

Maisons-Alfort, France, le jeudi 13 juillet 2006

## Séminaire de méthodologie en data mining statistique

**2 jours : mardi 10 et mercredi 11 octobre 2006 – Grand Hôtel Château Perrache, Lyon**

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. Mais qu'en est-il réellement dans l'atteinte de l'excellence opérationnelle ? Est-il valable de s'impliquer dans une telle démarche ou n'est-ce finalement que du « déjà vu » statistique ?

- La détection de facteurs influents dans l'amélioration des procédés de fabrication,
- la compréhension des comportements d'achat et la mise en œuvre de campagnes ciblées en marketing,
- la modélisation de données historiques et l'automatisation de processus d'alerte

sont autant d'exemples des moyens qu'offrent les techniques de data mining pour gagner en efficacité et en productivité. Cela-dit, est-il réellement rentable et utile de s'engager dans une telle démarche ? Comment qualifier la valeur ajoutée de ces techniques pour une entreprise ? Comment structurer une réflexion pour la mise en place de projets intégrant l'utilisation des techniques de data mining ?

Ce **séminaire de deux jours** propose une couverture méthodologique et pratique des techniques de data mining. L'approche pédagogique privilégiée pour ce séminaire consiste à passer de la théorie à des applications concrètes et vice-versa. Un large éventail d'outils est étudié en détail afin d'éviter le syndrome de boîte-noire si souvent perçu par les utilisateurs de techniques de data mining dont la mise en place aveugle a été facilitée par l'essor de d'outils informatiques très performants ces dernières années. L'accent est mis sur la manière dont ces outils doivent être utilisés dans l'esprit global du data mining.

Ce séminaire est le préalable à la mise en place d'un projet de data mining, l'étape suivante étant la pratique des techniques de data mining sur des cas concrets, puis l'accompagnement sur des cas spécifiques à l'entreprise.

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

Ce séminaire est destiné aux décideurs et professionnels intrigués par le potentiel des techniques de data mining (directeurs et responsables marketing, qualité, méthodes, études, stratégie, finance, ...), souhaitant structurer leur réflexion pour évaluer le bien-fondé de ces techniques dans le cadre de leurs problématiques.

Vous trouverez ci-joint un descriptif détaillé et un formulaire d'inscription auprès de la société organisatrice : StatSoft France, à retourner avant le vendredi 8 septembre 2006.

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

Dans l'attente de discuter des enjeux du data mining pour votre organisation, je vous prie d'agréer, Madame, Monsieur, l'expression de mes salutations distinguées.



Dr Diego Kuonen  
CEO  
Statoo Consulting

# Séminaire de méthodologie en data mining statistique

**2 jours : mardi 10 et mercredi 11 octobre 2006**

Organisé par StatSoft France et animé par

**Dr Diego Kuonen**

Expert en méthodologie et en data mining, Statoo Consulting, Suisse

## 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

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 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.

## Séminaire de méthodologie en data mining statistique

### Contenu

- 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)
- 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)
  - 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, 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

## 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, formation professionnelle et conseil en entreprise. 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. Le Dr Diego Kuonen est intervenu auprès de grands groupes industriels et de services en Europe.

## Pré-requis

Un pré-requis en statistiques élémentaires est souhaitable, y compris la régression linéaire multiple. Les personnes sans connaissances statistiques 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 vous sera remise lors de ce séminaire.

## Date et horaires

Mardi 10 octobre de 10 heures à 18 heures.  
Mercredi 11 octobre de 9 heures 30 à 17 heures 30.  
Durée totale : 12 heures.

## Lieu de formation

Lyon – Grand Hôtel Château Perrache.

## Tarif

Par participant

- 2 jours: 1350 € H.T.

Ce tarif comprend l'inscription au cours, une copie sur papier du support de formation 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.

## Organisation et inscription

La société StatSoft France est l'organisateur de ce séminaire ; voir bulletin d'inscription ci-joint. Un acompte de 30%, à l'ordre de StatSoft France, est demandé à l'inscription.

## Information

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

## SEMINAIRE DE METHODOLOGIE EN DATA MINING STATISTIQUE

*Animé par Dr Diego Kuonen, expert en data mining*

**2 jours** : mardi 10 et mercredi 11 octobre 2006  
(Durée : 12 heures)

Mardi 10 octobre de 10 heures à 18 heures  
Mercredi 11 octobre de 9 heures 30 à 17 heures 30

**Lieu** : Lyon – Grand Hôtel Château Perrache

*Comment les techniques de data mining permettent-elles de gagner en efficacité et en rentabilité? Le data mining apporte-t-il une réelle valeur ajoutée à l'organisation ou est-ce du « déjà vu » ? Comment structurer une démarche d'évaluation de l'intérêt économique de ces techniques dans l'organisation ?*

*Ce séminaire a pour objectifs de :*

- *fournir une présentation détaillée, méthodologique et pratique des techniques de data mining ;*
- *mettre en lumière les applications concrètes pour l'identification de sources d'informations et l'exploitation des données ;*
- *expliquer des notions conceptuelles pour motiver les applications spécifiques de data mining statistique.*

*Un large éventail d'outils est étudié en détail. L'accent est mis sur la compréhension et 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 en Europe.*

---

<b>Tarif du séminaire de méthodologie</b>	<b>2 jours</b>	<b>1350 € H.T. (durée : 12 heures)</b>
---	----------------	--

---

### Conditions de paiement

Un acompte de 30% est demandé à l'inscription (soit un chèque de 484.38 € T.T.C correspondant à 405 €H.T.). Chèque à l'ordre de la société organisatrice, prestataire de formation : StatSoft France, 31 cours des Juilliottes, 94700 Maisons-Alfort, France. Les conventions de formation seront envoyées à réception du bulletin d'inscription. RCS CRETEIL : B410 406 565 (97B00367) ; TVA INTRA : FR65 410 406 565 ; SIREN : 410 406 565 – Tel : +33(0)1 45 185 999.

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 (soit un chèque de 1130.22 € T.T.C correspondant à 945 € H.T.) sera demandé à réception de facture.

### Contact

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

Bulletin d'inscription

**Séminaire de méthodologie en data mining statistique**

**Du mardi 10 au mercredi 11 octobre 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.  
A retourner avant le vendredi 8 septembre 2006

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