



Utiliser le logiciel libre R en entreprise

François Pelletier

15/03/2017

Introduction

Cet atelier présentera un survol des usages du logiciel libre R dans le secteur financier et des assurances. Les domaines explorés seront les statistiques, le marketing et l'analyse géospatiale.

Ce logiciel permet de partager l'expertise et de favoriser la collaboration entre les différents secteurs d'affaires, dont la gestion des risques, le marketing et les opérations.

Il permet aussi une intégration rapide des nouveaux employés aux projets existants en ne nécessitant pas l'apprentissage d'une technologie propriétaire

Nouveaux employés

- Utiliser rapidement les connaissances acquises aux études ou dans un emploi précédent.
- Intégration plus rapide dans des projets multi-disciplinaires.
- CRAN contient au moins un package de qualité sur votre domaine d'expertise.
- Accessible librement à tous les étudiants et professionnels, multi-plateforme, simple et rapide à installer.

Technologie libre

- Accessible librement à tous les étudiants et professionnels, multi-plateforme, simple et rapide à installer.
- Économie en formation: pas d'apprentissage de technologie propriétaire spécifique à un employeur.
- Communauté active et de plus en plus diversifiée.

Fine pointe de la recherche

- Apport de nouvelles techniques rapidement dans l'entreprise.
- Permet de résoudre de nouveaux problèmes avant que les solutions commerciales n'apparaissent.

Développement et production

- RStudio Server (Interface web, Présentations et documents avec R intégré, Applications Shiny)
- Approche "In database" dans les bases de données commerciales offertes par:
 - Oracle SQL
 - Microsoft
 - IBM
 - PostgreSQL
- Intégration à Hadoop avec SparkR

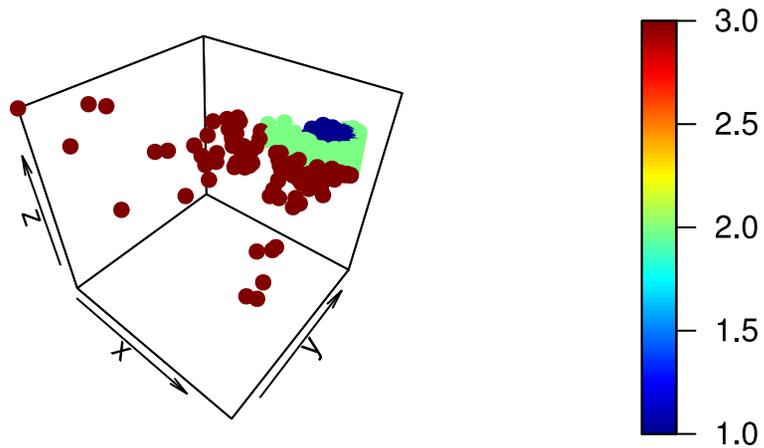
Plus proches voisins

- Famille d'algorithmes servant à identifier les voisins d'un point.
- Très efficace lorsque le nombre de points et de dimensions est très élevé.
- Permet de choisir plusieurs types de mesures de distances entre les points
- Utilisé dans plusieurs projets:
 - Détection de réclamations avec des caractéristique similaires
 - Identifier des clients ayant des caractéristiques similaires à un groupe d'intérêt

Plus proches voisins

```
library(RANN);library(readr);library(dplyr);library(plot3D)
```

130 caract. proj. sur 3 dim avec ACP; ANN; 20%+, 20%-

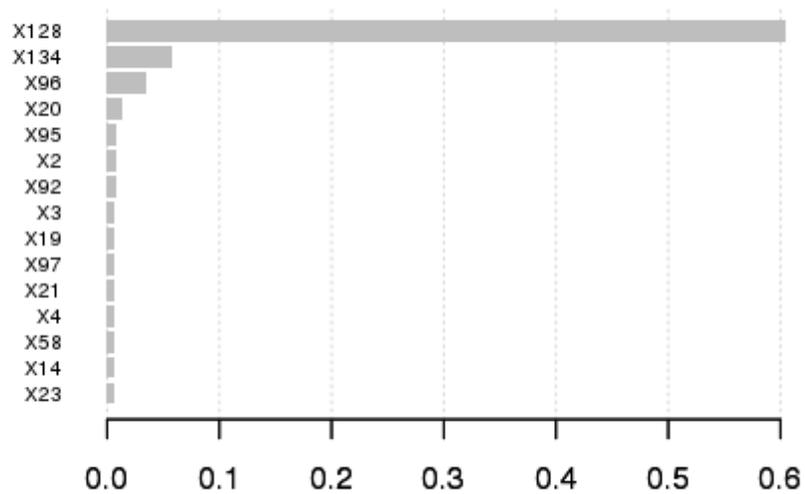


Modèles XGBoost

- Modéliser et prédire la réponse à une question simple: Est-ce que les maisons dans ce quartier valent plus que la moyenne nationale ?
- Arbres de décisions et méthodes d'ensemble (quorum, vote): meilleur modèle possible.
- Utilisé dans beaucoup de projets de modélisation:
 - Aide à la décision dans le système de règlement des réclamations
 - Ciblage publicitaire
 - Preuves de valeur avant l'achat de banques de données de fournisseurs externes.

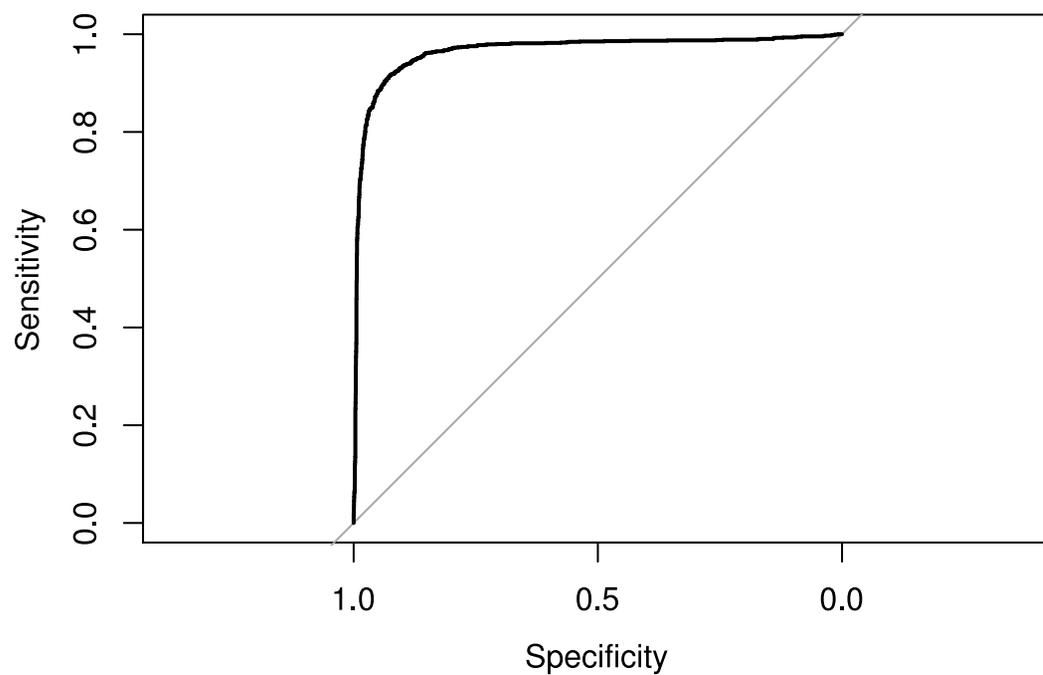
Modèles XGBoost

```
library(xgboost);library(caret);library(pROC)
```



Modèles XGBoost

Courbe ROC



Analyse marketing

Déterminer quel type de publicité motive le plus les clients potentiels.

Modèles

```
library(mgcv); library(scam)
```

```
library(rgenoud); library(broom)
```

```
library(modelr)
```

Calcul parallèle

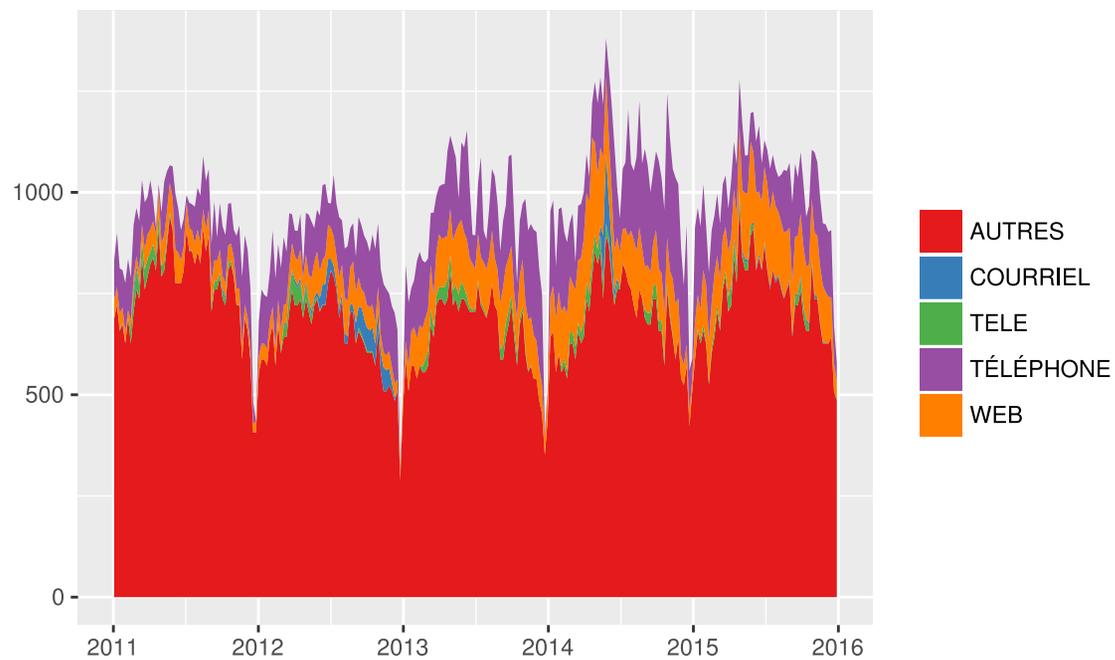
```
library(parallel); library(snow)
```

Graphiques

```
library(ggplot2); library(RColorBrewer)
```

Analyse marketing

Origine des soumissions



Données externes

Alertes d'Environnement Canada

```
library(readr);library(DT);library(lubridate);library(dplyr)
```

```
## # A tibble: 5 × 3
```

```
##   parsedAlert.getIdentifieur parsedAlert.getSender parsedAlert.getSent  
##           <chr>                <chr>                <dtm>  
## 1 2.49.0.1.124.cea07ad0.2014    cap-pac@ec.gc.ca 2014-12-01 19:25:00  
## 2 2.49.0.1.124.cea07ad0.2014    cap-pac@ec.gc.ca 2014-12-01 19:25:00  
## 3 2.49.0.1.124.427451ed.2014    cap-pac@ec.gc.ca 2014-12-01 19:25:00  
## 4 2.49.0.1.124.427451ed.2014    cap-pac@ec.gc.ca 2014-12-01 19:25:00  
## 5 2.49.0.1.124.f4c1c67f.2014    cap-pac@ec.gc.ca 2014-12-01 00:59:00
```

Données spatiales

Données vectorielles (shapefile) et matricielles (raster)

```
library(sp);library(rgdal)
```

```
regions_meteo_shp <- readOGR(dsn = path_shapefile_meteo,  
                             layer="land_CLCBaseZone_detail_proj",  
                             encoding = "latin1",  
                             stringsAsFactors = FALSE) %>%  
  spTransform(CRS("+init=epsg:4326"))
```

```
## OGR data source with driver: ESRI Shapefile
```

```
## Source: "meteo/land_CLCBaseZone_detail_proj/land_CLCBaseZone_detail_proj.shp", l
```

```
## with 1403 features
```

```
## It has 15 fields
```

Fusion de données

Fusion de jeux de données avec composantes spatiales

```
regions_meteo_shp_data_alertes <- regions_meteo_shp@data %>%  
  sp::merge(mes_alertes_subset_min,  
            all.x=TRUE,  
            by.x = "CLCBASEZA",  
            by.y = "geocode.getValue") %>%  
  sp::merge(pif_zone,all.x=TRUE,by="CLCBASEZA")
```

Visualisation

Visualisation interactive de données spatiales

```
library(leaflet)
```

