

# PLDAC 2016/2017 - Développement

## Fouille de grandes masses de données temporelles

### Encadrants :

- Sébastien Lefort (DAPA, LIP6)
- Marie-Jeanne Lesot (DAPA, LIP6)

### Sujet :

Un bot a collecté trois ans durant le classement des 100 livres les plus vendus sur Amazon. Les données contiennent les descriptions des livres ainsi que leurs statistiques de vente à différentes échelles temporelles (semaines, mois).

L'objectif de ce projet est de mettre en œuvre des méthodes de fouille de données/data mining sur des grandes masses de données (big data) temporelles réelles pour extraire de nouvelles connaissances, selon deux axes principaux :

- *Prédiction dynamique de ventes* : partant d'une description d'un livre, prédire quel va être le comportement de ses ventes. Par exemple, un livre de cuisine qui va bien se vendre en fin d'année et moins bien le reste du temps, etc.
- *Clustering pour déterminer les profils d'évolution des ventes*. Par exemple, certains livres montent et déclinent lentement dans le classement (dû au bouche à oreille par exemple), d'autres connaissent un pic de ventes éphémère (dû à l'actualité par exemple), etc.

Il pourra être nécessaire de proposer des adaptations et des extensions des algorithmes selon les caractéristiques de ces données et des résultats obtenus. D'autres pistes pourront également être explorées pour enrichir le traitement de ces données.

Le développement est envisagé en Python. Quelques connaissances en bases de données sont appréciées mais pas nécessaires.