



STATISTICA™ Data Miner

POURQUOI CHOISIR STATISTICA DATA MINER ?

Une solution puissante et personnalisable :

- Évolutive avec des algorithmes de pointe pour les arbres de décision, boosting, réseaux de neurones et de nombreux autres modèles prédictifs
- Outils de modélisation analytique classiques, notamment par régression logistique et modèles linéaires
- Compatibilité avec les standards de l'industrie :
 - Se connecte à toutes les bases de données et formats de fichiers
 - Big Data
 - Hadoop Distributed File System (HDFS™)
- Interface de programmation complète en API :
 - S'intègre dans votre architecture informatique existante
 - Planificateurs de tâches

Outils et solutions reconnus et appréciés par les experts, les utilisateurs expérimentés comme les débutants :

- Le meilleur indice de satisfaction des utilisateurs dans les grandes enquêtes indépendantes réalisées au cours de ces 3 dernières années ⁽¹⁾
- Double victoire au " Predictive Analytics Industry Quadrant " ⁽²⁾
- Choix des journalistes spécialisés : STATISTICA offre un rapport tarif/performances supérieur ⁽³⁾
- Utilisé par des centaines de milliers d'utilisateurs à travers le monde ; un réseau de 30 filiales réparties sur tous les continents pour accompagner vos projets

STATISTICA Data Miner est la plate-forme analytique avec le meilleur ROI, donnant accès à tous les outils d'analyse prédictive et de data mining.

- Gestion et préparation des données
- Modélisation et gestion des modèles, contrôle et historique des versions
- Déploiement dans des bases de données ou sous forme de code ; ex : Java, C#
- Solutions très performantes de scoring en temps réel et/ou par batch

STATISTICA Data Miner offre la gamme la plus riche du marché en termes de solutions de data mining, avec une interface-utilisateur extrêmement simple à utiliser, basée sur des icônes. Intègre une série de solutions prédictives automatisées et prêtes au déploiement pour une grande diversité d'applications-métier.



STATISTICA Data Miner est ultra-personnalisable. Il peut s'adapter à des spécifications fonctionnelles et techniques exigeantes grâce à son architecture ouverte. Également disponible, STATISTICA Data Miner Plus, un outil en français, guidant l'utilisateur pas-à-pas pour créer les modèles et processus analytiques désirés. Grâce aux configurations par défaut, vous pouvez élaborer des modèles en quelques clics.

STATISTICA jouit d'une reconnaissance inégalée depuis sa première version en 1993, et a obtenu les meilleurs scores dans TOUTES les études comparatives qui l'ont cité.

SPÉCIFICITÉS FONCTIONNELLES

Solution complète

STATISTICA Data Miner accroît vos chances d'améliorer la compréhension des phénomènes étudiés grâce à plus de 14 000 fonctions analytiques, graphiques, d'ETL/gestion des données et la gamme la plus large d'algorithmes de data mining et d'intelligence artificielle du marché. Il offre notamment :

- Sélection et Filtrage Meilleurs Prédicteurs
- Réseaux de Neurones Automatisés
- Analyse de Données Séquentielles, Règles d'Association et Analyse de Liens
- Arbres de Classification et de Régression (GTrees)
- Modèles Linéaires Généralisés
- EM, k-Means, Clustering
- Arbres CHAID, C&RT
- Modèles Additifs Généralisés
- Analyse en Composantes Indépendantes
- Régression PLS (PLS, NIPALS, ACP)
- Modélisation Multivariée (notamment MARSplines)
- K plus Proches Voisins
- Forêts Aléatoires (Bagging)
- Boosting d'Arbres de Décision
- Séparateurs à Vaste Marge (SVM - Support Vector Machines)
- Outils Avancés de Validation et de Vérification des Données
- Remplacement Automatique des Valeurs Manquantes
- Simulation Univariée et Multivariée de Données Normales ou Non-Normales
- Prédiction Inverse, Stratégies d'Optimisation Généraliste de Modèles Prédicatifs

Data Miner Plus

Cet Assistant convient parfaitement aux débutants ou pour enregistrer des processus analytiques aisément réutilisables.

Automatisation

Grâce à la plate-forme Entreprise, les processus analytiques du *Data Miner* peuvent devenir des modèles analytiques ultra-personnalisés que vous partagez avec tous les collaborateurs habilités, afin d'offrir des modèles prédictifs, des outils automatisés de calibration de modèles, ou une bibliothèque complète de procédures analytiques, exécutables d'un seul clic.

Gestion des Modèles

La plate-forme Entreprise offre tous les outils de gestion intégrée des modèles, le contrôle des versions et le déploiement rapide des modèles, depuis l'environnement de développement jusqu'à celui de production, sans nécessiter aucune programmation.

Règles Intégrées

La Plate-Forme Décisionnelle de *STATISTICA* (une extension facultative de la plate-forme Entreprise) est totalement intégrée aux outils de gestion des modèles, pour réaliser du scoring par batch ou en temps réel sur des problématiques de segmentation (pré-scoring) avec des règles-métier, des règles conditionnelles ou reposant uniquement sur des règles.

Scoring par Batch et Temps Réel

Les modèles et les processus analytiques générés par *STATISTICA Data Miner* peuvent être déployés pour du scoring par batch au travers de la plate-forme Entreprise ou en temps réel au travers de la plate-forme Live Score®

Personnalisation

STATISTICA Data Miner s'intègre parfaitement dans votre architecture existante ; la personnalisation de toutes vos analyses et interfaces s'effectue par les méthodes conventionnelles à la souris (en glisser-déplacer) ou à l'aide du langage de programmation universel en Visual Basic. Construisez rapidement et simplement des outils spécifiques pour les utilisateurs individuels ou vos applications.

Outil Complet de Gestion des Données

La plate-forme *STATISTICA* offre de nombreuses procédures de nettoyage et préparation des données, comme la suppression, le remplacement ou l'interpolation des valeurs manquantes, le filtrage des données, la détection et la suppression des valeurs aberrantes, le regroupement optimal des modalités, des méthodes de sélection et d'extraction automatique des meilleurs prédicteurs, la transformation par lot en utilisant des fonctions-date, des fonction de décalage, des opérateurs logiques et des fonctions d'agrégation statistique, ainsi que de nombreuses autres fonctionnalités pour vous aider dans la phase critique de préparation des données. Tous ces outils sont optimisés pour traiter de très grosses volumétries de données, et comptent parmi les plus rapides du marché.

Data Mining Visuel

STATISTICA Data Miner est réputé pour ses fonctionnalités avancées de représentation graphique (ex : Computerwoche 2011) ; partie intégrante de la plate-forme *STATISTICA*, solution la plus riche du marché pour l'exploration graphique des données et le data mining visuel, il permet d'accélérer la phase critique de préparation des données (identification des valeurs aberrantes, intervalle de validité des données, mise en évidence de phénomènes structurels).

Performance et Technologie

STATISTICA Data Miner intègre des algorithmes de pointe, une architecture multi-threading et multitâches (Client-Serveur) ultra-optimisée, une version native pour les plates-formes 64 bits, une interface directe avec les bases de données, la possibilité d'exécuter l'application en client léger et une parfaite intégration avec vos bases de données et autres systèmes de fichiers.

Options de Déploiement Flexibles, STATISTICA Live Score

Déploiement des modèles prédictifs en XML/PMML, en code C/C++, Java ou SAS, et sous forme de procédures stockées. *STATISTICA Data Miner* s'intègre parfaitement avec la plate-forme *STATISTICA Entreprise* pour un déploiement et une mise à jour efficaces et centralisés des modèles prédictifs dans toute l'entreprise ; les modèles peuvent également être envoyés au serveur *STATISTICA Live Score*, une solution client-serveur de scoring automatique en temps réel ayant été déployée avec succès dans certaines des plus grandes banques au monde.

Compatibilité et Architecture Ouverte

STATISTICA Data Miner est compatible avec tous les formats de fichiers courants ainsi qu'avec les protocoles et conventions ODBC, OLE DB et SQL ; *STATISTICA* est certifié SAP et propose des connecteurs pour certaines bases de données spécifiques (OSI PI) ; aucun format propriétaire de base de données/data warehouse, ni de langage de programmation/environnement à apprendre ; toutes les fonctionnalités sont exposées au modèle-objet (et au protocole SOAP dans la version client-serveur), et vous pouvez aisément intégrer des procédures tierces (ex : R, SAS).

Convivialité

STATISTICA Data Miner (l'indice de satisfaction le plus élevé dans les grandes enquêtes indépendantes réalisées auprès des utilisateurs au cours de ces 3 dernières années⁽¹⁾) offre les méthodes les plus avancées, avec des interfaces-utilisateurs simples et intuitives ; l'interface du *Data Miner Plus* se présente sous forme d'un Assistant qui vous guide pas-à-pas pour créer et automatiser efficacement vos processus analytiques ; créez des représentations graphiques de qualité en quelques clics ; tous les éléments de l'interface sont personnalisables à l'aide d'outils classiques ; vous pouvez analyser de grosses volumétries de données, même depuis un ordinateur portable.

Également disponible :



La Plate-Forme Décisionnelle de *STATISTICA* (une extension facultative de la plate-forme Entreprise) est totalement

intégrée aux outils de gestion des modèles, pour réaliser du scoring par batch ou en temps réel sur des problématiques de segmentation (pré-scoring) avec des règles-métier, des règles conditionnelles ou reposant uniquement sur des règles. Adapté à un déploiement des modèles prédictifs à grande échelle dans les secteurs réglementés, banques et assurances.



STATISTICA Text Miner (une extension facultative de *STATISTICA Data Miner*), offre une large gamme de procédures

d'extraction de texte, de préparation, d'analyse et d'exploration de données textuelles non-structurées (notamment des pages Web), avec de nombreuses options pour convertir ces données textuelles non-structurées en données numériques (typologies, segmentation, data mining prédictif) ; algorithmes de lemmatisation en français et d'autres langues ; possibilité d'analyser d'autres types d'informations non-structurées comme des images ou des sons.



STATISTICA Optimisation de Processus est une solution puissante qui intègre l'ensemble des cartes de contrôle qualité,

procédures d'analyse de capacité du processus, plans d'expériences, et méthodes Six Sigma avec une bibliothèque complète de techniques modernes et performantes de data mining exploratoire/prédictif ainsi que d'optimisation (des modèles de data mining). Propose différents outils généralistes d'optimisation de modèles prédictifs (par exemple, optimisation de campagne marketing) par Simplex, GA et/ou optimisation de grille



STATISTICA Analyse de Données Séquentielles, Règles d'Association et Analyse de Liens (SAL) est une application

complète permettant d'extraire des règles d'association à partir de jeux de données complexes.

¹ Enquêtes Rexer Analytics 2009, 2010 et 2011 sur le Data Mining

² Index des Vainqueurs décerné par Hurwitz dans le domaine de l'analyse prédictive

³ Computerwoche, Mars 2011



StatSoft France: 2, rue Louis Pergaud • 94700 Maison-Alfort • France • +33(0)1-45-185-999 • info@statsoft.fr • www.statsoft.fr

Afrique du Sud : StatSoft S. Africa (Pty) Ltd.
Allemagne : StatSoft GmbH
Australie : StatSoft Pacific Pty Ltd.
Brésil : StatSoft South America
Bulgarie : StatSoft Bulgaria Ltd.

Chili : StatSoft South America
Chine : StatSoft China
Égypte : StatSoft Middle East
Espagne : StatSoft Iberica Lda
France : StatSoft France

Hongrie : StatSoft Hungary Ltd.
Inde : StatSoft India Pvt. Ltd.
Israël : StatSoft Israel Ltd.
Italie : StatSoft Italia srl
Japon : StatSoft Japan Inc.

Norvège : StatSoft Norway AS
Pologne : StatSoft Polska Sp. z o.o.
Pays-Bas : StatSoft Benelux BV
Portugal : StatSoft Iberica LDA
Rép. Tchèque : StatSoft Czech Rep. s.r.o.

Royaume-Uni : StatSoft Ltd.
Russie : StatSoft Russia
Suède : StatSoft Scandinavia AB
Taiwan : StatSoft Taiwan
USA : StatSoft, Inc.