

## Fiche Stage

### CRD/Laboratoire/URD

TECH/ASAP/PROF

### Intitulé du Stage (1-2 lignes)

Personnalisation de score de campagnes à l'aide de la technique des k-moyennes

### Mission: (5-6 lignes)

Lorsqu'on désire contacter un client pour lui proposer un produit on calcule au préalable son appétence à ce produit. Il s'agit là de calculer la probabilité qu'il achètera ce produit. Cette probabilité est en fait calculée à l'aide d'un modèle prédictif (dans notre cas un classifieur naïf de Bayes) pour un ensemble de clients (le périmètre de la campagne). Les clients sont ensuite triés dans l'ordre décroissant de leur probabilité d'appétence. Le service marketing ne contacte ensuite que les plus appétant, i.e. ceux ayant la plus forte probabilité d'acheter le produit. En parallèle et avant le contact commercial il peut être intéressant de réaliser une typologie des clients qui seront contactés. L'idée étant de proposer des campagnes différenciées par segment.

Aujourd'hui, cette typologie est réalisée de manière non-supervisée à l'aide de la technique des k-moyennes. La valeur de k est prédéfinie et la métrique utilisée ne tient pas compte du prédicteur (le modèle délivrant la probabilité d'appétence) ce qui pose le problème de la stabilité des segments. Concrètement, si d'un mois sur l'autre on recalcule un score (avec les nouvelles données) et qu'on refait une typologie non-supervisée sur les clients topés par ce score, on est susceptible d'obtenir des segments de nature différente. L'idée est donc de réaliser une typologie à l'aide d'une technique des k-moyennes supervisée pour stabiliser les segments (grâce à une typologie supervisée par le prédicteur naïf de Bayes).

Le cœur du stage est donc la réalisation d'une typologie supervisée à l'aide d'une technique des k-moyennes supervisée. L'algorithme de k-moyennes supervisée sera à réaliser en partie.

### Profil: (3-4 lignes)

- Le profil souhaité est BAC + 5, Master Recherche ou Master Industriel informatique et/ou statistiques ou école d'ingénieur.
- Intérêt équilibré entre les aspects applicatifs et théoriques du sujet.

### Compétences (3-4 lignes)

- Le stagiaire devrait avoir une bonne connaissance des statistiques et des mathématiques.
- Les connaissances en Matlab (ou équivalent) seraient les bienvenues.
- Des connaissances en apprentissage statistique sont un plus.

### Modalités (durée, période, localisation)

5 mois, printemps-été 2011, Lannion (Bretagne)

### Le plus de l'offre (décrire le plus du stage ou de l'entité)

Proche de la mer, vous serez dans l'équipe de traitement des données d'Orange Labs directement en lien avec des problématiques opérationnelles du groupe Orange sur le CRM et l'Audience. Le stagiaire évoluera dans un contexte très recherche sur un sujet très porteur. Il sera intégré dans l'URD au sein d'une équipe recherche.

### Contacts

Vincent Lemaire - 02 96 05 31 07 - vincent.lemaire@orange-ftgroup.com

The English-speaking candidacies are accepted