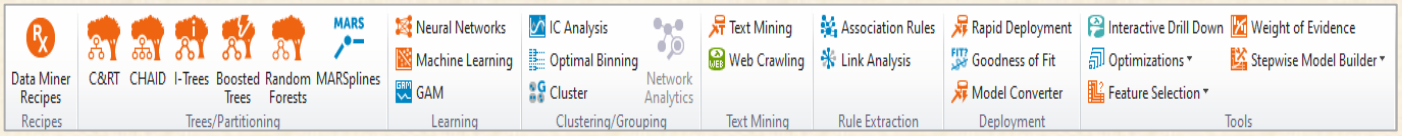
MTH8302 - Analyse de régression et analyse de variance

Chapitre 7 – Data Mining = Machine Learning

Chapitre 6 : Multivariate Adaptative Regression Splines (MARS)
Chapitre 7 : Introduction au Data Mining
Chapitre 8 : Classification and Regression Trees
Chapitre 9 : Random Forest (RF)
Chapitre 10 : Artificial Neural Networks (ANN)

Chapitre 7 — Introduction

STATISTICA data mining



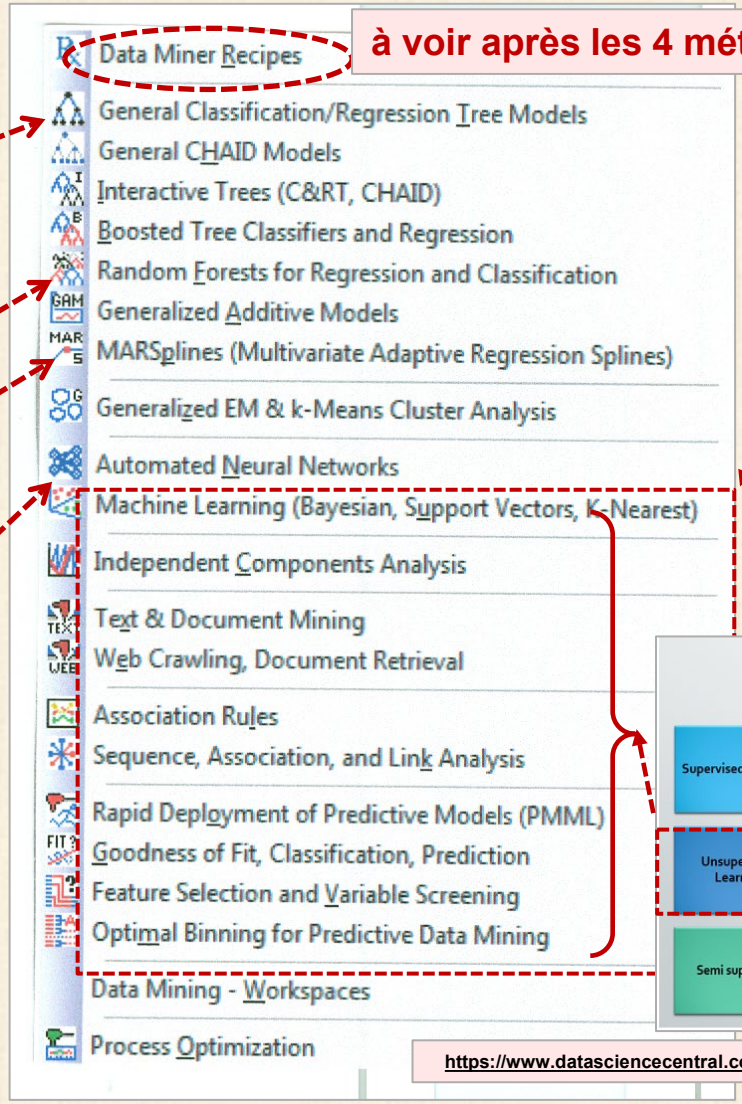
vues dans ces notes

CRT
Classification Regression Tree

RF Random Forest

MARS Multivariate Adaptive Regression Splines

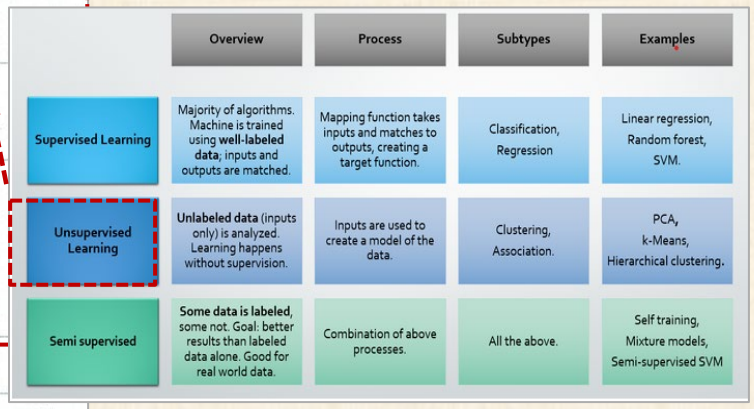
ANN Artificial Neural Networork



à voir après les 4 méthodes : Data Miner Recipes

non vues car
apprentissage non supervisé

MTH8302 = apprentissage supervisé



https://www.datasciencecentral.com/profiles/blogs/supervised-learning-vs-unsupervised-in-one-picture

DÉFINITIONS du DATA MINING

Data Mining = Machine Learning

fouille de données, extraction de connaissances
KDD = Knowledge Data Discovery

nouveau champ d'application à l'interface de la statistique et des technologies de l'information
(bases de données, intelligence artificielle, apprentissage ,etc.)

U.M.Fayyad, G.Piatetski-Shapiro

« the nontrivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data »

D.J.Hand « the discovery of interesting, unexpected, or valuable structures in large data sets »

Statistique, Science des données, Intelligence artificielle : liens, différences, convergence

Bernard CLÉMENT, PhD

Société Statistique de Montréal : 26 avril 2018

https://cours.polymtl.ca/mth6301/mth8302-Cours&Plus/Clement/Clement-Statistique_ScienceDonnees_IntelligenceArtificielle.pdf

DÉFINITIONS du DATA MINING

- Exploration d'une grande quantité de données
- (centaines de variables/milliers d'observations) en vue de rechercher des modèles relationnels entre des variables et ensuite de valider ces modèles en les appliquant sur de nouvelles données.
- Art et la science d'obtenir de la connaissance à partir des données
- **O B J E C T I F S**
  - identifier des structures, groupes, clusters, strates, ou dimensions dans les données qui ne semblent pas avoir de structures évidentes non supervisé
  - identifier des facteurs qui sont reliés à un résultat d'intérêt
  - (recherche d'un système de causes) non supervisé
  - prédire des variables d'intérêt (variables de réponse): nouveaux clients, nouveaux appliquants, etc

    data mining prédictif ou supervisé (var output Y)

Processus de «torture des données» jusqu'à la «confession»

Data Mining = forage (fouille) des données

Introduction to Data Mining

Knowledge Discovery vs. Statistical Analysis

StatSoft STATISTICA

- Statistical Analysis
  - Focuses on "hypothesis testing" and "parameter estimation"
  - Fits "parsimonious statistical models" with the goal to "explain" complex relationships with fewer parameters
  - Examples: Regression, nonparametric statistics, factor analysis, traditional quality control
- Data Mining
  - Focuses on knowledge discovery, detection of patterns, clusters, and so on; we only have data and no (or few) expectations and hypotheses
  - Fits simple models (such as regression) or complex models (such as neural nets) to enable valid prediction
  - Examples: Neural nets, stochastic gradient boosting of tree classifiers, random forests, support vector machines

LES DONNÉES SONT PARTOUT ! : on est dans l'air du BIG DATA

- Base données relationnelles — commodité de toute entreprise
- Construction d'immense entrepôt de données (data warehouses)
- Base de données transactionelles : point de vente (Point Of Sale)
- Base de données orientées objet, relationnelles, distribuées, hétérogènes et historiques
- Base de données spatiales (GIS), remote sensing
- Base de données scientifiques / ingénierie
- Données temporelles (e.g., transactions boursières)
- Text (documents, emails), base de données multimedia
- WEB: immense, hyperliens, dynamique, système d'information global

Web Mining

- Miner ce que les engins de recherche trouvent
- Classification automatique des documents Web
- Découvertes de pages Web de référence autoritaire
- Analyse des structures Web et réseaux

QUELQUES ÉLÉMENTS DISTINCTIFS

- La métaphore du Data Mining signifie qu'il y a des trésors ou pépites cachées sous des montagnes de données que l'on peut découvrir avec des outils spécialisés.
- Le Data Mining analyse des **données recueillies à d'autres fins**: c'est une analyse secondaire de bases de données, souvent conçues pour la gestion de données individuelles.
- Le Data Mining ne se préoccupe pas de collecter des données de manière efficiente et efficace (sondages, plans d'expériences).
- Recherche de modèles ou patterns (comportements)
  exemple : - niche de consommateurs à forte valeur
  - consommateurs à haut risque (domaine bancaire)
- Prédictif (supervisé) ou exploratoire (non supervisé)
- Pas d'estimation / tests mais découverte à l'aide d'algorithmes:
  arbre de décision, réseaux de neurones, SVM, réseaux bayesiens, classification, cartes de Kohonen, règles d'association, etc

QUELQUES ÉLÉMENTS DISTINCTIFS

- **L'inférence statistique classique ne fonctionne plus pour les très grands ensembles de données: toute hypothèse nulle est rejetée.**
- **Il faut remplacer les tests de signification par de la validation croisée : on testera si une structure reste valable dans une autre partie des données que celle qui a été explorée pour la définir. Critères définis plus loin.**
- **Les structures sont-elles valides?**
- **Vérifier l'utilité de ce que l'on découvre:**

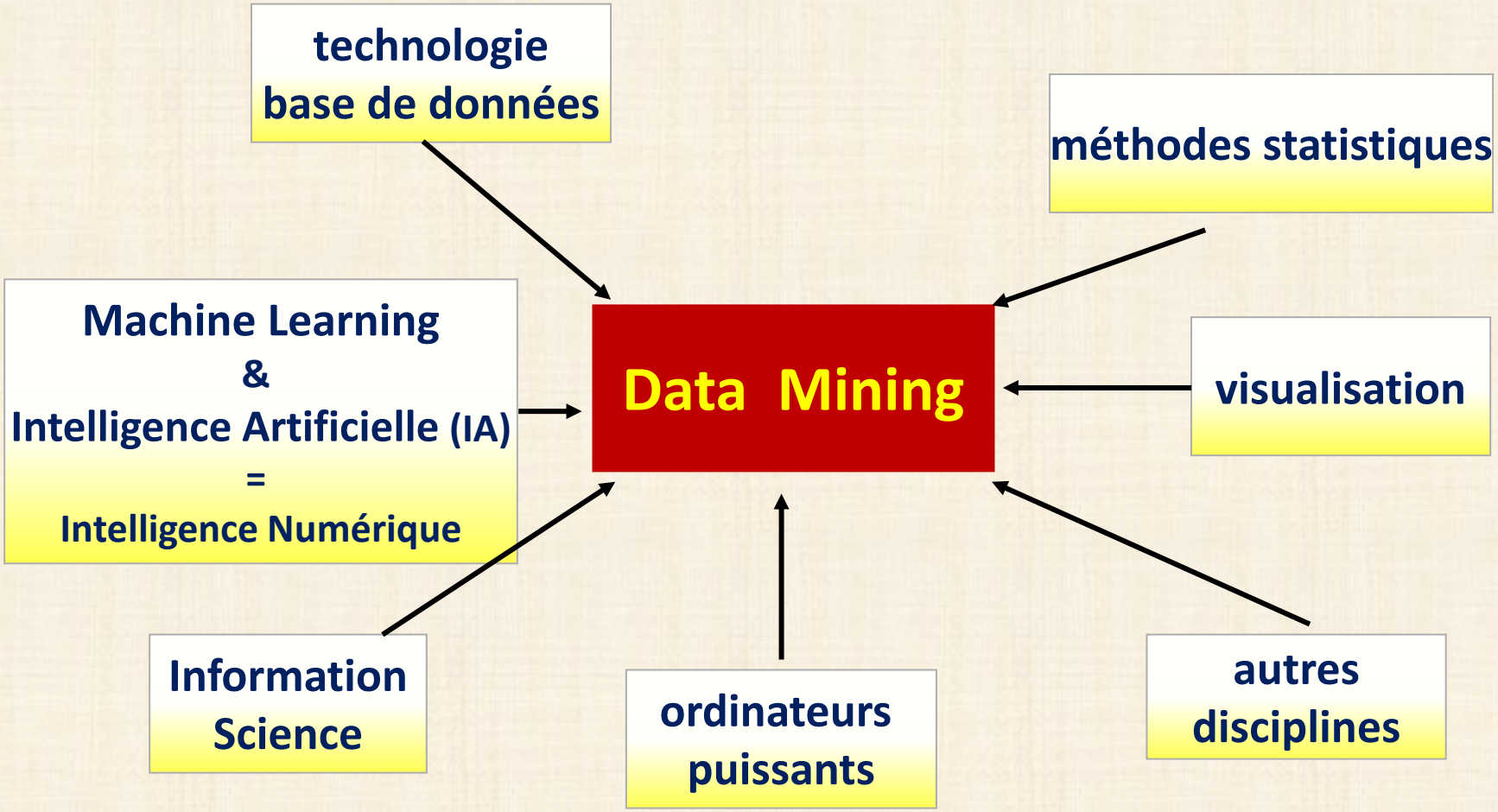
  corrélation n'est pas causalité !

- **Enjeu majeur: la qualité des données, données manquantes, données aberrantes (outliers) , biais , . . .**

NAISSANCE du DATA MINING

- L'évolution des SGBD vers l'informatique décisionnelle avec les entrepôts de données (Data Warehouse).
- La constitution de giga bases de données : transactions de cartes de crédit, appels téléphoniques, factures de supermarchés : terabytes de données recueillies automatiquement.
- Développement de la Gestion de la Relation Client (CRM) marketing client au lieu de marketing produit, attrition, satisfaction, fidélisation, efficacité des campagnes de promotion etc.
- Recherches en Intelligence artificielle (IA) apprentissage, extraction de connaissances

Data Mining convergence de plusieurs disciplines



DATA SCIENCE

Data Mining : convergence de plusieurs disciplines

Data Mining	Méthodes statistiques	Intelligence artificielle
Recherche de règles de classement	- Méthodes de discrimination - Réseaux de neuronnes - Segmentation	Apprentissage **supervisé**/ex. - règles - d'arbre de décision - raisonnement à base de cas
Régression	- Méthodes de régression - Réseaux de neuronnes	—
Classification automatique	- Classification automatique hiérarchique - Partitionnement - Réseaux de neuronnes	Apprentissage **non supervisé** -Classification°conceptuelle
Description synthétique	Stat. Élémentaire (histogramme, moy, écart-type) Outils d'interprét° de classes Méthodes factorielles (ACP)	Apprentissage **non supervisé** -Généralisation
Recherche de dépendances	Corrélations Analyse factorielles des corr. (AFC) Réseaux bayésiens	Apprentissage **non supervisé** -Généralisation -Recherche d'associations
Détection de déviations	Test stat sur les écarts	—

Exemples d'applications

- Détecter des patterns frauduleux dans les transactions sur carte de crédit
- Analyser les comportements de clients afin de proposer des achats potentiels (ex. achat de couches et de bière !)
- Identifier des stratégies pour acquérir de nouveaux clients
- Optimiser la performance de procédés manufacturiers complexes
- Déterminer des relations dans les banques de données de toute organisation de production de biens ou services

APPLICATIONS

Un site le plus reconnu en Data Mining :
http://www.kdnuggets.com

applications recensées (2003)

- banking
- bioinformatics / biotech
- direct marketing
- e-commerce / web
- entertainement
- fraud detection
- assurance
- investisement / stocks
- manufacturing
- medical / pharmaceuticals
- retail
- scientific data
- security
- supply chain analysis
- telecommunication
- travel

Kantardzic et Zurada (2005)

applications récentes du Data Mining
autres domaines que marketing / ventes

- Mining Wafer Fabrication (puces élect.)
- Damage Detection
- Sensor Array Data Processing
- Car Driver Assessment
- Discovery of Patterns in Earth Science
- Detection in Digital Imagery
- Experiences in Mining from Computer Simulation
- Gene Mapping
- Microarray Data Analysis
- Gene Expression Profiles for the Diagnosis of Diseases
- Pattern Recognition for Biomarker Discovery
- Mining the Cystic Fibrosis Data
- Learning Strategies for Web Crawling
- Data Mining for Crime Fighting
- Data Mining for Intrusion Detection
- Using Fractals in Data Mining

- Robotics
- Pattern recognition
- Image and speech analysis
- Medical diagnostics and monitoring
- Loan or credit solicitations

Besoins décisionnels :
exemple d'application : domaine bancaire

Interrogation rapports

Requête sur des données de détail

Visualisation

Combien de mouvements chaque client a-t-il effectué au cours du dernier mois?

On Line Application Process

OLAP

Analyse, détection de problèmes et opportunités

Analyse

Quelle est l'évolution sur 5 ans du nombre mensuel de mouvements pour chaque catégorie de clients?

Data Mining

Découverte de tendances cachées, règles significatives

Connaissance et prévision

Quels clients clôtureront leur compte au cours des 6 prochains mois? «churning»

Besoins décisionnels :
exemple d'application domaine bancaire

- **Entreprise :** banque
- **Activité :** prêts hypothécaires
- **Problème :** accepter ou refuser une demande de crédit?
- **Solution actuelle**

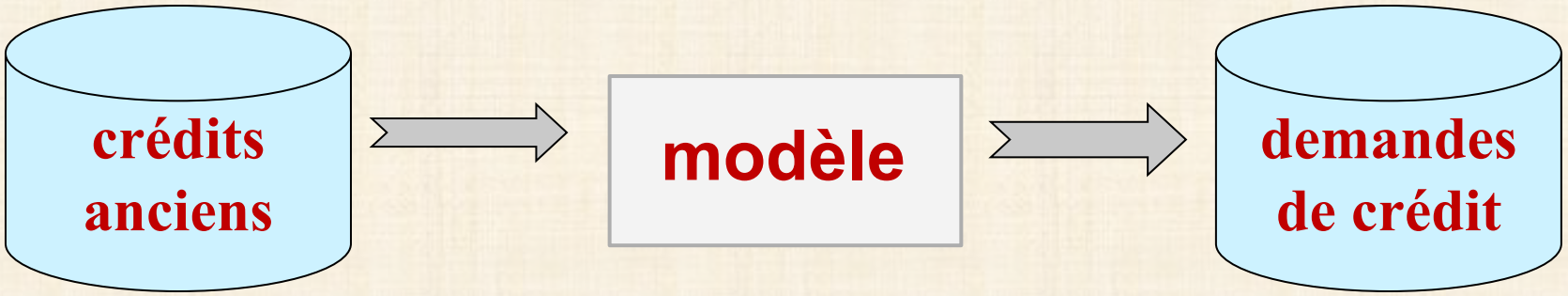
évaluation de la solvabilité du client

sur la base de critères définis par des

gestionnaires expérimentés

Une autre alternative:
gestionnaires expérimentés ⟶ Data Mining

Analyser les données historiques:
solvabilité observée lors des anciens crédits



pour prévoir la solvabilité des demandeurs de crédit

solvabilité des demandeurs de crédit

données historiques

Montant crédit	Taux cédit (%)	profession	État civil	revenus	solvabilité
100 000	7,5	enseignant	Marié	98 000	Oui
200 000	9,4	employé	Marié	108 000	Non
250 000	8,1	ouvrier	Célibat	120 000	Oui
220 000	5,3	cadre	Marié	160 000	Oui
300 000	8,1	ouvrier	Marié	150 000	Non
190 000	6,1	prof. libérale	Décédé	210 000	Oui
420 000	6,9	cadre	Marié	180 000	Oui

nouvelles données

210 000	8,2	employé	Célibat	120 000	
190 000	7,4	employé	Marié	170 000	
330 000	6,9	prof.lib.	Célibat	190 000	
170 000	7,0	cadre	Marié	205 000	
310 000	7,3	ouvrier	Marié	120 000	
240 000	6,9	fonction	Marié	110 000	
400 000	7,1	cadre	Mari.	190 000	

solvabilité des demandeurs de crédit

Exemple - fichier CreditScoring de Statistica : 1000 obs. x 20 var

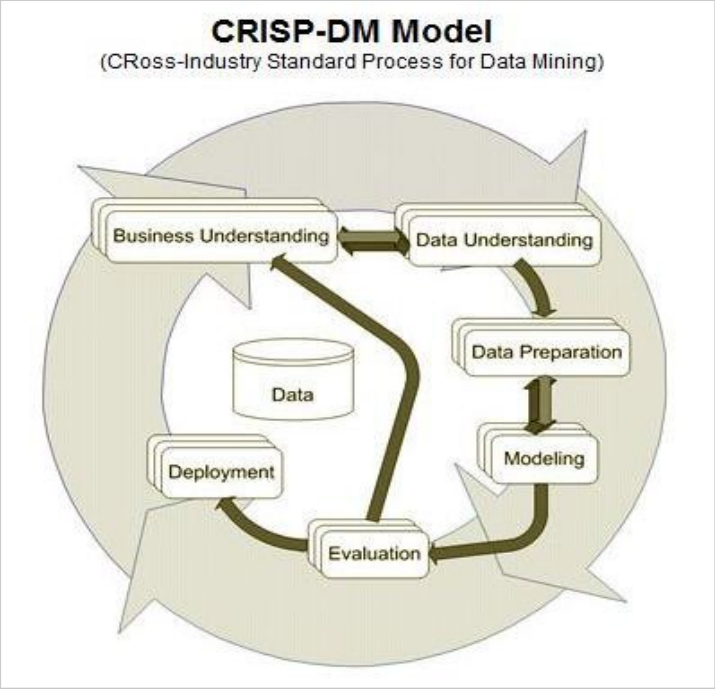
1 ID	2 Credit scoring	3 Balance of Current Account	4 Duration of Credit	5 Payment of Previous Credits	6 Purpose of Credit	7 Amount of Credit	8 Value of Savings	9 Employed by Current Employer for	10 Installment in % of Available Income	11 Marital Status	12 Gender	13 Living in Current Household for	14 Most Valuable Assets	15 Age	16 Further running credits	17 Type of Apartment	18 Number of previous credits at this bank	19 Occupation	20 TrainTest
1	bad	no running account	36	no problems with current credits	retraining	$3 003,00	no savings	5-8 years	25-35	single	male	< 1 year	life insurance	22	no further running credits	rented	2- 4	skilled employee	Test
2	good	no balance	48	hesistant	retraining	$17 085,60	>1400	1-5 years	25-35	single	male	1-5 years	life insurance	46	at other banks	rented	one	executive/self-employed	Train
3	bad	>$300	36	no previous credits	used car	$15 363,60	no savings	unemployed	< 15	divorced/living apart/mar	female	1-5 years	life insurance	24	no further running credits	rented	2- 4	executive/self-employed	Train
4	good	no running account	24	paid back	new car	$8 986,60	no savings	> 8 years	25-35	divorced/living apart/mar	female	>8 years	ownership of house or land	42	no further running credits	owned	2- 4	executive/self-employed	Test
5	good	>$300	24	no previous credits	retraining	$1 761,20	no savings	5-8 years	< 15	single	male	< 1 year	no assets	23	no further running credits	rented	one	skilled employee	Test
6	good	no balance	12	no previous credits	retraining	$1 451,80	<140	5-8 years	15- 25	single	male	>8 years	no assets	37	no further running credits	rented	one	unskilled with permanant re	Train
7	bad	no running account	30	no previous credits	used car	$4 351,20	no savings	<1 year	25-35	divorced/living apart	male	>8 years	car	29	no further running credits	rented	one	unskilled with permanant re	Train
8	good	no balance	15	paid back	furniture	$2 151,80	>1400	> 8 years	< 15	single	male	>8 years	no assets	48	no further running credits	rented	2- 4	skilled employee	Test
9	good	>$300	15	paid back	furniture	$2 059,40	no savings	1-5 years	< 15	single	male	>8 years	ownership of house or land	33	no further running credits	owned	2- 4	skilled employee	Train
10	bad	no balance	27	paid back	furniture	$3 528,00	140-700	1-5 years	< 15	single	male	1-5 years	car	21	no further running credits	rented	2- 4	unskilled with permanant re	Train
11	bad	no balance	24	no previous credits	used car	$5 679,80	no savings	5-8 years	15- 25	divorced/living apart	male	5-8 years	life insurance	41	no further running credits	rented	one	skilled employee	Train
12	bad	no running account	18	no previous credits	repair	$1 050,00	no savings	unemployed	< 15	divorced/living apart/mar	female	< 1 year	no assets	25	no further running credits	rented	one	unskilled with no permanant	Train
13	bad	no balance	36	no problems with current credits	retraining	$6 237,00	no savings	1-5 years	25-35	divorced/living apart	male	1-5 years	no assets	28	at department store	rented	2- 4	executive/self-employed	Train
14	good	>$300	6	no problems with current credits	retraining	$2 440,20	<140	1-5 years	> 35	single	male	1-5 years	no assets	32	no further running credits	rented	2- 4	unskilled with permanant re	Test
15	good	no running account	12	no previous credits	other	$2 650,20	no savings	1-5 years	< 15	divorced/living apart/mar	female	>8 years	car	27	no further running credits	rented	one	skilled employee	Train
16	good	>$300	42	no previous credits	furniture	$10 032,40	>1400	5-8 years	25-35	married/widowed	male	>8 years	car	27	no further running credits	free	one	skilled employee	Test
17	good	no running account	48	no previous credits	new car	$6 703,20	no savings	5-8 years	< 15	single	male	5-8 years	car	24	no further running credits	rented	one	skilled employee	Train

. . . . . .

980	good	no running account	24	no previous credits	used car	$5 836,60	no savings	1-5 years	< 15	single	male	>8 years	car	26	no further running credits	rented	one	skilled employee	Test
981	good	no balance	18	no previous credits	furniture	$1 821,40	no savings	> 8 years	< 15	married/widowed	male	1-5 years	no assets	30	no further running credits	rented	one	unskilled with permanant re	Test
982	bad	no balance	15	problematic running accounts	other	$1 769,60	<140	1-5 years	25-35	married/widowed	male	1-5 years	car	23	no further running credits	free	one	skilled employee	Train
983	good	no balance	7	no previous credits	furniture	$3 260,60	no savings	<1 year	> 35	divorced/living apart/mar	female	< 1 year	no assets	43	no further running credits	rented	one	skilled employee	Test
984	good	<= $300	24	no previous credits	furniture	$1 927,80	<140	> 8 years	< 15	divorced/living apart/mar	female	1-5 years	ownership of house or land	45	no further running credits	owned	one	skilled employee	Train
985	good	no running account	8	paid back	business	$1 629,60	no savings	> 8 years	15- 25	single	male	>8 years	ownership of house or land	49	at other banks	owned	2- 4	executive/self-employed	Test
986	good	>$300	18	paid back	furniture	$880,60	140-700	> 8 years	< 15	single	male	5-8 years	car	30	at other banks	rented	2- 4	executive/self-employed	Train
987	good	<= $300	24	no previous credits	used car	$5 248,60	no savings	<1 year	25-35	divorced/living apart/mar	female	>8 years	life insurance	24	no further running credits	rented	one	skilled employee	Train
988	bad	no balance	48	paid back	used car	$7 134,40	no savings	1-5 years	25-35	divorced/living apart/mar	female	5-8 years	life insurance	28	no further running credits	rented	one	executive/self-employed	Train
989	good	no balance	24	no problems with current credits	furniture	$8 964,20	no savings	<1 year	> 35	single	male	1-5 years	life insurance	31	no further running credits	rented	one	skilled employee	Train
990	bad	no running account	48	no previous credits	furniture	$9 798,60	no savings	5-8 years	> 35	married/widowed	male	< 1 year	no assets	32	no further running credits	rented	2- 4	skilled employee	Test
991	good	no balance	24	paid back	new car	$10 861,20	700-1400	> 8 years	25-35	divorced/living apart/mar	female	>8 years	ownership of house or land	27	no further running credits	free	one	skilled employee	Test
992	bad	>$300	15	no previous credits	repair	$6 472,20	<140	1-5 years	15- 25	single	male	1-5 years	car	38	no further running credits	rented	one	executive/self-employed	Train
993	bad	no running account	24	no previous credits	other	$2 018,80	no savings	5-8 years	< 15	divorced/living apart/mar	female	>8 years	life insurance	21	no further running credits	free	2- 4	skilled employee	Train
994	good	>$300	24	paid back	repair	$2 697,80	>1400	1-5 years	15- 25	divorced/living apart/mar	female	1-5 years	life insurance	31	no further running credits	rented	2- 4	skilled employee	Train
995	bad	no balance	60	no previous credits	repair	$8 803,20	no savings	1-5 years	< 15	single	male	>8 years	ownership of house or land	40	no further running credits	owned	one	skilled employee	Train
996	bad	>$300	21	hesistant	other	$7 004,20	>1400	1-5 years	> 35	divorced/living apart/mar	female	>8 years	car	27	at other banks	rented	2- 4	skilled employee	Train
997	good	no balance	30	no previous credits	used car	$5 364,80	no savings	<1 year	25-35	married/widowed	male	< 1 year	car	20	no further running credits	rented	one	skilled employee	Train
998	good	no balance	9	paid back	furniture	$1 615,60	no savings	> 8 years	25-35	single	male	>8 years	no assets	35	no further running credits	rented	5- 6	unskilled with permanant re	Test
999	good	>$300	24	paid back	usehold applian	$2 881,20	no savings	1-5 years	< 15	divorced/living apart	male	1-5 years	no assets	31	no further running credits	rented	2- 4	skilled employee	Train
1000	good	no balance	36	no problems with current credits	other	$4 006,80	<140	> 8 years	< 15	single	male	5-8 years	ownership of house or land	28	no further running credits	owned	one	skilled employee	Train

MODÈLES D'IMPLANTATION

- ***CRISP*** **European consortium of companies standard process model for data mining**



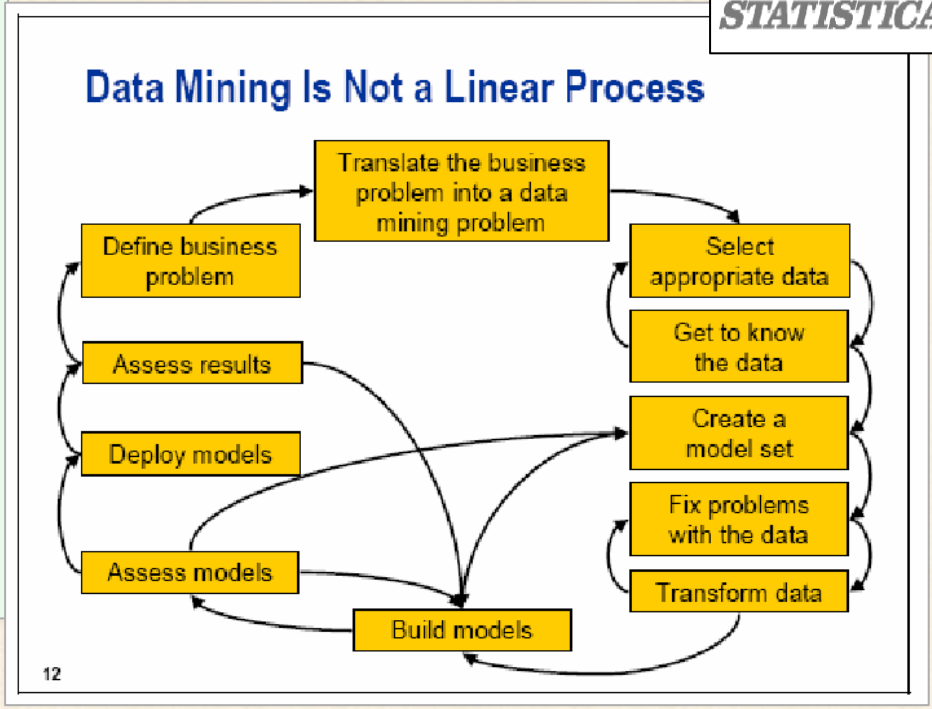
- ***SEMMA*** **(SAS Institute)**

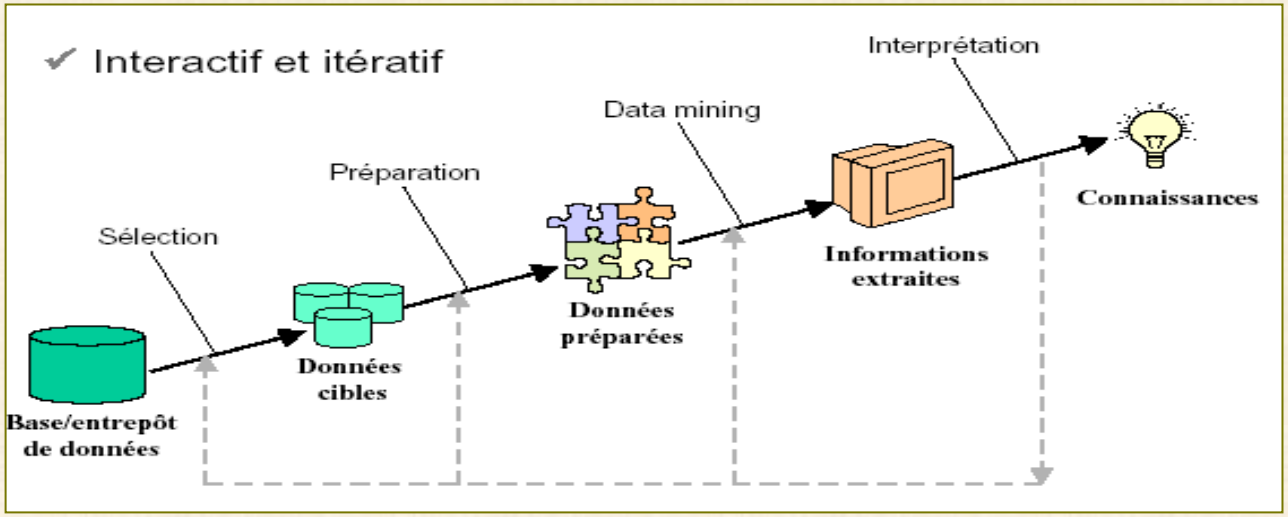
Sample → Explore → Modify → Model → Assess

Étapes du processus DATA MINING

1. **Définir le problème**
2. **Identifier et préparer les données (80% de l'activité du DM)**
3. **Construire le modèle et le tester**
4. **Évaluer le modèle et choisir la technique optimale**

PUIS
appliquer (déployer) le modèle aux données récentes et
Interpréter les résultats



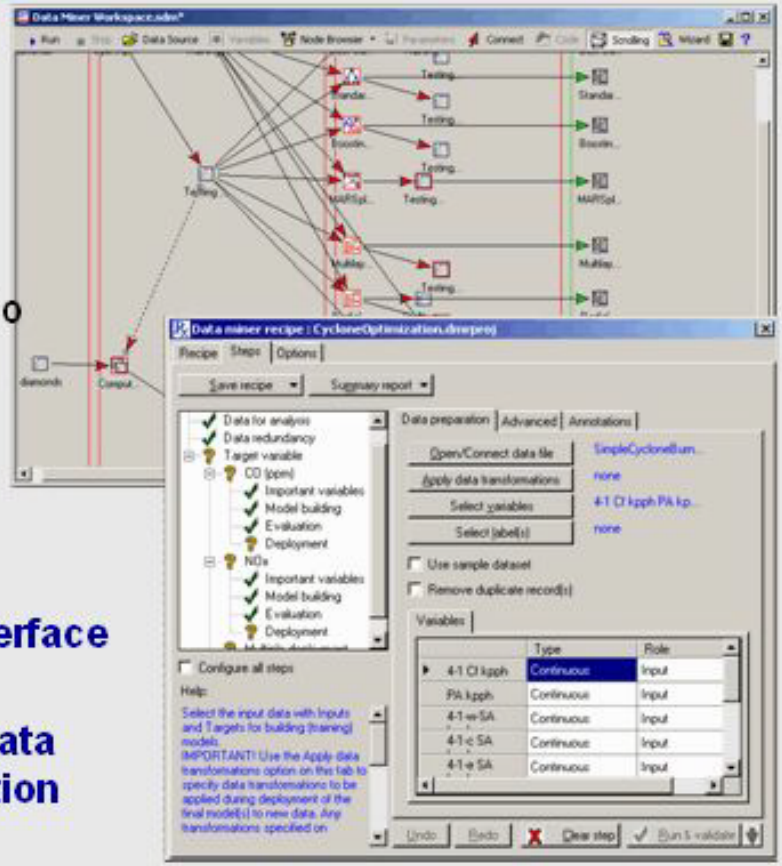


DataMiner Recipes de Statistica

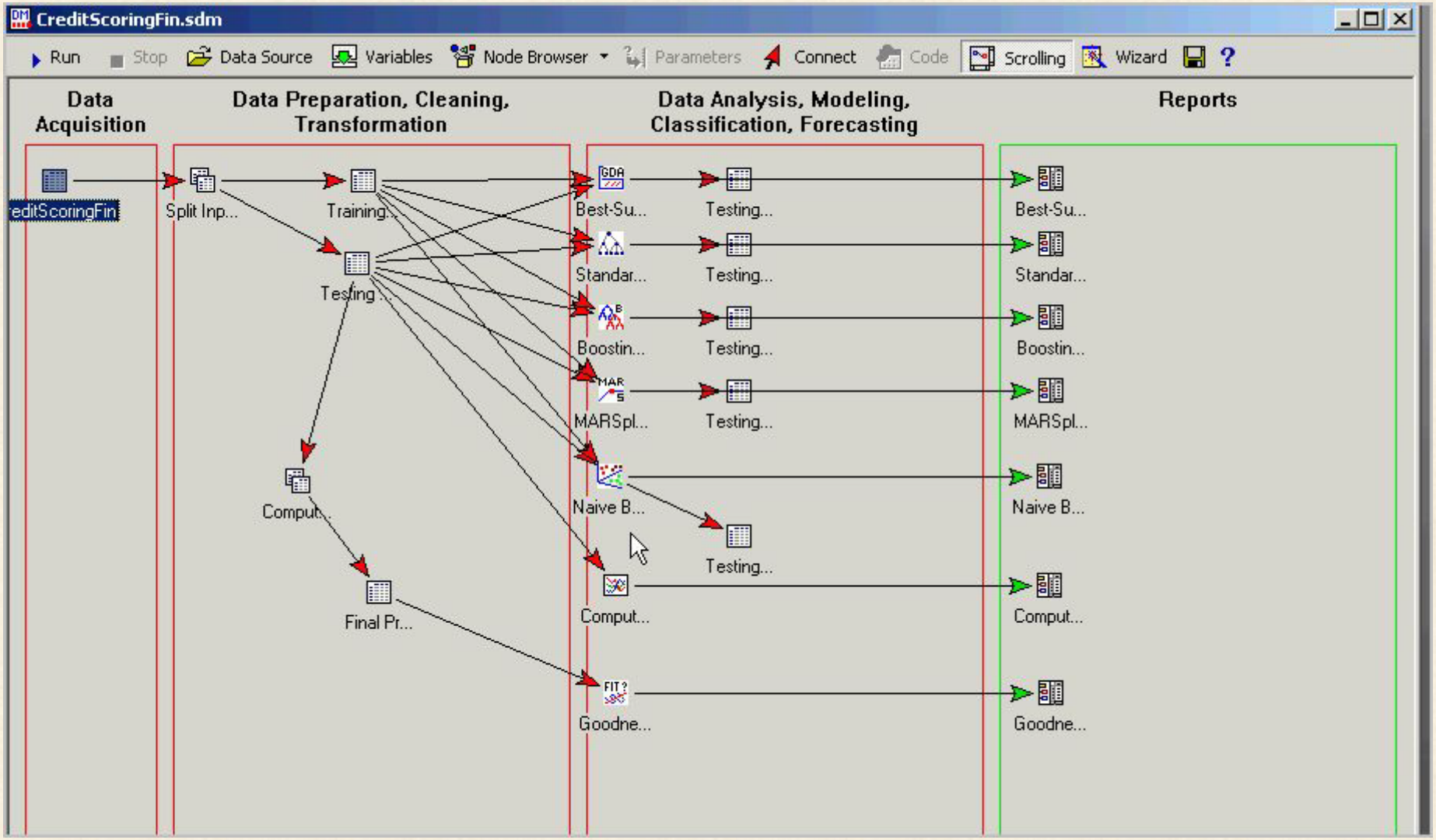
Data Miner Workspaces, Recipes

StatSoft STATISTICA

- Efficient UI solutions for effective data mining
- Data Miner Workspaces
  - Custom data mining workflows or templates (best practices)
  - Use *all STATISTICA* functionality to build work flows that go beyond traditional data mining (e.g., perform Predictive QC-Mining,..)
- **Run on desktop, offload to server**
- **Data Miner Recipes**
  - **Revolutionary, efficient user interface for both novices and experts**
  - **Single-click data mining, from data definition through model validation and deployment/ scoring**



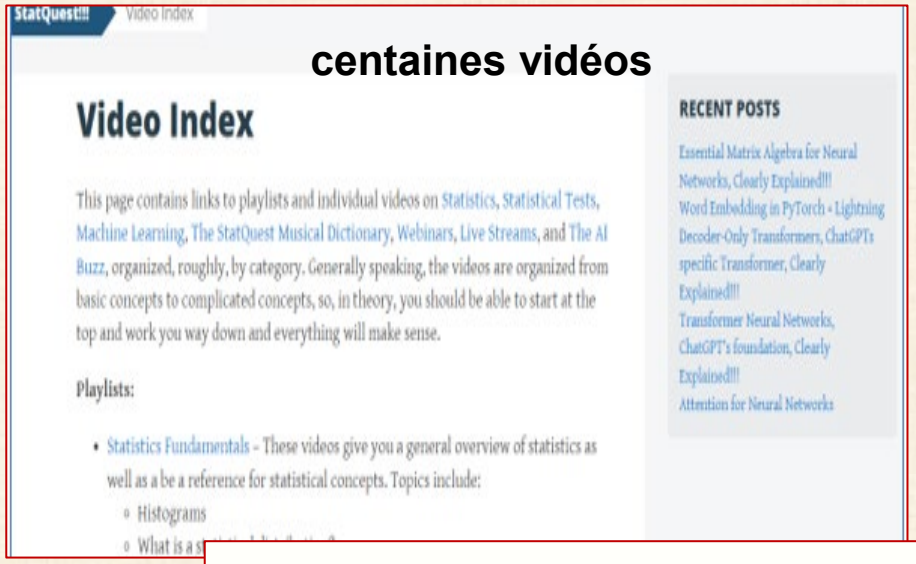
DataMiner Recipes de Statistica



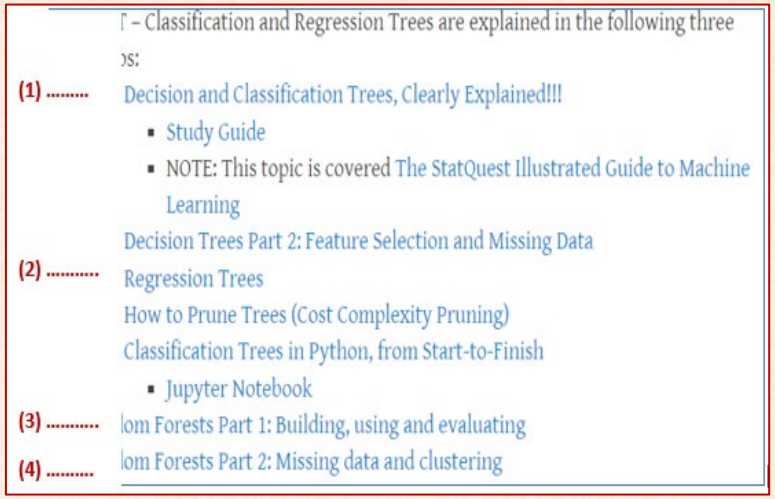
Un site recommandé : STATQUEST
site de Joshua Starmer

STATQUEST!!!
An epic journey through statistics and machine learning

https://statquest.org/video-index/



centaines vidéos



(1)	https://www.youtube.com/watch?v=_L39rN6gz7Y	Arbre classification : Y catégorique
(2)	https://www.youtube.com/watch?v=g9c66TUylZ4	Arbre de régression : Y continue
(3)	https://www.youtube.com/watch?v=J4Wdy0Wc_xQ	Forêts aléatoires 1
(4)	https://www.youtube.com/watch?v=sQ870aTKqiM	Forêts aléatoires 2

https://www.youtube.com/watch?v=_L39rN6gz7Y vidéo arbre classification

Données exemple du vidéo : arbre classification chapitre 9

	1 ID	2 X1_loves Popcorn	3 X2_loves Soda	4 X3_Age	5 Y_Loves Cool as Ice
1	1	yes	yes	7	no
2	2	yes	no	12	no
3	3	no	yes	18	yes
4	4	no	yes	35	yes
5	5	yes	yes	38	yes
6	6	yes	no	50	no
7	7	no	no	83	no

Monographies & Articles (pas à jour ….)

Berry, M., J., A., & Linoff, G., S., (2000). *Mastering Data Mining*. New York: Wiley

D.Hand (1999): *Why data mining is more than statistics write large*, ISI,Helsinki, http://www.stat.fi/isi99/index.html

D.Hand (2000): *Methodological Issues in Data Mining*, in Compstat 2000, Physica-Verlag, 77-85, 2000

Edelstein, H., A. (1999). *Introduction to Data Mining and Knowledge Discovery (3rd ed)*. Potomac, MD: Two Crows Corp.

Fayyad, U. M., Piatetsky-Shapiro, G., Smyth, P., & Uthurusamy, R. (1996). *Advances In Knowledge Discovery & Data Mining*. Cambridge, MA: MIT Press.

Friedman J. (1997): *Data Mining and Statistics, What's the Connection?* http://www-stat.stanford.edu/~jhf/ftp/dm-stat.ps

Friedman J. (1999): *The role of Statistics in Data Revolution*, ISI, Helsinki, http://www.stat.fi/isi99/index.html

Friedman J. (2009): première heure du cours STAT315B (Stanford Univ.) sur le Data Mining (donné à l'hiver 2009)
ttp://myvideos.stanford.edu/player/slplayer.aspx?course=STATS315B&p=true

Gaudard, M. Ramsey, P., Stephens, M. ((2006). *Interactive Data Mining and Design of Experiments: The JMP Partition and Custom Design Platforms*. North Haven Group, LLC

Giudici, P. (2003*). Applied Data Mining: Statistical Methods for Industry*, John Wiley & Sons.

Han, J., Kamber, M. (2000). *Data Mining: Concepts and Techniques*. New York: Morgan-Kaufman.

Hastie, T., Tibshirani, R., & Friedman, J. H. (2001). *The elements of Statistical Learning : Data Mining, Inference, and Prediction*. New York: Springer.

Kantardzic, M.M., Zurada, J. (editors) (2005). *Next Generation of Data-Mining Applications.* John Wiley & Sons, Copyright the Institute of electrical and Electronic Engineers (IEEE).

Larose, Daniel T, (2005*) Discovering Knowledge in Data : An Introduction to Data Mining* . John Wiley & Sons.

Nisbet R., Elder, J., Miner, G. (2009*) Handbook of Statistical Analysis & Data Mining Apllications*, Academic Press. ISBN 978-0-12-374765-5

Pregibon, D. (1997). *Data Mining*. Statistical Computing and Graphics, 7, 8.

StatSoft : 35 vidéos de 8-10 minutes sur YouTube http://www.statsoft.com/support/download/video-tutorials/

Tufféry, S. (2007). *Data Mining et statistique décisionnelle*, Éditions TECHNIP, Paris.

Weiss, S. M., & Indurkhya, N. (1997). *Predictive Data Mining: A practical Guide*. New York: Morgan-Kaufman.

Westphal, C., Blaxton, T. (1998). *Data Mining Solutions*. New York: Wiley.

Witten, I. H., & Frank, E. (2000). *Data Mining*. New York: Morgan-Kaufmann.

Monographies sur les réseaux de neurones

Bishop, C. (1995). *Neural Networks for Pattern Recognition*. Oxford: University Press.

Carling, A. (1992). *Introducing Neural Networks*. Wilmslow, UK: Sigma Press.

Fausett, L. (1994). *Fundamentals of Neural Networks*. New York: Prentice Hall.

Haykin, S. (1994). *Neural Networks: A Comprehensive Foundation*. New York: Macmillan Publishing.

Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, 43:59-69.

Patterson, D. (1996). *Artificial Neural Networks*. Singapore: Prentice Hall.

Ripley, B.D. (1996). *Pattern Recognition and Neural Networks*. Cambridge University Press.

Rumelhart, D.E., and J.L. McClelland (1986), *Parallel Distributed Processing*, Volume 1. The MIT Press. Foundations.

Sites Internet

http://www.kdnuggets.com/
http://www.ccsu.edu/datamining/
http://www.math.ccsu.edu/dm/dm%20resources.htm
http://www.dmreview.com
http://www.scd.ucar.edu/hps/GROUPS/dm/dm.html
http://www.infogoal.com/dmc/dmcdwh.htm