I

Votre progression ?

 Annonces

---

**Cours**

 Organisation cours	<input type="checkbox"/>
 Introduction à la science des données	<input type="checkbox"/>

---

**Notebooks**

 Environnement	<input type="checkbox"/>
 Utilisation de pandas	<input type="checkbox"/>
 Visualisation de dataframes	<input type="checkbox"/>
 Ingénierie des données	<input type="checkbox"/>
 Premières Classifications	<input type="checkbox"/>



# Les informations sont disponibles sous Moodle

- Contenu des cours, des TP
- Devoir à rendre via Moodle
- Informations diverses via Moodle
- Des ressources disponibles : les notebooks



# Les intervenants

- Salim Hafid, Doctorant LIRMM
- Pascal Poncelet, Prof UM
- Konstantin Todorov, MCF UM



# Le planning

- Les cours ont lieu le lundi
- Attention l'EDT Prose n'est pas forcément à jour
- En cas de modification, nous vous prévenons par un message via Moodle



# Le programme

- Le processus d'extraction de connaissances
- Comprendre les données
  - *ingénierie des données, statistiques descriptives, visualisation, choix des dimensions importantes, normalisation*
- Trouver les meilleurs modèles et mesures
  - *classification supervisée, classification non supervisée, pattern mining*
- Evaluation des modèles
  - *precision, rappel, F-measure, AUC,*
- Mise en place d'un pipeline complet pour utiliser les modèles appris sur de nouvelles données



# Un projet

- Thématique : en cours de définition - données textuelles
- Les données sont fournies pour faciliter les traitements et se focaliser sur la partie fouille
- Les dernières années :
  - détection d'opinions - Phrase positive ou négative ?
  - Fake news - True or False ?



# Un projet

- Travail en groupe
- Trouver la meilleure classification
- Un challenge ?
- Evaluation :
  - Qualité du rapport,
  - Les prétraitements, choix des représentations, choix des descripteurs, choix des classifieurs, ...
  - Explication des résultats du challenge



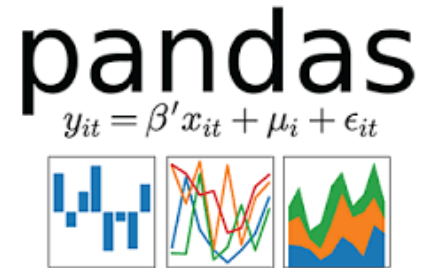


# Contactez les enseignants

- Les adresses mails :  
Salim.Hafid@lirmm.fr  
Pascal.Poncelet@lirmm.fr  
Konstantin.Todorov@lirmm.fr
- S'il vous plait mettre dans le sujet :
- [HAI817I] ....

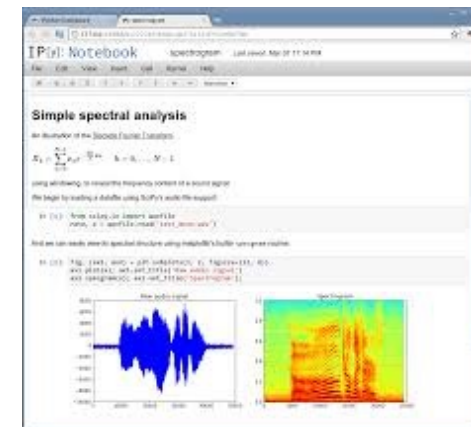
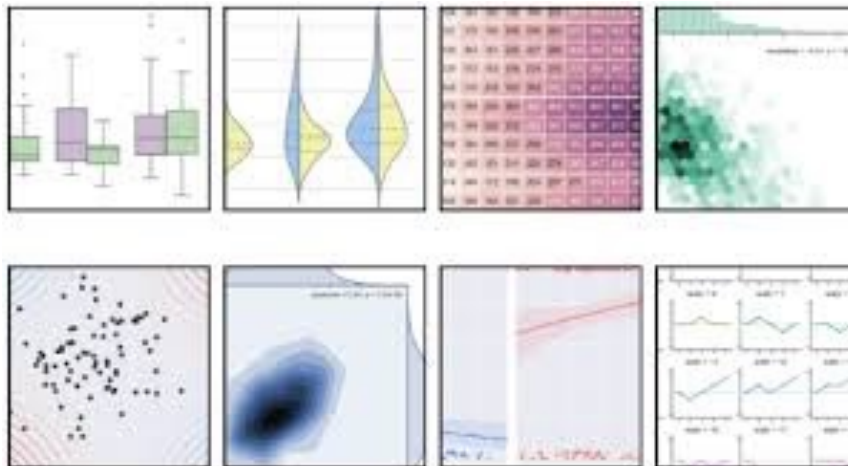


# Environnement de travail



- Les outils

- Sickit learn : bibliothèque libre Python destinée à l'apprentissage automatique
- Pandas, SciPy : pour manipuler les données
- Seaborn pour visualiser les statistiques sur les données
- Jupyter Notebook : pour faire de la science des données



# Une démo de notebook



- A faire rapidement :
  - S'inscrire sur Moodle
- Des questions ?

