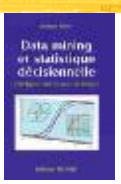
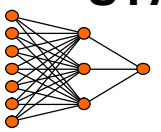
Stéphane Tufféry

Statisticien - Data Miner - Enseignant

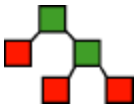


DATA MINING - SCORING

STATISTIQUE DÉCISIONNELLE



APPLICATION AU CRM



Plan du cours

- Qu'est-ce que le data mining ?
- A quoi sert le data mining ?
- Les 2 grandes familles de techniques
- Le déroulement d'un projet de data mining
- Coûts et gains du data mining
- Facteurs de succès - Erreurs à éviter
- Informatique décisionnelle et de gestion
- La préparation des données
- Techniques descriptives de data mining
- Techniques prédictives de data mining
- Logiciels et consultants
- Le text mining
- Le web mining
- *CNIL et limites légales du data mining*

Les limites légales de l'utilisation des données

« Informatique et libertés »

- Comme tout traitement informatique de données sur des personnes physiques, le data mining obéit en France à un certain nombre de règles, édictées dans ces textes :
  - la loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés
  - les délibérations de la Commission Nationale de l'Informatique et des Libertés (CNIL)
  - la Convention n° 108 du Conseil de l'Europe du 28/1/1981
  - la loi n° 2004-801 du 6 août 2004 relative à la protection des personnes physiques à l'égard des traitements de données à caractère personnel, transposant en droit français la directive européenne 95/46/CE, en modifiant la loi n° 78-17
- On trouve ces règles sur le site www.cnil.fr de la CNIL

Champ d'application des textes

- Informations nominatives (dites encore *données à caractère personnel*), c'est-à-dire celles concernant les personnes physiques identifiées ou identifiables
  - ne sont pas concernés les fichiers de personnes physiques rendues anonymes par la suppression de tout identifiant
  - sont concernés les fichiers d'entreprises unipersonnelles
- Traitements informatiques de ces informations
  - même ceux qui se bornent à les collecter et les enregistrer, en dehors de toute mise en œuvre ou de toute exploitation (article 5 de la loi 78-17).

Données légalement utilisables

- Ne doivent être, sauf cas particulier, ni traitées ni même collectées, les informations sur :
  - les origines raciales
  - les opinions politiques, philosophiques ou religieuses
  - les appartenances syndicales
  - les mœurs
  - la santé
  - les condamnations pénales
  - NIR, no d'inscription au répertoire national d'identification des personnes physiques

Droits des personnes physiques 1/3

- Les personnes physiques ont le droit que les informations nominatives les concernant soient :
  - légalement utilisables
  - loyalement collectées
  - stockées en sécurité
  - communiquées aux seuls tiers autorisés
  - rectifiées si besoin est
  - enregistrées pour des finalités déterminées et légitimes, par rapport auxquelles elles sont pertinentes et non excessives
  - effacées au bout d'un certain délai (correspondant à la finalité du traitement déclaré).

Droits des personnes physiques 2/3

- De façon générale, les personnes physiques :
  - doivent avoir connaissance des destinataires des informations nominatives qu'elles fournissent, de leur *droit d'accès et de rectification*, et, le cas échéant, de la possibilité de ne pas répondre aux questions facultatives : *droit à l'information*
  - doivent avoir connaissance de la cessibilité d'informations nominatives les concernant avec une finalité identique à celle d'origine
  - doivent avoir connaissance de la cessibilité d'informations nominatives les concernant avec une finalité différente de celle d'origine (prospection commerciale par exemple) et doivent les accepter expressément.

Droits des personnes physiques 3/3

- De façon générale, les personnes physiques :
  - peuvent avoir connaissance (à leur demande) des informations nominatives mémorisées les concernant, de l'existence et de la finalité d'un traitement informatique les concernant (articles 22 et 34) : *droit d'accès*
  - peuvent s'opposer, pour des raisons légitimes, à un traitement informatique d'informations nominatives les concernant (article 26) : *droit d'opposition*
  - ne peuvent pas exiger d'avoir connaissance du détail du traitement, à moins (article 3) que ces traitements fondent une décision qu'elles contestent.

Déclarations de traitements

- La déclaration à faire à la CNIL *préalablement à la mise en œuvre d'un nouveau traitement* automatisé de données à caractère personnel (un traitement de data mining, par exemple) est :
  - soit une ***déclaration simplifiée***, qui n'exige qu'un minimum d'informations, mais l'engagement que la déclaration soit strictement conforme à l'une des normes simplifiées en vigueur
  - soit une ***déclaration ordinaire***, dans les autres cas.
- Les déclarations de sites Web peuvent être faites en ligne.
- La CNIL a reçu 69 352 déclarations de traitement en 2003, et son « fichier des fichiers » recensait 941 076 traitements le 31/12/2003.

La loi 2004-801
(transposant la directive européenne 95/46/CE)

- Abolit la distinction entre secteurs public et privé
  - hormis les traitements publics liés à la sécurité
  - le secteur public n'est plus le seul à devoir requérir l'autorisation préalable de la CNIL dans certains cas
- Instaure une distinction entre traitements sensibles ou non
  - traitements sensibles : demande d'autorisation préalable
  - autres traitements : déclaration
    - voire exonération de déclaration (ex : paie du personnel)
- Un traitement peut être dit sensible en raison de :
  - la nature des données (NIR, données biométriques, génétiques, sensibles, relatives aux condamnations...)
  - l'ampleur des traitements (totalité de la population française)
  - la finalité des traitements (scoring, exclusion du bénéfice d'un droit, « listes noires », interconnexion de fichiers...)

La loi 2004-801
(transposant la directive européenne 95/46/CE)

- Crée les « correspondants à la protection des données » (CPO) dans les entreprises (article 22)
  - chargés de tenir le registre des traitements mis en œuvre et d'assurer le respect des obligations légales
  - non obligatoires pour l'entreprise
  - dispensent l'entreprise des déclarations mais non des autorisations préalables de traitements sensibles
  - nommés par l'entreprise sans accréditation de la CNIL
  - pourront être choisis au sein ou à l'extérieur de l'entreprise
  - devront jouir d'une certaine indépendance dans l'entreprise

Nouveaux pouvoirs de la CNIL
dans la loi de 2004

- Accéder à tout local professionnel servant à l'exploitation d'un fichier
- Rendre publics ses avertissements
- Infliger des amendes jusqu'à 150 000 € (300 000 € en cas de récidive)
  - au lieu de se limiter à dénoncer les infractions au Parquet
- Retirer une autorisation déjà donnée
- Interdire un traitement pendant une durée max de 3 mois

Spécificités du scoring de risque

- Un score de risque doit faire l'objet d'une *déclaration ordinaire* et non simplifiée.
- Cette déclaration doit indiquer les variables utilisées, les *paramètres du score* et les grilles de pondération.
- Aucune décision accordant ou refusant un crédit *ne peut avoir pour seul fondement un traitement automatisé* d'informations donnant une définition du profil ou de la personnalité de l'intéressé.
- Toute personne à laquelle un refus de crédit est opposé bénéficie du *droit d'accès* aux informations utilisées lors de l'examen de sa demande (y compris sa note de score) et peut, le cas échéant, en exiger la rectification.

Spécificités de la segmentation

- La CNIL admet l'affectation des clients en segments de clientèle, sous les réserves suivantes :
  - informations collectées « adéquates, pertinentes et non excessives »
  - droit d'accès aux informations
  - mise à jour périodique de l'affectation à un segment
  - non-automaticité et non-inéluctabilité des décisions en découlant
  - non-cession de ces informations à des tiers non autorisés.
- Les segments ne doivent pas comporter de qualificatifs péjoratifs, défavorables ou subjectifs sur les catégories d'individus, tels que « tempérament de jouisseur » ; sont en revanche admises les catégories : « vivant à crédit », « clients aisés et âgés », « petits épargnants ».

Fichiers de crédit

- Fichiers négatifs : ne contiennent que les emprunteurs ayant des incidents de remboursement
  - fichier FICP en France
- Fichiers positifs : contiennent tous les emprunteurs, avec les montants et échéances des endettements contractés
  - voire des données sur les charges, les revenus, le patrimoine, le logement, l'emploi... (aux USA)
  - fichiers utilisés par les professionnels du crédit, avec ouverture à la téléphonie, la VPC, le secteur du logement
  - fichiers interdits en France, mais autorisés en Allemagne, Autriche, Italie, Espagne, Portugal, Pays-Bas, Belgique, Royaume-Uni...
  - débat relancé fin 2003 par le secrétaire d'État aux PME mais la CNIL réaffirme son opposition le 13/5/2004 et le CNCT le 20/7/2004