

## Séminaire de Méthodologie en Data Mining

**2 jours : Mercredi 05 et Jeudi 06 novembre 2008**

Organisé par StatSoft France et animé par

**Dr Diego Kuonen**

Expert en data mining, Statoo Consulting

### Description

Ce séminaire de formation professionnelle propose une couverture méthodologique et pratique des techniques les plus actuelles utilisées en data mining. L'objectif est de permettre l'identification des sources d'informations, intuitives, chiffrées, connues ou complètement inconnues, et leur exploitation maximale. Ce séminaire comprend des illustrations des concepts et des méthodes.

### Approche

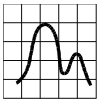
Le séminaire commence par une brève présentation du rôle et des applications du data mining afin d'inciter les entreprises à exploiter les informations contenues dans leurs bases de données (développement de médicaments, portefeuilles de produits, études, comportements de consommateurs, historiques de production, ...).

Ensuite, est dispensé un aperçu global du data mining et des méthodes pour apprendre à partir des données.

À l'issue de ces deux étapes, chaque technique est présentée en détail ainsi que la manière dont elle s'intègre dans l'esprit global du data mining.

À partir de là, dans la partie principale du séminaire, les terminologies, méthodes, ressources et problématiques de data mining vous sont présentées dans une approche méthodologique, résolument indépendante des éditeurs de logiciels. Pour toutes les techniques considérées, la méthodologie de base sera expliquée et agrémentée d'exemples.

En conclusion, ce séminaire sera organisé sous la forme d'un cours théorique couvrant les aspects méthodologiques et les applications pratiques du data mining statistique.



## Concernant l'animateur

Diego Kuonen, PhD en statistiques, est fondateur et directeur général de Statoo Consulting ([www.statoo.ch](http://www.statoo.ch)). Statoo Consulting est une société de conseil, indépendante des éditeurs de logiciels, spécialisée en conseil statistique, formation, analyse de données et data mining.

Le Dr Diego Kuonen intervient depuis de nombreuses années auprès de grands groupes industriels et de services en Europe. Il enseigne également dans diverses Universités en Suisse. Il est actuellement vice-président de la Société Suisse de Statistique, dont il préside également la section Statistique dans les Affaires et l'Industrie.

## Pré-requis

Un pré-requis en statistiques élémentaires est souhaitable, y compris la régression linéaire multiple. Les personnes sans connaissance statistique pourront toutefois suivre l'exposé de l'animateur dans ses présentations techniques.

## Présentation

La présentation sera en français avec des transparents en anglais. Une copie sur papier du support de cours en anglais sera remise lors de ce séminaire.

## Date et horaires

Mercredi 05 nov. de 10 heures à 18 heures ; Jeudi 06 nov. 2008 de 9 heures 30 à 17 heures 30.  
Durée totale : 12 heures.

## Lieu de formation

StatSoft France (Métro 8 – cours des Juilliottes)  
2 rue Louis Pergaud - 94700 – Maisons-alfort– France Tel : (+33)1/45 185 999

## Tarif

2 jours : 1350 € H.T Par participant

Ce tarif comprend l'inscription au cours, une copie sur papier du support de formation en anglais remise à l'issue de celui-ci, le repas de midi et les boissons pendant les pauses. Tarifs de groupe sur demande. Tous les participants recevront un certificat de participation.

## Organisation et inscription

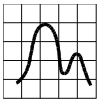
La société StatSoft France est l'organisateur de ce séminaire ; voir bulletin d'inscription ci-joint. Pour toute question relative à ce séminaire, veuillez contacter Samuel Dodé par e-mail [sdode@statsoft.fr](mailto:sdode@statsoft.fr) ou par téléphone au +33(0)1 45 185 132

## Conditions de paiement

Après réception du bulletin d'inscription, StatSoft France vous fera parvenir une convention de stage validant votre engagement à participer à la formation.  
A l'issue de cette formation une facture sera émise avec règlement de cette facture à réception.

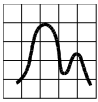
## Condition d'annulation

En cas d'annulation par l'entreprise **30 jours** avant le début du séminaire, aucune facture ne sera émise. En cas d'annulation par l'entreprise entre **30 et 15 jours** avant la date du séminaire, StatSoft France facturera 50 % du montant total du séminaire.  
En cas d'annulation à moins de **15 jours** du séminaire, StatSoft France facturera 100% de la prestation.



## Programme de la formation

- Introduction
- Rôles et applications du data mining
- Qu'est-ce que le data mining ?
  - Le data mining est-il du « déjà vu » statistique ?
  - Pourquoi a-t-on besoin de la statistique ?
  - Qu'est-ce qui distingue le data mining de l'analyse statistique ?
  - Deux approches analytiques du data mining
  - Le processus de découverte de connaissances dans les bases de données
  - Data mining et Business Intelligence
- Le processus de data mining et ses étapes
- Les données et leur préparation
  - Sources de données
  - Pourquoi préparer les données ?
  - Principaux aspects de la préparation des données (p.ex. intégration, nettoyage, transformation, réduction et discrétisation)
- Tâches et techniques de data mining
- Description et visualisation
- Caractérisation des données multivariées
- Dissimilarités et mesures de distance
- Les méthodes non-supervisées
  - Analyse en composantes principales
  - Analyse de proximité
  - Clustering (p.ex. algorithmes hiérarchiques, algorithmes de segmentation, mise en pratique)
  - Cartes auto-organisatrices de Kohonen
  - Groupes d'affinités et règles d'associations
  - Pour aller plus loin...
- Les méthodes supervisées
  - Introduction (p.ex. biais induit et complexité de modèle, fonctions de score, validation interne, validation externe)
  - Modélisation par classification (p.ex. analyse discriminante, Support Vector Machines, classification par plus proches voisins, classification Bayes-naïf).
  - Modélisation par régression (p.ex. modèles de régression linéaire multiple, modèles linéaires généralisés, modèles de régression non-paramétriques, modèles additifs généralisés)
  - Réseaux de neurones
  - Arbres de décision (p.ex. CART, C4.5 et C5.0, CHAID)
  - Ensemble d'apprentissage (p.ex. bagging, forêts aléatoires, boosting)
  - Le problème de surdimensionnalité (p.ex. extraction de prédicteurs, sélection de prédicteurs : filtres, wrappers, méthodes internes)
  - Evaluation comparative des méthodes de classification
  - Comparaison des méthodes de régression
  - Pour aller plus loin...
  - Expériences récentes – qu'a-t-on appris ?
- Data mining : mythes et réalités
- Critères de réussite en data mining
- Conclusion
- Références et ressources



## BULLETIN D'INSCRIPTION

### Séminaire de Méthodologie en Data Mining

Mercredi 05 et Jeudi 06 novembre 2008

Organisé par StatSoft France  
Animé par Dr Diego Kuonen, Statoo Consulting, Suisse

Merci de renvoyer par fax au +33 (0)1 45 185 285 ou par courrier électronique à [sdode@statsoft.fr](mailto:sdode@statsoft.fr).  
A retourner avant le vendredi 27 Juin 2008

Nom	
Prénom	
Société	
Département / Service	
Adresse	
Code postal	
Ville	
Pays	
Téléphone	
Email	
Date et Signature	
Nom et coordonnées téléphoniques du responsable formation (pour l'envoi de la convention)	

Tarif : 1350 € H.T. pour 2 jours.

Horaires : Mercredi 05 nov. de 10 heures à 18 heures ; Jeudi 06 nov. 2008 de 9 heures 30 à 17 heures 30.