



Maisons-Alfort, le mercredi 26 octobre 2005

## Séminaire de méthodologie en data mining statistique

**3 jours : du lundi 6 au mercredi 8 mars 2006 de 9 heures 30 à 17 heures 30**

Organisé par StatSoft France et animé par Dr Diego Kuonen, expert en data mining

Madame, Monsieur,

On parle aujourd'hui de l'utilisation des techniques de data mining comme d'une approche majeure pour l'amélioration des processus en entreprise et en recherche. Qu'en est-il réellement dans l'atteinte de l'excellence opérationnelle ? Est-ce que cela « vaut le coup » de s'impliquer dans cette démarche ou est-ce du « déjà vu » statistique ?

Ce **séminaire de trois jours** propose une couverture méthodologique et pratique des techniques de data mining et met en lumière les applications pour l'analyse de volumétries importantes de données historiques. L'approche pédagogique privilégiée pour ce séminaire consiste à passer de la théorie aux applications et vice-versa : Ainsi, des exemples tangibles sont étudiés pour déboucher sur la théorie et des notions conceptuelles sont expliquées pour montrer les applications dans des cas concrets. Globalement, l'équivalent d'une journée sera consacrée au suivi complet des démarches du data mining au moyen d'exercices appliqués permettant aux participants de mettre en place et de parfaire leur pratique des méthodes et concepts présentés au cours.

Un large éventail d'outils est étudié en détail. L'accent est mis sur la manière dont ces outils doivent être utilisés dans l'esprit global du data mining.

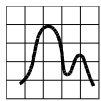
Le Dr Diego Kuonen travaille depuis de nombreuses années auprès de grandes entreprises et d'universités. Il dirige Statoo Consulting, une société de conseil en data mining en Suisse ([www.statoo.ch](http://www.statoo.ch)).

Ce séminaire est un « must » pour les personnes intriguées par les outils d'aide à la décision, par la modélisation et l'automatisation. Vous trouverez ci-joint un descriptif détaillé et un formulaire d'inscription.

N'hésitez pas à nous contacter si vous avez des questions.

Dans l'attente de pouvoir discuter avec vous de vos problématiques d'analyse de données.

Marie-Pierre CARARO  
[mcararo@statsoft.com](mailto:mcararo@statsoft.com)  
Tel direct : +33 (0)1-45-185-406



## Séminaire de méthodologie en data mining statistique

**3 jours : du lundi 6 au mercredi 8 mars 2006**

organisé par StatSoft France et animé par

**Dr Diego Kuonen**

Expert en data mining, Statoo Consulting, Suisse

### Description

Ce séminaire de formation professionnelle propose une couverture méthodologique et pratique des techniques utilisées en data mining. L'objectif consiste à identifier les informations intuitivement connues ou complètement inconnues contenues dans des sources volumineuses d'informations. Des cas concrets permettront aux participants de parfaire leur pratique des concepts et méthodes de data mining illustrées durant le cours.

### Approche

Vous passerez alternativement de la pratique à la théorie – de « l'esprit global » au détail – et de la compréhension conceptuelle à la capacité d'aborder les tâches spécifiques au data mining statistique.

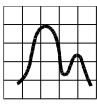
Le cours commence par une brève présentation du rôle et des applications du data mining afin d'inciter les entreprises et les centres de recherche à exploiter les informations contenues dans leurs entrepôts de données (développement de médicaments, portefeuilles de produits, études, comportements de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

Ensuite, est dispensé un aperçu global du data mining et de l'art d'apprendre à partir des données.

À l'issue de ces deux étapes, des techniques sont présentées en détail. L'animateur explique comment elles s'intègrent dans l'esprit global du data mining.

Ainsi, dans la partie principale du séminaire, les terminologies, méthodes, ressources et problématiques de data mining vous sont présentées par un spécialiste indépendant des éditeurs de logiciels. Pour toutes les techniques considérées, la méthodologie de base sera expliquée et agrémentée d'exemples. Au final, ce cours vous permettra une mise en application des méthodologies présentées, ainsi que de leurs philosophies sous-jacentes, à vos propres données ou à des données génériques.

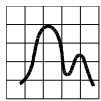
En conclusion, ce séminaire de trois jours se partage en périodes de cours couvrant respectivement les aspects méthodologiques et pratiques du data mining statistique, indépendamment de tout logiciel, et des exercices où vous pourrez, par vous-même, appliquer les méthodes vues au cours à vos propres données ou à des données génériques.



## Séminaire de méthodologie en data mining statistique

### Contenu de la partie méthodologique

- Introduction
- Rôle 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
- Un 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 des données)
- Tâches et techniques de data mining
- Description et visualisation
- Caractérisation des données multivariées
- Dissimilarités et mesures de distance
- Méthodes non-supervisées
  - Analyse en Composantes Principales
  - Analyse de Proximité
  - Clustering (p.ex. algorithmes hiérarchiques, algorithmes de segmentation, mise en pratique du clustering)
  - Cartes Auto-Organisatrices de Kohonen
  - Groupes d'affinités et règles d'associations
  - Pour aller plus loin...
- Méthodes supervisées
  - Introduction (p.ex. biais induit et complexité de modèle, fonction de score, validation interne, validation externe)
  - Modélisation par classification (p.ex. analyse discriminante, Support Vector Machines, classification par les 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, subbagging, 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
  - Pour aller plus loin...
- Le mythe du data mining face aux réalités
- Conclusion
- Références et ressources



## Concernant l'animateur

Diego Kuonen, PhD en statistique, est fondateur et directeur général de Statoo Consulting, Suisse ([www.statoo.ch](http://www.statoo.ch)). Statoo Consulting est une société de conseil suisse, indépendante des éditeurs de logiciels, spécialisée en conseil statistique, formation, analyse de données, data mining, Gestion analytique de la Relation Client et services de bioinformatique.

Le Dr Diego Kuonen a de nombreuses années d'expériences en enseignement et formation professionnelle. Il enseigne dans les Ecoles Polytechniques Fédérales de Lausanne et Zurich et dans d'autres universités suisses. 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, en particulier en régression linéaire multiple. Les personnes sans connaissances statistiques pourront toutefois suivre l'exposé de l'animateur dans ses présentations techniques.

Un ordinateur portable (MS-Windows XP/2000) muni du logiciel *STATISTICA Data Miner* sous licence-cours de 30 jours est indispensable. StatSoft France fournira, si nécessaire, la licence avant le début du cours.

## Présentation

La présentation sera en français avec des **transparents en anglais**. Une copie-papier des notes de cours en anglais vous sera remise lors de ce séminaire.

## Date et horaires

Lundi 6 au mercredi 8 mars 2006, de 9h30 à 17h30.

## Lieu de formation

Paris-Est (Près de la Gare de Lyon)

## Tarif

Par participant

- 3 jours: 1800 € H.T

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

## Inscription

Voir bulletin d'inscription ci-joint.

Un acompte de 30% est demandé à l'inscription.

## Information

Pour toute question relative à ce séminaire, veuillez contacter Marie-Pierre CARARO [mcararo@statsoft.com](mailto:mcararo@statsoft.com) au +33 (0)1-45-185-406.



## SÉMINAIRE DE MÉTHODOLOGIE EN DATA MINING STATISTIQUE *Animé par Dr Diego Kuonen, expert en data mining*

**3 jours** : du lundi 6 au mercredi 8 mars 2006  
de 9 heures 30 à 17 heures 30

**Lieu** : Paris-Est (près de la gare de Lyon)

*Comment les techniques de data mining peuvent-elles permettre d'atteindre l'excellence opérationnelle ? Est-ce que le data mining permet de gagner en productivité ou est-ce du « déjà vu » ?*

*Ce séminaire a pour objectifs :*

- *fournir une présentation détaillée, méthodologique et pratique des techniques de data mining met en lumière les applications concrètes pour la fouille d'entrepôts de données ou de sources d'informations volumineuses.*
- *comprend l'étude d'exemples concrets pour déboucher sur la théorie.*
- *explique des notions conceptuelles pour déboucher sur les applications à des tâches spécifiques de data mining statistique.*
- *mettre en place et parfaire par des exercices appliqués la pratique des méthodes et concepts présentés au cours.*

*Un large éventail d'outils est étudié en détail. L'accent est mis sur la manière dont ces outils doivent être utilisés dans l'esprit global du data mining.*

---

Tarif	3 jours	1800 € H.T.	(durée : 18 heures)
-------	---------	-------------	---------------------

---

### Conditions de paiement

Un acompte de 30% est demandé à l'inscription (soit un chèque de 645,84 € T.T.C correspondant à 540 €H.T.) Toute annulation reçue par écrit 30 jours avant le début du séminaire sera remboursée. Les annulations reçues 15 jours avant la date du séminaire seront remboursées à hauteur de 50%. Aucun remboursement ne sera fait pour des annulations à moins de 15 jours avant la date du cours. Le solde du paiement (1260 € H.T. soit un chèque de 1506,96 € T.T.C) sera demandé sur place aux participants.

### Contact

Pour toute question relative à ce séminaire, veuillez contacter Marie-Pierre CARARO par mcararo@statsoft.com ou au +33 (0)1-45-185-406.



Bulletin d'inscription

## **Séminaire de méthodologie en data mining statistique**

**6 au 8 mars 2006**

Organisé par StatSoft France

et 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 [mcararo@statsoft.com](mailto:mcararo@statsoft.com)

A retourner avant le vendredi 3 février 2005

Nom	
Prénom	
Société	
Département/Service	
Adresse	
Code postal	
Ville	
Pays	
Téléphone	
Fax	
Email	
Date et Signature	
Commentaires	

Tarif : 1800 € H.T. (3 jours; 18 heures). Voir modalités de paiement ci-jointes.